

基于改进 YOLOv8 的果园环境蟠桃果实识别方法

王 琛，马春悦，郭秀茹，王志军^{*}，孙 博，郭旭超

(1. 山东农业大学信息科学与工程学院，泰安 271018；2. 山东省苹果技术创新中心，泰安 271018)

摘 要：针对常见果园环境下复杂背景干扰、小目标检测困难以及果实密集分布导致蟠桃果实识别精度不佳的问题，该研究结合 MetaFormer 与卷积门控线性单元模块 (convolutional gated linear unit module, CGLU)，提出一种基于改进 YOLOv8 架构的蟠桃果实识别方法 CCGs-YOLO。首先，为解决复杂背景干扰下信息感知不足问题，引入融合卷积与注意力机制的 C2f_ConvFormer 模块和具备通道建模能力的 C2f_CaFormer 模块，以提升复杂背景下的识别性能。其次，为解决小目标检测困难问题，结合门控线性单元提升模型在复杂背景中的鲁棒性。另外，为优化果实密集分布下的定位精度，采用基于 Inner-IoU 的回归损失函数，以改善果实密集分布情景下的果实定位效果。最后，在自建数据集上进行对比试验。结果表明，CCGs-YOLO 相较于 YOLOv8 准确率 P 提高 3.30 个百分点，模型大小缩减至 4.88 MB，参数量降至 2.42 M。此外，CCGs-YOLO 模型在边缘设备上的 mAP50-95 指标提升至 88.82%。综上表明该模型在自然果园环境中具有良好的适应性和果实识别能力，可为蟠桃果实自动化采摘的快速识别提供方法支持。

关键词：蟠桃；YOLOv8；果实识别；目标检测；ConvFormer；CaFormer；CGLU

doi: 10.11975/j.issn.1002-6819.202509024

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2026)-12-0186-09

王琛, 马春悦, 郭秀茹, 等. 基于改进 YOLOv8 的果园环境蟠桃果实识别方法[J]. 农业工程学报, 2026, 42(12): 186-194. doi: 10.11975/j.issn.1002-6819.202509024 <http://www.tcsae.org>

WANG Chen, MA Chunyue, GUO Xiuru, et al. Identifying flat peach fruits in orchard environments using an improved YOLOv8 network model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2026, 42(12): 186-194. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202509024 <http://www.tcsae.org>

0 引 言

蟠桃作为常见的水果，凭借其独特的扁平外形、细腻口感以及丰富的营养价值在水果市场中广受欢迎。然而，随着种植规模的不断扩大与人工成本的持续上升^[1]，传统方法主要依赖人工方式进行果实识别与采摘，易受个人经验影响，在大规模种植条件下存在人工成本高、耗时费力等问题。

为此，研究者们尝试引入基于机器学习的图像识别方法，通过提取图像特征来识别果实区域，从而为实现自动采摘作业提供方法支撑。NGUYEN 等^[2]在遮光条件下使用 RGB-D 相机对果园苹果进行图像采集，同时提出一种基于颜色和形状特征的识别与定位算法，以实现苹果果实的检测。试验结果表明，对未遮挡的苹果识别率可达到 100.0%、部分遮挡苹果的正确识别率可达 82.0%。刘晓洋等^[3]针对自然果园中着色不均匀的果实，设计了基于超像素特征的苹果果实分割方法，分割准确率达 92.14%，具有较好的分割效果。尽管传统机器学习

方法在果实图像识别中取得了一定成效，但其依赖手工设计特征而难以提取高层语义信息，在复杂环境适应性和实时性等方面存在限制，因而难以满足果园环境下对果实进行高效稳定识别的实际需求。

近年来，卷积神经网络等深度学习模型因具备强大的图像特征学习与表达能力而被广泛运用于果实目标检测，在应对果园场景下复杂背景与多样目标，提升果实识别的精度与速度等方面优势明显。例如，为精准高效进行荔枝品种识别，黄铭杰等^[4]提出的 SCL-YOLO11 模型通过引入 C2f_SimAM 注意力模块与 CMUNeXt 大核深度可分离卷积，将识别精度提高至 99.61%，同时降低了网络规模。李茂等^[5]提出的 YOLOv8-PBi 模型在引入 PConv (partial convolution)、BiFPN (bidirectional feature pyramid network) 与 WIoU 机制后，将模型权重减少了 46.22%，检测帧率提升至 43 帧/s，有效增强了板栗果实自然环境下的识别能力与边缘部署性能，为智能化采收提供了技术支撑。YANG 等^[6]提出的 CA-YOLOv5 结合双向特征金字塔结构，在不同包装状态下的苹果检测中达到 89.8% 的 mAP。为提升对特殊形态果实的检测能力，GAI 等^[7]将 YOLOv4 中的矩形标注框替换为圆形，在樱桃检测中使 mAP 提升 15%；DU 等^[8]提出 DSW-YOLO 模型，引入可变形卷积，在草莓遮挡检测中实现 86.7% 的 mAP。孙俊等^[9]提出改进的 MRW-YOLOv5s 模型对葡萄进行目标识别，通过引入 MobileNetv3、坐标注意力模块、RepVGG Block 及 WIoU

收稿日期: 2025-09-02 修订日期: 2026-04-06

基金项目: 国家自然科学基金项目 (32401702); 山东省自然科学基金项目 (ZR2023MF098); 山东省自然科学基金青年项目 (ZR2023QF016); 山东省高等学校优秀青年创新团队 (2024KJH091)

作者简介: 王琛, 研究方向为人工智能与农业信息化应用。

Email: wangchensdau@163.com

※通信作者: 王志军, 博士, 教授, 研究方向为计算机应用、农业信息化。

Email: wzjsdau@163.com

损失函数，实现了仅 7.56 M 参数量、10.03 ms 检测时延和 97.74% 的 mAP，相较原 YOLOv5s 模型精度提升 2.32 个百分点，速度提升 6.13 ms。张志远等^[10]通过改进 YOLOv5 结构并引入 Transformer 与 BiFPN 模块，使樱桃果实识别精度提升 2.9 个百分点，并成功部署于 Android 平台。JIN 等^[11]提出 CO-YOLO 模型用于油茶果实姿态识别，引入 MMA 与 DPW 模块，在模型大小减少 77.1% 的同时，mAP@0.5 提升 4.8%，优于 YOLOv7-tiny、YOLOv9s 等主流模型。SUN 等^[12]提出基于 YOLO-FHLD 的枣果实成熟度检测方法，采用 C2f 模块和 HS-FPAN 特征融合，结合 Focal Loss 与知识蒸馏，该方法使 F1 值提升至 79.48%，mAP 提高至 85.40%，模型体积减至 3.51 MB。REN 等^[13]基于 YOLO 架构融合 GhostNet 轻量网络与 EMA（exponential moving average）注意力机制，构建 YOLO-GEW 模型并在梨果实检测上达到 88.83% 的精度。LIU 等^[14]在 YOLOv7 中引入坐标注意力模块与 EIou 损失函数，在小黄桃识别中 mAP 和召回率方面取得了良好的结果，分别达到 80.4% 和 73%，增强了复杂环境下的适应能力。在应对小目标和密集果实方面，LV 等^[15]通过增设小目标检测层与旋转框机制，实现幼桃姿态 93.04% 的识别准确率；FAN 等^[16]结合 YOLOv8 与加权 BiFPN 结构，在幼桃疏果检测中取得 98.4% 的 mAP。WU 等^[17]提出的 D-YOLO 模型采用 MobileNetv3 作为主干网络，通过融合 BiFPN 结构在草莓果实识别中取得了 89.6% 平均精度。

尽管上述主流目标检测方法在多种果实识别任务中取得了显著进展，然而对于自然果园环境下的蟠桃果实识别任务，现有深度网络模型在复杂背景干扰、小目标检测及果实密集下的检测等方面仍存在一定的局限，导致难以直接适配实际的蟠桃自动采摘任务。为此，本文提出一种基于改进 YOLOv8 网络模型的果园环境中蟠桃果实识别方法，以在复杂自然场景下实现更精确、鲁棒的果实检测。

1 材料与方法

1.1 数据集获取与处理

1.1.1 数据集建立

本研究所用图像数据于 2024 年 7 月至 9 月在山东省肥城市桃园镇某蟠桃种植基地采集，果实品种为瑞蟠 21 号。采集过程中，团队采用 SONY Alpha 7 IV 搭配 FE4/24-70G 镜头相机获取原始图像样本，共采集 1 305 幅单分辨率为 7 008×4 672 像素的高清图像。为增强数据的环境多样性，拍摄时间为 9 时、13 时及 17 时，涵盖自然光下的顺光、逆光与侧光等条件。如表 1 所示，结合果园技术人员意见后记录果实不同密度（稀疏、密集）及遮挡状态（枝叶覆盖、果实交叠）下的分布情况。蟠桃果实图像采集示例如图 1 所示。在数据预处理阶段，针对因枝干位移与手持抖动导致的图像模糊问题，开展人工质量筛选，最终保留 1 243 幅质量合格的 JPG 格式图像，作为基础图像。

表 1 数据集特征分布

Table 1 Distribution of dataset characteristics

类别 Category	划分标准 Definition	数量 Quantity/张	占比 Proportion/%
稀疏场景 Sparse scene	目标<4 个	131	10.54
中等密度场景 Moderately scene	4~10 个	964	77.53
密集场景 Dense scene	>10 个	148	11.90
轻度遮挡 Slight occlusion	IoU<0.25	834	67.08
中度遮挡 Moderately occlusion	IoU 0.25~0.5	377	30.31
重度遮挡 Severe occlusion	IoU >0.5	32	2.57

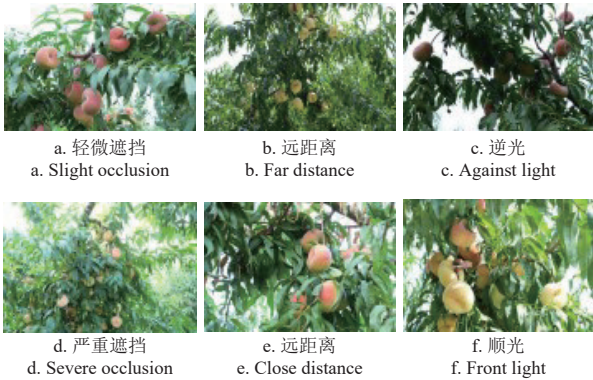


图 1 果园环境中蟠桃果实图像示例

Fig.1 Example image of flat peach fruits in an orchard environment

1.1.2 数据处理

为确保标注精度与效率，采用 Labellmg 工具对图像中的蟠桃果实进行人工矩形框标注，并生成包含目标中心点坐标与宽高尺寸信息的 TXT 格式标签文件，所有目标统一分类为 peach。首先，按照 7:2:1 比例划分为训练姐、验证集和测试机；然后，使用 Roboflow 平台对图像数据进行了增强处理，主要包括旋转、裁剪、亮度调整以及噪声扰动等操作。增强效果如图 2 所示，最终训练集图片 2 610 张，验证集图片 249 张，测试集图片 124 张，对应标签数量分别为 18 624、1 739、861 个。

1.2 基于 CCGs-YOLO 的网络模型改进

1.2.1 YOLOv8 目标检测模型

YOLO 系列模型在作物监测^[18]、病虫害检测^[19]、果实评估^[20]中广泛应用并取得显著成果。YOLOv8 采用创新的无锚框设计和任务对齐分配机制，使得阳性和阴性样本的动态分配更加高效，实现了速度与精度之间的最佳平衡^[21]。YOLOv8 由 4 个主要部分组成：输入层、主干网络、颈部和头部。输入层负责数据预处理和增强，主干网络进行特征提取，颈部部分融合不同尺度的特征，头部则负责预测目标标签的置信度。YOLOv8 提供多个版本（n、s、m、l、x），其中 YOLOv8n 具有较少的层数和较低的计算复杂度，模型结构最为轻量化，检测速度最快，在计算资源、实时性方面有着显著优势。近年 YOLOv9、YOLOv10 及 YOLOv11 等版本相继推出，但其结构复杂度相对较高，且在农业果实检测等实际应用

场景中的验证仍相对有限。综合考虑模型的成熟性、稳定性,本研究将 YOLOv8n 为基础模型。

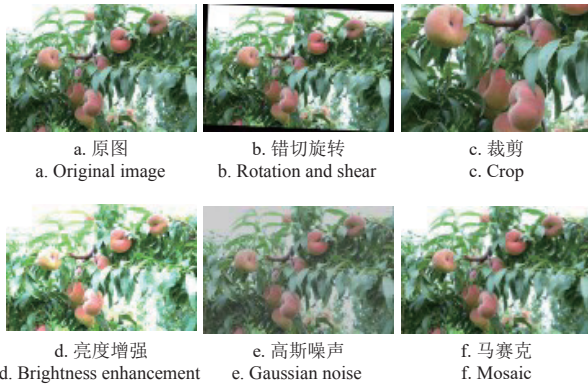
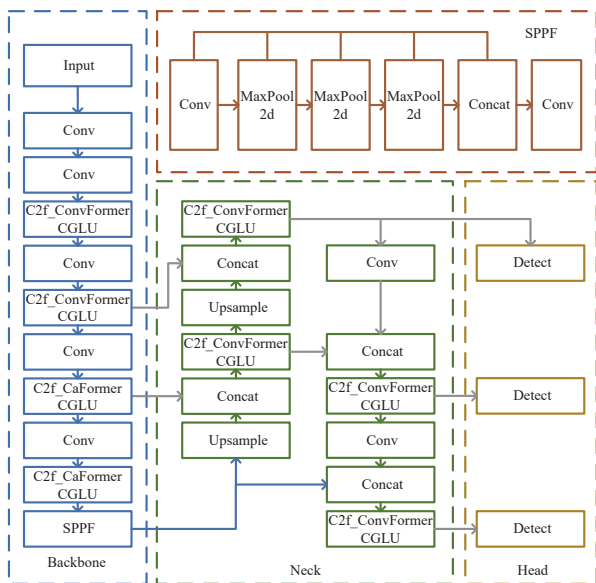


图 2 蟠桃果实图像数据增强效果示例

Fig.2 Example of image data enhancement effect on flat peach fruits

1.2.2 CCGs-YOLO 目标检测模型

YOLOv8 主干网络中采用 C2f 模块进行特征提取,能够在大部分情况下有效提取特征,但在处理如蟠桃果实遮挡、重叠及小目标检测等复杂情况时,存在一定的局限性。CCGs-YOLO 是基于 YOLOv8 架构的改进,其网络结构如图 3 所示。



注: Backbone 表示主干网络; Neck 表示颈部网络; Head 表示头部网络; SPPF 为改进的空间金字塔池化模块。
Note: Backbone denotes the backbone network; Neck denotes the neck network; Head denotes the head network; SPPF refers to the improved Spatial Pyramid Pooling Fast module.

图 3 改进后的 CCGs-YOLO 网络结构示意图

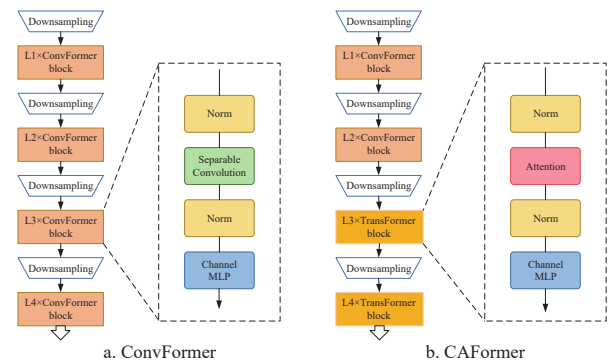
Fig.3 Schematic diagram of the improved CCGs-YOLO network structure

与原始 YOLOv8 相比,CCGs-YOLO 在保持整体架构一致的前提下,对主干网络与检测头部分的关键模块进行了优化,以增强特征表达能力和检测精度。CCGs-YOLO 采用更具建模能力的 C2f_ConvFormer 模块和 C2f_CaFormer 模块进行替换^[22]。在此基础上,CCGs-YOLO 进一步引入具备局部感知能力的门控机制,即在上述模块中融合了 CGLU 模块^[23],构成 C2f_ConvFormer-

CGLU 与 C2f_CaFormerCGLU 模块,以提升特征通道间的动态建模能力。CGLU 能够更精准地聚焦于小目标本身及其关键上下文区域,有效抑制冗余信息,在不显著增加计算复杂度的前提下增强了模型对小目标的表达能力和复杂背景下的鲁棒性。尽管检测头结构与 YOLOv8 保持一致,均采用三层特征输出,但由于前期特征提取与融合模块的优化,CCGs-YOLO 所传递的特征语义更为丰富,整体检测性能得到显著提升,特别是在面向小目标、密集分布和复杂背景的蟠桃果实识别任务中展现出更强的适应能力。

1) 融合 ConvFormer 与 CAFormer 的 MetaFormer 模型

传统视觉 Transformer 虽然能处理长距离依赖和高维特征,但计算复杂度较高。为了解决这一问题,本文基于 MetaFormer 框架,并结合 ConvFormer 与 CAFormer (channel attention transformer),以实现特征建模能力和推理效率提升,ConvFormer、CAFormer 架构如图 4 所示。



注: Downsampling 为下采样; ConvFormer block 与 Transformer Block 为基于 MetaFormer 框架的核心计算单元; Norm 为归一化处理; Separable Convolution 为可分离卷积; Attention 为建模长距离依赖; Channel MLP 为通道多层感知机。

Note: Downsampling refers to downsampling; ConvFormer block and Transformer Block are core computational units based on the MetaFormer framework; Norm denotes normalization; Separable Convolution refers to depthwise separable convolution; Attention is used to model long-range dependencies; Channel MLP denotes the channel-wise multilayer perceptron.

图 4 ConvFormer 和 CAFormer 结构示意图

Fig.4 Schematic diagram of ConvFormer and CAFormer structures

MetaFormer 是一种通用的视觉模型架构,旨在将视觉 Transformer 抽象为“令牌混合器+通道 MLP”模块的组合,且不限令牌混合器的类型。ConvFormer 作为全卷积方案,在各阶段使用深度可分离卷积作为令牌混合器降低计算复杂度。令牌混合器公式如式(1)所示。

$$\text{Convolutions}(X) = \text{Conv}_{\text{pw}2}(\text{Conv}_{\text{dw}}(\sigma(\text{Conv}_{\text{pw}1}(X)))) \quad (1)$$

式中 X 为输入特征图, Conv_{pw} 为 Pointwise 卷积, Conv_{dw} 为 Depthwise 卷积。

ConvFormer 在推理效率方面表现优秀,但由于缺乏全局建模能力,其在复杂任务中的表现仍存在一定局限。为进一步提升特征建模能力,本研究引入 CAFormer 作为 YOLOv8 主干网络的替代方案,浅层与中层采用融合卷积与注意力机制的 C2f_ConvFormer 模块,用于提升模

型对局部与上下文信息的感知能力；在深层则引入具备通道建模能力的 C2f_CaFormer 模块，进一步增强对高语义特征的理解能力。CAFormer 在前两个阶段采用深度可分离卷积以保持计算效率，在后两个阶段引入标准自注意力机制，以增强模型对长距离依赖的建模能力。

2) 融合轻量深度可分离卷积的门控线性单元

传统的门控线性单元 (gated linear unit module, GLU) 作为一种高效的通道混合器，通过动态门控方式实现信息流控制，有助于增强网络对关键特征的建模能力并抑制冗余信息。门控机制通过两个线性投影的逐元素相乘操作，生成介于 0 和 1 之间的门控信号，从而选择性地保留或抑制另一分支的信息，实现更灵活的特征调控。然而，传统 GLU 结构未能充分利用局部空间信息，限制了其在复杂任务中的感知能力。如图 5 所示，为了克服传统 GLU 的局限性，CGLU 特征增强模块在门控分支中引入了轻量级的 3×3 深度可分离卷积，使得门控信号不仅具备动态调控特性，还能融合局部上下文信息，有效缓解了全局平均池化粒度过粗、缺乏位置信息的问题。

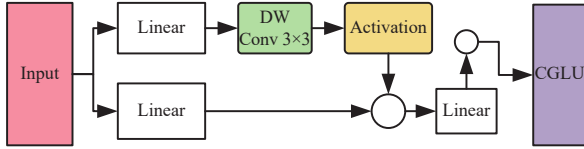
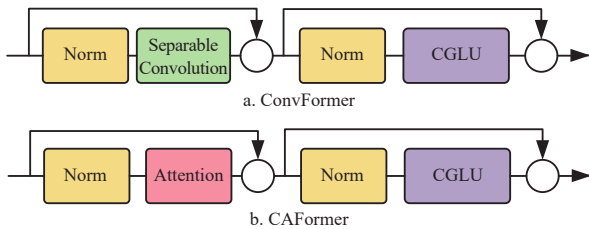


图 5 CGLU 模块结构示意图

Fig.5 Schematic diagram of the CGLU

如图 6 所示，本文在原始主干网络结构中，将 Transformer Block 和 ConvFormer Block 中的 Channel MLP 模块替换为 CGLU 模块。该模块由两个并行的线性分支组成，一支作为主信息通路，另一支经过深度可分离卷积与激活函数处理后作为门控信号，二者在通道维度上进行逐元素相乘，从而实现了对通道特征的动态调控。该结构在保持较低计算复杂度的同时，有效提升了特征表达能力与局部结构建模能力。在小目标检测方面，CGLU 通过引入具有局部感知能力的门控机制，有效提升了模型对微小目标的特征感知与表达能力。相比传统的通道混合模块，CGLU 增强了模型在复杂背景下的小目标检测鲁棒性。



注：Norm 为归一化处理；Separable Convolution 为可分离卷积；Attention 为建模长距离依赖；CGLU 为结合卷积与门控机制结构。
Note: Norm denotes normalization; Separable Convolution refers to depthwise separable convolution; Attention is used to model long-range dependencies; CGLU refers to a structure that combines convolution and gating mechanisms.

图 6 改进后的 ConvFormer 和 CAFormer 结构示意图

Fig.6 Schematic diagram of the improved ConvFormer and CAFormer structures

3) 引入 Inner-IoU 的损失函数

在目标检测任务中，交并比 (intersection over union,

IoU) 是衡量预测框与真实框重叠程度的常用指标，如式

(2) 所示。然而，传统的 IoU 在高重叠区域下可能难以有效地区分框与框之间的微小差异，特别是在目标边界接近时，IoU 可能无法充分反映框内细节的差异。为提升定位精度，YOLOv8 默认引入了包含中心点距离与长宽比一致性的 CIoU 损失函数。然而，在蟠桃复杂果园环境中，受枝叶遮挡、果实重叠等因素影响，预测框与真实框往往难以完全重合，CIoU 在非重叠阶段梯度较小，优化效果有限。

对此，本文在 CIoU 基础上引入 Inner-IoU 作为辅助监督，改进边框回归损失函数，从而增强模型对非完全重叠目标的判别能力^[24]，Inner-IoU 通过度量预测框与真实框内部区域的重合程度，能够在遮挡严重或重叠复杂的情况下提供更稳定、细致的梯度信息，有助于提升模型在边缘模糊、边界接近等场景下的定位精度^[25]。Inner-CIoU 引入尺度因子比 ratio 来控制计算，损失的辅助边界框的比例大小，ratio 的取值范围为 [0.5, 1.5]，本文设置为 0.7，计算式如下。

$$L_{\text{Inner-CIoU}} = L_{\text{CIoU}} + (\text{IoU} - \text{IoU}^{\text{Inner}}) \quad (2)$$

式中 L_{CIoU} 为 CIoU 损失，IoU 为实际交并比， $\text{IoU}^{\text{Inner}}$ 为通过缩放系数获得的边界框与目标边界框的交并比。

1.3 试验平台及评价指标

1.3.1 试验平台及参数设置

本文中试验所用操作系统为 Windows10，CPU 型号为 Intel i9-12900KF，GPU 型号为 NVIDIA GeForce RTX 4090(24 GB)，深度学习框架为 Pytorch2.1.0，编程语言为 Python3.11，GPU 加速库为 CUDA12.6。在参数设置方面，Batch size 为 32，输入图像分辨率为 640×640 像素，Epoch 为 200，Optimizer 为 AdamW，Lr 为 0.002，Momentum 为 0.9，Decay 为 0.000 5，权重初始化方法使用预训练权重加载 backbone，学习率衰减策略为线性衰减。

1.3.2 模型评价指标

本研究采用精确率 (precision, P)、F1、IoU 阈值为 0.5 时的平均精度均值 (mAP50)、IoU 阈值为 0.5 至 0.95 时的平均精度均值 (mAP50-95)、参数量 (parameters, Params)、模型大小 (model size) 以及每秒帧数 (frames per second, FPS) 等指标，对模型进行评估。

2 结果与分析

2.1 消融试验结果

为验证 CCGs-YOLO 模型结构改进的有效性，本文基于 YOLOv8 模型设计了 8 组消融试验，逐步引入 CaFormer、CGLU 及 Inner-CIoU 回归损失函数，以评估不同模块组合对模型检测性能的影响。

表 2 结果表明，单独引入 ConvFormer 模块后，模型 F1 值由 89.96% 提升至 90.44%，mAP50 达到 96.40%，每秒帧数为 286.54，说明 ConvFormer 在增强局部与全局特征提取能力的同时，提高了目标检测精度。继续引入 CaFormer 模块后，模型的 F1 值从基线模型的 89.96% 提升至 90.63%，平均精度均值 mAP50 由 95.89% 提升至

96.12%，表明该模块在增强特征建模能力方面具有积极作用，每秒帧数保持在 357.78，显示出较高的实时性。在仅结合 CGLU 机制的情况下，准确率为 89.94%，F1 值为 90.20%，mAP50 达到 96.12%，每秒帧数为 266.83，说明门控线性单元在提升关键通道特征表达能力的同时，也对模型检测性能有一定提升。在 CaFormer 模块结合 CGLU 机制后，准确率由基础模型的 89.07% 提升至 91.47%，F1 值升至 90.77%，mAP50 进一步提高至 96.24%，表明门控机制强化了关键通道的表达能力。在

回归损失方面，将 CIoU 替换为本文提出的 Inner-CIoU 后，模型 mAP50 由 95.89% 提升至 96.02%，mAP50-95 从 75.81% 上升至 75.84%，显示出更优的目标定位精度和收敛稳定性。当同时引入 Inner-CIoU、CaFormer 与 CGLU 三者后，准确率提升至 93.07%，较初始模型增加 3.31 个百分点，F1 值达到 90.87%，提升 0.91 个百分点，mAP50 提升为 96.24%。同时，模型大小从 5.97 MB 缩减至 4.88 MB，参数量由 3.01 M 减少至 2.42 M，每秒帧数为 362.07，在性能、轻量化和推理速度实现了良好的结合。

表 2 消融试验结果

Table 2 Results of ablation experiments

改进策略 Improvement strategy					模型大小 Model size/MB	P/%	F1/%	mAP50/%	mAP50-95/%	参数量 Params /M	FPS/(帧·s ⁻¹)
CIoU	Inner-CIoU	ConvFormer	CaFormer	CGLU							
√	-	-	-	-	5.97	89.76	89.96	95.89	75.81	3.01	515.86
√	-	√	-	-	5.51	90.32	90.44	96.40	76.34	2.75	286.54
√	-	-	√	-	5.47	89.07	90.63	96.12	76.42	2.73	357.78
√	-	-	-	√	4.93	89.94	90.20	96.12	76.44	2.44	266.83
√	-	-	√	√	4.88	91.47	90.77	96.24	76.42	2.42	328.38
-	√	-	-	-	5.97	89.21	89.77	96.02	75.84	3.01	436.82
-	√	-	√	-	5.46	90.75	90.02	95.91	75.66	2.73	325.98
-	√	-	√	√	4.88	93.07	90.87	96.24	76.35	2.42	362.07

注：“-”表示未添加该模块，“√”表示添加该模块。

Note: “-” indicates that the module is not added, and “√” indicates that the module is added.

上述结果表明，本文所提出模型能够有效平衡准确性与效率，增强了复杂自然环境下果实识别能力。

2.2 损失函数对比结果

为验证 CCGs-YOLO 模型所使用的 Inner-CIoU 回归损失函数在目标定位精度上的有效性，本文选取 GIoU、DlIoU、CIoU、ElIoU、SIoU、ShapeIoU 及 PIoU 等主流系列损失函数作为对比对象，在相同模型结构与训练策略下进行试验。

如表 3 所示，Inner-CIoU 在准确率达到 93.07%，高于其他损失函数，体现出其在复杂自然场景中对目标的判别能力更强。Inner-CIoU 的 F1 为 90.87%，相比 CIoU 的 90.77% 和 ShapeIoU 的 90.58% 均有所提升，表明其在兼顾检测精度与召回能力方面表现更为均衡。在 mAP 指标上，Inner-CIoU 在 mAP50 达到 96.24%，优于 ShapeIoU 的 76.09%。

表 3 不同损失函数试验结果对比

Table 3 Comparison of test results for different loss functions %

损失函数 Loss function	P	F1	mAP50	mAP50-95
CIoU	91.47	90.77	96.24	76.42
GIoU	90.17	90.44	96.44	76.43
DlIoU	89.64	90.32	96.21	76.43
ElIoU	90.13	89.94	96.23	76.30
SIoU	91.90	90.37	96.54	76.85
ShapeIoU	90.48	90.58	96.15	76.09
PIoU	90.87	89.77	96.03	75.94
Inner-CIoU	93.07	90.87	96.24	76.35

如图 7 所示，结合训练过程中各损失函数的准确率，可进一步观察到 Inner-CIoU 在模型 Epoch 160 之后的训练后期仍保持稳定提升趋势，精度曲线波动小且于收敛，展现出极强的训练稳定性与边界拟合能力。综合试验结果分析，引入 Inner-CIoU 后的 CCGs-YOLO 模型具备更优的目标定位能力和更强的泛化能力，更适合果园环境蟠桃果实的目标检测任务。

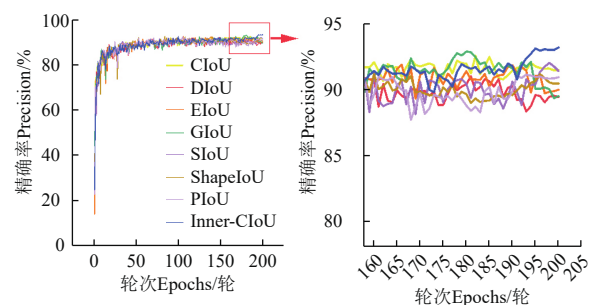


图 7 不同损失函数精确率对比

Fig.7 Comparison of precision using different loss functions

2.3 主流模型对比结果

为评估所提出 CCGs-YOLO 模型性能，本文将其与当前主流的 YOLO 系列（YOLOv12^[26]、YOLOv11^[27]、YOLOv10^[28]、YOLOv8^[29]）、Faster-RCNN^[30] 及 SSD^[31] 模型进行对比，试验结果如表 4 所示。

表 4 与其他主流模型试验结果对比

Table 4 Comparison with other mainstream model test results

模型 Model	模型大小 Model size/MB	P/%	F1/%	mAP50/ %	mAP50-95/ %	参数量 Params/ M
YOLOv12	5.21	90.91	90.32	96.10	75.68	2.51
YOLOv11	5.24	91.35	89.90	96.16	76.94	2.58
YOLOv10	5.50	88.91	89.79	95.84	76.51	2.27
YOLOv8	5.97	89.76	89.96	95.89	75.81	3.01
Faster-RCNN	108.20	60.97	73.12	89.34	41.29	28.28
SSD	90.61	83.37	87.89	93.67	64.95	11.39
CCGs-YOLO	4.88	93.07	90.87	96.24	76.35	2.42

由表 4 可见，CCGs-YOLO 准确率达到 93.07%，较 YOLOv11 提升 1.72 个百分点，较 YOLOv12 提升 2.16 个百分点，展现出更强的识别精度。在 F1 值方面，CCGs-YOLO 达到 90.87%，优于所有对比模型，较 YOLOv12 提升 0.55 个百分点，较 YOLOv11 提升 0.97 个百分点。

在平均精度指标上, CCGs-YOLO 的 mAP50 为 96.24%, 高于 YOLOv11 和 YOLOv12, 展现出更为稳定的检测能力。CCGs-YOLO 的 mAP50-95 为 76.35%, 略低于 YOLOv11 的 76.94%, 但其模型体积最小, 仅为 4.88 MB, 参数量仅为 2.42 M。在与 Faster-RCNN 和 SSD 的对比中, CCGs-YOLO 在各项指标上均优于二者, 同时模型体积和参数量远小于二者。综合来看 CCGs-YOLO 在检测精度、模型规模与效率之间展现出较好的综合性能。

在此基础上, 如图 8 所示本文将准确率较高的主流模型预测结果进行对比, 展示了各模型在实际图像中的 peach 类目标检测的结果。具体而言, 图 8a 展示了存在

边缘目标与复杂背景的场景, CCGs-YOLO 检测出 8 个目标, 置信度最高为 0.91, 位于图像边缘区域的目标置信度分别为 0.45 和 0.41, 能够有效识别目标, 相较其他模型体现出良好的边界感知能力; 图 8b 为存在枝叶遮挡与尺度差异的场景, 改进后模型检测数量最多, 达到 10 个; 图 8c 为果实遮挡分布场景, CCGs-YOLO 模型检测最高置信度为 0.92, 多个小尺寸目标置信度稳定; 图 8d 展示了在果实密集场景下弱特征目标情况, 改进后模型检测出 8 个目标, 置信度最高值为 0.92, 相较于 YOLOv10、YOLOv11 与 YOLOv12, 在漏检方面具有较强的感知能力。



图 8 不同模型图像检测对比

Fig.8 Comparison of image detection between different models

综上所述, CCGs-YOLO 在检测精度、F1 值及平均精度等指标上表现良好。同时, 通过实际图像的目标检测结果分析可见, CCGs-YOLO 在遮挡、多目标、小尺寸和边缘目标等复杂场景下具备较强的检测稳定性与特征感知能力。

2.4 可视化分析

为更直观地分析改进后的 CCGs-YOLO 模型性能, 本文引入 Grad-CAM (gradient-weighted class activation mapping) 进行可视化分析。图 9 中展示了多种条件下的真实图像识别结果, 在轻微遮挡下, CCGs-YOLO 能够集中响应果实主体, 即使部分遮挡也能准确定位; 在严重遮挡下, 模型仍保持较强响应, 优于 YOLOv8; 在近距离和远距离场景中, CCGs-YOLO 对目标边缘及小目标的响应更为完整, 增强了多尺度目标识别能力; 在顺光条件下, 模型能突出果实主要区域, 减少对背景的影响;

在背光条件下, CCGs-YOLO 热力图响应稳定, 目标特征清晰。总的来说, 与 YOLOv8 相比 CCGs-YOLO 模型的注意力更加集中于目标的关键区域, 减少了扩散现象。CCGs-YOLO 模型展现出较强的鲁棒性与泛化能力, 进一步验证了其在复杂环境下的有效性。

2.5 边缘计算设备部署

为进一步验证所提出模型在实际果园场景中的应用可行性, 本文在联保科技边缘计算设备上对 YOLOv8 与改进后的 CCGs-YOLO 模型进行了对比测试。该设备型号为 EA-B310, 搭载了 NVIDIA Jetson Xavier NX 核心处理器, 其硬件架构包含 6 核 Carmel ARMv8.2 64 位 CPU 以及拥有 48 个 Tensor Core 的 384 核 Volta GPU。

结果如表 5 所示, 从检测精度来看, CCGs-YOLO 的准确率达到 93.02%, 较 YOLOv8 的 89.72% 提升 3.30

个百分点, F1 值提升至 90.83%, mAP50 提高至 96.24%, 优于原始模型。在边缘计算设备 Tensor 核心加速与浮点运算优化的作用下, CCGs-YOLO 在 mAP50-95 指标上

达到 88.82%。CCGs-YOLO 的模型文件大小为 4.88 MB, 较 YOLOv8 的 5.96 MB 减小 18%, 参数量从 3.01 M 降至 2.42 M。

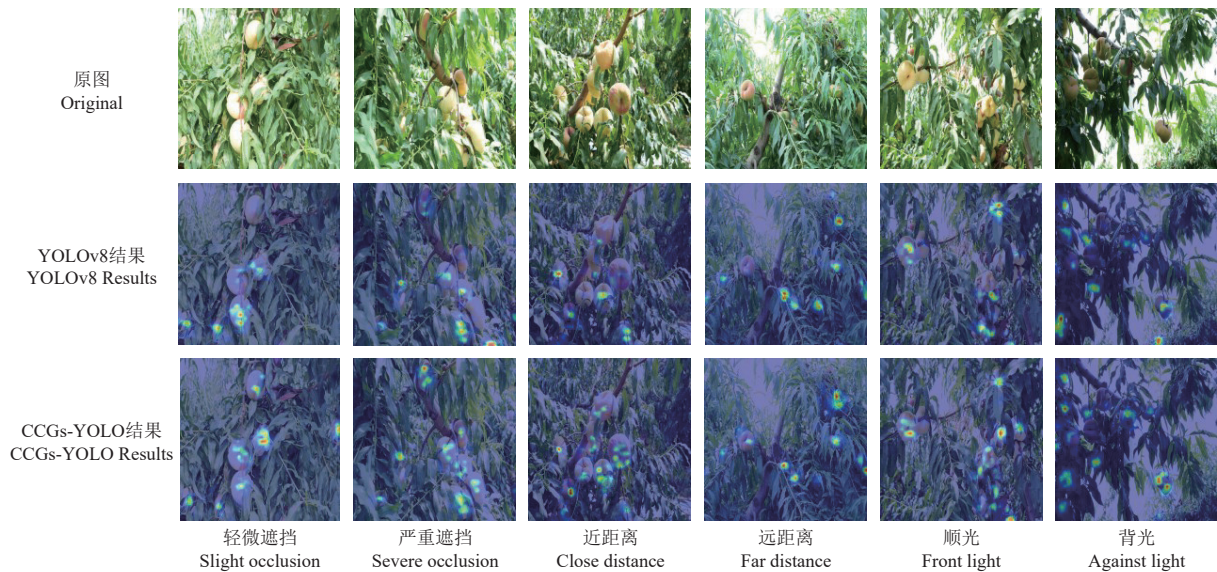


图 9 不同环境下模型可视化热力图对比

Fig.9 Comparison of model visualization heat maps in different environments

表 5 边缘计算设备部署试验结果

Table 5 Edge computing device deployment test results

模型 Model	模型大小 Model size/MB	P/%	F1/%	mAP50/ %	mAP50-95/ %	FPS/ (帧·s ⁻¹)	参数量 Params/M
YOLOv8	5.96	89.72	89.94	95.89	75.81	25.32	3.01
CCGs-YOLO	4.88	93.02	90.83	96.24	88.82	15.62	2.42

由于引入了更复杂的注意力机制和卷积增强模块, 其推理计算量有所增加, 因此在边缘设备上的帧率略低于 YOLOv8。综合来看, CCGs-YOLO 在精度、目标定位能力和轻量化方面实现了优化, 整体性能优于 YOLOv8。

3 结 论

针对自然果园环境中蟠桃果实识别中存在的复杂背景、小目标、密集遮挡等问题, 本文提出了一种基于改进 YOLOv8 的目标检测模型 CCGs-YOLO, 并围绕模型结构设计、损失函数改进、主流方法对比、可视化验证及边缘部署能力等方面展开系统研究, 主要结论如下:

1) 在自建数据集上, 改进后模型在准确率 P 、F1 值、mAP50 及 mAP50-95 指标上分别达到 93.07%、90.87%、96.24% 及 76.35%, 在复杂背景、遮挡及密集分布场景下表现出较高的检测准确性与稳定性, 同时在较低参数量与模型体积条件下仍保持良好的检测性能。

2) 通过引可视化分析方法验证改进后模型注意力机制的有效性, 结果显示改进后模型相较 YOLOv8 更专注于果实关键区域, 具备更强的特征提取能力与背景抑制效果。同时, 在边缘计算设备部署测试中, 改进后模型精确率 P 为 93.02%, 参数量与模型大小为 2.42 M 和 4.88 MB, 在保持较高检测精度的基础上压缩了模型大小与参数量, 具有良好的轻量化特性。

CCGs-YOLO 模型在自然果园环境中提升了蟠桃果实的识别能力, 为自动采摘提供了有效支持。然而, 由于引入了复杂的注意力机制和卷积增强模块, 模型在资源受限的边缘设备上帧率较低, 未来工作将重点优化模型的计算性能与推理效率, 以提升边缘设备上的实时性表现, 进而适配嵌入式应用。同时, 后续研究将进一步扩展数据采集范围, 构建多源、多环境的数据集。

[参 考 文 献]

- [1] 周慧娟, 苏明申, 杜纪红, 等. 上海桃产业信息现状与发展趋势[J]. 上海农业学报, 2024, 40(5): 107-114.
ZHOU Huijuan, SU Mingshen, DU Jihong, et al. Current situation and development trends of peach industry information in shanghai province[J]. Acta Agriculturae Shanghai, 2024, 40(5): 107-114. (in Chinese with English abstract)
- [2] NGUYEN T T, VANDEVOORDE K, WOUTERS N, et al. Detection of red and bicoloured apples on tree with an RGB-D camera[J]. Biosystems Engineering, 2016, 146: 33-44.
- [3] 刘晓洋, 赵德安, 贾伟宽, 等. 基于超像素特征的苹果采摘机器人果实分割方法[J]. 农业机械学报, 2019, 50(11): 15-23.
LIU Xiaoyang, ZHAO De'an, JIA Weikuan, et al. Fruits segmentation method based on superpixel features for apple harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 15-23. (in Chinese with English abstract)
- [4] 黄铭杰, 蔡伟强, 张子健, 等. 基于改进 YOLO11 的荔枝果实品种实时精准识别算法[J]. 农业工程学报, 2025, 41(11): 156-164.
HUANG Mingjie, CAI Weiqiang, ZHANG Zijian, et al. Real-time accurate recognition algorithm for litchi fruit varieties based on improved YOLO11[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE),

- 2025, 41(11): 156-164 (in Chinese with English abstract)
- [5] 李茂, 肖洋轶, 宗望远, 等. 基于改进 YOLOv8 模型的轻量化板栗果实识别方法[J]. 农业工程学报, 2024, 40(1): 201-209.
LI Mao, XIAO Yangyi, ZONG Wangyuan, et al. Detecting chestnuts using improved lightweight YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE) 2024, 40(1): 201-209. (in Chinese with English abstract)
 - [6] YANG R T, HE Y B, HU Z W, et al. CA-YOLOv5: A YOLO model for apple detection in the natural environment[J]. *Systems Science & Control Engineering*, 2024, 12(1): 2278905.
 - [7] GAI R L, 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.
 - [8] DU X Q, CHENG H C, MA Z H, et al. DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels[J]. *Computers and Electronics in Agriculture*, 2023, 214: 108304.
 - [9] 孙俊, 吴兆祺, 贾忆琳, 等. 基于改进 YOLOv5s 的果园环境葡萄检测[J]. 农业工程学报, 2023, 39(18): 192-200.
SUN Jun, WU Zhaoqi, JIA Yilin, et al. Detecting grape in an orchard using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(18): 192-200. (in Chinese with English abstract)
 - [10] 张志远, 罗铭毅, 郭树欣, 等. 基于改进 YOLO v5 的自然环境下樱桃果实识别方法[J]. 农业机械学报, 2022, 53(S1): 232-240.
ZHANG Zhiyuan, LUO Mingyi, GUO Shuxin, et al. Cherry fruit detection method in natural scene based on improved YOLOv5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(S1): 232-240. (in Chinese with English abstract)
 - [11] JIN S X, ZHOU L, ZHOU H P. CO-YOLO: A lightweight and efficient model for Camellia oleifera fruit object detection and posture determination[J]. *Computers and Electronics in Agriculture*, 2025, 235: 110394.
 - [12] SUN H X, REN R, ZHANG S J, et al. Maturity detection of 'Huping' jujube fruits in natural environment using YOLO-FHLD[J]. *Smart Agricultural Technology*, 2024, 9: 100670.
 - [13] REN R, SUN H X, ZHANG S J, et al. Intelligent Detection of lightweight Yuluxiang pear in non-structural environment based on YOLO-GEW[J]. *Agronomy*, 2023, 13(9): 2418.
 - [14] LIU P Z, YIN H. YOLOv7-Peach: An algorithm for immature small yellow peaches detection in complex natural environments[J]. *Sensors (Basel)*, 2023, 23(11): 5096.
 - [15] LV J D, NIU L L, XU L M, et al. A visual identification method of the growth posture of young peach fruits in orchards[J]. *Scientia Horticulturae*, 2024, 335: 113355.
 - [16] FAN J B, SAH B H, RONG X W, et al. Peach fruit thinning image detection based on improved YOLOv8 and data enhancement Techniques[J]. *IEEE Access*, 2024, 12: 191199-191218.
 - [17] WU E H, MA R J, DONG D M, et al. D-YOLO: A lightweight model for strawberry health detection[J]. *Agriculture*, 2025, 15(6): 570.
 - [18] 孙丰刚, 王建丽, 季英超, 等. 基于机器视觉的松林天牛远程智能监测系统[J]. 农业工程学报, 2023, 39(17): 190-198.
SUN Fenggang, WANG Jianli, JI Yingchao, et al. Remote intelligent monitoring system for pine forest longicorn beetles based on machine vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(17): 190-198. (in Chinese with English abstract)
 - [19] 杨锋, 姚晓通. 基于改进 YOLOv8 的小麦叶片病虫害检测轻量化模型[J]. 智慧农业 (中英文), 2024, 6(1): 147-157.
YANG Feng, YAO Xiaotong. Lightweighted Wheat Leaf Diseases and Pests Detection Model Based on Improved YOLOv8[J]. *Smart Agriculture*, 2024, 6(1): 147-157. (in Chinese with English abstract)
 - [20] 李志臣, 罗卫平, 凌秀军, 等. 基于改进 YOLOv8 的自然环境下板栗蓬检测方法[J]. 中国农机化学报, 2024, 45(12): 251-258.
LI Zhichen, LUO Weiping, LING Xiujun, et al. Detection method of Chinese chestnut in natural environment based on improved YOLOv8[J]. *Journal of Chinese Agricultural Mechanization*, 2024, 45(12): 251-258. (in Chinese with English abstract)
 - [21] TERVEN J, CoRDOVA-ESPARZA D, ROMERO-GONZÁLEZ J. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.
 - [22] YU W, SIC, ZHOU P, et al. MetaFormer baselines for vision[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(2): 896-912.
 - [23] SHI D. Transnext: Robust foveal visual perception for vision transformers[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, USA: IEEE, 2024: 17773-17783.
 - [24] ZHANG H, XU C, ZHANG S. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[EB/OL]. (2023-11-05)[2025-09-02]. <https://doi.org/10.48550/arXiv.2311.02877>.
 - [25] XIAO G, HOU S, ZHOU H. PCB defect detection algorithm based on CDI-YOLO[J]. *Scientific Reports*, 2024, 14(1): 7351.
 - [26] MENG X, BU F, WU J, et al. Research and application of tomato leaf disease recognition method based on improved YOLOv12[C]//2025 International Conference on Trustworthy Big Data and Artificial Intelligence (ICTBAI). Tokyo, Japan: IEEE, 2025: 67-72.
 - [27] WANG L, WANG S, WANG B, et al. Jujube-YOLO: A precise jujube fruit recognition model in unstructured environments[J]. *Expert Systems with Applications*, 2025, 291: 128530.
 - [28] YIN Y, SUN X, YU G, et al. CBFW-YOLOv8: Automated recognition method for fish body surface diseases in recirculating aquaculture systems[J]. *Computers and Electronics in Agriculture*, 2025, 236: 110612.
 - [29] HE Z, KARKEE M, ZHANG Q. Enhanced machine vision system for field-based detection of pickable strawberries: Integrating an advanced two-step deep learning model merging improved YOLOv8 and YOLOv5-clas[J]. *Computers and Electronics in Agriculture*, 2025, 234: 110173.
 - [30] YU H, SONG H, XU L, et al. SED-RCNN-BE: A SE-Dual channel RCNN network optimized binocular estimation model for automatic size estimation of free swimming fish in aquaculture[J]. *Expert Systems with Applications*, 2024, 255: 124519.
 - [31] XIONG S, FAN Z. Research on pedestrian detection technology based on SSD algorithm[C]//2021 International Conference on Digital Society and Intelligent Systems (DSInS). Chengdu, China: IEEE, 2021: 352-356.

Identifying flat peach fruits in orchard environments using an improved YOLOv8 network model

WANG Chen , MA Chunyue , GUO Xiuru , WANG Zhijun^{*} , SUN Bo , GUO Xuchao

(1. School of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China;

2. Apple Technology Innovation Center of Shandong Province, Tai'an 271018, China)

Abstract: Recognition accuracy of small targets is often required for dense fruit distribution in natural orchard environments. Therefore, a flat peach detection model based on an improved YOLOv8 architecture, termed CCGs-YOLO, was proposed in this study. The proposed model integrates the MetaFormer framework with a convolutional gated linear unit module. A hybrid module combining convolution and attention mechanisms was introduced to enhance spatial feature extraction and improve feature representation under complex background conditions. Meanwhile, a channel-aware module was incorporated to simulate inter-channel dependencies, thereby improving the discrimination capability between fruit targets and cluttered backgrounds. Specifically, the C2f_ConvFormer module was employed to simultaneously capture local and global contextual information, while the C2f_CaFormer module was introduced to enhance channel interaction and feature aggregation. In addition, a convolutional gated linear unit mechanism was embedded into the network to improve feature selection capability and robustness against background noise. To address the small object detection problem in densely distributed fruit scenarios, localization accuracy was further improved. An optimized regression loss function based on an inner-overlap constraint, named Inner-CIoU, was adopted to achieve more accurate bounding box regression and reduce localization errors caused by overlapping targets. Experimental results demonstrated that the ConvFormer module improved the F1-score to 90.44% and the mAP50 to 96.40%, indicating enhanced feature extraction capability. The CaFormer module increased the F1-score from 89.96% to 90.63%, while the mAP50 further improved to 96.12%, demonstrating effective channel modeling capability under relatively high inference efficiency. When only the convolutional gated linear unit mechanism was applied, the F1-score reached 90.20% and the mAP50 achieved 96.12%, verifying its effectiveness in enhancing feature representation. Furthermore, the combination of CaFormer and convolutional gated linear unit improved the Precision to 91.47%, the F1-score to 90.77%, and the mAP50 to 96.24%, demonstrating the complementary advantages of channel modeling and gated feature selection. In terms of localization performance, Inner-CIoU improved both mAP and model convergence stability compared with the conventional CIoU loss function. After integrating all improved components, the model achieved relatively better overall performance. Precision reached 93.07%, representing an increase of 3.31 percentage points compared with the baseline model. The F1-score reached 90.87%, while the mAP50 achieved 96.24%. Meanwhile, the model size was reduced from 5.97 MB to 4.88 MB, and the number of parameters decreased from 3.01 million to 2.42 million, indicating that the proposed model possesses favorable lightweight characteristics while maintaining relatively high inference speed of 362.07 FPS. In addition, comparative experiments were conducted with several mainstream models, including different versions of the YOLO series. Under challenging scenarios such as occlusion, small targets, dense distribution, and edge targets, the proposed model achieved relatively superior performance in terms of Precision, F1-score, and mAP, demonstrating improved detection stability and feature perception capability. Visualization analysis further indicated that the improved model could focus more accurately on fruit regions and suppress background interference to a certain extent. Furthermore, deployment experiments on edge computing devices demonstrated that the proposed model could still maintain relatively high detection accuracy and stable performance under practical application conditions, with the mAP50-95 reaching 88.82%, indicating potential for real-world applications. Overall, the proposed model effectively balanced detection accuracy, model lightweight characteristics, and computational efficiency, demonstrating good robustness and adaptability in complex orchard environments. The proposed approach can provide a feasible technical solution for rapid and accurate fruit recognition in flat peach harvesting.

Keywords: flat peach; YOLOv8; fruit recognition; object detection; ConvFormer; CaFormer; CGLU