

基于改进 YOLOv5s 模型的自然场景中绿色柑橘果实检测

吕强^{1,2}, 林刚^{1,2}, 蒋杰^{1,2}, 王明之^{1,2,3}, 张皓杨^{1,2}, 易时来^{1,2}

(1. 西南大学柑桔研究所, 重庆 400712; 2. 国家柑橘工程技术研究中心, 重庆 400712;
3. 西南大学工程技术学院, 重庆 400715)

摘要: 针对未成熟柑橘果实智能检测存在精度低、模型大的问题, 该研究提出了一种基于 YOLOv5s 改进的绿色柑橘检测算法模型 YOLO-GC, 实现对复杂自然环境中果实的实时准确检测。首先, 针对 YOLOv5s 网络模型较大且难以部署的问题, 以轻量级 GhostNet 网络替换原始的骨干网络, 同时为减小模型轻量化后精度下降的影响和提高对绿色柑橘特征的关注度, 嵌入全局注意力机制 (global attention mechanism, GAM) 增强网络在复杂环境下对果实特征的提取能力; 其次, 为了改善密集与小目标果实的检测效果, 引入 BiFPN (bi-directional feature pyramid network) 架构进行多尺度的加权特征融合; 最后, 为减少果实与枝叶遮挡、重叠造成的漏检, 采用 GIoU (generalized intersection over union) 损失函数结合 Soft-NMS (soft-non-maximum suppression) 算法优化边界框回归机制。试验结果表明: 相较于 YOLOv5s, YOLO-GC 的权重模型内存减小了 53.9%, 参数量减少了 55.2%, 平均精度 AP_{0.5} 提升了 1.2 个百分点, 平均推理时长减少 46.2%。YOLO-GC 模型的综合检测性能优于 YOLOv8 等 7 种常用网络模型, 在安卓手机 APP 中检测准确率达到 97.2%, 推理时长减少了 85.8%。研究表明, 该研究模型为复杂环境中绿色果实检测及模型部署应用提供了技术支撑。

关键词: 柑橘; 图像识别; YOLO; 绿色果实; 轻量化模型; 移动部署

doi: 10.11975/j.issn.1002-6819.202404190

中图分类号: TP391.4; S24

文献标志码: A

文章编号: 1002-6819(2024)-18-0147-08

吕强, 林刚, 蒋杰, 等. 基于改进 YOLOv5s 模型的自然场景中绿色柑橘果实检测[J]. 农业工程学报, 2024, 40(18): 147-154. doi: 10.11975/j.issn.1002-6819.202404190 <http://www.tcsae.org>

LYU Qiang, LIN Gang, JIANG Jie, et al. Detecting green citrus fruit in natural scenes using improved YOLOv5s model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 147-154. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404190 <http://www.tcsae.org>

0 引言

柑橘是中国种植面积和产量第一的水果, 已成为中国最重要的经济作物之一^[1]。快速准确地检测未成熟的果实生成果园的产量预测图^[2], 在此基础上进行水肥管理、花果管理、树体管理等是果园智能化生产的重要环节^[3]。自然环境中未成熟的绿色柑橘因与冠层枝叶背景颜色相近^[4]、生长环境复杂多变^[5]等问题, 其识别效果不稳定。人工智能和深度学习推动了精准农业的智能化技术研究与应用, 其中基于深度学习的目标检测技术广泛应用于果实成熟度判别^[6]、产量预估^[7]、机械化采摘^[8]。随着智能手机的普及和性能日益增强, 研究人员研发移动端平台 APP 开展果实识别、估产等农业信息应用技术研究^[9]。因此, 为提升复杂自然场景下对绿色柑橘果实网络模型的识别稳定性, 研究一种平衡推理速度与精度, 适合部署到边缘移动设备的轻量化网络模型, 成为柑橘智能化生产亟需解决的重要问题之一。

目前, 针对果实的识别方法主要分为传统的机器视

觉算法和基于深度学习的检测算法^[10]。传统的机器视觉技术主要利用果实的表面特征区分果实与背景^[11], 实现对果实的识别。HE 等^[12]提出深度边界盒回归森林方法, 通过描述不同尺度绿色柑橘的形状、纹理和颜色特征实现对果实的识别, 检测精度仅为 87.6%。相比之下, 基于深度学习的方法能够利用卷积神经网络提取丰富的浅层特征与深层次的语义特征, 逐渐取代了传统的图像处理算法, 并得到广泛应用^[13-14]。为解决自然场景中早期的绿色柑橘尺寸较小, 背景条件复杂的问题, 熊俊涛等^[15]提出 Faster R-CNN 模型检测绿色柑橘, 平均检测精度仅为 85.49%, 且双阶段的检测模型较大。为了推动深度学习模型在农业中的应用, 许多学者开始将轻量化模型部署到边缘设备端^[16]。ZHOU 等^[17]将 MobileNetV3 部署到 Android 平台, 利用开发的 APP 实现猕猴桃的产量估算, 单幅图像检测时间为 103 ms, 但安卓端的精度仅有 90.3%。LYU 等^[18]提出一种将 YOLOv5-CS 网络与跟踪算法相结合的方法, 在边缘设备 Nvidia Jetson Xavier NX 上实现绿色柑橘的检测计数, 检测精度有 98.23%, 但在光线变化、遮挡等复杂环境下的检测效果仍然不够理想。ZHENG 等^[19]提出一种 YOLO-BP 的多尺度特征融合网络识别绿色柑橘, 平均精度为 91.55%, 检测速度为每秒 18 帧, 但在一些有部分遮挡和重叠场景的果实上存在错检。龙燕等^[20]为实现不同环境下疏果期苹果的检

收稿日期: 2024-04-26 修订日期: 2024-08-05

基金项目: 重庆市技术创新与应用发展专项重点项目 (cstc2019jcsx-gksbX0095); 中央高校基本科研业务费项目 (XDJK2019C072)

作者简介: 吕强, 博士, 副研究员, 研究方向为农业信息技术、图像处理。
Email: qlu@swu.edu.cn

测, 在 YOLOv7 模型的基础上引入 Swin Transformer 小目标注意力模块, 增加了网络对小目标果实的关注度, 同时使用 SIOU (shape intersection over union) 损失函数, 幼果检测准确率达到 92.7%, 但模型内存占用量高达 81MB。此外, XU 等^[21] 实现准确且快速地检测复杂环境中的柑橘, 采用 GhostNet 网络改进 YOLOv4 骨干网络, 引入 ECA (efficient channel attention) 注意力机制增强多尺度特征权重, 并使用 DIOU (distance intersection over union) 损失函数, 模型的准确率达到 93.45%, 但模型内存为 43.5MB, 在不同程度的模糊、遮挡和光照变化情况下仍存在漏检。

目前绿色柑橘果实检测的网络模型在遮挡、重叠、密集、小目标等复杂多变自然环境下鲁棒性不足, 而且网络模型的参数量和计算量较大, 不利于实际生产场景中的应用。针对自然场景中绿色柑橘果实检测存在的问题, 本文提出了 YOLO-GC 网络模型, 实现模型的轻量化, 提高复杂多变环境下对果实特征提取和分类的能力。同时, 在安卓手机端开发了 APP 进行绿色柑橘识别性能的测试, 为算法模型在边缘移动端的部署与产量预估提供技术支持。

1 材料与方法

1.1 材料

1.1.1 数据采集

本研究于 2023 年的 7—9 月采集了果园环境中的绿色柑橘图像, 包括沃柑 (*Citrus reticulata* Blanco cv. Orah mandarin)、纽荷尔脐橙 (*Citrus sinensis* (L.) Osbeck. cv. Newhall navel orange)、不知火 (*Citrus reticulata* Blanco cv. Shiranui mandarin)、春香 (*Citrus tamurana* Hort. Ex Tanka cv. Haruka tangelo)、冰糖橙 (*Citrus sinensis* (L.) Osbeck cv. Bingtangcheng sweet orange)、塔罗科血橙 (*Citrus sinensis* (L.) Osbeck cv. Tarocco blood orange) 共 6 个柑橘品种。图像拍摄设备包括 iPhone XR、华为 Mate30 Pro、小米 11Ultra、红米 10X 四种型号手机, 图片分辨率为 3 024×3 024、3 060×3 060 像素。在拍摄过程中, 手持设备在距离柑橘冠层 0.3~2.5 m 的范围内, 采集了不同光照条件、不同视角和不同尺度的柑橘冠层图像, 共有 2 400 幅。

1.1.2 数据扩增与数据集制作

统一将图片的尺寸调整为 640×640, 以 3:1:1 的比例将 2 400 幅绿色柑橘图像随机划分为训练集、验证集和测试集。由于实际应用中数据采集时存在光照强度变化、噪声和模糊等多种情况, 为增加样本复杂性, 通过数据扩增让网络模型充分学习复杂自然场景中目标果实的特征, 提升其在复杂场景下的泛化能力。本研究对训练集和验证集采用了减小亮度、增大亮度、高斯模糊、椒盐噪声以及雾化处理 5 种方法进行数据扩增^[22], 如图 1 所示。本文数据集扩增前后的训练集、验证集和测试集分布情况见表 1。

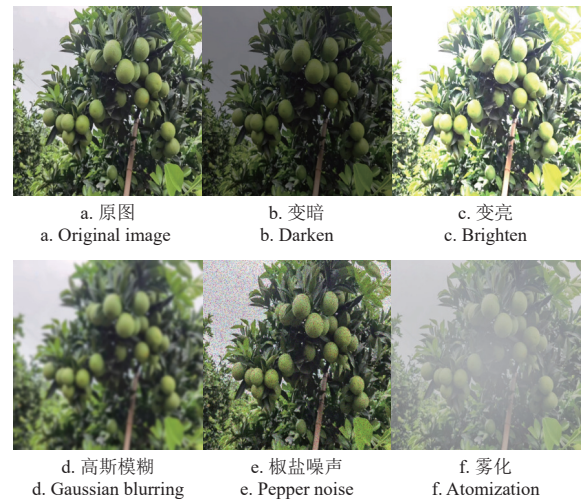


图 1 样本数据扩增

Fig.1 Sample data augmentation

表 1 绿色柑橘图像的样本数量分布

Table 1 Distribution of sample size for green citrus images

样本 Sample	训练集 Train set	验证集 Validation set	测试集 Test set	总计 Total
原样本 Original sample	1 440	480	480	2 400
扩增后 After augmentation	8 640	2 880	480	12 000

1.2 方法

YOLOv5s 作为单阶段目标检测器, 无需繁琐的候选区域生成和分类步骤, 具有更为简洁的网络结构, 平衡模型性能与速度, 保证了对目标检测的实时性与准确性, 是深度学习在农业领域应用较为广泛的算法^[9]。为了在复杂自然环境中实现果实准确检测并适应边缘移动设备的推理, 本文基于 YOLOv5s 目标检测网络, 提出了 YOLO-GC 网络模型, 该网络做了如下改进: 1) 使用 GhostNet 网络代替原始的骨干网络, 实现网络结构轻量化; 2) 在骨干网络和特征融合层嵌入 GAM 注意力机制增强对绿色柑橘特征的感知能力, 提升网络模型轻量化后的精度; 3) 在特征融合层使用特征金字塔融合网络 BiFPN 对特征图进行双向加权, 改善不同尺寸和密集场景下的果实识别效果; 4) 使用 GIoU 损失函数结合 Soft-NMS 算法, 减少绿色柑橘在遮挡和重叠场景中的错检和漏检。基于上述改进得到的 YOLO-GC 网络结构如图 2 所示。

1.2.1 网络模型轻量化

骨干网络 (Backbone) 作为 YOLO-GC 网络的重要组成部分, 本研究使用 GhostNet 作为 YOLO-GC 骨干网络, 高效提取图像中有效的果实特征。GhostNet 网络采用简单线性变换生成幻影特征取代传统卷积生成的特征, 从而解决网络特征图冗余和计算成本高的问题。该网络的 Ghost 模块由普通的卷积模块和线性模块构成, 首先通过少量普通卷积操作生成部分特征图, 然后利用大量轻量级线性变换处理冗余特征图, 最后通过恒等映射方式拼接组合两组特征图, 以获得有效的目标特征图。

本文提出的 YOLO-GC 采用 GhostNet 作为主干网络，通过引入 Ghost 模块取代传统的卷积模块，旨在减少主干

网络的参数数量和额外的计算负荷，实现参数的高效利用，并提升模型的推理速度。

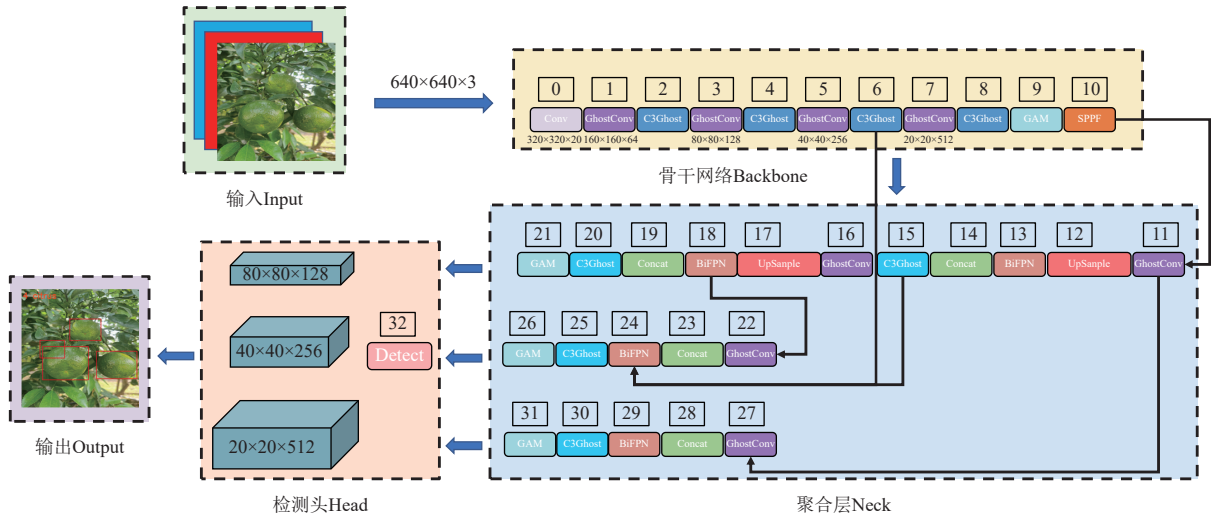


图 2 YOLO-GC 网络模型结构
Fig.2 YOLO-GC network model architecture

1.2.2 注意力机制

GhostNet 网络舍弃冗余特征减小模型的复杂度，会导致图像中部分重要特征的损失。为提高对绿色柑橘特征的感知能力和改善轻量化后网络的识别精度，在骨干网络和特征融合层引入 GAM 注意力机制。GAM 通过引入多层感知器，以通道注意力、空间注意力的顺序组合子模块放大全局特征的交互，减少信息扩散，突出重要信息的表达^[24]。GAM 输入特征映射图 $F_1 \in R^{L \times H \times W}$ ，其中 L 表示特征图的通道数， H 和 W 分别表示特征图的高度和宽度。中间状态特征 F_2 和 F_3 如式 (1) 和 (2) 所示：

$$F_2 = M_c(F_1) \otimes F_1 \quad (1)$$

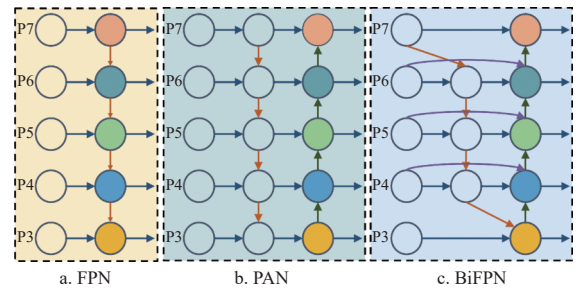
$$F_3 = M_s(F_2) \otimes F_2 \quad (2)$$

式中 M_c 表示通道注意力， M_s 表示空间注意力。

1.2.3 特征融合层

由于 YOLO 使用的 FPN (feature pyramid network) 与 PAN (path aggregation network) 特征融合架构对图像特征利用率较低。本文采用 BiFPN 进行多尺度特征融合，整合不同尺度的特征信息，并提供更好的上下文感知能力^[25]，以提升柑橘果实在密集以及小目标场景下的检测准确率。FPN、PAN 和 BiFPN 三种特征融合的原理如图 3 所示。BiFPN 在同一尺度的特征图上建立双向连接后删除对特征融合贡献不大且冗余的节点，降低模型的计算复杂度^[26]。此外，它在原始的输入节点和输出节点中有一条单独的通道，进行不同尺度特征信息的融合，提高多尺度目标特征的融合程度。YOLO-GC 网络的特征融合层利用自适应的上下采样方式，通过多尺度的双向加权高效准确地融合高层语义和底层图像的信息，在不同尺度的特征图上实现了信息的交互融合。因此，YOLO-GC 能更好地理解不同尺寸和分辨率的绿色果实，从而提高小尺寸果实检测的精度，并增强对密集环境下

绿色柑橘的检测性能。



注：P3~P7 指经过骨干网络的不同分辨率的特征图。
Note: P3-P7 refer to feature maps of different resolutions passing through the backbone network.

图 3 三种多尺度特征融合框架
Fig.3 Three multi-scale feature fusion frameworks

1.2.4 优化损失函数和非极大值抑制

为了解决自然环境中绿色柑橘果实与枝叶、果实之间不同程度的遮挡和重叠问题，本文采用 GIoU 损失函数结合 Soft-NMS 算法，通过优化检测框与预测框之间的相似度，并结合低分检测框阈值剔除算法，提升模型在这些复杂场景下的目标检测效果。

GIoU 通过优化目标框的位置回归，考虑预测框的位置、大小、重叠和遮挡情况。GIoU 通过结合 IoU (intersection over union) 的全局信息^[27]，加入非重叠部分的边界框，衡量预测边界框和真实边界框的相似度。IoU 和 GIoU 分别如式 (3) 和 (4) 所示：

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \quad (4)$$

式中 A 表示预测的边界框， B 表示真实的边界框， C 表示两个边界框 A 和 B 的最小闭合区域。

传统的非极大值抑制 (non-maximum suppression, NMS) 算法会删除与较高 IoU 值边界框相似的先验框, 导致生成的候选框的置信度分数下降, 造成重叠的目标漏检。Soft-NMS 算法^[28]通过引入衰减权重函数, 根据重叠部分的大小、中心点距离和纵横比为相邻边界框设置权重。与 NMS 不同, 它不会直接删除低分阈值的框, 而是降低遮挡目标的置信度。Soft-NMS 算法通过式 (5) 实现:

$$S_i = \begin{cases} S_i, \text{IoU}(M, B_i) < N_i \\ S_i(1 - \text{IoU}(M, B_i)), \text{IoU}(M, B_i) \geq N_i \end{cases} \quad (5)$$

式中 M 表示得分较高的候选先验框, B_i 表示要处理的候选先验框, N_i 为设定的阈值, IoU 是衰减高分的候选先验框与前一帧候选先验框的重叠分数。目标果实重叠程度越高, 分数的衰减程度越大, S_i 表示剩余的第 i 个候选先验框的分数。

1.3 试验平台与评价指标

试验过程中使用服务器进行模型的训练和优化, 操作系统: Ubuntu 20.04, 显卡: NVIDIA Geforce RTX 3090, 处理器: 13 th Intel(R) Core(TM) i9-13900K, 显存: 24GB。编程语言 Python 3.9, 深度学习框架: Pytorch。APP 编码平台: Android Studio, JDK 1.8, 开发语言: Java。安卓手机端 APP 使用的测试设备: 红米 10X, 运行内存: 8G, Android 13。为评估提出方法的试验效果, 本文使用以下评价指标: 精确率 (precision, P), 召回率 (recall, R), IoU 阈值为 0.5 时的平均精度 ($\text{AP}_{0.5}$), IoU 阈值从 0.5 到 0.95 且步长为 0.05 时的平均精度 ($\text{AP}_{0.5\sim 0.95}$), 准确率和召回率的调和平均数 (F_1); 一秒内执行浮点运算数量 (FLOPs); 参数量 (Parameters); 权重模型内存 (weights size), 单幅图像的平均推理时长 (inference time)^[9]。

2 结果与分析

2.1 网络模型训练结果

本研究使用 COCO 数据集训练的模型作为预训练权重模型, 训练参数设置如下: 图片 640×640 像素, 每批次处理图像数量 16; 迭代 300 轮; 优化器为 SGD; 初始学习率为 0.001; 动量值为 0.937; Moasic 数据增强方法强度为 0.7; Mixup 数据增强方法强度为 0.5。其中图片分辨率、Moasic 和 Mixup 数据增强方法强度的参数值是依据本文制作的数据集特性和采取的数据增强方式设定^[29]; 每批次处理图像数量根据训练模型硬件设备的算力和显存决定; 迭代轮数是通过观察网络模型在训练过程中损失函数的收敛情况设置。优化器类型、学习率、动量值参数的设定参照了官方在 COCO 公共数据集上训练 YOLOv5s 模型时的最佳配置^[30], 通过移植原始模型的训练参数可以为本文模型的训练提供一个较好的起点, 加速训练过程, 从而提升模型在绿色柑橘检测任务上的性能。

YOLOv5s 和改进后模型利用本文自制的绿色柑橘数据集训练, 损失函数的变化如图 4 所示。对比模型改进前后的位置和置信度损失值变化趋势可知: 改进后的算法模型相较于 YOLOv5s, 其位置损失函数值和置信度损失函数值在收敛时表现出更小的损失值, 说明改进后的算法模型生成的预测框与真实框之间的误差较小, 预测框的位置更精准, 同时反映改进后的模型在正确预测目标方面具有更强的能力。

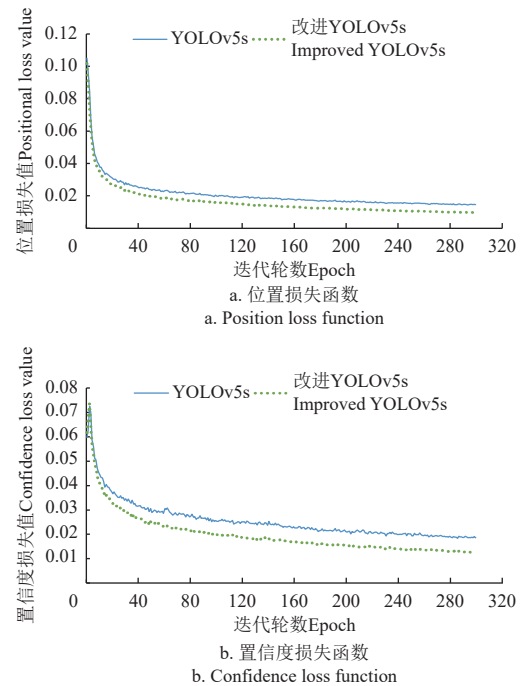


图 4 模型改进前后损失函数变化曲线
Fig.4 Loss function change curve before and after model improvement

2.2 消融试验结果

为验证 YOLO-GC 绿色柑橘果实检测网络模型对原始 YOLOv5s 网络改进的效果, 并探究改进后的各个模块的功能, 本文利用 5 组消融试验证明提出的改进方法的有效性。消融试验中基准网络以及经各个模块改进后的网络对比结果如表 2 所示。首先为实现模型的轻量化, 使用 GhostNet 替换原始骨干网络和卷积层较多的 C3 模块, 精度略有下降, 但模型内存与 FLOPs 分别减小 8.0 MB 和 7.8 G, 参数量减少 59.2%。其次, 在骨干网络和特征融合层嵌入 GAM 注意力模块以及使用 BiFPN 改进原始的金字塔特征融合网络后, 在模型内存、参数量和计算量增加较小的情况下, 精确率、召回率以及平均精度 $\text{AP}_{0.5}$ 分别提升了 1.3、1.5 和 1.2 个百分点。最后使用 GIoU 损失函数结合 Soft-NMS 算法改进 CIoU 损失函数和 NMS 算法, 精度略有提升, 但没有增加模型的大小和参数量。最终 YOLO-GC 模型的精确率与召回率分别达到 97.6% 和 90.3%, 平均精度 $\text{AP}_{0.5}$ 与 $\text{AP}_{0.5\sim 0.95}$ 分别达到 97.8% 和 89.6%, 模型计算量仅为 8.3 G, 参数量为 3.57×10^6 , 权重内存为 6.69 MB。

表 2 YOLOv5s 模型消融试验结果
Table 2 Results of the YOLOv5s model ablation experiments

GhostNet	GAM	BiFPN	Soft-NMS +GIoU	精确率	召回率	平均精度 Mean precision/%		浮点计算量	参数量	权重内存
				Precision P/%	Recall R/%	AP _{0.5}	AP _{0.5~0.95}	Floating point operations FLOPs/G	Parameters /10 ⁶	Weights size/MB
×	×	×	×	96.5	89.4	96.6	88.9	15.8	7.96	14.50
√	×	×	×	96.1	88.1	96.3	88.5	8.0	3.25	6.50
√	√	×	×	96.8	89.0	96.9	89.0	8.1	3.32	6.56
√	√	√	×	97.4	89.6	97.5	89.4	8.3	3.57	6.69
√	√	√	√	97.6	90.3	97.8	89.6	8.3	3.57	6.69

注: √表示使用该模块, ×表示不使用该模块。

Note: √ indicates that the module is used, × indicates that the module is not used.

2.3 不同方法检测试验

为对 YOLO-GC 模型进行性能测试, 本研究选择与 YOLOv5s^[25]、YOLOv7^[14]、YOLOX^[31]、YOLOv8^[32]、CenterNet^[33]、Faster R-CNN^[15]、RetinaNet^[26] 等 7 种常用的目标检测模型比较, 平均精度 AP0.5 分别提升了 1.2、1.3、1.4、0.9、1.7、4.6、3.9 个百分点, 检测对比结果如表 3 所示。从表 3 可得, 在绿色柑橘识别任务的准确性方面, 相较于其它单阶段和双阶段算法, YOLO-GC 模型表现出了更高的识别精确率、召回率和 F1 得分, 相较于 YOLOv5s 分别提升了 1.1、0.9 和 2.0 个百分点, 其

中 F1 得分为 97.0%。此外, 本文提出网络的模型比原始权重模型的内存减小了 53.9%, 参数量减少了 55.2%, 计算量减少了 47.5%, 推理时长也减少了 46.2%, 而且相较于 YOLOv8 等单阶段算法模型权重内存更小, 推理时长也更短。由此, 本文提出的 YOLO-GC 网络模型在检测准确性和检测速度方面均优于其他 7 种常用网络模型, 并且具有较小的权重内存, 更适合部署到边缘设备中。而其他几种网络模型由于检测精度相对较低, 权重模型大, 推理时间较长, 则不适宜部署到算力较小的边缘设备上。

表 3 不同模型与 YOLO-GC 网络的性能对比

Table 3 Performance comparison of different models with YOLO-GC network

模型 Model	P/%	R/%	AP _{0.5} /%	F ₁ 分数 F ₁ score/%	推理时长 Inference time/ms	权重内存 Weights size/MB
YOLOv5s	96.5	89.4	96.6	95.0	6.5	14.50
YOLOv7	96.6	89.3	96.5	95.0	6.3	11.70
YOLOX	96.3	89.5	96.4	95.0	7.9	19.40
YOLOv8	96.8	89.8	96.9	96.0	5.8	22.50
CenterNet	96.2	88.9	96.1	95.0	26.9	124.60
Faster R-CNN	94.4	87.6	93.2	94.0	36.4	108.40
RetinaNet	94.8	87.5	93.9	94.0	48.6	138.80
YOLO-GC	97.6	90.3	97.8	97.0	3.5	6.69

本文对 YOLOv5s 与 YOLO-GC 模型在测试集中进行了比较, 通过分析发现在遮挡、重叠、密集、小目标、弱光、逆光等因素或多因素复合的复杂场景中, YOLOv5s 模型检测出的果实数量低于 YOLO-GC 模型, 尤其是在严重遮挡、较密集、小目标果实、整株以及极端光照等场景下果实检测结果差异明显, 如图 5 所示。主要原因是 YOLOv5s 模型对绿色柑橘特征的关注度不足, 难以从复杂变化的背景中有效区分果实与背景; 对于小目标和密集的果实, 特征融合网络对不同尺度的特征利用不充分, 导致部分有效特征丢失; 模型的损失函数和回归框剔除机制对于低分检测框直接剔除, 造成了大量遮挡和重叠果实的漏检。YOLO-GC 模型利用 GAM 注意力模块提升了对绿色柑橘特征的提取能力, 引入 BiFPN 结构解决了小目标和密集情况下的漏检, 同时通过 GIoU 损失函数和 Soft-NMS 算法减少了遮挡和重叠的影响。最终 YOLO-GC 模型的检测能力得到提升, 检测到的果实数量更多, 检测精度更高。

2.4 误检和漏检分析

为了对后续模型应用提供指导和算法优化, 本文针对 YOLO-GC 模型检测果实时存在的错检和漏检情况进行了分析, 如图 6 所示。图 6a 中果实受逆光的光照条件影响, 导致叶片与果实的区分度较低, 容易把与果实颜色、形状轮廓等特征相似的叶片误识别为果实。图 6b 中果实漏检的主要原因是绿色柑橘果实与周围的冠层枝

叶背景相似, 且受到与果实颜色一致枝叶的严重遮挡, 导致模型无法提取绿色柑橘果实的特征, 误以为是除柑橘以外的其它目标, 从而出现漏检。针对上述错检和漏检问题, 未来将进一步增强网络模型在复杂环境下对果实各个维度有效特征的学习能力, 以更好地从复杂自然环境背景中区分果实。

2.5 移动端部署与测试

为验证本文的 YOLO-GC 算法模型在边缘设备端的检测性能, 本研究将模型移植到安卓手机, 并在开发的手机 APP 上进行绿色柑橘果实检测与计数。为了提升在手机端的模型运行效果, 首先将完成训练的 PyTorch 权重模型导出为 ONNX 格式, 然后转换为适合手机移动端的 NCNN 格式。为了进一步加快推理速度, 本研究对 NCNN 格式的网络模型进行了 FP16 量化, 将参数和激活值从 32 位浮点数转换为 16 位浮点数格式, 以减少存储需求和计算成本, 同时保持高效的推理性能。在边缘安卓移动手机 APP 端, YOLOv5s 与 YOLO-GC 网络模型分别使用本研究测试集的 480 张图片进行测试, 检测对比结果如表 5 所示。本文提出的算法模型在移动端的综合检测性能优于 YOLOv5s, 检测的准确率相较于 YOLOv5s 提高 2.4 个百分点, 高达 97.2%, 平均推理时长减少了 655 ms, 缩短了 85.8%, 仅为 108 ms, 漏检和错检的情况得到改善。结果表明, 本文提出的 YOLO-GC 网络模

型平衡了在移动手机端设备的检测精度与速度, 更适合 部署到边缘移动端设备中进行绿色柑橘果实的检测。

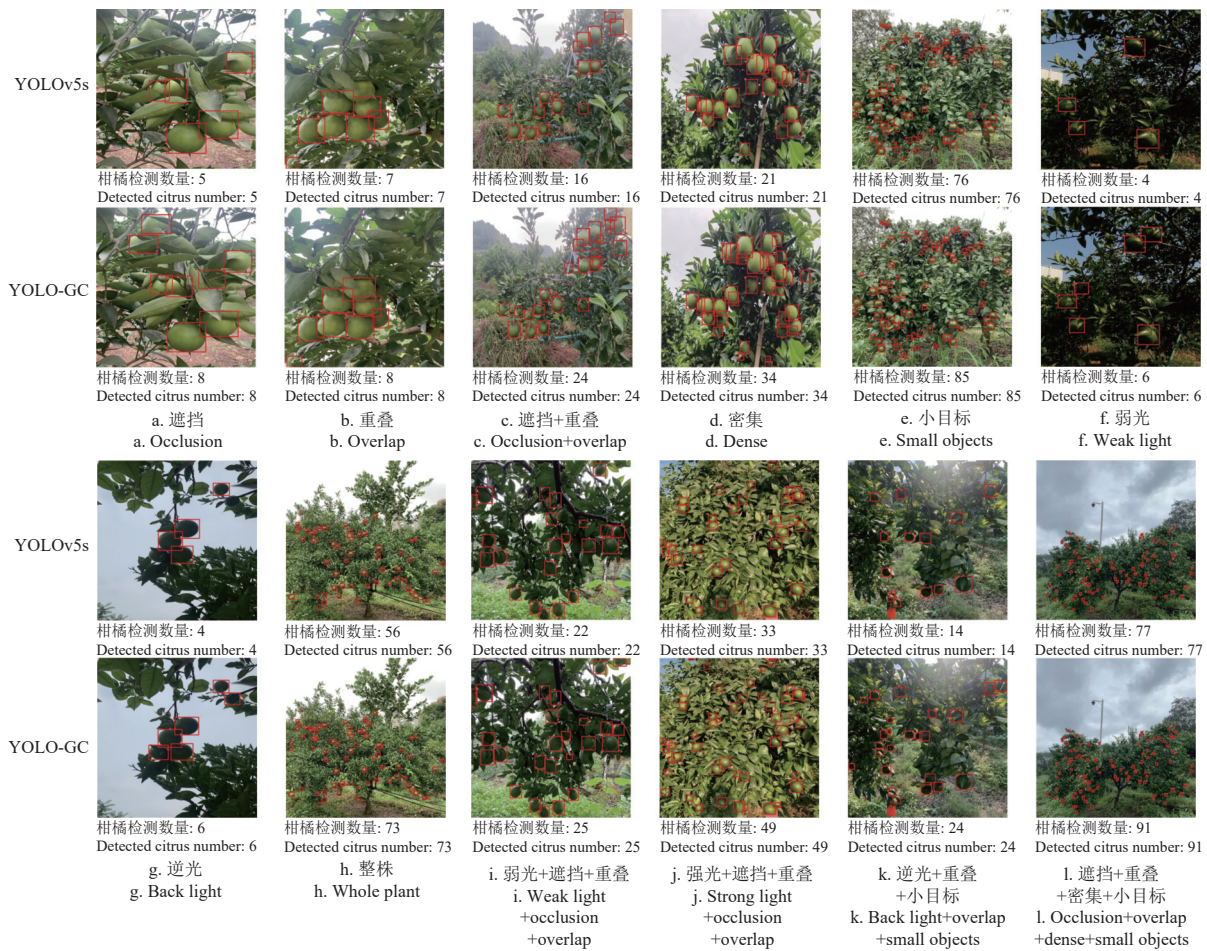


图 5 YOLOv5s 与 YOLO-GC 在不同环境下检测的柑橘数量
Fig.5 Number of citrus detected by YOLOv5s and YOLO-GC in different environment

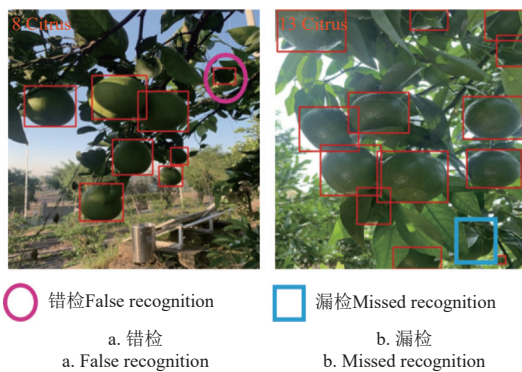


图 6 错检和漏检实例
Fig.6 False and missed cases

表 4 移动端检测性能对比

模型 Models	准确率 Accurate rate/%	漏检率 Missed rate%	误检数 False count	推理时长 Inference time/ms
YOLOv5s	94.8	5.2	36	763
YOLO-GC	97.2	2.8	12	108

YOLO-GC 算法模型在安卓移动端 APP 检测的功能界面和效果如图 7 所示, 针对小目标、遮挡、重叠等绿色柑橘果实均表现出不错的检测效果。为评估本文提出的 YOLO-GC 模型在安卓移动端 APP 中的检测性能, 本

研究在测试集中随机选取 300 张图片在智能手机上进行了绿色柑橘果实的检测与计数试验。通过对 Python 脚本计算图片中人工标注的实际果实数量与算法模型检测到的果实数量进行回归拟合, 结果如图 8 所示, 决定系数 R^2 为 0.974, 均方根误差 RMSE 为 1.33。



图 7 安卓手机 APP 端检测功能界面与效果
Fig.7 The interface and effect of the detection function on the APP side of the Android phone

拟合结果表明, YOLO-GC 算法在移动手机端 APP 中对绿色柑橘检测与计数具有较高的准确性和可靠性, 能够满足复杂自然环境中绿色柑橘果实检测与计数的需求。

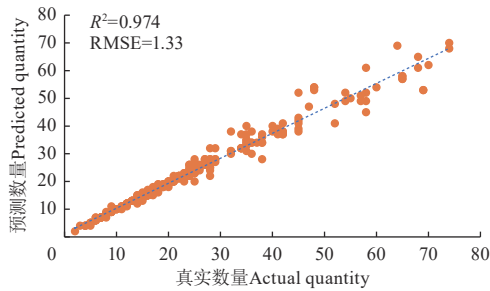


图 8 实际数量与模型预测数量的回归结果

Fig.8 Regression results between predicted and actual quantities

3 结 论

本文针对复杂自然环境中绿色柑橘果实检测模型识别精度不高和较大不利于部署应用的问题,提出了兼顾推理速度与精度的 YOLO-GC 网络模型,得到主要结论如下:

1) 与 YOLOv5s 相比,本文提出的 YOLO-GC 模型精确率、召回率、平均精度 $AP_{0.5}$ 分别提升了 1.1、0.9 和 1.2 个百分点,参数量、计算量、权重模型的内存分别减少了 55.2%、47.5%、53.9%。

2) 与 YOLOv5s、YOLOv7、YOLOX、YOLOv8、CenterNet、Faster R-CNN、RetinaNet 目标检测网络相比, YOLO-GC 模型的平均精度分别提升了 1.2、1.3、1.4、0.9、1.7、4.6、3.9 个百分点,仅占用 6.69 MB 的权重内存,精确率、召回率、平均精度和 F_1 得分分别达到了 97.6%、90.3% 和 97.8% 和 97.0%。

3) 边缘安卓移动端 APP 的绿色柑橘检测试验, YOLO-GC 模型准确率提升 2.4 个百分点,达到 97.2%,平均推理时长减少了 85.8%,仅用时 108 ms。

基于 YOLOv5s 改进的 YOLO-GC 算法模型对于与背景颜色相近、生长环境、以及作物形态类似的作物是通用的,但在不同作物上的表现需要根据作物特点进一步参数微调 and 模型优化,以确保模型在其他特定作物上的准确性和稳定性。本文提出的 YOLO-GC 网络模型为复杂环境中与背景相似的果实的实时检测提供方案,对早期的果实估产、精准水肥管理、挂果量调控提供算法支撑。

【参 考 文 献】

- [1] 陈海兵, 吴群, 徐斌, 等. 柑橘信息系统的探索与实践[J]. 中国果树, 2024(1): 122-127.
CHEN Haibing, WU Qun, XU Bin, et al. Exploration and practice of citrus information system[J]. China Fruits, 2024(1): 122-127. (in Chinese with English abstract)
- [2] LI H, LEE W S, WANG K. Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images[J]. *Precision Agriculture*, 2016, 17(6): 678-697.
- [3] WANG C L, LEE S W, ZOU X J, et al. Detection and counting of immature green citrus fruit based on the Local Binary Patterns (LBP) feature using illumination-normalized images[J]. *Precision Agriculture*, 2018, 19(6): 1062-1083.
- [4] 梁晓婷, 庞琦, 杨一, 等. 基于 YOLOv4 模型剪枝的番茄缺陷在线检测[J]. 农业工程学报, 2022, 38(6): 283-292.
LIANG Xiaoting, PANG Qi, YANG Yi, et al. Online detection of tomato defects based on YOLOv4 model pruning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(6): 283-292. (in Chinese with English abstract)
- [5] ZHAO R N, GUAN Y J, LU Y Q, et al. FCOS-LSC: A novel model for green fruit detection in a complex orchard environment[J]. *Plant Phenomics*, 2023, 5: 0069.
- [6] YOUSSEF O, REED D, KHALID E. Yield estimation and visualization solution for precision agriculture[J]. *Sensors (Basel)*, 2021, 21(19): 6657.
- [7] 刘洁, 李燕, 肖黎明, 等. 基于改进 YOLOv4 模型的橙果识别与定位方法[J]. 农业工程学报, 2022, 38(12): 173-182.
LI Jie, LI Yan, XIAO Liming, et al. Recognition and location method of orange based on improved YOLOv4 model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(12): 173-182. (in Chinese with English abstract)
- [8] HAMZEH M, MOHSEN S, ABBAS A, et al. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions[J]. *Computers and Electronics in Agriculture*, 2021, 191: 106533.
- [9] WANG L L, ZHAO Y J, XIONG Z J, et al. Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model[J]. *Frontiers in Plant Science*, 2022, 13: 965425.
- [10] 宋怀波, 尚钰莹, 何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报, 2023, 54(1): 1-19.
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(1): 1-19. (in Chinese with English abstract)
- [11] SENGUPTA S, LEE S W. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions[J]. *Biosystems Engineering*, 2014, 117: 51-61.
- [12] HE Z L, XIONG J T, CHEN S M, et al. A method of green citrus detection based on a deep bounding box regression forest[J]. *Biosystems Engineering*, 2020, 193: 206-215.
- [13] 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.
- [14] 贾雪莹, 赵春江, 周娟, 等. 基于改进 YOLOv7 模型的柑橘表面缺陷在线检测[J]. 农业工程学报, 2023, 39(23): 145-151.
JIA Xueying, ZHAO Chunjiang, ZHOU Juan, et al. Online detection of citrus surface defects using improved YOLOv7 modeling[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(23): 145-151. (in Chinese with English abstract)
- [15] 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. (in Chinese with English abstract)
- [16] ZHAO J F, DU C F, LI Y, et al. YOLO-Granada: a lightweight attentioned Yolo for pomegranates fruit detection[J]. *Scientific Reports*, 2024, 14(1): 16848.
- [17] ZHOU Z X, SONG Z Z, FU L S, et al. Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation[J]. *Computers and Electronics in Agriculture*, 2020, 179: 105856.
- [18] LYU S L, LI R Y, ZHAO Y W, et al. Green citrus detection and counting in orchards based on YOLOv5-CS and AI edge system[J]. *Sensors (Basel)*, 2022, 22(2): 576.
- [19] ZHENG Z H, XIONG J T, LIN H, et al. A method of green citrus detection in natural environments using a deep convolutional neural network[J]. *Frontiers in Plant Science*, 2021, 12: 705737.
- [20] 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 (Transactions of the CSAE), 2023, 39(14): 191-199. (in Chinese with English abstract)
- [21] XU L J, WANG Y H, SHI X S, et al. Real-time and accurate detection of citrus in complex scenes based on HPL-YOLOv4[J]. *Computers and Electronics in Agriculture*, 2023, 205: 107590.
- [22] WU Y X, CHEN J N, WU S K, et al. An improved YOLOv7 network using RGB-D multi-modal feature fusion for tea shoots detection[J]. *Computers and Electronics in Agriculture*, 2024, 216: 108541.
- [23] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 1580-1589.
- [24] ZHANG H C, WANG Y J, RU L Y. Detection method of wheat spike improved YOLOv5s based on the attention mechanism[J]. *Frontiers in Plant Science*, 2022, 13: 993244.
- [25] XIONG Z J, WANG L L, ZHAO Y J, et al. Precision detection of dense litchi fruit in UAV images based on improved YOLOv5 model[J]. *Remote Sensing*, 2023, 15(16): 4017.
- [26] 孙俊, 钱磊, 朱伟栋, 等. 基于改进 RetinaNet 的果园复杂环境下苹果检测[J]. *农业工程学报*, 2022, 38(15): 314-322.
- SUN Jun, QIAN Lei, ZHU Weidong, et al. Apple detection in complex orchard environment based on improved RetinaNet[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(15): 314-322. (in Chinese with English abstract)
- [27] REZATOFI G H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 658-666.
- [28] MBOUEMBE T L P, LIU G X, PARK S, et al. Accurate and fast detection of tomatoes based on improved YOLOv5s in natural environments[J]. *Frontiers in Plant Science*, 2024, 14: 1292766.
- [29] FU X L, ZHAO S L, WANG C H, et al. Green fruit detection with a small dataset under a similar color background based on the improved YOLOv5-AT[J]. *Foods*, 2024, 13(7): 1060.
- [30] MA J, LU A, CHEN C, et al. YOLOv5-lotus an efficient object detection method for lotus seedpod in a natural environment[J]. *Computers and Electronics in Agriculture*, 2023, 206: 107635.
- [31] ZHANG Y C, ZHANG W B, YU J Y, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107062.
- [32] XIAO B J, MINH N, YAN W Q. Fruit ripeness identification using YOLOv8 model[J]. *Multimedia Tools and Applications*, 2023, 83(9): 28039-28056.
- [33] 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 (Transactions of the CSAE)*, 2021, 37(18): 161-169. (in Chinese with English abstract)

Detecting green citrus fruit in natural scenes using improved YOLOv5s model

LYU Qiang^{1,2}, LIN Gang^{1,2}, JIANG Jie^{1,2}, WANG Mingzhi^{1,2,3}, ZHANG Haoyang^{1,2}, YI Shilai^{1,2}

(1. Citrus Research Institute, Southwest University, Chongqing 400712, China; 2. National Citrus Engineering and Technology Research Center, Chongqing 400712, China; 3. College of Engineering and Technology, Southwest University, Chongqing 400715, China)

Abstract: Accurate detection of immature green citrus fruits in trees is one of the most critical steps for production decisions, such as early yield prediction, precise water and fertilizer management, and regulation of number of fruits hanging. However, the large model has been confined to identifying the immature citrus, due to the similar green citrus to the canopy background. Great challenges have been brought to rapidly and accurately detect the fruits, even to deploy the model. In this study, an improved model, YOLO-GC (you only look once-green citrus) was proposed to detect the green citrus fruits using YOLOv5s. The improved model was also deployed into the edge mobile devices, in order to achieve the real-time and convenient detection of green citrus fruits in trees. Firstly, the original backbone network was replaced with a lightweight GhostNet one, because the YOLOv5s network model was large and difficult to deploy. Meanwhile, the attention was then improved to the green citrus features. Accuracy degradation was reduced after the model was lightweight. Global Attention Mechanism (GAM) was embedded in the backbone network and the feature fusion layer, in order to extract the fruit features in complex environments. Secondly, a BiFPN (Bi-directional Feature Pyramid Network) architecture was introduced into the feature fusion layer for the multi-scale weighted feature fusion, in order to improve the dense and small-targeted fruits. Finally, the GIoU (Generalized Intersection over Union) loss function combined with Soft-NMS (Soft-Non-Maximum Suppression) was used to optimize the bounding box regression, in order to reduce the omission caused by the occlusion and overlapping of fruits and branches. The experimental results showed that the weight model memory of YOLO-GC was reduced by 53.9%, compared with the YOLOv5s. The number of parameters and the average inference time were reduced by 55.2% and 46.2%, respectively, whereas, the average precision (AP_{0.5}) was improved by 1.2 percentage points. There was a lower amount of fruit leakage and misdetection in a variety of complex natural environments. The comprehensive performance of the YOLO-GC model was superior to that of CenterNet and seven commonly used network models, such as the YOLOv5s, YOLOv7, YOLOX, YOLOv8, CenterNet, Faster R-CNN, and RetinaNet target networks. The average accuracy of the YOLO-GC model was improved by 1.2, 1.3, 1.4, 0.9, 1.7, 4.6, and 3.9 percentage points, respectively, only 6.69 MB of weighted memory, thus achieving 97.6%, 90.3%, 97.8%, and 97.0% for the precision, recall, average precision, and F₁ score, respectively. The YOLO-GC model was then deployed to the Android mobile App for testing. The detection accuracy reached 97.2%, which was 2.4 percentage points higher than that of the YOLOv5s. Furthermore, the inference duration (1 038 ms) was reduced by 85.8%. The IC model fully met the requirements of high-accuracy recognition and real-time inference of green citrus on the Android phone side. The finding can provide technical support to detect the green-like fruits in complex environments. The improved model was also deployed in the edge intelligent devices.

Keywords: citrus; image recognition; YOLO; green fruit; lightweight model; mobile deployment