DOI:10.13878/j.cnki.jnuist.20240410002

岳有军[1]　漆潇[1]　赵辉[2]　王红君[2]

# 基于改进 YOLOv8 的果园复杂环境下苹果检测模型研究

**摘要**

为了使采摘机器人能够在果园复杂环境下(如不同光照条件、叶子遮挡、密集的苹果群和超远视距等场景)对成熟程度各异的苹果果实进行快速且精确的检测,本文提出一种基于改进 YOLOv8 的苹果果实检测模型.首先,将 EMA 注意力机制模块集成到 YOLOv8 模型中,使模型更加关注待检测果实区域,抑制背景和枝叶遮挡等一般特征信息,提高被遮挡果实的检测准确率;其次,使用提取特征更加高效的三支路 DWR 模块对原始 C2f 模块进行替换,通过多尺度特征融合方法提高小目标检测能力;同时结合 DAMO-YOLO 的思想,对原始 YOLOv8 颈部进行重构,实现高层语义和低层空间特征的高效融合;最后,使用 Inner-SIoU 损失函数对模型进行优化,提高识别精度.在复杂的果园环境中,以苹果作为检测对象,实验结果表明:本文所提算法在测试集下的查准率、召回率、$mAP_{0.5}$、$mAP_{0.5\sim0.95}$以及 F1 分数分别达到 86.1%、89.2%、94.0%、64.4%和 87.6%,改进后的算法在大部分指标上均优于原始模型.在不同数量果实场景下的对比实验结果表明,该方法具有优异的鲁棒性.

**关键词**

模式识别;深度学习;目标检测;YOLOv8

**中图分类号** TP391.4

**文献标志码** A

**收稿日期** 2024-04-10

**资助项目** 天津市科技支撑计划(19YFZCSN00360,18YFZCNC01120)

**作者简介**

岳有军,男,博士,教授,研究方向为复杂系统建模及智能控制、机器人导航与控制技术.17709321913@163.com

漆潇(通信作者),男,硕士生,研究方向为目标检测、农业机器人.2807210522@qq.com

1 天津理工大学 电气工程与自动化学院,天津,300384

2 天津理工大学 天津市复杂系统控制理论及应用重点实验室,天津,300384

## 0 引言

目前,水果采摘机器人已在农业生产中得到广泛应用[1],它能够有效地解决人工采摘成本高、效率低以及劳动力不足等问题[2].由于果实识别的速度、准确率以及在不同复杂环境下的适应能力直接影响机器人的工作效率和稳定性,因此,研究在复杂环境下准确检测果实的视觉模型具有重要意义[2].

早期针对水果图像检测的研究主要分为传统特征观察分析法和基于机器视觉的浅层模型算法两种.它们主要通过观察果实的颜色、形状和边缘纹理等特征,或利用一些特征增强技术使果实表现出与背景不同的显著特征来检测果实.例如:Wu 等[3]基于模型分割和聚类的方法提出一种利用 RGB-D 相机将 3D 轮廓特征和颜色数据相结合用来进行水蜜桃的检测,检测准确率达 88.68%,但单个水果处理时间较长;Feng 等[4]提出一种基于多光谱动态成像技术的算法来定位水果区域,但该算法只对完整果实识别精度较高,对不完整果实区域的识别精度只有 72%;Lv 等[5]提出一种基于压缩感知和离散小波变换-非下采样轮廓波变换多源图像融合的识别方法进行果园环境下绿色葡萄识别,其平均识别准确率达 92%.传统特征观察分析法主要通过水果颜色或形态进行图像特征的额外设计,因此模型检测速度较慢,且通用性不强.初广丽等[6]针对球状类水果,采用边缘跟踪算法和图像分割算法从图像中分离出完整目标,将分离出的完整目标利用最小二乘法进行识别,准确率达 95%以上.基于机器视觉的浅层模型算法促进了水果目标检测的发展和研究,但受光照强度、树叶枝干遮挡等自然环境因素的影响较大,导致其识别率较低、模型通用性较弱.

近年来,基于深度学习的目标检测已成为水果识别的一个热门研究方向.例如:赵德安等[7]为解决不同光照下果实遮挡、套袋以及黏连导致识别精度不高等问题,提出一种基于 YOLOv3 网络模型的苹果检测算法,经过训练的模型在验证集下的准确率为 97%,但该算法并未在密集苹果群中取得很好的检测效果;王金鹏等[8]提出一种轻量化 YOLOv4-LITE 目标检测算法用来检测真实果园环境下的火龙果,该模型检测准确率为 96.48%,但训练模型数据集较少,且火龙果背景较为单一,通用性不强;易诗等[9]提出一种基于特征递归融合 YOLOv4 模型,在果园环境下对春见柑橘的平均检测精度为 94.6%,但该模型并未考虑有遮挡物的情况;龙燕等[10]设计了一种适用于近

景色小目标检测的深度学习网络,旨在实现疏果期苹果目标准确检测,其平均精度为 95.2%,检测准确率为 92.7%,召回率为 91.0%,与原模型相比具有更强的鲁棒性和更好的检测效果,但该模型内存占用较大、实时性较差.现有果实检测模型存在以下问题待解决:1)在果园真实环境下远距离检测,以及被测部分有遮挡情况时出现精度下降、置信度偏低、漏检等问题;2)已有算法主要针对成熟水果,对未成熟水果并未做进一步研究,故大多数水果检测模型通用性不强.

针对上述问题,本文以苹果作为研究对象,提出一种基于改进 YOLOv8 的果园复杂环境下的苹果检测方法,在保证检测效果的同时,能够满足实时性需求,并适应果园环境中近远距离、不同光照、不同成熟程度的苹果以及不同程度遮挡情况下的检测挑战.主要改进思路如下:1)通过融合 EMA[11](Efficient Multi-scale Attention)注意力模块和三支路 DWR[12](Dilation-Wise Residual)模块对主干网络进行改进;2)利用 Rep-GFPN[13] 颈部网络对原始 YOLOv8 颈部网络进行重构,约束计算量使其更加轻量化,在不增加太多计算成本的同时能够容纳更多的特征;3)通过引入 Inner-SIoU[14](Inner Spatial Intersection over Union)内部标度交并比损失函数,增强模型检测效果的泛化能力.

## 1 研究方法

### 1.1 YOLOv8 网络结构

YOLOv8[15] 由 Ultralytics 开发团队在 2023 年 1 月提出,是目前主流的目标检测算法.YOLOv8 模型由 Input、Backbone、Neck 和 Head 四部分组成.在 Input 中存在 Mosaic 数据增强方法,并且针对大小不一的模型,部分超参数也会随之改变,例如大模型会采取 MixUp 和 CopyPaste 方法进行数据增强,此操作能够有效提升模型的泛化能力和鲁棒性.Backbone 将图片信息提取之后供给 Neck 和 Head 使用,由若干个 Conv、C2f 模块以及末端的 SPPF 组成.Conv 模块用于提取特征并同时实现特征图的整理,它仅由 Conv2d 和 BatchNorm2d 两部分构成,通过借鉴 YOLOv7[16] 和 C3 模块的残差结构,设计出 YOLOv8 独有的 C2f 结构,在确保轻量的条件下仍能获取大量的梯度流信息,同时其通道数可随模型尺度大小发生相应变化,实现模型性能有效提升.SPPF 通过空间金字塔池化操作,可将不同尺度的特征进行融合.Neck 部分采用 FPN(Feature Pyramid Network)+PAN(Path Aggregation Network)的结构,可将骨干网络提取的特征充分利用从而实现特征融合,该结构可提高多尺度的定位能力以及语义表达能力.Head 部分通过 Backbone 和 Neck 处理之后的特征来获得检测对象的种类以及位置信息并做出识别,其结构也升级为当前主流的解耦头结构,实现了分类和检测头的分离,有效地解决了分类和定位关注侧重点不一的问题,同时采取无锚框(Anchor-Free)的目标检测方法来提升检测速度.

### 1.2 改进的 YOLOv8 苹果检测模型

#### 1.2.1 注意力机制模块

注意力机制对局部重要信息有着优异的侧重效果,可以使模型更加关注检测对象的相关特征,已广泛应用于各种领域的计算机视觉任务.EMA 是一种高效多尺度的注意力机制,之所以能够在降低计算成本的同时保留各个通道的信息,是因为其可以将部分通道进行重塑从而成为批量维度,有效地避免通道降维情况的发生.EMA 相较于其他注意力机制有以下两大优点:一是其能够通过全局信息编码实现并行子网络通道权重的调整;二是其 2 个并行子网络的输出特征可以利用跨纬度交互的方式实现.

EMA 总体结构如图 1 所示.对于任意输入特征 $\boldsymbol{X} \in \mathbf{R}^{C\times H\times W}$,EMA 为学习不同种类的语义,将 $\boldsymbol{X}$ 划分为跨通道方向的 $G$ 个子特征,输入特征 $\boldsymbol{X}=[\boldsymbol{X}_0,\boldsymbol{X}_i,\cdots,\boldsymbol{X}_{G-1}]$,$\boldsymbol{X}_i \in \mathbf{R}^{C//C\times H\times W}$.为满足不损失一般性的前提,令 $G \ll C$,并假设每个子特征中感兴趣区域的特征用注意力权重符来增强.输入张量形状定义为 $C//G\times H\times W$,其中包括 2 条 1×1 分支的并行支路和 1 条 3×3 分支路径.在 1×1 支路中采用水平/垂直 2 个维度方向的一维全局平均池化操作,即可在减轻计算量的同时获取所有通道间的依赖关系,数据特征处理操作如下:

$$Z_C^H(H)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}\boldsymbol{X}_C(H,i), \tag{1}$$

$$Z_C^W(W)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}\boldsymbol{X}_C(j,W). \tag{2}$$

式中:$C$ 代表通道数;$H$、$W$ 为输入特征空间的维度.

在不降低维度情况下完成特征提取后,通过 2 个 Sigmoid 非线性函数拟合线性卷积后的二项分布,再通过乘法运算实现每个组内通道注意力图的聚合,然后对 1×1 分支输出利用二维全局平均池化组归一化(group norm)对全局空间信息进行编码.在另一支路

中采用3×3卷积获取局部跨通道交互特征来扩大特征空间.EMA通过采用1×1和3×3的卷积模块使上下文信息最大限度地保留,将其聚合在特征之间.

通过对2条分支输出的全局信息编码进行二维全局平均池化,其输出被转换成对应的维度形状,即$\mathbf{R}_1^{(1\times C//G)}\times\mathbf{R}_2^{(C//G\times H\times W)}$.通过Softmax非线性函数完成线性变换的拟合,在2个分支规模一致的输出后进行连接使之转换成$\mathbf{R}^{(1\times H//W)}$的格式,再通过矩阵乘法(matmul)将连接后的结果相乘,得到空间注意力图和空间位置信息注意力图,2个生成空间注意力权重值的集合由每组输出的特征图进行计算.最终通过激活函数Sigmoid来捕获像素级成对关系,同时凸显全部像素的全局上下文.由于EMA最终输出的大小较输入的大小并未发生任何变化,因此将EMA添加到YOLOv8网络中非常方便.经过多次实验之后,本文选择在Backbone末端即SPPF模块后添加EMA注意力.

### 1.2.2 改进C2f模块

YOLOv8骨干网络使用了4个C2f模块以保证图像特征提取的优质性.为确保在网络高层可扩展感受野中进行特征提取,因此将C2f模块替换为提取特征更加高效的三支路DWR模块.DWR模块以残差方式设计,其原理如图2所示.图2中:RR和SR分别表示区域剩余化和语义剩余化;Conv表示卷积;DConv表示深度方向卷积;D-$n$表示扩张率为$n$的扩张卷积;圆圈中的+表示加法运算;$c$表示特征图通道的基数.

在残差内部,采用两步法有效提取多尺度上下文信息.首先,对任意输入特征进行残留特征的生成,再通过一个3×3卷积结合归一化(BN)层和激活函数(ReLU)层生成一系列不同大小的特征图,其中3×3卷积被用作初步特征提取;之后,将特征图分成若干组,对每组进行不同扩张率的扩张深度卷积操作,在捕获更多复杂语义的同时对特征图进行简单的形态滤波,以获得期望感受野,此过程将深度方向扩张卷积的作用简化为形态滤波使学习过程变得有序,从而更有效地保护多尺度上下文信息,也被称为语义剩余化;最后,将多尺度上下文信息通过BN层进行聚合,然后利用1×1逐点卷积合并特征,与输入特征形成残差结构以实现更全面的特征表示.

基于DWR模块对原始C2f模块进行替换,得到改进模块C2f_DWR,用此模块代替原有骨干网络中的最后2个C2f.

### 1.2.3 改进颈部

尽管YOLOv8有较强的多尺度性能,但对小物体检测精度仍然不高,且存在漏检情况,其主要原因是随着网络的加深,一些较为浅层的信息将不再被网络保留,而小目标信息主要存在于网络浅层中.为确保高层网络仍能实现浅层特征提取,从而提高大视场下极小苹果目标检测的准确率,仅仅依靠改进的C2f_DWR模块显然不够,因此可以通过蒸馏增强和融合多尺度特征的方法,将小目标检测性能上升到更高的水平.虽然YOLOv8将FPN和PAN出色地结合在一起,但不同尺度特征共享相同通道数,导致

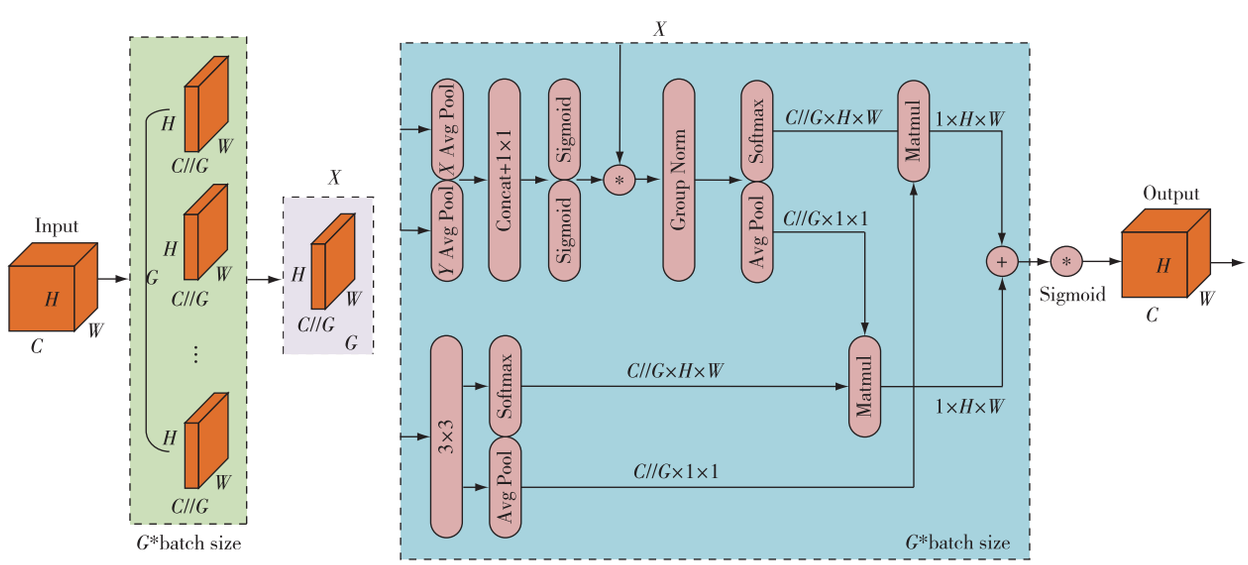


图1 EMA结构

Fig. 1 EMA structure

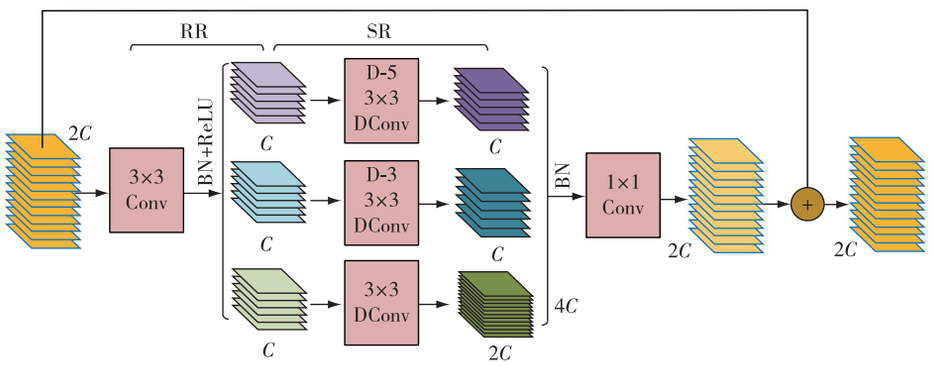


图 2　DWR 结构

Fig. 2　DWR structure

难以给出一个最优通道数来保证高层低分辨率特征和低层高分辨率特征具有同样丰富的表达能力,因此,本文在 YOLOv8 的基础上,结合 DAMO-YOLO 利用不同通道数表示不同尺度特征的思想,约束计算量使其更加轻量化,实现对低层特征与高层特征表达能力的灵活控制.利用 Rep-GFPN 对原始 YOLOv8 颈部网络进行重构,由此设计出一种新的具有跨尺度连接方式的特征融合网络 Rep-YOLOv8,在不增加太多计算成本的同时能够容纳更多的特征,从而实现高层语义和低层空间特征的高效融合.改进后的特征融合网络如图 3 所示.

图 3 中,红色圆圈、绿色圆圈以及橙色圆圈分别代表小型、中型和大型目标检测器.图 3a 中,FPN 结构是由上到下的特征金字塔,只对高层的语义特征向下传递,因此它对定位信息的传递效果尤为不佳;图 3b 中,为解决 FPN 的信息传递局限性,特在 FPN 之后添加了一个由下到上的特征金字塔,可实现底层定位特征的向上传递,可在具有特征信息的同时满足定位信息的保留;图 3c 中,YOLOv8 基于 PAN 的思想对网络结构进一步优化,具体为消除无特征融合的节点,使网络结构更为简便.但上述结构仍然未能很好地解决小目标的定位以及识别能力较差的问题,其原因在于特征提取过程中,小目标抗干扰能力不强,易受中大型目标的影响,常被误认为干扰信息而删除,因此,小目标信息会随着网络的加深而不断减少,最终导致网络对小目标检测效果大打折扣.

本文采用 Rep-YOLOv8 的颈部结构(图 3d)不仅实现了特征信息由下向上的传递,还实现了特征信息的由上到下的传递,这种双向传递代替了原来简单相加的特征融合方式,可在 3 层目标检测器中最大限度地保留更多浅层特征.Rep-YOLOv8 保留了 PAN 中的特征金字塔,并在节点同一层前后均添加 Fusion Block 模块,将原始基于卷积的特征融合改进为 CSPNet 连接,在加权特征融合的策略下,在提升模型精度的同时不增加较多计算量.

Fusion Block 模块结构如图 4 所示:将多个特征图通过不同的路径输入到 Fusion Block 模块中,通过采用多个简单的 1×1 和 3×3 卷积块有效地降低计算复杂性,捕获更为复杂的空间信息;利用 3×3 卷积重复块(Rep 3×3)和 3×3 卷积块重复多次(×$N$)深化

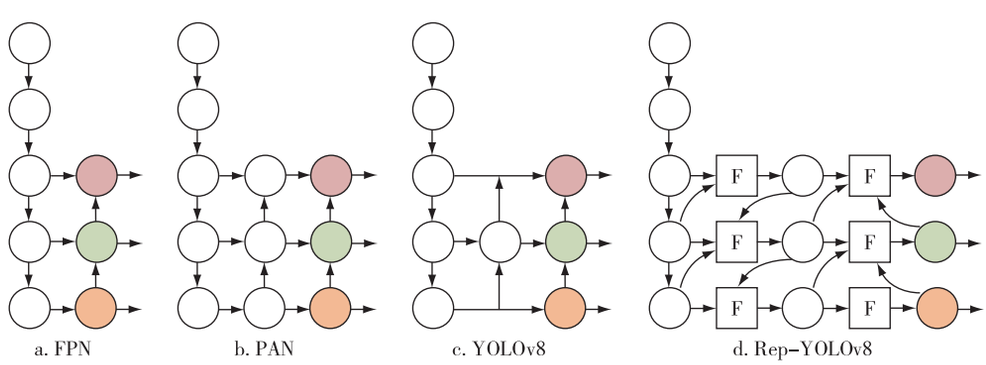


图 3　FPN、PAN、YOLOv8、Rep-YOLOv8 网络结构

Fig. 3　Network structures of FPN,PAN,YOLOv8,and Rep-YOLOv8

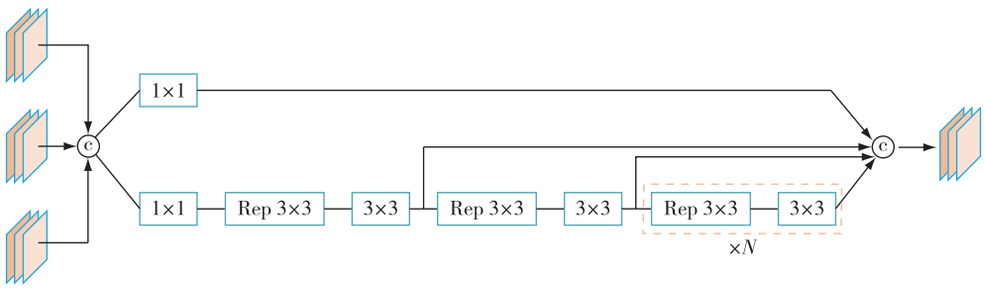


图 4　Fusion Block 结构

Fig. 4　Fusion Block structure

模型对特征的学习,从而增强特征的表达能力;最后一个融合操作将所有处理后的特征再次整合,输出最终特征图.

### 1.2.4　改进损失函数

由于 YOLOv8 采用的 CIoU Loss 完全交并比损失函数对边界框损失的拟合能力有一定程度的增强,对训练数据的质量要求较高,当部分训练数据存在质量问题或质量低于平均水平时,其泛化能力和收敛能力均会受到影响.为此,本文选择 Inner-SIoU Loss 代替 CIoU Loss.在 Inner-IoU 中加入了比例因子比率,它可以控制辅助边界框的比例大小.通过对不同的数据集和检测器使用不同尺度的辅助包围盒,可以克服现有方法泛化能力弱的局限性.

如图 5 所示,目标框(Target Box,TB)和锚框分别表示为 $b^{gt}$ 和 $b$,TB 框和内部 TB 框的中心点由 $(x_c^{gt},y_c^{gt})$ 表示,而 $(x_c,y_c)$ 表示锚框和内部锚框的中心点.TB 框的宽度和高度分别表示为 $w^{gt}$ 和 $h^{gt}$,而锚框的宽度和高度分别由 $w$ 和 $h$ 表示,对于比例因子 $k_{ratio}$ 通常在[0.5,1.5]之间,当 $k_{ratio}$ 小于 1 时,辅助边框尺寸小于实际边框,其回归的有效范围小于 IoU 损失,但其梯度绝对值大于 IoU 损失所得的梯度,能够加速高 IoU 样本的收敛.与之相反,当 $k_{ratio}$ 大于 1 时,较大尺度的辅助边框扩大了回归的有效范围,对于低 IoU 的回归有所增益.Inner-IoU 定义如下:

$$b_l^{gt}=x_c^{gt}-\frac{w^{gt}\times k_{ratio}}{2},b_r^{gt}=x_c^{gt}+\frac{w^{gt}\times k_{ratio}}{2},\quad(3)$$

$$b_t^{gt}=y_c^{gt}-\frac{h^{gt}\times k_{ratio}}{2},b_b^{gt}=y_c^{gt}+\frac{h^{gt}\times k_{ratio}}{2}.\quad(4)$$

式中:$b_l^{gt}$、$b_r^{gt}$、$b_t^{gt}$ 和 $b_b^{gt}$ 分别表示目标框的左边界、右边界、顶部边界和底部边界.

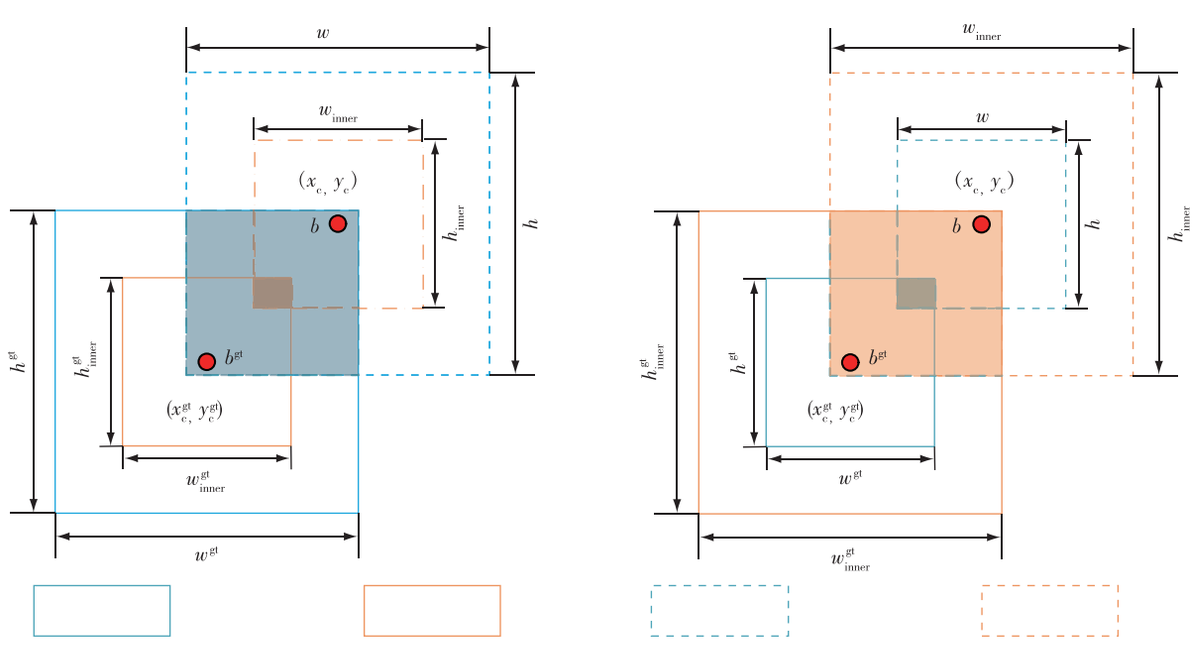


图 5　Inner-IOU 参数定义

Fig. 5　Inner-IoU parameter definitions

$$b_{\mathrm{l}}=x_{\mathrm{c}}-\frac{w\times k_{\mathrm{ratio}}}{2},b_{\mathrm{r}}=x_{\mathrm{c}}+\frac{w\times k_{\mathrm{ratio}}}{2},\tag{5}$$

$$b_{\mathrm{t}}=y_{\mathrm{c}}-\frac{h\times k_{\mathrm{ratio}}}{2},b_{\mathrm{b}}=y_{\mathrm{c}}+\frac{h\times k_{\mathrm{ratio}}}{2}.\tag{6}$$

式中:$b_{\mathrm{l}}$、$b_{\mathrm{r}}$、$b_{\mathrm{t}}$ 和 $b_{\mathrm{b}}$ 分别表示锚框的左边界、右边界、顶部边界和底部边界.

目标框与锚框的交集面积和并集面积分别用 $S_{\mathrm{inter}}$ 和 $S_{\mathrm{union}}$ 表示,其定义如下:

$$S_{\mathrm{inter}}=(\min(b_{\mathrm{r}}^{\mathrm{gt}},b_{\mathrm{r}})-\max(b_{\mathrm{l}}^{\mathrm{gt}},b_{l}))\times(\min(b_{\mathrm{b}}^{\mathrm{gt}},b_{\mathrm{b}})-\max(b_{\mathrm{t}}^{\mathrm{gt}},b_{\mathrm{t}})),\tag{7}$$

$$S_{\mathrm{union}}=(w^{\mathrm{gt}}\times h^{\mathrm{gt}})\times k_{\mathrm{ratio}}{}^{2}+(w\times h)\times k_{\mathrm{ratio}}{}^{2}-S_{\mathrm{inter}}.\tag{8}$$

最后利用式(9)求得 Inner-IoU 交并比:

$$\mathrm{IoU}^{\mathrm{inner}}=\frac{S_{\mathrm{inter}}}{S_{\mathrm{union}}}.\tag{9}$$

SIoU Loss 在 CIoU Loss 的基础上考虑了锚框和 GT 框之间的角度对包围盒回归的影响,并将角度损失引入包围盒回归损失函数.其定义如下:

$$L_{\mathrm{SIoU}}=1-\mathrm{IoU}+\frac{(\Delta+\Omega)}{2}.\tag{10}$$

IoU 用来预测边界框和真实边界框之间的交集与并集的比值来衡量两个边界框的重叠程度,取值范围为[0,1],其定义如下:

$$\mathrm{IoU}=\frac{b\cap b^{\mathrm{gt}}}{b\cup b^{\mathrm{gt}}}.\tag{11}$$

角度损失 $\Lambda$ 表示 TB 框和锚框的中心点连接之间的最小角度:

$$\Lambda=\sin\left(2\sin^{-1}\frac{\min(|x_{\mathrm{c}}^{\mathrm{gt}}-x_{\mathrm{c}}|,|y_{\mathrm{c}}^{\mathrm{gt}}-y_{\mathrm{c}}|)}{\sqrt{(x_{\mathrm{c}}^{\mathrm{gt}}-x_{\mathrm{c}})^{2}+(y_{\mathrm{c}}^{\mathrm{gt}}-y_{\mathrm{c}})^{2}+\epsilon}}\right).\tag{12}$$

式(12)旨在将锚框带到最近的坐标轴上,并根据角度的变化考虑优先接近 $X$ 轴还是 $Y$ 轴,当角度为 45°时,$\Lambda=1$;当中心点在 $X$ 轴或 $Y$ 轴上对齐时,$\Lambda=0$.

在考虑角度成本后,重新定义距离损失如下:

$$\Delta=\frac{1}{2}\sum_{t=w,h}(1-\mathrm{e}^{-\gamma\rho_t}),\gamma=2-\Lambda,\tag{13}$$

$$\begin{cases}\rho_x=\left(\dfrac{b_x-b_x^{\mathrm{gt}}}{w^{\mathrm{c}}}\right)^{2},\\ \rho_y=\left(\dfrac{b_y-b_y^{\mathrm{gt}}}{h^{\mathrm{c}}}\right)^{2}.\end{cases}\tag{14}$$

形状损失主要描述 TB 框和锚框之间的尺寸差异,定义如下:

$$\Omega=\frac{1}{2}\sum_{t=w,h}(1-\mathrm{e}^{\omega_t})^{\theta},\theta=4,\tag{15}$$

$$\begin{cases}\omega_w=\dfrac{|w-w_{\mathrm{gt}}|}{\max(w,w_{\mathrm{gt}})},\\ \omega_h=\dfrac{|h-h_{\mathrm{gt}}|}{\max(h,h_{\mathrm{gt}})}.\end{cases}\tag{16}$$

式中,$\theta$ 的值决定了形状代价的重要性,该参数的取值范围为 2~6,一般情况下其值为 4.

最后,将 Inner-IoU 损失应用于现有的 SIoU 边界框回归损失函数:

$$L_{\mathrm{Inner\text{-}SIoU}}=L_{\mathrm{SIoU}}+\mathrm{IoU}-\mathrm{IoU}^{\mathrm{inner}}.\tag{17}$$

综上,本文实现了对特征提取主干网络、颈部结构以及损失函数的改进,改进模型的网络结构如图 6 所示.

## 2 实验数据集

本文所用的苹果数据集除实地拍摄苹果图像以外,其余部分均选自昵图网(www.nipic.com)、汇图

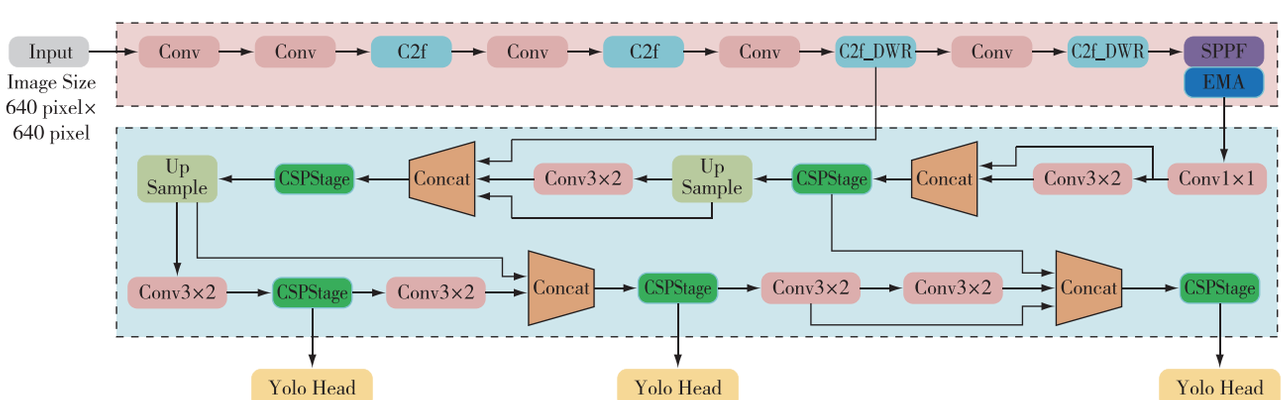


图 6 改进后 YOLOv8 网络结构

Fig. 6 Network structure of the improved YOLOv8

网(www.huitu.com)和红动中国图片网站(www.redocn.com).在获得大量图片数据后,筛选出具有果园真实环境下的3 036张图片以供实验,其中包括:多云和晴天等天气状况图像;顺光拍摄、侧光拍摄、逆光拍摄和夜间人工补光拍摄图像;剩余图像满足不同颜色、大小、光照条件、背景以及果实遮挡和枝叶遮挡等情况.利用LabelImg软件手动对目标果实所在区域进行矩形框的标注,标注完成后,将数据集按7∶2∶1的比例分为训练集、测试集和验证集.图7展示了数据集不同条件下的部分图像.

## 3 实验结果与分析

### 3.1 实验环境参数配置及模型测试指标

本文全部实验基于Ubuntu18.04操作系统,实验环境为Python3.9、CUDA11.0及Pytorch1.7.实验所用计算机CPU为Intel Xeon(R) E5-2650 v4@2.20 Hz×48,12 GB的NVIDIA GeForce RTX1080Ti×2,运行内存为64 GB.相关参数设置:训练轮数为200、批量大小为16、输入图片尺寸为640×640、随机种子为0且固定随机数、优化器SGD.通过梯度下降方法实现学习率的调整,初始学习率0.01,最小学习率0.000 1,学习率随训练轮数不断下降.损失函数比例因子 $k_{\text{ratio}}$ 设为1.4,该值为实验中得到的最优解.

为更加直观地展示本文对YOLOv8的改进效果,采用每秒检测帧数(FPS)、查准率($P$)、召回率($R$)以及平均精度均值(mAP)等评价指标评估模型性能.

$$P=\frac{\text{TP}}{\text{TP}+\text{FP}},\tag{18}$$

$$R=\frac{\text{TP}}{\text{TP}+\text{FN}}.\tag{19}$$

其中:TP表示预测框与标签框正确匹配的数量;FP表示实际为假,预测为真;FN表示实际为真,预测

a. 顺光　b. 侧光　c. 逆光
d. 夜间人工光照　e. 成熟套袋　f. 未成熟套袋
g. 未成熟重叠遮挡　h. 成熟重叠遮挡　i. 成熟大视场

图7 复杂场景下的苹果图像

Fig. 7 Apple images in complex scenarios

为假.

模型在每个类别上的性能用 AP 来衡量,mAP 为所有类别 AP 的平均值.mAP 为 IoU 取值 0.5 时的总类别平均精度,$mAP_{0.5\sim0.95}$表示 IoU 取值不同的平均精度.

$$mAP = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 P(R)\,dR. \tag{20}$$

为综合考虑查准率 $P$ 和召回率 $R$ 对检测算法的影响,引入 F1 评价指标,其定义如下:

$$F1 = \frac{2PR}{P+R}. \tag{21}$$

## 3.2 结果评估

### 3.2.1 重构 C2f_DWR 不同数量与位置实验结果与分析

YOLOv8 主干网络共有 4 个 C2f 模块.为更好地了解重构 C2f 模块对模型整体性能影响,进行简单的不同数量与位置实验:YOLOv8-$n$ 表示将第 $n$ 个 C2f 模块进行重构;YOLOv8-$n$+$m$ 表示对第 $n$ 个和第 $m$ 个 C2f 模块同时替换.实验结果如表 1 所示.

**表 1 添加不同数量 C2f_DWR 的实验结果**

Table 1 Experimental results of adding different amounts of C2f_DWR

| 模型 | P/% | R/% | F1/% | $mAP_{0.5}$/% | $mAP_{0.5\sim0.95}$/% | FPS/(帧/s) |
|---|---|---|---|---|---|---|
| YOLOv8-1 | 86.0 | 87.8 | 86.89 | 93.2 | 63.3 | 61.35 |
| YOLOv8-2 | 85.7 | 88.1 | 86.89 | 93.3 | 63.4 | 64.25 |
| YOLOv8-3 | 86.1 | 87.8 | 86.94 | 93.5 | 63.7 | 65.76 |
| YOLOv8-4 | 85.9 | 88.0 | 86.93 | 93.4 | 63.5 | 68.96 |
| YOLOv8-3+4 | 86.3 | 88.2 | 87.24 | 93.5 | 63.5 | 63.71 |

由表 1 可知:重构多个 C2f 模块会导致 FPS 下降,且重构 C2f 模块的位置越接近 SPPF 模块,性能提升越为明显.因此,本文算法为了在实现检测性能提升的同时不影响其检测速度,利用 C2f_DWR 对最后 2 个 C2f 模块进行替换,使其能够更好地平衡检测速度与精度.

### 3.2.2 消融实验结果与分析

为展示本文对苹果检测的各项改进效果,选取 YOLOv8n 为基准算法,并将本文提出改进措施在此基准上加以实现,进行消融实验,实验结果如表 2 所示.

表 2 中:A 表示在主干网络中引入 C2f_DWR 模块和 EMA 注意力机制;B 表示采用本文提出的 Rep-YOLOv8 结构替换原始 YOLOv8 颈部;C 表示采用 Inner-SIoU 作为边界框回归函数;D 表示在改进主干的基础上进行颈部重构;E 表示在改进颈部同时引入 Inner-SIoU 损失函数;F 表示在改进主干网络的同时引入 Inner-SIoU 损失函数;Ours 表示对 YOLOv8n 添加上述所有改进措施.由表 2 可知:A 使得大部分检测指标均有大幅提升;B 实现 $R$ 和 mAP 两方面的提升;C 与原算法相比,除 FPS 外其余指标全面提升;D 与原始算法相比,$mAP_{0.5}$和 $mAP_{0.5\sim0.95}$分别提升 0.9 和 0.8 个百分点;E 使 $mAP_{0.5}$和 $mAP_{0.5\sim0.95}$各提升 0.8 和 0.7 个百分点;F 使 $mAP_{0.5}$和 $mAP_{0.5\sim0.95}$分别提升 1 和 0.9 个百分点,实现除 FPS 外其余指标均有提升;Ours 在保证检测实时性的前提下,$P$、$R$、$mAP_{0.5}$、$mAP_{0.5\sim0.95}$ 和 F1 值分别为 86.1%、89.2%、94.0%、64.4%和 87.6%,大部分指标均优于原始模型,能够出色地满足实时检测任务.

### 3.2.3 不同算法对比实验结果与分析

为展示本文算法的改进效果,采用另外 5 种目标检测算法对本文制作的数据集进行训练,得到最优模型后在测试集进行测试,结果如表 3 所示.

**表 2 消融实验结果**

Table 2 Results from ablation experiment

| 模型 | 改进主干 | 改进颈部 | 改进损失 | P/% | R/% | F1/% | $mAP_{0.5}$/% | $mAP_{0.5\sim0.95}$/% | FPS/(帧/s) |
|---|---|---|---|---|---|---|---|---|---|
| YOLOv8n | | | | 86.5 | 87.2 | 86.8 | 92.7 | 63.1 | 70.63 |
| A | ✓ | | | 86.3 | 88.5 | 87.4 | 93.6 | 63.8 | 61.35 |
| B | | ✓ | | 85.5 | 87.7 | 86.6 | 93.1 | 63.5 | 61.73 |
| C | | | ✓ | 86.4 | 87.5 | 86.9 | 93.4 | 63.6 | 70.92 |
| D | ✓ | ✓ | | 86.4 | 87.7 | 87.0 | 93.6 | 63.9 | 53.76 |
| E | | ✓ | ✓ | 87.2 | 87.1 | 87.1 | 93.5 | 93.8 | 59.52 |
| F | ✓ | | ✓ | 87.2 | 87.5 | 87.3 | 93.7 | 64.0 | 61.35 |
| Ours | ✓ | ✓ | ✓ | 86.1 | 89.2 | 87.6 | 94.0 | 64.4 | 51.55 |

**表 3　对比实验结果**

Table 3　Results from comparative experiment

| 模型 | $P$/% | $R$/% | F1/% | $mAP_{0.5}$/% | FPS/(帧/s) |
|---|---|---|---|---|---|
| Faster-RCNN | 58.8 | 93.7 | 72.0 | 88.2 | 17.96 |
| SSD | 87.8 | 54.9 | 68.0 | 82.6 | 25.62 |
| YOLOv3 | 86.1 | 86.0 | 86.0 | 91.8 | 21.73 |
| YOLO5n | 85.2 | 88.3 | 86.7 | 92.4 | 65.77 |
| YOLOv7-tiny | 84.9 | 88.2 | 86.5 | 92.6 | 46.29 |
| 文献[17]模型 | 85.1 | 88.5 | 86.8 | 93.0 | 58.82 |
| 文献[18]模型 | 86.0 | 88.6 | 87.3 | 93.2 | 58.14 |
| YOLOv8n | 87.2 | 87.5 | 87.3 | 92.7 | 70.63 |
| Ours | 86.1 | 89.2 | 87.6 | 94.0 | 51.55 |

由表 3 可知:本文算法的 $R$ 值高于除 Faster-RCNN 以外的所有算法,但 Faster-RCNN 的 F1 值较低;本文算法在 F1 和 $mAP_{0.5}$两项指标优势明显;与文献[17-18]相比,除 FPS 指标外,本文算法也有比较优势;相较于原模型,本文算法在保证检测速度的同时仍实现了 $R$、F1 和 $mAP_{0.5}$指标的提升.结果表明,本文算法在进行不同的检测任务时能够保持良好的检测性能,极大地减少漏检、误检等情况的发生.

### 3.2.4　不同苹果数量的对比实验结果与分析

由于农场的苹果树木一般都是成行、成列有序种植,为模拟采摘机器人更加真实的工作环境,验证本文改进算法在不同果实数量场景的检测效果和鲁棒性,需要建立不同果实数量测试集.由于单个苹果检测不易突出本文改进算法与原始算法之间的差异,因此该测试集分为 3 类,分别为多个苹果场景(每张图片包含 2~10 个苹果)、密集苹果场景(每张图片包含 11~30 个苹果)、大视场苹果场景(远距离拍摄),并且将上述 3 类场景下的苹果按苹果成熟程度分为成熟苹果和未成熟苹果,由此得到 3 类不同数量场景下的成熟苹果和未成熟苹果.比较改进前后 2 种方法的不同检测效果,部分测试结果如图 8 所示.

将上述含有不同数目苹果图片按测试集进行测试后,得到表 4 各项性能指标.由表 4 可分析出,当检测多个苹果和密集苹果时,无论是成熟或未成熟苹果,2 种方法检测效果差异甚微,而随着果实数目不断增加,本文算法的优势才得以体现,特别是在成熟大视场情况下,本文算法的所有指标全部提升.在成熟大视场场景中,本文算法的 $P$、$R$、$mAP_{0.5}$ 和 $mAP_{0.5\sim0.95}$分别提升 3.5、8.4、1.0 和 2.8 个百分点,而在未成熟大视场场景中,由于果实颜色与叶片颜色相近,从而使检测任务变得困难,尽管本文方法在查准率方面略有不足,但其他指标与原始算法相比提升效果显著,尤其 $R$、$mAP_{0.5}$和 $mAP_{0.5\sim0.95}$分别提升 18.6、8.5 和 4.7 个百分点.综合结果表明,本文算法更适用于复杂环境下的苹果检测,尤其是针对数量庞大环境下的苹果检测,性能有显著提升,验证了其在不同苹果数目下的鲁棒性.

**表 4　YOLOv8 改进前后对不同数量苹果的识别结果**

Table 4　Recognition results of YOLOv8 before and after improvement for apples with varying fruit counts　%

| 苹果 | 检测算法 | $P$ | $R$ | $mAP_{0.5}$ | $mAP_{0.5\sim0.95}$ |
|---|---|---|---|---|---|
| 成熟多个 | YOLOv8 | 97.6 | 83.6 | 93.3 | 80.8 |
| | Improved YOLOv8 | 99.1 | 83.7 | 95.4 | 83.9 |
| 成熟密集 | YOLOv8 | 95.5 | 79.5 | 92.6 | 76.1 |
| | Improved YOLOv8 | 96.5 | 90.1 | 92.8 | 76.4 |
| 成熟大视场 | YOLOv8 | 82.0 | 73.7 | 86.3 | 50.4 |
| | Improved YOLOv8 | 85.5 | 82.1 | 87.3 | 53.2 |
| 未成熟多个 | YOLOv8 | 87.4 | 91.8 | 90.9 | 82.8 |
| | Improved YOLOv8 | 84.3 | 87.5 | 91.5 | 81.8 |
| 未成熟密集 | YOLOv8 | 85.3 | 76.7 | 86.2 | 60.9 |
| | Improved YOLOv8 | 87.3 | 78.0 | 88.3 | 63.3 |
| 未成熟大视场 | YOLOv8 | 89.1 | 54.7 | 68.5 | 43.2 |
| | Improved YOLOv8 | 85.7 | 73.3 | 77.0 | 47.9 |

a. 改进前　　b. 改进后

图 8　YOLOv8 改进前后对不同数量苹果的检测效果

Fig. 8　Detection performance of YOLOv8 before and after improvement for apples with varying fruit counts

3.2.5 公开数据集对比实验

为进一步验证本文所改进算法的鲁棒性和适用性,特在 kaggle 开源数据库中下载 apple-single-object-detection 公开数据集进行实验.该数据集包含 4 733 张不同数量、不同环境、不同成熟度等不同条件下的苹果果实,并将该数据集按照 8∶1∶1的比例进行训练集、测试集和验证集进行随机划分.实验参数与上述实验保持一致,实验结果如表 5 所示.表 5 结果表明,本文算法具有广泛适用性.

**表 5 公开数据集对比实验结果**

Table 5 Comparative experiment results on public datasets

%

| 方法 | $mAP_{0.5}$ | $mAP_{0.5\sim0.95}$ | $P$ | $R$ | F1 |
|---|---|---|---|---|---|
| YOLOv8 | 90.3 | 72.8 | 81.5 | 84.9 | 83.2 |
| Ours | 91.6 | 74.0 | 86.3 | 83.9 | 85.1 |

## 4 结论

本文针对传统目标检测算法在复杂环境下对苹果识别率低、鲁棒性较差等问题,提出一种基于 YOLOv8 的检测算法,该方法在测试集下的 $P$、$R$、$mAP_{0.5}$、$mAP_{0.5\sim0.95}$和 F1 值分别为 86.1%、89.2%、94.0%、64.4%和 87.6%,检测速度可达 51.55 帧/s,整体性能优于对比算法,能够在保证实时性的前提下仍保持识别苹果的准确性,有效地减少漏检、误检等情况的发生.

为进一步验证本文算法的鲁棒性,设计不同果实数目场景下的改进前后算法性能对比实验.结果表明:当检测多个苹果与密集苹果时,无论成熟或不成熟果实,其 $mAP_{0.5}$、$mAP_{0.5\sim0.95}$均有小幅度上升;在大视场环境中,本文算法的优势更加凸显,特别是在未成熟大视场环境下,其检测性能实现大幅度增强,表明本文算法的鲁棒性明显优于原始算法.在公开数据集上进行实验结果表明,本文算法具有广泛适应性.

## 参考文献

References

[1] Jia W K,Tian Y Y,Luo R,et al.Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot[J].Computers and Electronics in Agriculture,2020,172:105380

[2] 赵辉,乔艳军,王红君,等.基于改进 YOLOv3 的果园复杂环境下苹果果实识别[J].农业工程学报,2021,37(16):127-135
ZHAO Hui,QIAO Yanjun,WANG Hongjun,et al.Apple fruit recognition in complex orchard environment based on improved YOLOv3[J].Transactions of the Chinese Society of Agricultural Engineering,2021,37(16):127-135

[3] Wu G,Zhu Q B,Huang M,et al.Automatic recognition of juicy peaches on trees based on 3D contour features and colour data[J].Biosystems Engineering,2019,188:1-13

[4] Feng J,Zeng L H,He L.Apple fruit recognition algorithm based on multi-spectral dynamic image analysis[J].Sensors,2019,19(4):949

[5] Lv J D,Lv X J,Ma Z H.A fruit recognition method of green grape images in the orchard[J].New Zealand Journal of Crop and Horticultural Science,2022,50(1):1-16

[6] 初广丽,张伟,王延杰,等.基于机器视觉的水果采摘机器人目标识别方法[J].中国农机化学报,2018,39(2):83-88
CHU Guangli,ZHANG Wei,WANG Yanjie,et al.A method of fruit picking robot target identification based on machine vision[J].Journal of Chinese Agricultural Mechanization,2018,39(2):83-88

[7] 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 Dean,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

[8] 王金鹏,高凯,姜洪喆,等.基于改进的轻量化卷积神经网络火龙果检测方法[J].农业工程学报,2020,36(20):218-225
WANG Jinpeng,GAO Kai,JIANG Hongzhe,et al.Method for detecting dragon fruit based on improved lightweight convolutional neural network[J].Transactions of the Chinese Society of Agricultural Engineering,2020,36(20):218-225

[9] 易诗,李俊杰,张鹏,等.基于特征递归融合 YOLOv4 网络模型的春见柑橘检测与计数[J].农业工程学报,2021,37(18):161-169
YI Shi,LI Junjie,ZHANG Peng,et al.Detecting and counting of spring-see citrus using YOLOv4 network model and recursive fusion of features[J].Transactions of the Chinese Society of Agricultural Engineering,2021,37(18):161-169

[10] 龙燕,杨智优,何梦菲.基于改进 YOLOv7 的疏果期苹果目标检测方法[J].农业工程学报,2023,39(14):191-199
LONG Yan,YANG Zhiyou,HE Mengfei.Recognizing apple targets before thinning using improved YOLOv7[J].Transactions of the Chinese Society of Agricultural Engineering,2023,39(14):191-199

[11] Ouyang D L,He S,Zhang G Z,et al.Efficient multi-scale attention module with cross-spatial learning[J].arXiv e-Print,2023,arXiv:2305.13563

[12] Wei H R,Liu X,Xu S C,et al.DWRSeg:dilation-wise residual network for real-time semantic segmentation[J].arXiv e-Print,2022,arXiv:2212.01173

[13] Xu X Z,Jiang Y Q,Chen W H,et al.DAMO-YOLO:a report on real-time object detection design[J]. arXiv e-Print,2022,arXiv:2211.15444
[14] Zhang H,Xu C,Zhang S J.Inner-IoU:more effective intersection over union loss with auxiliary bounding box [J].arXiv e-Print,2023,arXiv:2311.02877
[15] Wang G,Chen Y F,An P,et al.UAV-YOLOv8:a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J].Sensors,2023,23(16):7190
[16] Yang H,Liu Y,Wang S,et al.Improved apple fruit target recognition method based on YOLOv7 model[J].Agriculture,2023,13(7):1278
[17] 庞超,王传安,苏煜,等.基于改进 YOLOv8 的水稻病害检测方法[J].内蒙古农业大学学报(自然科学版),2024,45(2):62-68
PANG Chao,WANG Chuanan,SU Yu,et al.Rice disease detection method based on improved YOLOv8[J].Journal of Inner Mongolia Agricultural University (Natural Science Edition),2024,45(2):62-68
[18] 李慧琴,宋赵铭,刘存祥,等.基于 YOLOv8n 的番茄果实检测模型改进[J/OL].河南农业大学学报:1-14[2024-05-16]. https://doi.org/10.16445/j.cnki.1000-2340.20240511.002
LI Huiqin,SONG Zhaoming,LIU Cunxiang,et al.Improvement of tomato fruit detection model based on YOLOv8n[J/OL].Journal of Henan Agricultural University:1-14[2024-05-16]. https://doi.org/10.16445/j.cnki.1000-2340.20240511.002

# Apple detection in complex orchard environments based on improved YOLOv8

YUE Youjun[1] QI Xiao[1] ZHAO Hui[2] WANG Hongjun[2]
1 School of Electrical Engineering and Automation,Tianjin University of Technology,Tianjin 300384,China
2 Tianjin Key Laboratory for Control Theory & Application in Complicated Industry Systems,Tianjin University of Technology,Tianjin 300384,China

**Abstract** To enable harvesting robots to quickly and accurately detect apples of varying maturity levels in complex orchard environments (including different lighting conditions,leaf occlusion,dense apple clusters,and ultra-long-range vision scenarios),we propose an apple detection model based on improved YOLOv8.First,the Efficient Multi-scale Attention (EMA) module is integrated into the YOLOv8 to enable the model to focus on the region of interest for fruit detection and suppress general feature information such as background and foliage occlusion,thus improving the detection accuracy of occluded fruits.Second,the original C2f module is replaced by a more efficient three-branch Dilation-Wise Residual (DWR) module for feature extraction,which enhances the detection capability for small objects through multi-scale feature fusion.Simultaneously,inspired by the DAMO-YOLO concept,the original YOLOv8 neck is reconstructed to achieve efficient fusion of high-level semantics and low-level spatial features.Finally,the model is optimized using the Inner-SIoU loss function to improve the recognition accuracy.In complex orchard environments with apples as the detection target,experimental results show that the proposed algorithm achieves $P_{\text{recision}}$, $R_{\text{ecall}}$, $\text{mAP}_{0.5}$, $\text{mAP}_{0.5\text{-}0.95}$, and F1 score of 86.1%, 89.2%, 94.0%, 64.4%, and 87.6%, respectively on the test set.The improved algorithm outperforms the original model in most indicators,and demonstrates excellent robustness through comparative experiments with varying fruit counts,offering practical value for applications in addressing the precise identification challenge faced by fruit harvesting robots in complex environments.

**Key words** pattern recognition;deep learning;object detection;YOLOv8