

基于改进 YOLOv8s 的柑橘果实成熟度检测

李莹, 刘梦莲, 何自芬, 娄杨伟

(昆明理工大学机电工程学院, 昆明 650500)

摘要: 为了解决柑橘果实被枝叶和其他果实遮挡以及不同果实之间的成熟情况不同等导致识别果实成熟度困难的问题, 该研究提出一种改进 YOLOv8s 模型的柑橘果实成熟度检测方法。首先, 在网络主干中添加混合注意力变换器模块 (hybrid attention transformer, HAT), 使模型利用更多的像素信息以提高对果实成熟度特征的代表能力; 然后, 将检测头替换为四头的自适应空间特征融合检测头 (four-adaptively spatial feature fusion, FASFF) 实现果实成熟度特征的尺度不变性。改进 YOLOv8s 模型在测试集上的精确度 P 为 94.8%, 召回率 R 为 88.6%, 平均精度均值 $mAP_{0.5}$ 和 $mAP_{0.5\sim0.95}$ 分别为 95.6% 和 88.9%, 模型参数量为 21.51 M, 检测速度为 52.9 帧/s。比 Faster-RCNN、YOLOv3-spp、YOLOv7 和原始 YOLOv8s 模型精确度分别提升 26.0、5.8、4.8 和 5.5 个百分点, 平均精度均值 $mAP_{0.5}$ 分别提升 0.5、1.7、1.6 和 1.6 个百分点。改进 YOLOv8s 模型能够为柑橘采摘机器人的选择性采收提供有效支持。

关键词: 柑橘; 成熟度; YOLOv8s; 混合注意力变换器 (HAT); 自动化采摘

doi: 10.11975/j.issn.1002-6819.202407206

中图分类号: S24; S233.4

文献标志码: A

文章编号: 1002-6819(2024)-24-0157-08

李莹, 刘梦莲, 何自芬, 等. 基于改进 YOLOv8s 的柑橘果实成熟度检测[J]. 农业工程学报, 2024, 40(24): 157-164. doi:

10.11975/j.issn.1002-6819.202407206 <http://www.tcsae.org>

LI Ying, LIU Menglian, HE Zifen, et al. Detecting citrus fruit maturity using improved YOLOv8s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(24): 157-164. (in Chinese with English abstract)

doi: 10.11975/j.issn.1002-6819.202407206 <http://www.tcsae.org>

0 引言

传统柑橘果实采摘主要是依靠人工辨别成熟度进行采摘, 收获采摘约占整个生产作业工作量的 40%^[1], 不仅效率低而且人工成本高, 此外, 人工采摘很容易造成果皮损伤以及对果实成熟度的误判, 不能保证在柑橘的最佳采摘时期对其进行采摘。因此, 有必要使用采摘机器人对柑橘进行智能化采摘, 而研究自然环境下对柑橘果实成熟度具有高识别率的视觉识别系统, 对于实现柑橘的自动化采摘具有重要意义^[2-3]。

目前, 传统的图像处理技术已被广泛应用于苹果^[4]、香蕉^[5]、葡萄^[6]、西瓜^[7]、番茄^[8]等水果的成熟度检测。莫松涛等^[9]构建了一种基于遗传算法与 SVM 的香蕉果实成熟度判别模型, 用手工测量的香蕉菱角特征作为判断成熟度的特征, 最终模型预测的平均准确率为 86.20%。邹伟^[10]将柑橘图像的 RGB 颜色空间转换为 HSV 颜色空间, 根据 H 通道颜色直方图判断柑橘的成熟度等级, 该方法准确度达到 90% 以上。以上通过人手动提取目标的棱角、颜色等特征的方法, 存在试验成本高、模型表达特征能力有限等问题, 难以胜任复杂环境下的果实成熟度检测任务。

深度学习技术因其在处理复杂视觉信息和在目标特

征融合方面具有独特优势而被广泛应用^[11-12], 目前, 其已被应用于蓝莓^[13]、枣^[14]、油茶果实^[15]等水果的成熟度检测。苗荣慧等^[16]针对樱桃番茄相邻成熟度特征差异不明显以及果实之间多存在相互遮挡, 修改原模型主干为 MobileNetV3, 并将全局注意力机制加入特征融合网络中, 改进后的 YOLOv7 模型在测试集上的 mAP 达到 98.2%, 模型大小降低 55.7 M。陈锋军等^[17]基于挂枝状态下无损检测的需求以及国际油橄榄理事会标准制定了油橄榄 6 级成熟度, 改进后的 EfficientDet 模型在测试图像中的 mAP 达到 94.60%, 平均检测时间为 0.337 s。以上研究表明, 深度学习技术在复杂环境下的水果成熟度检测任务中效果显著, 但仍需针对不同的应用对象做出针对性的改进才能取得较优异的检测效果。YOLOv8 模型检测精度高、检测速度快, 已广泛应用于缺陷检测^[18-19]、无人机航拍目标检测^[20-21]等领域。谭厚森等^[22]通过对原始 YOLOv8n 模型进行轻量化操作, 使原模型在 GPU 设备上的检测速度提升了 34.0%, $mAP_{0.5\sim0.95}$ 达到 88.3%, 提高了非结构环境下香梨的识别准确率和检测速度。李旭等^[23]针对自然环境下的辣椒果实分布密集且枝叶遮挡严重导致模型识别率低的问题, 提出一种改进模型 YOLOX_Pepper, 该模型对辣椒的平均检测精度达到 93.30%, 单张图片检测平均用时 0.026 s。上述研究表明, YOLOv8 模型针对非结构化环境且目标密集的场景具有优秀的检测潜力, 所以本文选择 YOLOv8 模型作为该研究的基准模型。

柑橘果实具有如下特点: 1) 柑橘果树枝叶茂盛, 极

收稿日期: 2024-07-23 修订日期: 2024-11-13

基金项目: 国家自然科学基金项目 (62171206)

作者简介: 李莹, 博士, 副教授, 研究方向为智能制造和目标检测。

Email: 7114091@qq.com

易对果实形成遮挡,果实体积也较大,果实之间多存在重叠遮挡;2)柑橘果实在生长过程中由于果树的地理位置、果实在整棵树上的相对位置差异导致接受光照情况不同,果实不会同时成熟,同一棵树上的果实成熟程度也不完全相同;针对上述问题,YOLOv8 模型对柑橘果实成熟度检测效果较差,利用深度学习模型预测柑橘果实成熟度的研究较少。本文以不同成熟情况的柑橘果实为研究对象,通过针对性地改进 YOLOv8s 模型结构,重点提升模型对柑橘果实成熟度的检测精度,同时保持检测精度与检测速度之间的平衡,使其能够胜任低算力设备的实时检测任务,为柑橘果实的自动化采摘提供理论依据。

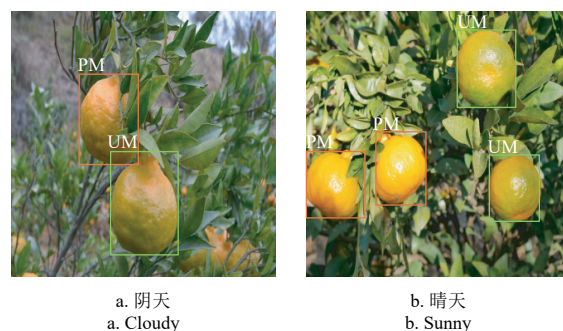
1 试验数据

1.1 图像采集

本次研究所用数据集采集于云南省昆明市晋宁区卧龙山庄,柑橘品种为碰柑。碰柑属芸香科木本植物,果皮呈橙红或橙黄色,因其果皮疏松,内果皮极易分离,所以又被列入宽皮柑橘之中。在柑橘果实果熟期内,使用智能手机红米 Note 12 Turbo 以及松下 GF10 相机分别在阴天和晴天进行拍摄,采集日期为 2023 年 12 月 7 日和 2023 年 12 月 28 日。拍摄时距离柑橘果树 1.0~1.3 m,从果树的东南西北四个方位分别进行拍摄,图像包括顺光、逆光、单目标、多目标、有枝叶遮挡、无枝叶遮挡和重果等。拍摄完成后筛选掉模糊、重复等无效图像,最后得到原始柑橘图像 1 048 幅,为了增加柑橘数据集的多样性,提高模型对不同拍摄角度、不同遮挡程度以及因抖动造成图片模糊等情况下的柑橘果实识别的鲁棒性,本文使用旋转、缩放以及高斯模糊等方式对单张原始图片分别进行数据增强,扩充后的柑橘数据集图像共 2 096 幅,筛选掉部分质量较差的图像,最终得到的图像为 2 085 幅。

1.2 果实成熟度划分

根据农业部标准 NY/T 716-2003^[24] 以及宽皮柑橘产后贮藏物流操作规程 GH/T 1336-2021^[25] 明确规定的柑橘采摘成熟度标准,即果实转色面积转成柑橘固有色泽的比例应大于 2/3 才能进行采摘。本次采集的数据为正常生长且表面均匀着色的柑橘果实,参考杨长辉等^[26] 在研究柑橘果实是否可采摘时对树叶遮挡果实的定义,同时考虑果实远距离销售运输的需求,本次研究将果面成熟度占比缩放至可见视野的一半,即当果实转色面积与整个柑橘果实拟合圆面积之比大于 2/3 时便可进行采摘。因此,本文从柑橘采摘时的实际需求出发,将柑橘按照是否可以采摘划分为两类成熟度:1)未达到可采摘成熟度的柑橘果实 UM (unpickable maturity),即果面转色面积小于等于 2/3;2)达到可采摘成熟度的柑橘果实 PM (pickable maturity),即果面转色面积大于 2/3。数据集中不同天气条件下的柑橘果实图像如图 1 所示。



注: UM 为未达到可采摘成熟度的柑橘果实, PM 为达到可采摘成熟度的柑橘果实。

Note: UM is for citrus fruit that has not reached pickable maturity and PM is for citrus fruit that has reached pickable maturity.

图 1 不同天气条件下的柑橘果实图像

Fig.1 Images of citrus fruits under different weather conditions

1.3 数据集制作

根据划分好的分类标准,采用 LabelImg 软件对柑橘图像数据集进行标注,保存为 YOLO 标注格式的 txt 类型文件。由于柑橘果树枝叶茂盛且果实之间也存在重叠遮挡的情况,如果受遮挡面积过大,则人工难以判断果实成熟情况。因此,在进行数据集标注时遵循以下原则:1)对于果实表面被遮挡面积与整个柑橘果实拟合圆面积之比>50%的柑橘果实,人工难以判断其成熟情况,对这类果实不予标注;2)考虑到机器人采摘时可能由于枝干遮挡柑橘果实造成果实损伤,或机械手、机械臂损坏,故对这类柑橘果实也不进行标注;3)对于受遮挡程度≤50%的果实,人工判断其果实成熟情况后标注。标注完成后,将数据集按照 8:1:1 的比例随机划分为训练集、验证集和测试集,最终训练集 1 668 幅图像,验证集 208 幅图像,测试集 209 幅图像。

2 柑橘果实成熟度检测模型

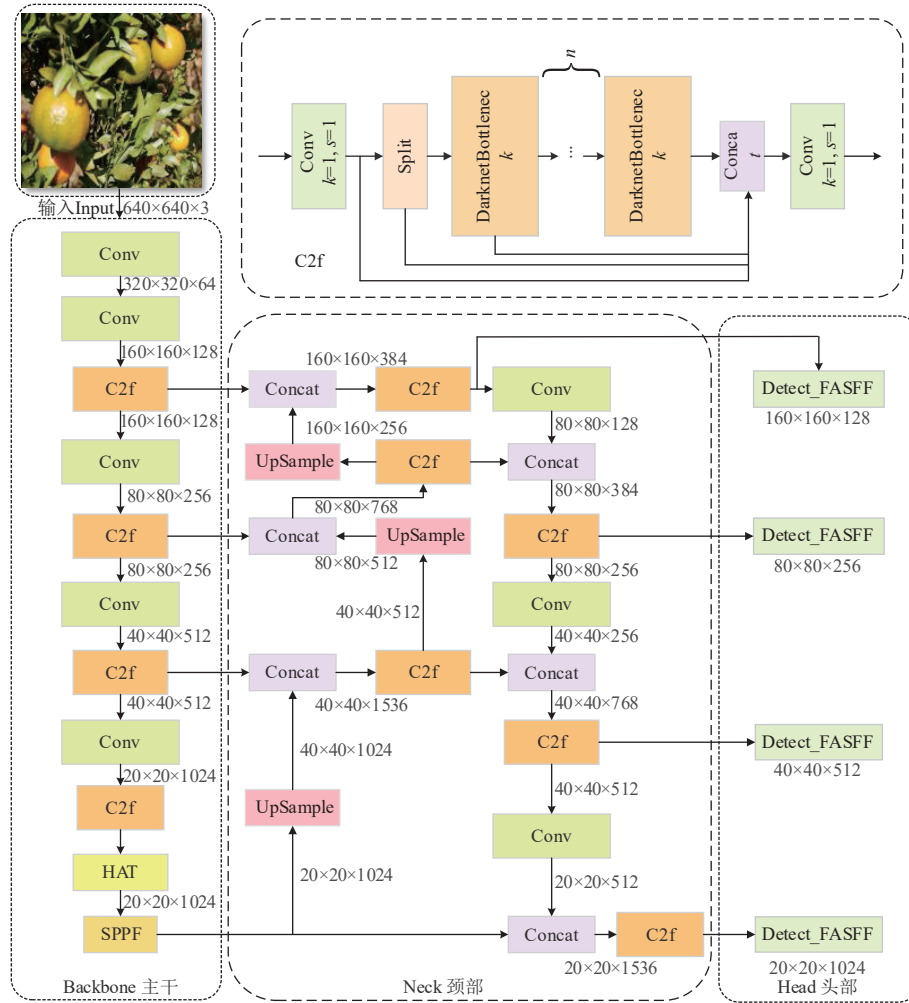
2.1 改进的 YOLOv8s 网络模型

YOLO 系列模型因其检测准确性高、实时性好而得到广泛的应用。与 YOLOv5s 模型相比,YOLOv8s 将 C3 模块替换成了 C2f 模块,实现了进一步的轻量化,同时在损失计算上,YOLOv8s 将 DFL Loss 和 CIoU Loss 作为回归损失,进一步提升了性能。自然环境下的柑橘果实极易受到茂盛的枝叶遮挡,果实之间也多存在重叠遮挡的情况,进而导致区分柑橘果实成熟度的关键特征被遮挡,使得模型识别果实成熟度时容易存在漏检的问题。同时柑橘果实在自然环境中并不是同时成熟,同一棵树上的果实成熟程度也会有差异,模型在准确识别柑橘果实成熟度上较为困难。因此,为了解决自然环境下柑橘果实多受枝叶遮挡以及果实之间重叠遮挡导致模型难以准确识别果实成熟度的问题,本文选择了 YOLOv8s 模型并对其进行了改进。

首先,为了提高模型对果实成熟度的检测能力,在特征提取网络中加入混合注意力变换器模块 (hybrid attention transformer, HAT),它通过在图像中激活更多的像素,使模型能够利用更多的信息捕捉柑橘果实的成

熟度特征, 从而提升模型识别的准确性。其次, 将检测头替换为四头的自适应空间特征融合检测头 (four-adaptively spatial feature fusion, FASFF), FASFF 使模型

学习特征的自适应空间融合权重, 解决柑橘果实不同特征尺度间的不一致性问题, 有助于提高检测的准确性。改进后的 YOLOv8s 网络结构如图 2 所示。



注: Conv 为卷积模块; C2f 为跨阶段部分特征融合模块; HAT 为混合注意力变换器模块; SPPF 为空间金字塔池化层; UpSample 为上采样层; Concat 为特征连接模块; Detect_FASFF 为修改过后的自适应空间特征融合检测头; Split 为特征分层操作; k 表示卷积核大小; s 表示步距; n 表示 DarknetBottleneck 模块的个数。

Note: Conv is the convolution module; C2f is the cross stage partial feature fusion module; HAT is the hybrid attention transformer module; SPPF is the spatial pyramid pooling layer; UpSample is the UpSampling layer; Concat is the feature connection module; Detect_FASFF is the modified adaptive spatial feature fusion detection head; Split is the feature layering operation; k denotes convolution kernel size; s denotes stride; n denotes the number of DarknetBottleneck modules.

图 2 改进后的 YOLOv8s 网络结构图
Fig.2 Structural diagram of the improved YOLOv8s model

2.1.1 HAT 模块

高分辨率和低分辨率图像的区别在于单位面积内的像素量的多少, 前者更加清晰且图像包含的细节更多。HAT 模块^[27]通过融合通道注意力和自注意力机制来提升单图像超分辨率 (single image super-resolution, SR) 重建的性能。自然环境下的柑橘果实会由于风吹或者采摘时机械臂的影响存在抖动, 造成图像质量模糊的情况。HAT 模块通过将这些低质量的柑橘果实图像重建为高分辨率图像, 让模型能够利用更多的信息捕捉果实的成熟度特征, 从而提高检测的性能。

HAT 模块在对图像进行深层特征提取时, 残差混合注意力组 (residual hybrid attention group, RHAG) 发挥了主要作用。RHAG 由若干个混合注意力模块 (hybrid

attention block, HAB)、一个重叠交叉注意力模块 (overlapping cross-attention block, OCAB) 和一个带残差连接的 3×3 卷积层组成。其中, 混合注意力模块 HAB 的结构如图 3 所示, 在第一个 LayerNorm 层之后, 通过在标准的 Swin Transformer 模块中插入基于通道注意力的卷积块 (channel attention block, CAB) 增强网络的特征表示能力, 它与基于窗口的多头自注意力模块 (S)W-MSA 并行。采用通道注意力时, 因为计算通道注意力权重会涉及到图像的全局信息, 所以柑橘果实图像中会有更多的像素被激活, 因此网络能利用这些信息来更细致地表示果实成熟度特征。此外, W-MSA 模块采用较大的窗口尺寸计算自注意力, 扩大了像素的使用范围, 从而模型能够使用更多的输入像素来构建果实的成熟度特征。

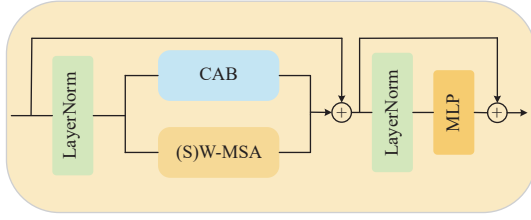
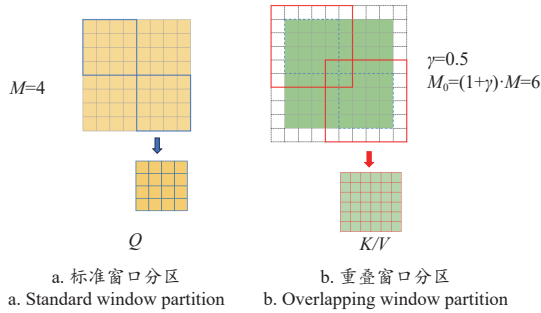


图3 混合注意力模块
Fig.3 Hybrid attention block (HAB)

OCAB 模块用于建立跨窗口连接, 增强基于窗口自适应注意力的表征能力, 可用于弥补标准 Swin Transformer 模型中移动窗口机制无法完美实现跨窗口信息交互的不足, 重叠交叉注意力层 (overlapping cross-attention, OCA) 的重叠窗口分区如图4所示。



注: M 表示标准窗口分区的窗口大小, M_0 表示重叠窗口分区的窗口大小, γ 表示控制重叠大小的常数, Q 表示查询, K 和 V 表示键和值。
Note: M denotes the window size of the standard window partition, M_0 denotes the window size of the overlapping window partition, γ denotes the constant that controls the overlap size, Q denotes the query, K and V denote the key and value.

图4 OCA 模块的重叠窗口分区

Fig.4 Overlapping window partitions for OCA (overlapping cross-attention) module

标准窗口分区可视为一个滑动分区, 其卷积核大小和步长均为窗口大小 M ($M=4$), 重叠窗口分区采用与标准窗口分区不同的窗口大小 M_0 来划分投影特征。重叠窗口分区的卷积核大小为 M_0 , 步长为 M , 使用大小为 $\frac{\gamma M}{2}$ ($\gamma=0.5$) 的填充来确保重叠窗口大小的一致性。由图4可知, OCA 从更大的窗口特征中计算 K/V (键/值), 从而可以利用更多有用的信息增强跨窗口目标信息的交互, 使模型获得关于柑橘果实成熟度的中间特征更加清晰。

2.1.2 FASFF 检测头

自适应空间特征融合策略由 LIU 等^[28] 提出, 用于解决不同特征尺度间的不一致性问题。本次拍摄的柑橘果实由于远近差异而存在大小不同, 但柑橘果实特征具有尺度不变性。为了充分利用目标的特性, 提高果实成熟度检测的准确性, 本文将检测头替换为了四头的自适应空间特征融合检测头, 其原理如下:

对于某一层次的特征, 首先将其他层次的特征整合并调整到相同的分辨率, 然后进行训练以找到自适应空间融合权重, 最后不同层次的特征被自适应地融合在一起。有些特征可能因为在这个空间位置上携带相互矛盾的信息而被过滤, 有些特征可能因更具区别性而占据主

导地位。

FASFF 检测头对第 l 层的特征融合如下:

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} + \delta_{ij}^l \cdot x_{ij}^{4 \rightarrow l} \quad (1)$$

式中 $x_{ij}^{l \rightarrow l}$ 表示在特征图上位置 (i, j) 处的特征向量从第 l 层调整到第 l 层, 依此类推; α_{ij}^l 、 β_{ij}^l 、 γ_{ij}^l 和 δ_{ij}^l 是指网络自适应学习到的第 l 层 4 个不同层次特征图的空间重要性权重; y_{ij}^l 表示在第 l 层上生成最终用于预测的融合特征。FASFF 检测头通过在每个空间位置上过滤冲突信息以压制不同尺度特征间的不一致性, 学习自适应空间融合权重, 从而将不同层次的柑橘果实特征自适应地融合在一起, 提高了特征的尺度不变性。

2.2 试验环境配置与网络参数设置

试验用计算机处理器型号为 AMD Ryzen 55500, 运行内存为 16 GB, 显卡型号为 NVIDIA GeForce RTX 3060, 显存为 12 GB; 运行系统为 window10 操作系统, 深度学习框架为 Pytorch1.12.1, 开发环境为 python3.9.7, 运行架构为 CUDA11.3。通过试验发现, 当 epoch 设置为 200 时, 模型在验证集上的损失曲线趋于平稳, 平均精度均值 $mAP_{0.5}$ 和 $mAP_{0.5 \sim 0.95}$ 的值基本保持不变, 故将训练轮数设置为 200。每批次输入图片数目为 16, 输入图片分辨率为 640×640 像素, 初始学习率设置为 0.01, 权重衰退系数为 0.0005, 模型训练时使用 SGD 优化器进行迭代优化。

2.3 评价指标

对自然环境下的柑橘果实成熟度进行检测需要同时考虑模型检测精度和实际应用能力。为了评价模型对柑橘果实成熟度检测的性能, 本文选取精确度 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、参数量 (parameters) 和每秒传输帧数 FPS (frames per second) 对本文所提模型进行有效性评价。

利用取不同置信度值下的 Precision 和 Recall 的值进行组合, 可以画出 P-R 曲线, 曲线下方的面积, 即为某一类别的平均精度 (average precision, AP)。mAP 则用于综合衡量模型在所有类别上的检测效果。IoU (intersection over union) 用于衡量模型生成的预测框与真实框之间的重叠程度, $mAP_{0.5}$ 表示在 IoU 阈值等于 0.5 时的平均精度均值, $mAP_{0.5 \sim 0.95}$ 计算 IoU 阈值从 0.5 到 0.95 (步长为 0.05) 范围内的 mAP 的平均值。

3 结果与分析

3.1 改进 YOLOv8s 模型消融试验结果

为了进一步验证改进 YOLOv8s 模型的检测性能, 建立了 4 组网络进行消融试验, 进一步验证每个改进模块相对原始 YOLOv8s 模型的改进效果。从测试集中随机挑选 4 幅图像进行测试, 网络测试结果如图5所示。

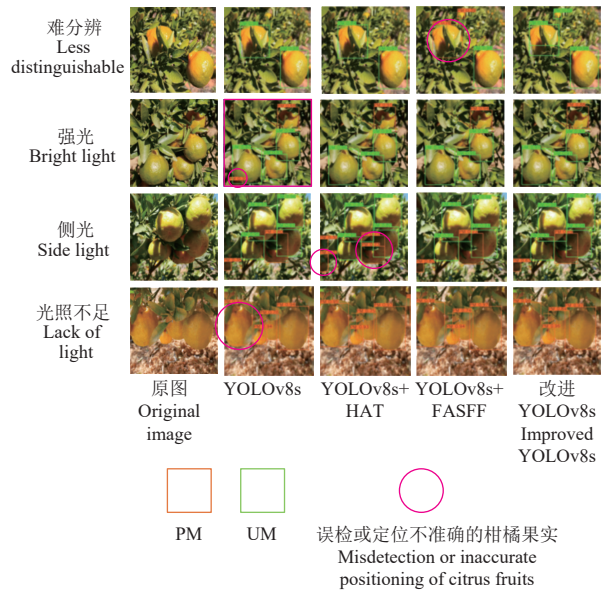


图 5 不同光照条件下 YOLOv8s 及其改进模型对柑橘果实的检测

Fig.5 Detection of citrus fruits by YOLOv8s and its improved model under different light conditions

由图 5 可知，原始 YOLOv8s 模型对柑橘果实的特征提取能力不足，如在强光中，将左下角的 1 个非目标类识别为待检测目标，同时在光照不足中，对最左边的柑橘果实的定位不够准确。原始 YOLOv8s 模型中添加 HAT 模块之后，使模型利用更多的像素信息以更加清晰地表示柑橘果实的成熟度特征，提高了识别的准确性，未出现原始 YOLOv8s 模型在强光中的误检问题，但对目标的定位精度较差，如在侧光中，未能准确定位左边和右下方的 2 个可采摘成熟度的柑橘果实 PM。与原始 YOLOv8s 模型相比，将检测头替换为 FASFF 检测头之后，模型通过学习不同空间位置的自适应空间融合权重，提高了对受遮挡程度较大的果实的定位准确性和特征提取能力，未出现原模型在强光中所示的误检问题，对侧光中两个可采摘成熟度的柑橘果实 PM 定位也更加准确。但单独添加 FASFF 检测头的模型仍然存在误检的情况，如在难分辨图像中，将左上方的 1 个不可采摘柑橘 UM 错误识别为可采摘目标 PM。由图 5 可知，改进 YOLOv8s 模型不仅对受遮挡果实的定位更准确，同时还提高了对柑橘果实成熟度判别的准确性。

为了定量分析 4 组网络的检测性能，使用测试集中的 209 幅柑橘果实图像进行测试，消融试验对比结果如表 1 所示。

经测试，原始 YOLOv8s 模型检测的精确度 P 和平均精度均值 $mAP_{0.5}$ 分别为 89.3% 和 94.0%，模型参数量为 11.13 M，在主干添加 HAT 模块后，与原模型相比召回率 R 和 $mAP_{0.5}$ 分别提升 1.6 和 1.0 个百分点，说明 HAT 通过激活图像中更多的像素有效地整合了全局的像素信息，使模型不仅能更细致地表示被遮挡柑橘果实的成熟度特征，同时减轻了枝叶遮挡以及果实重叠遮挡导致的漏检问题；将原始的 YOLOv8s 检测头替换为

FASFF 检测头时，模型在精确度 P 和 $mAP_{0.5\sim0.95}$ 上比原始模型分别提高了 4.0 和 2.2 个百分点，说明 FASFF 的自适应空间特征融合结构使模型能更广泛地利用不同空间位置的有用信息，从而提高了对不同尺度的柑橘果实检测的准确性，而且模型的参数量仅增加了 5.01 M，有利于模型部署于移动端。综上所述，本文所提出的两个改进点均能够提高原始 YOLOv8s 模型的检测性能。当同时引入混合注意力变换器 HAT 模块和自适应空间特征融合检测头 FASFF 时，改进 YOLOv8s 模型相较于原模型在精确度 P 、平均精度均值 $mAP_{0.5}$ 以及 $mAP_{0.5\sim0.95}$ 上分别提升 5.5、1.6 和 3.0 个百分点，从而使原模型的检测精度有较大的提升，而模型的每秒传输帧数 FPS 为 52.9 帧/s，满足柑橘自动采摘时实时检测的需求。

表 1 改进 YOLOv8s 模型消融试验结果
Table 1 Results of ablation tests with improved YOLOv8s model

HAT	FASFF	精确度		召回率		$mAP_{0.5}/\%$	$mAP_{0.5\sim0.95}/\%$	参数量 Parameters/ M	每秒传输 帧数 Frames per second FPS/ (帧·s ⁻¹)
		Precision P/%	Recall R/%	Precision P/%	Recall R/%				
×	×	89.3	89.0	94.0	85.9	11.13	112.5		
√	×	90.9	90.6	95.0	87.5	16.49	83.8		
×	√	93.3	89.4	95.1	88.1	16.14	59.0		
√	√	94.8	88.6	95.6	88.9	21.51	52.9		

注：“×”表示未使用此改进策略，“√”表示使用此改进策略； $mAP_{0.5}$ 表示取 IoU 阈值为 0.5 时，模型在类别 PM 和 UM 上的平均精度均值， $mAP_{0.5\sim0.95}$ 表示 IoU 阈值取 0.5 到 0.95 时（步长为 0.05），分别计算每个 IoU 阈值下的 mAP 值，然后再取平均。
Note: “×” indicates that this improvement strategy is not used and “√” indicates that it is used; The $mAP_{0.5}$ indicates the mean average precision of the model over the categories PM and UM when the IoU threshold is taken to be 0.5, and $mAP_{0.5\sim0.95}$ indicates that when the IoU threshold is taken to be from 0.5 to 0.95 (in steps of 0.05), the mAP value is calculated separately for each IoU threshold and then averaged.

3.2 改进 YOLOv8s 模型对比试验结果

为定性评价改进 YOLOv8s 模型的检测效果，分别使用 YOLOv3-spp^[29]、YOLOv7^[30]、Faster-RCNN^[31]、YOLOv8s 和改进的 YOLOv8s 模型对测试集中的部分柑橘图像进行检测，检测结果如图 6 所示。

由图 6 可知，改进 YOLOv8s 模型的检测效果明显优于其他检测模型。YOLOv3-spp 模型将远景中最下方可采摘成熟度的柑橘果实 PM 错误识别为不可采摘类 UM，且在强光中容易将 3 个遮挡情况较严重的果实识别为待检测目标，这些说明模型对目标的特征提取能力不足，导致模型易出现误检的情况。YOLOv7 模型情况同 YOLOv3-spp 相似，如强光 3 个柑橘果实由于受遮挡程度较大而不在待识别目标范围之内，模型将其检测出来，密集遮挡中左上角的叶片由于反光模型将其检测为 UM，表明 YOLOv7 对柑橘果实的特征提取能力较弱，容易将非目标检测为目标。Faster-RCNN 虽然能将待检测柑橘果实的成熟度类别较为准确地判断出来，但果实预测框的定位精度较差，如远景最下方的 1 个柑橘果实以及强光最右方的 1 个柑橘果实。YOLOv8s 模型虽然在预测框的定位精度方面比 Faster-RCNN 效果要好，但是模型对区分柑橘果实成熟度的关键特征提取不充分，容易将柑

橘果实成熟度误判, 如将图 6a 中最下方的 1 个柑橘果实 实 PM 错误识别为 UM。

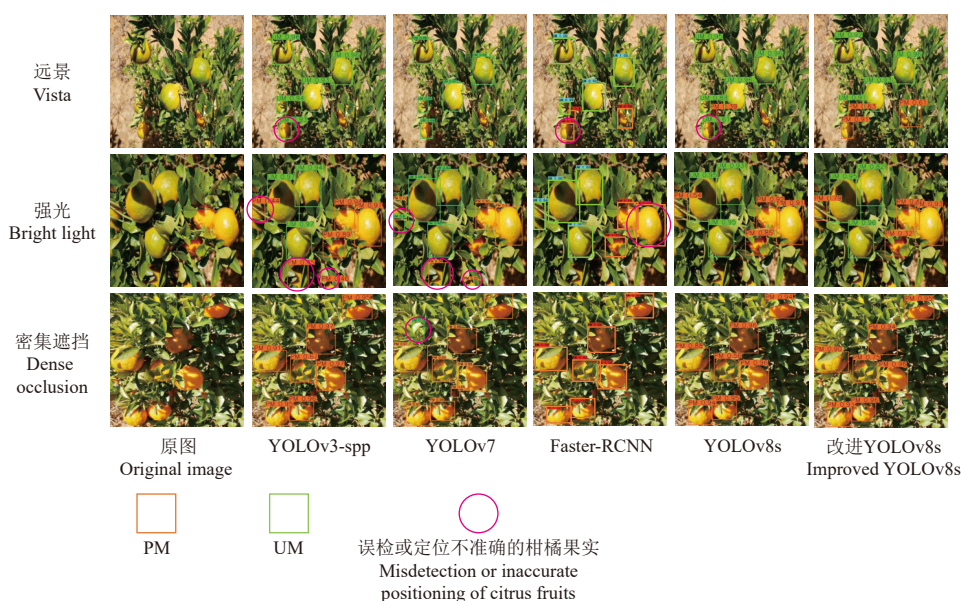


图 6 不同模型检测结果对比

Fig.6 Comparison of detection results of different models

为定量评价改进 YOLOv8s 模型的检测性能, 在相同实验条件下, 分别对上述 5 种模型进行训练, 使用测试集中的 209 幅柑橘图像对训练后的模型进行性能测试, 评估结果如表 2 所示。

表 2 不同检测模型在测试集中的对比试验结果

Table 2 Comparison test results of different detection models in the test set

模型 Model	P/%	R/%	mAP _{0.5} /%	mAP _{0.5~0.95} /%	参数量 Parameters/ M	FPS/ (帧·s ⁻¹)
Faster-RCNN	68.8	96.2	95.1	62.5	136.71	22.7
YOLOv3-spp	89.0	90.4	93.9	87.0	104.71	23.5
YOLOv7	90.0	87.0	94.0	84.4	36.49	41.8
YOLOv8s	89.3	89.0	94.0	85.9	11.13	112.5
改进 YOLOv8s						
Improved YOLOv8s	94.8	88.6	95.6	88.9	21.51	52.9

由表 2 可知, 改进 YOLOv8s 模型在精确度、平均精度均值 mAP_{0.5} 以及 mAP_{0.5~0.95} 上, 均优于其他 4 种模型, 分别达到 94.8%、95.6% 和 88.9%。其中精确度上比 Faster-RCNN、YOLOv3-spp、YOLOv7 以及原始 YOLOv8s 模型分别提高 26.0、5.8、4.8 和 5.5 个百分点, 平均精度均值 mAP_{0.5} 分别提升 0.5、1.7、1.6 和 1.6 个百分点, 平均精度均值 mAP_{0.5~0.95} 分别提升 26.4、1.9、4.5 和 3.0 个百分点。改进 YOLOv8s 模型的召回率为 88.6%, 比原始 YOLOv8s 模型降低 0.4 个百分点, 比 YOLOv7 模型高 1.6 个百分点。

在模型参数量上, 通过增加 HAT 模块加强了模型利用更多像素信息来表示特征的能力, 利用 FASFF 检测头提高了柑橘果实成熟度特征的尺度不变性, 因此改进 YOLOv8s 模型的参数量增至 21.51 M, 虽然比原始 YOLOv8s 模型多 10.38 M, 但检测的精确度、平均精度均值 mAP_{0.5} 和 mAP_{0.5~0.95} 比原始 YOLOv8s 模型分别提

高 5.5、1.6 和 3.0 个百分点。在模型检测速度方面, 改进 YOLOv8s 模型每秒传输帧数为 52.9 帧/s, 比 Faster-RCNN、YOLOv3-spp、YOLOv7 模型分别高 30.2、29.4、11.1 帧/s, 满足农业自动化采摘场景下的实时检测需求。

通过定性和定量分析, 原始 YOLOv8s 模型的主干网络中添加混合注意力变换器模块 HAT, 使模型在对柑橘果实进行特征提取时扩大了像素使用范围, 能利用更多的信息以更细致地表示果实的成熟度特征, 从而使模型检测性能得到提升。将原始 YOLOv8s 模型的检测头替换为四头的自适应空间特征融合检测头 FASFF, 通过在每个空间位置上抑制不同尺度特征间的不一致性, 只保留有用信息用于组合, 有效提高了果实成熟度特征的尺度不变性。因此, 改进后的 YOLOv8s 模型可以对自然环境下的柑橘果实成熟度进行快速且准确的检测。

4 结 论

本文针对自然环境下的柑橘果实易受树枝叶遮挡、果实之间的相互遮挡以及不同果实之间存在成熟快慢不同等情况, 导致现有模型对柑橘果实成熟度识别率较低的问题, 提出一种基于改进 YOLOv8s 模型的柑橘果实成熟度检测方法。在网络主干中添加 HAT 模块以及替换原始 YOLOv8s 模型检测头为 FASFF 检测头。

与 Faster-RCNN、YOLOv3-spp、YOLOv7 和 YOLOv8s 等目前主流的目标检测模型相比, 改进 YOLOv8s 模型对柑橘果实成熟度检测的综合效果最优, 在精确度上的优势最为明显, 比 Faster-RCNN、YOLOv3-spp、YOLOv7 以及原始 YOLOv8s 分别提高 26.0、5.8、4.8 和 5.5 个百分点, 在平均精度均值 mAP_{0.5} 上分别提升 0.5、1.7、1.6 和 1.6 个百分点。改进 YOLOv8s 模型参数量为 21.51 M, 模型每秒传输帧数 FPS 为 52.9 帧/s, 可为柑橘采摘机器人在

自然环境下自动准确地识别果实成熟度提供有效支持。

需要指出的是, 本方法对柑橘果实成熟度检测仍然存在漏检、误检的情况, 后续可以尝试改进传统的非极大值抑制算法 NMS (non-maximum suppression, NMS) 以避免重叠果实由于相互遮挡而造成的漏检问题。同时, 可以通过模型剪枝和轻量化操作来进一步减少模型参数量, 提高检测速度, 使其易于部署在移动端。此外, 本文所研究的柑橘品种为碰柑, 仅为宽皮柑橘中的一种, 后续还可以将研究对象拓展到其他的宽皮柑橘品种。

[参 考 文 献]

- [1] 李会宾, 史云. 果园采摘机器人研究综述[J]. 中国农业信息, 2019, 31(6): 1-9.
LI Huibin, SHI Yun. Review on orchard harvesting robots[J]. China Agricultural Informatics, 2019, 31(6): 1-9. (in Chinese with English abstract)
- [2] 王勇, 陶兆胜, 石鑫宇, 等. 基于改进 YOLOv5s 的不同成熟度苹果目标检测方法[J]. 南京农业大学学报, 2024, 47(3): 602-611.
WANG Yong, TAO Zhaosheng, SHI Xinyu, et al. Apple target detection method with different ripeness based on improved YOLOv5s[J]. Journal of Nanjing Agricultural University, 2024, 47(3): 602-611. (in Chinese with English abstract)
- [3] 周剑, 徐中贵, 谢知音. 基于空间金字塔池化的 YOLOv3 的柑橘多分级研究[J]. 现代农业装备, 2023, 44(2): 35-43, 97.
ZHOU Jian, XU Zhonggui, XIE Zhiyin. Citrus multigrading study based on spatial pyramid pooling with YOLOv3[J]. Modern Agricultural Equipment, 2023, 44(2): 35-43, 97. (in Chinese with English abstract)
- [4] MULYANI E D S, SUSANTO J P. Classification of maturity level of fuji apple fruit with fuzzy logic method[C]//2017 5th International Conference on Cyber and IT Service Management(CITSM). Denpasar, Indonesia: IEEE, 2017: 1-4.
- [5] MAZEN F M A, NASHAT A A. Ripeness classification of bananas using an artificial neural network[J]. Arabian Journal for Science and Engineering, 2019, 44: 6901-6910.
- [6] FERRARA G, MARCOTULI V, DIDONNA A, et al. Ripeness prediction in table grape cultivars by using a portable NIR device[J]. Horticulturae, 2022, 8(7): 613-613.
- [7] 邹小波, 张俊俊, 黄晓玮, 等. 基于音频和近红外光谱融合技术的西瓜成熟度判别[J]. 农业工程学报, 2019, 35(9): 301-307.
ZOU Xiaobo, ZHANG Junjun, HUANG Xiaowei, et al. Distinguishing watermelon maturity based on acoustic characteristics and near infrared spectroscopy fusion technology[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(9): 301-307. (in Chinese with English abstract)
- [8] 伍蕊芮, 张志勇, 韩小平, 等. 基于图像处理技术的番茄成熟度检测研究[J]. 农业技术与装备, 2021(4): 50-51, 54.
WU Yingrui, ZHANG Zhiyong, HAN Xiaoping, et al. Research on tomato ripening detection based on image processing technology[J]. Agricultural Technology and Equipment, 2021(4): 50-51, 54. (in Chinese with English abstract)
- [9] 莫松涛, 董涛, 赵汐璇, 等. 基于遗传算法与 SVM 的香蕉果实成熟度判别模型[J]. 果树学报, 2022, 39(12): 2418-2427.
MO Songtao, DONG Tao, ZHAO Xixuan, et al. Discriminant model of banana fruit maturity based on genetic algorithm and SVM[J]. Journal of Fruit Science, 2022, 39(12): 2418-2427. (in Chinese with English abstract)
- [10] 邹伟. 基于机器视觉技术的柑橘果实成熟度分选研究[J]. 农业与技术, 2023, 43(17): 41-44.
- [11] 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.
XV Yanwei, LI Jun, DONG Yuanfang. Survey of development of YOLO object detection algorithms[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(9): 2221-2238. (in Chinese with English abstract)
- [12] LI Y, FENG Q, LI T, et al. Advance of target visual information acquisition technology for fresh fruit robotic harvesting: A review[J]. Agronomy-Basel, 2022, 12(6): 1336.
- [13] 田有文, 覃上声, 闫玉博, 等. 基于改进 YOLOv8 的田间复杂环境下蓝莓成熟度检测[J]. 农业工程学报, 2024, 40(16): 153-162.
TIAN Youwen, QIN Shangsheng, YAN Yubo. Detecting blueberry maturity under complex field conditions using improved YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(16): 153-162. (in Chinese with English abstract)
- [14] 王菁, 范晓飞, 赵智慧, 等. 基于 YOLO 算法的不同品种枣自然环境下成熟度识别[J]. 中国农机化学报, 2022, 43(11): 165-171.
WANG Jing, FAN Xiaofei, ZHAO Zhihui, et al. Maturity identification of different jujube varieties under natural environment based on YOLO algorithm[J]. Journal of Chinese Agricultural Mechanization, 2022, 43(11): 165-171. (in Chinese with English abstract)
- [15] 陈锋军, 陈闯, 朱学岩, 等. 基于改进 YOLOv7 的油茶果实成熟度检测[J]. 农业工程学报, 2024, 40(5): 177-186.
CHEN Fengjun, CHEN Chuang, ZHU Xueyan. Detection of Camellia oleifera fruit maturity based on improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(5): 177-186. (in Chinese with English abstract)
- [16] 苗荣慧, 李志伟, 武锦龙. 基于改进 YOLOv7 的轻量化樱桃番茄成熟度检测方法[J]. 农业机械学报, 2023, 54(10): 225-233.
MIAO Ronghui, LI Zhiwei, WU Jinlong. Lightweight maturity detection of cherry tomato based on improved YOLO v7 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 225-233. (in Chinese with English abstract)
- [17] 陈锋军, 张新伟, 朱学岩, 等. 基于改进 EfficientDet 的油橄榄果实成熟度检测[J]. 农业工程学报, 2022, 38(13): 158-166.
CHEN Fengjun, ZHANG Xinwei, ZHU Xueyan, et al. Detection of the olive fruit maturity based on improved EfficientDet[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(13): 158-166. (in Chinese with English abstract)
- [18] 侯宪庆, 黎远松, 石睿, 等. 基于改进 YOLOv8s 的铝型材表面缺陷检测[J/OL]. (2024-06-17)[2024-11-25]. <https://link.cnki.net/urlid/50.1155.n.20240614.1739.002>.
HOU Xianqing, LI Yuansong, SHI Rui, et al. Aluminum profile surface defect detection based on improved YOLOv8s[J/OL]. (2024-06-17)[2024-11-25]. <https://link.cnki.net/urlid/50.1155.n.20240614.1739.002>.
- [19] 罗亮, 郎霄, 祖国庆, 等. 一种基于改进 YOLOv8 的气缸套缺陷检测方法[J/OL]. (2024-06-12)[2024-11-25]. <https://link.cnki.net/urlid/42.1294.th.20240611.1616.002>.
LUO Liang, LANG Xiao, ZU Guoqing, et al. A cylinder liner defect detection method based on improved YOLOv8[J/OL]. (2024-06-12)[2024-11-25]. <https://link.cnki.net/urlid/42.1294.th.20240611.1616.002>.
- [20] 蒋伟, 王万虎, 杨俊杰. AEM-YOLOv8s: 无人机航拍图像

- 的小目标检测[J]. 计算机工程与应用, 2024, 60(17): 191-202.
- JIANG Wei, WANG Wanhua, YANG Junjie. AEM-YOLOv8s: Small target detection algorithm for UAV aerial images[J]. Computer Engineering and Applications, 2024, 60(17): 191-202. (in Chinese with English abstract)
- [21] 沈学利, 王灵超. 基于 YOLOv8n 的无人机航拍目标检测[J]. 计算机系统应用, 2024, 33(7): 139-148.
- SHEN Xueli, WANG Lingchao. UAV aerial photography target detection based on YOLOv8n[J]. Computer Systems & Applications, 2024, 33(7): 139-148. (in Chinese with English abstract)
- [22] 谭厚森, 马文宏, 田原, 等. 基于改进 YOLOv8n 的香梨目标检测方法[J]. 农业工程学报, 2024, 40(11): 178-185.
- TAN Housen, MA Wenhong, TIAN Yuan, et al. Improved YOLOv8n object detection of fragrant pears[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(11): 178-185. (in Chinese with English abstract)
- [23] 李旭, 刘青, 匡敏球, 等. 基于改进 YOLOX 的自然环境下辣椒果实检测方法[J]. 农业工程学报, 2024, 40(21): 1-8.
- LI Xu, LIU Qing, KUANG Minqiu. Detecting chili pepper fruits in a natural environment using improved YOLOX[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(21): 1-8. (in Chinese with English abstract)
- [24] 农业农村部. 柑橘采摘技术规范: NY/T 716-2003[S]. 北京: 中国标准出版社, 2004.
- [25] 全国果品标准化技术委员会贮藏加工分技术委员会. 宽皮柑橘采收贮藏物流操作规程: GH/T 1336-2021[S]. 北京: 中华全国供销合作总社, 2021.
- [26] 杨长辉, 刘艳平, 王毅, 等. 自然环境下柑橘采摘机器人识别定位系统研究[J]. 农业机械学报, 2019, 50(12): 14-22, 72.
- YANG Changhui, LIU Yanping, WANG Yi, et al. Research on recognition and localisation system for citrus picking robot in natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 14-22, 72. (in Chinese with English abstract)
- [27] CHEN X, WANG X, ZHOU J, et al. Activating more pixels in image super-resolution transformer[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023: 22367-22377.
- [28] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-25)[2024-11-25]. <https://www.arxiv.org/abs/1911.09516v2>.
- [29] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2024-11-25]. <https://www.arxiv.org/abs/1804.02767v1>.
- [30] WANG C Y, BOCHKOVSKIY A, MARK LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023: 7464-7475.
- [31] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

Detecting citrus fruit maturity using improved YOLOv8s

LI Ying, LIU Menglian, HE Zifen, LOU Yangwei

(School of Electromechanical Engineering, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Accurate and automatic identification of citrus fruit ripeness can greatly contribute to the fruit flavor and quality during intelligent harvesting. However, citrus fruits in the natural environment can not ripen at the same time, leading to the varying degree of ripeness on the same tree. It is also difficult to identify the maturity of citrus fruits, due to the occlusion by branches and leaves, as well as the overlapping occlusion among fruits. In this study, the national standard was developed for the pickable maturity of citrus fruits, with reference to the national standard of the Ministry of Agriculture, NY/T 716-2003, and the operating procedure of wide-skinned citrus for post-harvest storage and logistics, GH/T 1336-2021, according to the phenotypic characteristics of citrus fruits during the ripening period. The citrus fruits to be picked were classified into two categories: 1) Citrus fruits beyond the harvestable maturity of UM (unpickable maturity, UM); and 2) Citrus fruits reaching pickable maturity, PM (pickable maturity, PM). An improved YOLOv8s model was also proposed to identify the ripeness degree of citrus fruits. Firstly, a hybrid attention transformer (HAT) module was added into the network backbone. The low-quality citrus fruit images were reconstructed into the high-resolution images. More information was utilized to capture the ripeness features of the fruits, thus improving the accuracy of the detection; Secondly, the detection head was replaced with a four-head adaptive spatial feature fusion (FASFF) detection head. Adaptive spatial fusion weights were learned at different spatial locations. Thus, different levels of citrus fruit features were effectively fused together to improve the scale invariance of features. The improved YOLOv8s model was finally tested by 209 citrus fruit test images. The precision P of improved YOLOv8s model on the test set was 94.8%, the recall R was 88.6%, the $mAP_{0.5}$ and $mAP_{0.5-0.95}$ were 95.6% and 88.9%, respectively. Comparing with the Faster-RCNN, YOLOv3-spp, YOLOv7, and original YOLOv8s models, the precision of the improved YOLOv8s model was improved by 26.0, 5.8, 4.8, and 5.5 percentage points, respectively, while the mean average precision $mAP_{0.5}$ was improved by 0.5, 1.7, 1.6, and 1.6 percentage points, respectively, and the $mAP_{0.5-0.95}$ was improved by 26.4, 1.9, 4.5, and 3.0 percentage points. In terms of the number of parameters of the model, the improved YOLOv8s model was 21.51 M, although it was 10.38 M higher than that of the original YOLOv8s. The precision P , $mAP_{0.5}$ and $mAP_{0.5-0.95}$, were higher than those of the original YOLOv8s by 5.5, 1.6, and 3.0 percentage points, respectively. The number of transmitted frames per second of the model was 52.9, which was fully met the real-time detection requirements of the mobile edge device. The improved YOLOv8s model shared the high detection accuracy and real-time detection speed. Therefore, the improved YOLOv8s model can be expected to detect the ripeness of citrus fruits in the natural environment. The finding can provide the effective support to the selective harvesting of automatic picking robots.

Keywords: citrus; ripeness; YOLOv8s; hybrid attention transformer (HAT); automated picking