

基于改进 YOLOv8 的苹果质量检测模型研究

王 亮¹ 肖小玲¹ 向琪琪¹

WANG Liang XIAO Xiaoling XIANG Qiqi

摘 要

为实现复杂场景下苹果质量检测的高准确性与自动化,包括自然生长环境、餐盘摆放、果篮盛载,以及混于其他果群等情况,文章提出一种基于改进 YOLOv8 的苹果质量检测模型。该模型引入三重注意力机制(triplet attention),通过交互计算减少空间和通道信息丢失,提升模型在识别和区分不同苹果质量方面的准确性;使用 DBB 重参数化模块替换原始的 C2f 模块,增强多尺度特征融合能力,解决损坏苹果表面特征差异大的问题。在苹果检测数据集上进行验证,实验结果表明,改进后的模型在苹果质量检测任务中检测精度为 71.3%,召回率为 77.5%,mAP0.5 为 79.7%,较 YOLOv8 原算法分别提高了 2.3%、1.4%、3.3%,提高了模型的检测效果。

关键词

苹果质量检测; YOLOv8; 三重注意力机制; 多尺度特征融合; 重参数化模块

doi: 10.3969/j.issn.1672-9528.2025.01.037

0 引言

在农业生产中,苹果作为极为常见且兼具高经济价值与食用价值的水果,在全球范围内广泛种植^[1]。其质量检测环节,于供应链管理及市场销售流程而言,有着举足轻重的意义。随着全球对食品安全和质量要求的提高^[2],传统人工检测方法已经无法满足高效、精准的需求。因此,研究在复杂环境下精准高效地检测苹果质量的模型具有重要意义。

早期苹果质量检测技术主要依赖人工检测和简单的图像处理技术。人工检测通常以视觉评估苹果的颜色、形状和表面缺陷来确定其质量,这一方法耗时耗力,效率低且容易受检验人员主观判断干扰,同时会因个人疲劳状态产生偏差;简单的图像处理技术,如颜色分析和形状分析^[3],通过测量苹果的颜色范围和形状来检测明显缺陷,但这些技术在处理复杂环境条件和细微瑕疵时往往不够准确。

随着科技的进步,早期监测方法逐渐被更先进的自动化和智能化检测技术所取代,这些新技术不仅提高了检测的准确性和效率,还能够处理更复杂的质量评估任务。如李林蔚^[4]在其研究中提出的 M2Det - ME 算法,有效平衡了检测准确率和检测速度,为苹果缺陷检测提供了更优的解决方案。武星等人^[5]通过改进 YOLOv3 网络结构,设计了轻量化的 Light - YOLOv3 网络,其在检测速度和准确率方面均有显著提高,为该领域提供了重要的参考。张嘉

琪^[6]在 YOLOv5s 基础上引入 BiFPN 和 NAM 注意力机制改进为 NAM - YOLO 网络以提高苹果缺陷检测性能,以及引入 CBAM 注意力机制、改进 ASFF 模块和增加苹果可测量区域检测网络改进为 MTL - YOLO 网络以提升苹果多特征检测能力。张震等人^[7]对 YOLO_v7 进行改进,引入 PConv 和 ECA 模块,并采用 SSA 算法优化学习率,提高模型的检测性能。

综上所述,苹果质量检测对于苹果产业的发展具有重要意义。然而,当前的检测方法在复杂环境下存在准确性不足和自动化程度低等问题。为了解决这些问题,本文提出了一种基于改进 YOLOv8 的苹果质量检测模型,通过引入三重注意力机制和 DBB 重参数化模块,旨在提高模型在复杂环境下对苹果质量检测的准确性和自动化水平。该研究有望为苹果产业的智能化发展提供有力支持,推动苹果质量检测技术的进步。

1 YOLOv8 算法

YOLOv8 是 YOLO 系列目标检测网络中的一个重要进展^[8],在汲取此前优势的基础上,进行改进和创新,YOLOv8 将模型分为 5 种不同的尺寸,分别为 n 、 s 、 m 、 l 和 x 。每种型号在网络深度、宽度和复杂度上逐步增加,从而在精度和速度之间提供了灵活的选择。

本文选择以 YOLOv8s 为基础,在此基础上进行改进。

YOLOv8 网络结构如图 1 所示,分为输入层、主干网络、颈部网络和检测头 4 部分。

1. 长江大学计算机科学学院 湖北荆州 434100

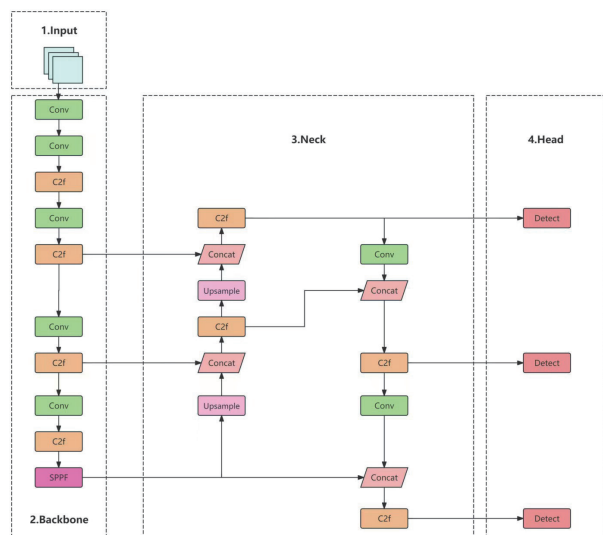


图1 YOLOv8 网络结构

2 改进 YOLOv8s 算法

2.1 改进 YOLOv8s 网络结构

为提高模型对苹果质量的检测精度, 本文以 YOLOv8s 为基线模型进行改进, 包括在主干网络与颈部网络间引入三重注意力机制和将主干网络与颈部网络的 C2f 模块替换为 DBB 模块。改进后的 YOLOv8s 网络结构如图 2 所示。

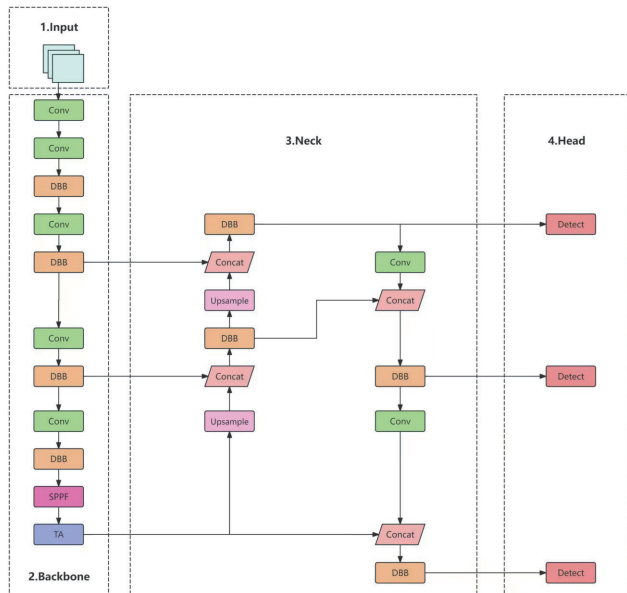


图2 改进后的 YOLOv8 网络结构

2.2 引入 Triple Attention 注意力机制

注意力机制能够建立通道或者空间位置之间的相互依赖关系, 使模型能够专注于重要的目标区域, 同时忽略不相关的背景信息。

传统的计算通道注意力的方法通过对输入张量在空间维

度上进行全局平均池化, 将其压缩为一个单一像素点来计算通道权值, 导致空间信息的大量丢失。Triple Attention 注意力机制是一种轻量化的三重注意力机制^[9], 通过交互计算减少了空间和通道信息丢失^[10]的同时几乎不增加模型计算量。其结构如图 3 所示。

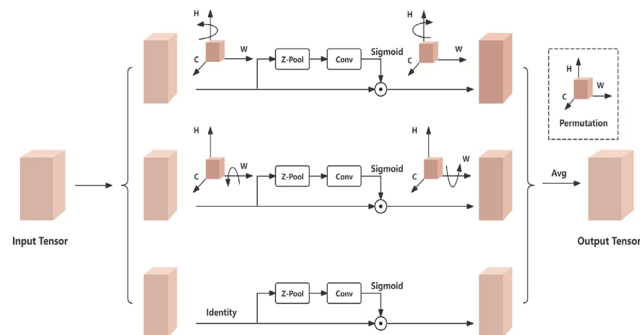


图3 Triple Attention 模块结构图

Triple Attention 注意力机制包含 3 个并行的分支结构, 每个分支负责捕获输入的空间维度和通道维度之间的交叉维度。

当张量 $X \in \mathbf{R}^{C \times H \times W}$ 进入第一个分支, 首先沿着 H 轴逆时针旋转 90° , 得到旋转张量 $\hat{x}_1$, 接着经过 Z-Pool 池化成形状为 $2 \times H \times C$ 的张量, 再进行卷积运算, 然后将张量传递到 Sigmoid 激活层 (σ) 生成注意力权值, 最后沿着 H 轴顺时针旋转 90° , 输出张量 $\hat{x}_1^*$, 建立通道 C 与维度 H 交互。

当张量 $X \in \mathbf{R}^{C \times H \times W}$ 进入第二个分支, 首先沿着 W 轴逆时针旋转 90° , 得到旋转张量 $\hat{x}_2$, 接着经过 Z-Pool 池化成形状为 $2 \times C \times W$ 的张量, 再进行卷积运算, 然后将张量传递到 Sigmoid 激活层 (σ) 生成注意力权值, 最后沿着 W 轴顺时针旋转 90° , 输出张量 $\hat{x}_2^*$, 建立通道 C 与维度 W 交互。

当张量 $X \in \mathbf{R}^{C \times H \times W}$ 进入第三个分支, 首先经过 Z-Pool 池化成形状为 $2 \times H \times W$ 的张量, 再进行卷积运算, 然后将张量传递到 Sigmoid 激活层 (σ) 生成注意力权值, 将权重应用于张量 X , 输出张量 $\hat{x}_3$, 建立通道 H 与维度 W 的交互。

最后将 3 个输出张量进行平均聚合, 生成最终的张量 $C \times W \times H$ 。最终张量的计算公式为:

$$y = \frac{1}{3} [\hat{x}_1 \sigma(\bar{\phi}_1(x_1^*)) + x_2 \sigma(\bar{\phi}_2(x_2^*)) + x \sigma(\phi_3(x_3))] \quad (1)$$

2.3 引入 DBB 重参数化模块

C2f 模块往往使用单一的卷积操作类型, 但输入的图像特征通常多样且复杂, 导致模块无法充分获取图像特征。而 DBB 模块^[11]在训练阶段, 采用多分支结构以丰富网络的特征表示能力, 这些分支包括平均池化、多尺度卷积等, 并采用更复杂的结构并行排列这些层; 在推理阶段, 多分支结构将会转换为单个卷积层, 如图 4 所示。

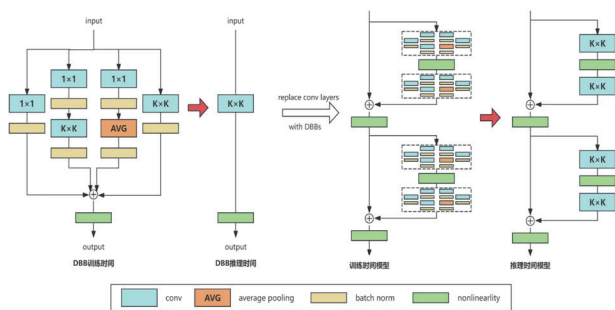


图 4 DBB 模块结构图

这一转换不仅提升了模型性能，还减少了模型在预测阶段的额外计算量，保证模型的预测速度。DBB 模块的转换共有 6 种方式^[12]，分别是：Conv-BN 合并、并行合并、串行合并、并行拼接、平均池化变换和多尺度卷积合并。如图 5 所示。

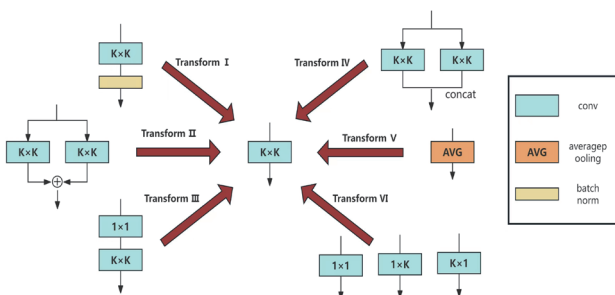


图 5 DBB 模块六种转换

3 实验及结果分析

3.1 实验环境及数据集

实验环境采用 Windows11 操作系统，中央处理器为 13th Gen Intel(R) Core(TM) i7-13620H，显卡为 Nvidia Geforce RTX 4050，显存 6 GB，内存大小为 16 GB 的电脑上进行，PyTorch 版本为 2.1.2，CUDA 版本为 12.3，Python 版本为 3.9.1。

训练图像输入采用 640 px×640 px 的统一大小，批次处理数为 4，迭代周期为 150，训练运行的设备为 GPU，以 Adamax 优化器作为调整模型参数的最小化损失函数，动量为 0.937，初始学习率设置为 0.01。

本实验数据图片来自 roboflow 网站，由 Arfiani Nur Sayidah 创建的 Apple Sorting Computer Vision Project。数据集共有 697 张图片，其中包含正常苹果图片和损坏苹果图片。

3.2 评价指标

本实验采用精确率 P 、召回率 R 、平均精度均值 mAP、每秒执行的浮点运算数 GFLOPs、参数量 Params 作为评价指标评估算法模型的性能，评价指标的计算公式分别为：

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 P(t) dt \quad (4)$$

$$mAP = \frac{\sum_{n=1}^N AP_n}{N} \quad (5)$$

式中：TP 表示分类为正的样本；FP 表示被分类为负的负样本；FN 表示被分类为负的正样本；AP 表示某一类别的平均准确率，等于以召回率 P 为横坐标、准确率 R 为纵坐标所获得 $P-R$ 曲线下的面积；mAP@0.5 表示 IoU 阈值取 0.5 时，所有类别 AP 平均值。

3.3 消融实验

本文基于 YOLOv8s 算法做出两点改进，分别是采用 TA 三重注意力机制和重参数化模块 DBB。为验证引入模块对算法模型性能的影响，本文通过逐步引入和组合不同模块设计了消融实验，以评估各模块在提升模型性能方面的效果。实验结果如表 1 所示。

表 1 消融实验结果

Model	$P/\%$	$R/\%$	mAP@0.5/%	Parameters	GFLOPs
YOLOv8s	0.69	0.761	0.764	11 126 358	28.4
YOLOv8s-TA	0.675	0.748	0.768	11 126 558	28.4
YOLOv8s-DBB	0.697	0.789	0.786	11 126 358	28.4
YOLOv8s-TA-DBB	0.713	0.775	0.797	11 126 558	28.4

由表 1 可知，本文的改进模型与原模型相比准确率、召回率、平均准确率分别提高了 2.3 个百分点、1.4 个百分点、3.3 个百分点，增加了极少的参数量和计算量。

当只增加 TA 注意力机制到主干网络与颈部网络间时，模型增加了极少的参数量和计算量，平均准确率也有所提升。

当只用 DBB 模块替换主干网络与颈部网络的 C2f 模块，模型召回率和平均准确率均有较大提升，并且参数量和计算量没有提升。

3.4 对比实验

为了客观评估改进模型的性能，本文将改进模型与 YOLOv3、YOLOv5s、YOLOv8s 以及 YOLOv9s 算法模型在苹果数据集上进行实验对比，实验结果如表 2 所示。

表 2 不同网络模型对比实验结果

Model	$P/\%$	$R/\%$	mAP@0.5/%	Parameters	GFLOPs
YOLOv3	0.684	0.717	0.75	103 665 782	282.2
YOLOv5s	0.682	0.715	0.745	9 112 310	23.8
YOLOv8s	0.69	0.761	0.764	11 126 358	28.4
YOLOv9s	0.729	0.715	0.759	7 167 862	26.7
YOLOv8s-TA-DBB	0.713	0.775	0.797	11 126 558	28.4

在精确率和召回率方面,改进模型的精确率为 71.3%,检测精度高于大多数模型;模型的召回率为 77.5%,目标捕获能力显著高于其他模型。YOLOv9s 虽然精确率略高于改进模型,但是其召回率明显偏低,模型在检测时容易漏检,从而降低了模型的整体性能。

在平均精度方面,改进模型的 mAP 为 79.7%,普遍高于其他模型,最小平均精度差距都为 3.3%,展现了改进模型在苹果质量检测中的鲁棒性和稳定性。

在参数量和计算量方面,改进模型与 YOLOv8s 模型几乎持平,相比 YOLOv5s 和 YOLOv9s 也仅有小幅提升,具有较低的计算成本和内存占比。

综上所述,改进模型在保持较低资源消耗的同时,取得了最高的平均检测精度,因此在苹果质量检测的应用中具有显著的优势。

为了更加直观地评价改进模型的性能,本文对苹果质量检测效果进行展示。如图 6 所示,改进模型准确分类出苹果质量,边界框精确定位苹果目标。通过可视化结果,可以清晰地看到改进模型在处理复杂场景检测时仍有较好的性能表现,表明了其在实际应用中的可靠性和实用性。



图 6 检测结果

4 结论

(1) 为实现复杂环境下苹果质量检测的准确性,本文提出了一种基于改进 YOLOv8 的检测算法,通过引入三重注意力机制 Triple Attention 增加通道间信息的关联性,然后引入 DBB 重参数化模块,有效提升了特征提取网络的性能。消融实验充分验证了改进策略在苹果质量检测的有效性。

(2) 改进模型与原模型相比,在几乎不增加参数量和计算量的前提下,平均检测精度提升 3.3%,具有良好的检测结果,且通过对比实验证明了改进模型综合性能优于 YOLOv3、YOLOv5s、YOLOv9s 等算法。可实现在复杂环境下实现苹果质量检测,为实现苹果质量智能分拣奠定了基础。

参考文献:

[1] 杨雪琴. 平凉市苹果全产业链增值路径分析[J]. 农业科技与信息, 2023 (4): 167-172.

- [2] 聂继云,李静,杨振锋,等. 苹果品质和质量安全问题与对策[J]. 中国果树,2007(3):60-62.
- [3] ZHANG H. Research on detection and classification of fruit based on computer vision technology[D]. Hohai University, 2011[2024-07-26]. <https://webofscience.clarivate.cn/wos/alldb/full-record/PQDT:67345789>.
- [4] 李林蔚. 基于深度学习的苹果缺陷检测技术研究[D]. 北京: 北京工业大学, 2020.
- [5] 武星,齐泽宇,王龙军,等. 基于轻量化 YOLOv3 卷积神经网络的苹果检测方法[J]. 农业机械学报, 2020, 51 (8): 17-25.
- [6] 张嘉琪. 基于深度学习的苹果品质检测与分拣分类研究[D]. 青岛: 青岛科技大学, 2023.
- [7] 张震,周俊,江自真,等. 基于改进 YOLO v7 轻量化模型的自然果园环境下苹果识别方法[J]. 农业机械学报, 2024, 55 (3): 231-242+262.
- [8] 徐彦威,李军,董元方,等. YOLO 系列目标检测算法综述[J]. 计算机科学与探索, 2024, 18 (9): 2221-2238.
- [9] ZHANG P, DENG H W, CHEN Z. RT-YOLO: a residual feature fusion triple attention network for aerial image target detection[J]. Computers, materials and continua,2023, 75 (1): 1411-1430.
- [10] 高瑞,熊彦平,魏辰峰,等. 基于三重注意力的轨道交通场景多任务感知算法[J]. 控制与信息技术, 2024(5):47-56.
- [11] 王海群,王炳楠,葛超. 重参数化 YOLOv8 路面病害检测算法[J]. 计算机工程与应用, 2024, 60 (5): 191-199.
- [12] 赵敏. 基于机器视觉的桥梁表面病害检测研究[D]. 太原: 太原科技大学, 2024.

【作者简介】

王亮(2002—),男,安徽六安人,硕士研究生,研究方向:目标检测与计算机视觉。

肖小玲(1973—),女,湖北荆州人,博士,教授、研究生导师,研究方向:智能信息处理与网络安全。

(收稿日期: 2024-10-13)