

基于EMA与Swin-Transformer的YOLOv8 苹果检测方法

刘子龙, 杨嘉炫

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 针对果园苹果分布形态复杂、密集且检测困难的问题, 提出基于EMA与Swin-Transformer的改进YOLOv8模型。首先, 通过可变核卷积替代YOLOv8骨干网络中的部分卷积, 实现灵活的特征采样; 其次, 引入Swin-Transformer模块, 增强全局与局部信息交互, 提升特征提取效率; 再次, 通过添加高效多尺度注意力机制, 获取不同通道维度的特征; 最终, 采用WIoU替代CIoU, 以增强模型泛化能力。实验采用两个真实果园数据集验证, 改进YOLOv8模型在数据集A上的精确率、召回率和mAP_0.5分别为91%、90.3%和94.8%, 较原模型提升了4.6、4.8和5.1个百分点, 单帧检测时间为8.2 ms; 在数据集B上, 三项指标分别为86.8%、79.3%和88.9%, 较原模型提升了5.8、4.1和8.5个百分点, 单帧检测时间为7.9 ms。改进算法在提升检测性能的同时, 保持了实时性, 利于移动端部署。

关键词: EMA; YOLO; 注意力机制; 小目标检测; 精准农业

DOI: 10.11907/rjtk.251235

扫描二维码阅读全文:



中图分类号: TP394.1

文献标识码: A

文章编号: 1672-7800(2026)006-0173-09

YOLOv8 Based on EMA and Swin-Transformer for Apple Detection

LIU Zilong, YANG Jiaxuan

(School of Optical-Electrical and Computer Engineering, University of Shanghai for science and technology, Shanghai 200093, China)

Abstract: To address the challenges of complex, dense apple distributions and detection difficulties in orchards, this study proposes an improved YOLOv8 model based on EMA and Swin-Transformer. The model replaces some of the convolutions in the YOLOv8 backbone network with alterable kernel convolution for flexible feature sampling, incorporates a Swin-Transformer module to enhance global-local information interaction and improve feature extraction efficiency, adds an efficient multi-scale attention mechanism to capture features across different channel dimensions, and replaces CIoU with WIoU to boost model generalization. Experiments were conducted on two real-world orchard datasets. On Dataset A, the improved model achieved precision, recall, and mAP_0.5 of 91%, 90.3%, and 94.8%, respectively, representing increases of 4.6, 4.8, and 5.1 percentage points over the original YOLOv8 model, with a single-frame detection time of 8.2 ms. On Dataset B, the three metrics are 86.8%, 79.3% and 88.9%, respectively, representing increases of 5.8, 4.1, and 8.5 percentage points over the original YOLOv8 model, with a single-frame detection time of 7.9 ms. The improved algorithm significantly enhances detection performance while maintaining real-time capability, facilitating its deployment on mobile devices.

Key Words: EMA; YOLO; attention mechanism; small target detection; precision agriculture

0 引言

我国是全球最大的苹果生产国, 其产量约占全球苹果总产量的50%, 并深受广大消费者的喜爱^[1]。随着人工智能时代的到来, 苹果种植逐渐向规模化、机械化和精准化

方向发展。通过采用机器人替代果农人工采摘, 可以大幅减少时间成本并提高采摘效率^[2]。然而, 苹果的自然生长环境较为复杂, 常出现果实间的重叠以及树枝和树叶等物体的遮挡, 这给采摘机器人的识别环节带来了较大挑战^[3]。

水果检测和识别研究最早通过分析水果的纹理、颜色和外观轮廓等特征进行判别。早期的传统方法包括色差

收稿日期: 2025-04-02

基金项目: 国家重大科研仪器研制项目(2020YFC2008704)

作者简介: 刘子龙(1972-), 男, 上海理工大学光电信息与计算机工程学院副教授、硕士生导师, 研究方向为机器学习、图像处理; 杨嘉炫(2000-), 男, 上海理工大学光电信息与计算机工程学院硕士研究生, 研究方向为深度学习、图像处理。本文通讯作者: 杨嘉炫。

法、模糊 C 均值方法、K 最近邻方法、支持向量机和 K-means 聚类方法^[4]。虽然这些方法能够识别出目标,但其稳定性均较差,容易受到光照和遮挡等因素的影响,导致采摘效率得不到保障^[5]。

近年来,深度学习已经在国内外的水果检测研究领域得到了广泛应用,国内外的研究人员也对此投入了大量精力^[6]。与传统方法相比,深度学习技术可以利用神经网络从训练数据中学习相应特征,从而实现对新数据的识别与检测,并获取其类别和坐标位置的信息^[7]。因此,基于深度学习的方法具有检测精度高、检测速度快以及抗干扰能力强的优点,适用于水果的检测与识别^[8]。Tian 等^[9]利用改进的 YOLOv3 算法,对不同生长阶段的苹果进行了识别。尽管优化了检测的准确性,但在识别的实时性上还有改进的空间。Kelman 等^[10]通过分析果实图像的多个强度分布实现了重叠苹果的定位,其准确率高达 93%,但仍存在模型检测过程耗时较长、检测效率不高的问题。Fan 等^[11]提出了一种基于优化 YOLOv4 算法的实时苹果缺陷检测方法。该方法采用基于 L1 范数的非极大值抑制法,在剪枝网络微调并去除冗余预测框后,使其模型的平均准确率提升至 93.74%,同时具有更好的时效性和更高的准确率。Yang 等^[12]设计了一种基于轻量化 YOLOv8 的西红柿检测方法,引入了深度卷积和双通道门注意力机制,最终使得 mAP_{0.5} 提升至 93.4%,但是其召回率的提升非常有限。

此前许多研究虽在理想化的实验室环境下取得了较高的准确率,但未能充分考虑现实果园的复杂环境因素,导致算法在实际应用中泛化性不足。在复杂的果园环境下,由于果实间的严重重叠和多种遮挡等问题,检测工作面临不小的挑战^[13]。如何克服果园环境的复杂性,在资源有限的移动设备上实现对苹果的精准检测,是实际应用中亟待解决的关键问题之一^[14]。针对上述问题,本文以 YOLOv8 为基础模型进行了一系列优化:①引入了 AKConv 卷积模块,代替 YOLOv8 骨干网络中的部分卷积模块;②使用了 Swin-Transformer 模块,促进全局与局部间的信息交流,以更高效地提取特征;③添加 EMA 注意力机制来获取不同通道维度的特征,同时,减小计算开销;④使用 Wise IoU 替换 CIoU,以提高模型的抗干扰能力。此外,本文使用的两个真实果园数据集包含了 5 000 张照片,涵盖多种干扰场景,能充分验证算法的优越性与泛化性能。

1 YOLOv8 算法与改进

1.1 YOLOv8 算法

YOLOv8 算法由 Ultralytics 公司于 2023 年发布,其网络结构如图 1 所示,主要由输入端、主干网络、颈部网络和检测端 4 个部分构成^[15]。与此前的 YOLO 系列相比(如 YOLOv5、YOLOv7),该算法最大的改进在于使用了中心检测

算法的无锚(Anchor-free)检测方式^[16]。这种检测方式可以直接检测目标的中心区域和边缘信息,且由于无需手动设计锚框,可以减少选择锚框数量和尺寸所带来的干扰,从而使得 YOLOv8 的检测精度在小目标检测方面得到提升^[17]。

本文对 YOLOv8 进行了如下优化:①主干部分使用 C2f 模块来替换此前 YOLO 系列中的 C3 模块,并将上采样前的单核卷积进行去除,从而使 C2f 模块在保证模型轻量化的同时,确保了梯度流信息的丰富性;②检测头部分使用 Decoupled-Head,它包含两个不同的分支,分别负责分类与回归任务,从而有效缓解了两类任务间的冲突^[18]。上述优化措施使得 YOLOv8 在检测精度方面表现出更高的优越性,因而许多研究人员更倾向于在其基础上进行改进以满足自己的需求。

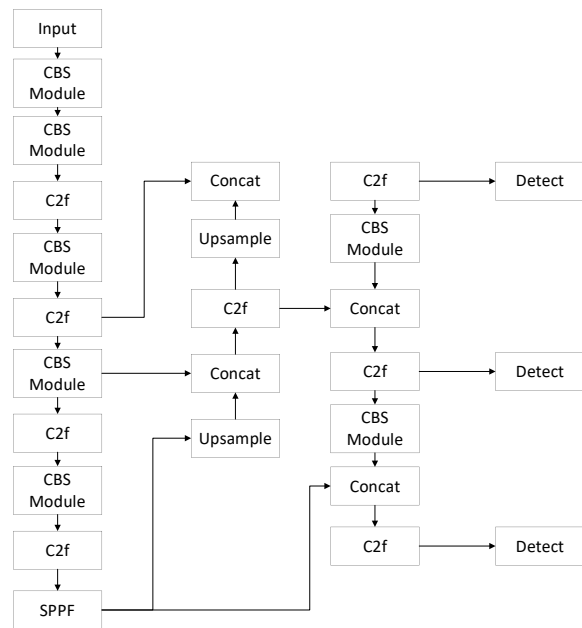


Fig. 1 The network structure of YOLOv8

图1 YOLOv8网络结构

1.2 YOLOv8 模型改进

1.2.1 EMA 注意力机制

EMA 注意力机制是一种高效的多尺度通道注意力机制,其网络结构如图 2 所示。它能优化特征提取过程,在保留不同通道特征信息的同时,有效减少计算开销。EMA 注意力机制首先将特征图 X 分为 G 组,利用 3 个子网络并行提取注意力权重,待结果拼接后通过跨通道信息交互进行建模;其次,将所得结果拆分为两个向量,运用 Sigmoid 函数提取注意力权重,并将这些注意力权重引入原始特征图中 3×3 分支的卷积操作,以捕捉部分跨通道信息,进而扩大特征空间^[19];再次,并行子网络负责提取不同尺度的特征信息,在 1×1 分支中提取注意力权重后,将其与 3×3 分支中的特征图进行融合,同时, 3×3 分支也采用相同操作提取注意力权重,并将其与 1×1 分支中经过 Group Normalization(组归一化)处理的特征图相结合;最

终,将融合后的特征进行相加,即可获得来自不同空间的多尺度信息。

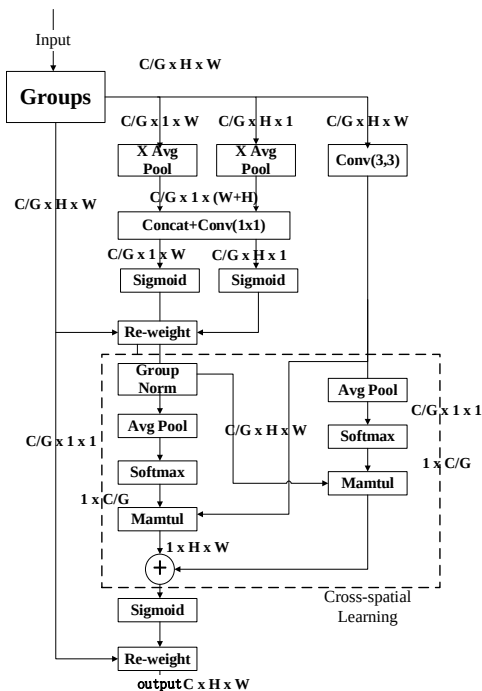


Fig. 2 EMA attention mechanisms

图2 EMA注意力机制

1.2.2 Swin-Transformer 模块

为了解决小目标间遮挡导致的识别困难以及目标与背景颜色相近引发的误检问题,本文引入了Swin-Transformer模块。通过捕捉目标不同尺度的特征,将YOLOv8中的C2f模块替换为Swin-Transformer Block,从而增强模型获取全局信息的能力,同时,提高YOLOv8的检测性能^[20]。该模块的结构如图3所示。

由图3可知,该模块首先对输入特征进行层归一化(LN)处理^[21];其次,通过基于窗口的多头自注意力(W_MSA),将残差连接融合不同层次的特征;再次,对融合后的特征进行LN处理;最后,通过多层感知机(MLP)进行非线性变换得到最终表示。移动窗口自注意力模块(SW_MSA)与W_MSA结构相似,均采用了滑动窗口的方式计算自注意力。这种方式可以更好的考虑到不同窗口之间的相互关系,从而更敏锐地捕捉到全局的信息。该模块中各组件的表达式如公式(1)一式(4)所示:

$$Y^l = W_MSA[LN(Y^{l-1})] + Y^{l-1} \quad (1)$$

$$Y^l = MLP[LN(Y^l)] + Y^l \quad (2)$$

$$Y^{l+1} = SW_MSA[LN(Y^l)] + Y^l \quad (3)$$

$$Y^{l+1} = MLP[LN(Y^{l+1})] + Y^{l+1} \quad (4)$$

其中, Y^{l-1} 代表输入, Y^l 表示第一个残差之后的输出, Y^l 表示第二个残差之后的输出, Y^{l+1} 表示第三个残差之后的输出, Y^{l+1} 表示最终的输出。

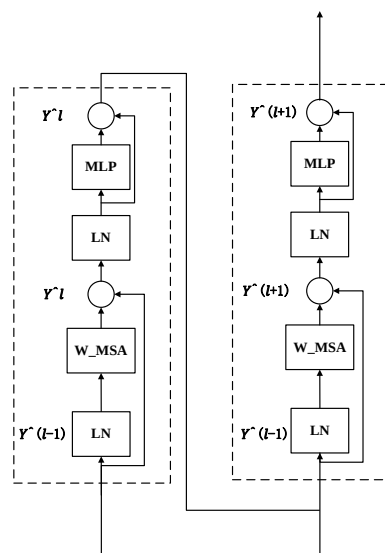


Fig. 3 Swin-Transformer block

图3 Swin-Transformer模块

1.2.3 AKConv 卷积模块

深度学习中的常规卷积存在着一些局限性,主要表现在卷积的操作局限于窗口,导致不能很好地获取全局信息,且通常采样模式是固定的。面对果园环境的复杂多变及果实形态的差异,固定不变的采样方式和正方形卷积核难以灵活应对这些变化带来的挑战。

针对上述情况,本文模型使用了可变核卷积(Alterable Kernel Convolution, AKConv)来代替传统卷积^[22]。AKConv旨在提供一种灵活的卷积机制。首先,通过输入的目标坐标确定卷积核初始采样位置;其次,通过2D卷积采样得到图像特征;最后,基于采样结果计算采样位置偏移量。与此同时,AKConv会根据偏移量调整采样位置,并对特征图进行重采样。重采样的结果经过重塑、再卷积和归一化后,最终通过SiLU激活函数得到输出结果。与初始采样形状相比,经过重采样后的每个位置的采样模式都会发生变化,这使得AKConv能够根据输入特征图的内容,动态调整其操作方式,从而为卷积网络提供了出色的灵活性与适应能力。通过引入可变核卷积,有助于模型处理多尺度的密集场景,提高对多尺度目标和小目标的检测性能,其结构如图4所示。

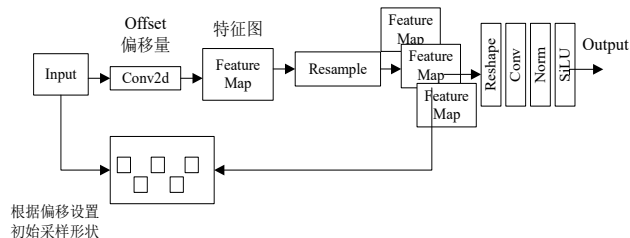


Fig. 4 The network structure of AKConv

图4 AKConv网络结构

1.2.4 Wise IoU 损失函数

在目标检测领域,模型通常使用预测框与真实框之间的交并比(Intersection over union, IoU)来衡量损失。该指标是目标检测损失函数的重要组成部分,一个优秀的损失函数能显著提升模型的检测性能。近年来,大多数研究都假设或者默认训练数据中的样本质量较高,并倾向于加强边界框损失的拟合能力^[23]。然而,实际情况往往并非如此。数据集制作中难以避免的低质量标注,会对模型的检测性能带来负面影响。为了解决这一问题,本文使用带有动态非单调聚焦机制的Wise IoU损失函数。

YOLOv8中使用的损失函数是CIoU,其表达式如公式(5)一式(6)所示:

$$L_{CIoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2} + \frac{4}{\pi^2} (\tan^{-1} \frac{w^{gt}}{h^{gt}} - \tan^{-1} \frac{h}{w}) \quad (5)$$

$$\nu = \frac{4}{\pi^2} (\tan^{-1} \frac{w^{gt}}{h^{gt}} - \tan^{-1} \frac{h}{w}) \quad (6)$$

其中, $p^2(b, b^{gt})$ 表示真实方框和预测方框中心点之间的欧氏距离, c^2 表示覆盖两个方框的最小封闭方框的对角线长度, w^{gt} 和 h^{gt} 分别表示真实方框的高度和宽度。CIoU虽然考虑了预测框和真实框之间的欧式距离,但忽略了高质量标注数据和低质量标注示例数据之间的比例关系。

然而,Wise IoU的核心思想在于结合注意力函数与非单调动态聚焦系数,旨在让模型更多地关注标注质量较差

的困难样本,进而提升模型的检测性能^[24]。Wise IoU的表达式如公式(7)一式(10)所示:

$$R_{WIoU} = e^{\frac{p^2(b, b^{gt})}{W_s^2 + H_s^2}} \quad (7)$$

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \quad (8)$$

$$\gamma = \frac{\beta}{\delta \theta^{\beta - \delta}} \quad (9)$$

$$L_{WIoU - \nu} = R_{WIoU} \times L_{IoU} \times \gamma \quad (10)$$

其中, R_{WIoU} 为注意力函数, β 表示离群度, L_{IoU}^* 指代IoU损失的梯度增益, L_{IoU} 表示归一化因子, θ 和 δ 是构造非单调系数的超参数, γ 为非单调动态聚焦系数。

1.2.5 改进后的YOLOv8网络结构

本文将优化模块引入YOLOv8模型,可以得到改进后的网络结构,如图5所示。

由图5可知,本文所提出的模型对骨干网络中的CBS卷积模块进行了替换,采用AKConv模块以提升模型对多尺度目标的适应性,进而增强对小目标的检测性能。同时,模型将原有的C2f模块替换为Swin-Transformer模块,此举有助于加强模型内部的全局信息交互,从而更有效地提取特征。在Neck部分,模型新增了EMA注意力机制,该机制能够在提取多尺度特征的同时,减少计算参数的数量。此外,模型还引入了Wise IoU损失函数以替代CIoU,这一改进显著增强了模型对不同样本的泛化能力。

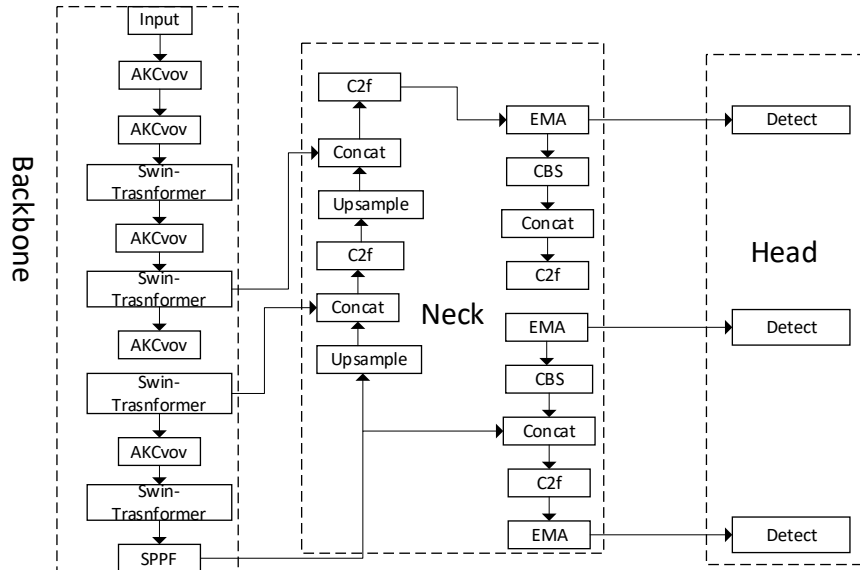


Fig. 5 The network structure of improved YOLOv8

图5 改进后YOLOv8网络结构

2 实验

2.1 实验配置与环境

本文模型的实验是在一台配备了Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz处理器和英伟达(NVIDIA) 3090-24G GPU的Linux服务器上进行的。实验采用了Py-

Torch 2.0作为深度学习框架,CUDA版本为11.8,并选择Adam作为优化器(训练持续时间为100个周期)。实验的参数设置如下:批量大小(batch size)为16,初始学习率为0.01,训练与验证阶段输入图像均采用固定的640×640像素分辨率。

所选实验平台与参数配置旨在充分发挥服务器CPU与GPU的计算性能,并借助PyTorch框架以实现高效的训

练与测试过程^[25]。

2.2 数据集

本文使用了两套数据集。其中,数据集A包含2 576张图片,数据集B包含2 538张图片。这两套数据集涵盖了苹果相互遮挡、枝叶遮挡果实以及单果与密集果实分布等多种场景,以此模拟真实的果园复杂环境,从而更有效地检验模型的检测性能。苹果图像数据集的标注采用LabelImg软件,并以txt文件形式保存,文件中包含目标的类别ID与坐标信息。

在实验设置中,数据集按照8:1:1的比例划分为训练集、验证集和测试集。其中,2 060幅图像被随机分配到训练集,剩余图像被平均分配到验证集和测试集中。部分数据集展示如图6和图7所示。



Fig. 6 Display of dataset A

图6 数据集A展示



Fig. 7 Display of dataset B

图7 数据集B展示

2.3 评价指标

为全面评估苹果检测模型的性能,本文采用了一系列评估指标,包括精确度、召回率、平均精确度(AP),以及IoU阈值为0.5时的mAP(mAP_0.5)和IoU阈值为0.5~0.95范围内的mAP(mAP_0.5:0.95)等^[26]。

在上述评估指标的公式中,包含了以下关键参数:TP(真阳性)是指预测为阳性样本且实际为阳性样本的数量;FP(假阳性)是指预测为阳性样本但实际为阴性样本的数量;FN(假阴性)是指预测为阴性样本但实际为阳性样本的数量;IoU表示预测的边界框与真实边界框之间的交集与并集的比率;Q代表数据集中的类别总数^[27-28]。这些指标的计算公式定义如公式(11)~式(15)所示:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{P} \quad (12)$$

$$mAP_{0.5} = \frac{\sum_{i=1}^Q AP_i(IoU \geq 0.5)}{Q} \quad (13)$$

$$mAP_{0.5:0.95} = \frac{\sum_{i=1}^Q AP_i(IoU \geq 0.5 \text{ to } 0.95)}{Q} \quad (14)$$

$$\sum_{i=1}^Q AP_i = \int_0^1 Precision(Recall) dRecall \quad (15)$$

2.4 实验结果与分析

2.4.1 改进前后的YOLOv8对比实验

将原始YOLOv8与本文改进的YOLOv8放置于相同的实验环境,设置如下:批量大小(batch-size)设置为16,学习率(learning rate)设置为0.01,模型的训练轮数(epoch)设置为100,并采用Adam优化器。两个数据集改进前后的性能指标对比如表1和表2所示。

表1和表2分别展示了在两个数据集上,改进前后YOLOv8的精度(P)、召回率(R)和mAP_0.5的具体数值。从表中数据可以看出,在数据集A中,本文模型精度提高了4.6个百分点,召回率R提高了4.8个百分点,mAP_0.5提高了5.1个百分点;在数据集B上,本文模型分别较原始YOLOv8在精度、召回率和mAP_0.5上分别提高了5.5、4.1和8.5个百分点。这些显著的性能提升进一步验证了本文改进算法对小目标检测的有效性。

Table 1 Performance comparison of YOLOv8 before and after improvements on dataset A

表1 数据集A上YOLOv8改进前后的性能指标对比

模型	P/%	R/%	mAP_0.5/%
YOLOv8s	86.4	85.5	89.7
本文模型	91.0	90.3	94.8

Table 2 Performance comparison of YOLOv8 before and after improvements on dataset B

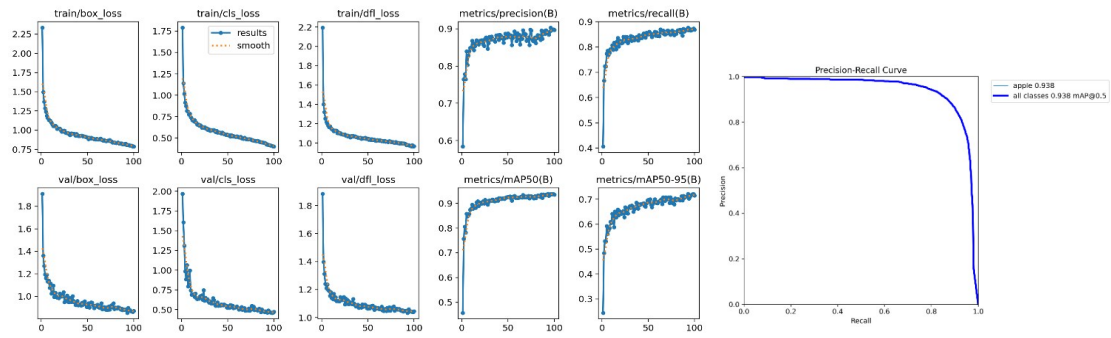
表2 数据集B上YOLOv8改进前后的性能指标对比

模型	P/%	R/%	mAP_0.5/%
YOLOv8s	81.0	75.2	80.4
本文模型	86.5	79.3	88.9

模型改进前后的曲线图如图8所示。在训练过程中,各项指标变化曲线包括训练分类损失(cls_loss)、训练目标框损失(box_loss)、召回率R(recall)、平均精度均值mAP_0.5(mean average precision)以及P-R曲线等。其中,图8(a)表示改进前各项指标的曲线,图8(b)表示改进后的指标曲线。

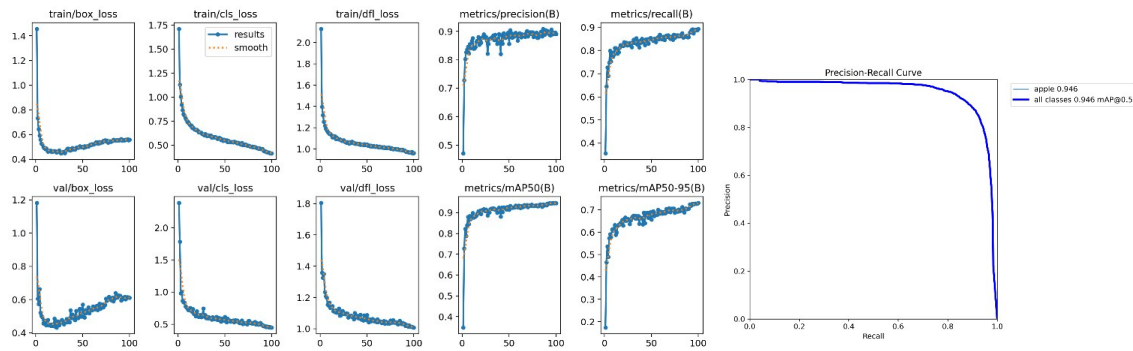
由图8可知,分类损失曲线基本呈现重合状态。由于本文仅针对单一类别目标(苹果)开展检测任务,分类损失曲线最终趋于0值。进一步分析图8中的两组回归损失曲线可知,改进后的模型在回归损失指标上表现出显著下降趋势,同时召回率与mAP_0.5指标均获得明显提升,其P-R曲线性能亦优于改进前模型。结合表1的量化数据对比可知,改进后的模型在各项性能指标上均展现出优于原始YOLOv8模型的检测效果。

YOLOv8模型改进前后的检测效果如图9所示。选取了数据集中的3组真实果园图像进行检测,如图9和图10所示,(a)与(b)分别表示原始算法和本文改进算法的检测结果。实验表明,原始YOLOv8模型在检测过程中存在明显的漏检和误检问题,而本文算法能够准确识别图中所有苹果。因此,可以看出本文算法在真实复杂果园背景下,



(a) Various performance curves before improvement

(a) 改进前的各项性能曲线图



(b) Various performance curves after improvement

(b) 改进后的各项性能曲线图

Fig. 8 Curves of model performance before and after improvement

图8 模型改进前后的曲线图



(a) Before improvement

(b) After improvement

(a) 改进前

(b) 改进后

Fig. 9 Detection effect comparison before and after improvement on dataset A

图9 数据集A改进前后的检测效果对比图



(a) Before improvement

(b) After improvement

(a) 改进前

(b) 改进后

Fig. 10 Detection effect comparison before and after improvement on dataset B

图10 数据集B改进前后的检测效果对比图

对于小目标和远距离遮挡苹果的检测效果显著提升,表现出更优异的综合性能。

2.4.2 消融实验结果和分析

为评估本文中每个改进点对原始 YOLOv8 模型的影响,在两个数据集上分别进行了消融实验。通过对各个改进模块分别进行验证和分析,结果如表 3 所示。

Table 3 Ablation experiments

表 3 消融实验

模型	数据集 A			数据集 B		
	P/%	R/%	mAP_0.5/%	P/%	R/%	mAP_0.5/%
Baseline	86.4	85.5	89.7	81.0	75.2	80.4
+EMA	87.9	85.9	93.1	83.2	75.3	85.2
+Swin-Transformer	87.4	87.5	92.2	82.0	74.6	85.5
+AKConv	88.3	88.5	91.5	84.9	78.3	83.9
+WIoU	86.6	85.2	91.1	80.4	77.6	85.8
Ours	91.0	90.3	94.8	86.8	79.3	88.9

由表 3 可知,在数据集 A 上,第一组 Baseline 模型是没用引入任何改进点的原始 YOLOv8 模型;第二组模型引入 EMA 模块,与 Baseline 模型相比,mAP_0.5 和 P 分别提高了 3.4 和 1.5 个百分点;第三组模型加入了 Swin-Transformer 模块,mAP_0.5 和 P 分别提高了 2.5 和 1 个百分点;第四组模型使用了 AKConv 卷积,使得 P 和 R 分别提高了 1.9 和 3 个百分点;第五组模型通过使用 WIoU 作为损失函数,虽然 R 小幅度下降了,但是 mAp_0.5 和 P 都有所提升;最后一组模型则是本文整体改进后的算法模型,三项指标(P、R 和 mAP_0.5)分别提高了 4.6、4.8 和 5.1 个百分点。

在数据集 B 上,第二组模型通过添加 EMA 注意力机制,三项指标都有小幅度的提升;第三组模型加入 Swin-Transformer 模块后,虽然召回率相比 Baseline 下降了 0.6 个百分点,但是精度和平均精度均有所提升;第四组模型引入 AK 卷积后,P、R 和 mAP_0.5 分别提升了 3.9、3.1 和 3.5 个百分点;第五组模型通过使用 WIoU 损失函数仅 P 有所下降,其余指标均为提升。通过引入上述 4 个改进点,本文改进模型的 3 项指标(P、R 和 mAP_0.5)分别提高了 5.8、4.1 和 8.5 个百分点。

实验表明,各改进点均能提升模型的检测性能。而当所有改进点被综合引入后,模型的整体检测性能达到了最优水平,这充分验证了各模块在两个数据集上的有效性。

2.4.3 与其他模型的对比实验结果分析

为了更全面地验证本文模型的性能优越性,本文将其与其他目标检测模型进行了对比实验,这些对比模型包括 SSD、Faster R-CNN、YOLOv6n、YOLOv8s、YOLOv10n、DETR,以及近两年针对 YOLOv8 的改进算法(SA-YOLOv8 和 YOLO-MS)。这些对比模型中既包含了经典的目标检测算法,也包含了当前主流的检测方法。实验结果如表 4 所示,其中各项指标的最高值已加粗。此外,图 11 以清晰的折线图形式,对比展示了不同模型的检测效果,以供进一

步分析。

Table 4 Results of comparison experiments

表 4 对比实验结果

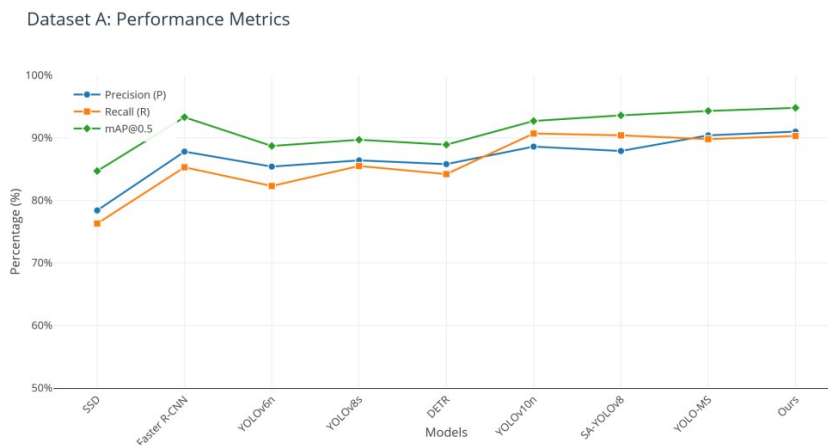
模型	数据集 A			数据集 B		
	P/%	R/%	mAP_0.5/%	P/%	R/%	mAP_0.5/%
SSD	78.4	76.3	84.7	70.4	69.8	73.2
Faster R-CNN	87.8	85.3	93.3	61.2	60.4	69.5
YOLOv6n	85.4	82.3	88.7	82.5	73.6	78.8
YOLOv8s	86.4	85.5	89.7	81.0	75.2	80.4
DETR	85.8	84.2	88.9	87.0	76.9	81.3
YOLOv10n	88.6	90.7	92.7	82.7	77.5	83.6
SA-YOLOv8	87.9	90.4	93.6	83.6	78.4	88.6
YOLO-MS	90.4	89.8	94.3	84.5	79.1	88.2
Ours	91.0	90.3	94.8	86.8	79.3	88.9

根据表 4 和图 11 在数据集 A 上的数据,对比分析可知:本文所提模型在多项指标上优势明显,其精确率 P(91.0%)与 mAP_0.5(94.8%)均为最高;仅有召回率一项(90.7%)YOLOv10n 模型取得最高值。在所有参与对比的模型中,推出时间最早的 SSD 模型实验结果最不理想。Faster R-CNN 模型虽然指标数据表现良好,但因其属于两阶段目标检测模型,在实时性方面存在明显短板。将对比赛场景转换至数据集 B,仅 DETR 模型在精确率 P 这一指标上略胜本文模型一筹,高出 0.2 个百分点,在其余指标方面,本文模型均展现出显著优势。综合两个数据集的实验结果,结论如下:本文改进后的 YOLOv8 模型在检测性能上表现最佳。

3 结语

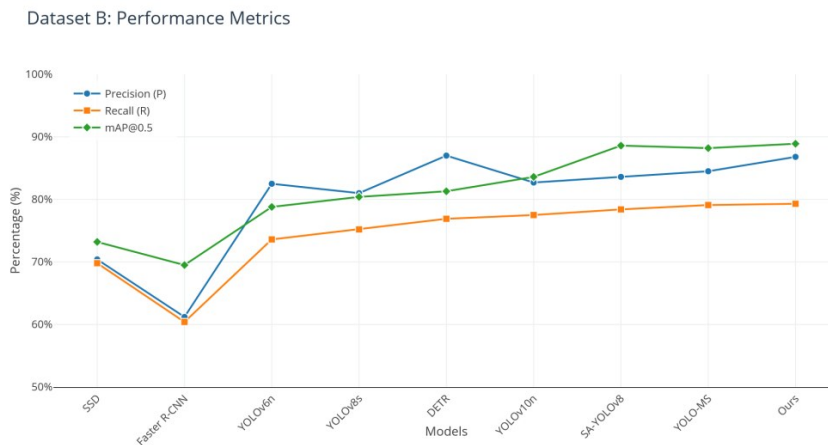
针对自然果园环境下苹果采摘过程中目标检测任务所面临的挑战,即遮挡物密集以及小目标检测精度不足的问题,本文提出了一种基于 YOLOv8 改进的模型架构。首先,通过集成 AKConv 卷积模块,对 YOLOv8 骨干网络中的部分标准卷积层进行替换,实现了更为灵活的特征采样与提取;其次,引入 Swin-Transformer 模块,以促进全局与局部特征之间的信息交互,从而提升特征提取的效率;再次,嵌入 EMA 注意力机制,旨在有效捕捉不同通道维度的特征信息,并降低计算复杂度;最终,采用 WIoU 损失函数替代原有的 CIoU,以增强模型的泛化性能。

实验结果表明,改进后的算法不仅显著提升了检测性能,还保持了优异的实时检测能力,为检测模型在移动设备上的部署提供了有力支持。在未来的研究中,模型还有进一步改进和优化的空间,例如,模型的鲁棒性和抗干扰能力可以进一步增强,以适应更多复杂的自然场景和检测样本。同时,还可以对模型进行更深入的轻量化设计,减少其计算参数,利于便携式设备在真实果园环境中的实际应用。



(a) Folded line graphs of comparison experiments on dataset A

(a) 数据集 A 的对比实验结果折线图



(b) Folded line graphs of comparison experiments on dataset B

(b) 数据集 B 的对比实验结果折线图

Fig. 11 Folded line graphs of comparison experiments**图 11 对比实验的折线对比图****参考文献:**

- [1] WANG N, WOLF J, ZHANG F S. Towards sustainable intensification of apple production in China—yield gaps and nutrient use efficiency in apple farming systems [J]. *Journal of Integrative Agriculture*, 2016, 15 (4): 710–729.
- [2] JI W, DING Y, XU B, et al. Adaptive variable parameter impedance control for apple harvesting robot compliant picking [J]. *Complexity*, 2020 (1): 1–15.
- [3] GONGAL A, AMATYA S, KARKEE M, et al. Sensors and systems for fruit detection and localization: a review [J]. *Computers and Electronics in Agriculture*, 2015, 116: 8–19.
- [4] LUO Z L, LI D, LIU X J. Lightweight detection algorithm for Tibetan medicinal plants based on improved YOLOv8 [J]. *Software Guide*, 2025, 24 (1): 86–92.
罗志凌, 李东, 刘晓静. 基于改进 YOLOv8 的轻量化藏药材植物检测算法 [J]. *软件导刊*, 2025, 24 (1): 86–92.
- [5] JIN Y, LIU J, XU Z, et al. Development status and trend of agricultural robot technology [J]. *International Journal of Agricultural and Biological Engineering*, 2021, 14 (4): 1–19.
- [6] GENÉ-MOLA J, VILAPLANA V, ROSELL-POLO J R, et al. Multi-modal deep learning for Fuji apple detection using RGBD cameras and their radiometric capabilities [J]. *Computers and Electronics in Agriculture*, 2019, 162: 689–698.
- [7] HAO Z X, WANG Q. Enhanced algorithm for small target detection in UAV aerial images based on YOLO-v7 [J]. *Software Guide*, 2024, 23 (1): 167–172.
郝紫霄, 王琦. 基于 YOLO-v7 的无人机航拍图像小目标检测改进算法 [J]. *软件导刊*, 2024, 23 (1): 167–172.
- [8] LI X, CHEN X, YANG J, et al. Transformer helps identify kiwifruit diseases in complex natural environments [J]. *Computers and Electronics in*

- Agriculture, 2022, 200: 107258.
- [9] 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.
- [10] KELMAN E, SMITH J, BROWN A. Apple detection in nighttime tree images using the geometry of light patches around highlights[J]. Computers and Electronics in Agriculture, 2015, 114: 154-162.
- [11] FAN S, LIANG X, HUANG W, et al. Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOv4 network[J]. Computers and Electronics in Agriculture, 2022, 193: 106715.
- [12] YANG G, WANG J, NIE Z, et al. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention[J]. Agronomy, 2023, 13(7): 1824-1837.
- [13] HAO Q, GUO X, YANG F, et al. Fast recognition method for multiple apple targets in complex occlusion environment based on improved YOLOv5[J]. Journal of Sensors, 2023(1): 1-12.
- [14] CHEN J, LIU H, ZHANG Y, et al. A multiscale lightweight and efficient model based on YOLOv7: applied to citrus orchard[J]. Plants, 2022, 11(23): 3260-3276.
- [15] HERMENS F. Automatic object detection for behavioural research using YOLOv8[J]. Behavior Research Methods, 2024, 56(12): 7307-7330.
- [16] ZHU W B, XIA L C, CHEN L, et al. Defect detection method of O-ring based on improved YOLOv5[J]. Journal of Shanghai University for Science and Technology, 2022, 44(5): 440-448.
朱文博,夏林聪,陈龙,等. 基于改进YOLOv5的O型密封圈缺陷检测方法[J]. 上海理工大学学报, 2022, 44(5): 440-448.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision (ECCV), 2018: 3-19.
- [18] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [19] WU Z, ZHANG Y, WANG X, et al. Algorithm for detecting surface defects in wind turbines based on a lightweight YOLO model[J]. Scientific Reports, 2023, 14: 24558.
- [20] FENG Z, WANG N, JIN Y, et al. Enhancing cotton whitefly (*Bemisia tabaci*) detection and counting with a cost-effective deep learning approach on the Raspberry Pi[J]. Plant Methods, 2024, 20: 161.
- [21] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [DB/OL]. <https://arxiv.org/abs/1704.04861>.
- [22] LIANG Z M, HUANG Y P, SONG Z H, et al. Survey on deep learning-based 3D object detection methods in autonomous driving[J]. Journal of Shanghai University for Science and Technology, 2024, 46(2): 103-119.
梁振明,黄影平,宋卓恒,等. 自动驾驶中基于深度学习的3D目标检测方法综述[J]. 上海理工大学学报, 2024, 46(2): 103-119.
- [23] TONG G, SHAO Y, PENG H. Learning local contextual features for 3D point clouds semantic segmentation by attentive kernel convolution[J]. Visual Computer, 2023, 39(6): 2623-2638.
- [24] YANG D, SOLIHIN M I, ARDIYANTO I, et al. A streamlined approach for intelligent ship object detection using EL-YOLO algorithm[J]. Scientific Reports, 2024, 14: 15254.
- [25] LU Y, YU J, ZHU X, et al. YOLOv8-Rice: a rice leaf disease detection model based on YOLOv8[J]. Paddy and Water Environment, 2024, 22: 695-710.
- [26] GAI R, CHEN N, YUAN H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model[J]. Neural Computing and Applications, 2023, 35(19): 13895-13906.
- [27] CHU P, LI Z, LAMMERS K, et al. Deep learning-based apple detection using a suppression mask R-CNN[J]. Pattern Recognition Letters, 2021, 147: 206-211.
- [28] KUMAR P S, SHWETAPADMA P, KUMARI S B, et al. On tree detection, counting & post-harvest grading of fruits based on image processing and machine learning approach—a review[J]. International Journal of Engineering and Technology, 2017, 9(2): 649-663.

(责任编辑:陈维捷)