

基于改进 YOLOv7 的苹果生长状态及姿态识别

陈 青^{1,2}, 殷程凯², 郭自良², 吴玄博², 王金鹏^{1,2}, 周宏平^{1,2}

(1. 南京林业大学林业资源高效加工利用协同创新中心, 南京 210037; 2. 南京林业大学机械电子工程学院, 南京 210037)

摘 要: 针对目前苹果在复杂环境下难以进行生长状态分类识别、姿态信息同步获取等问题, 该研究提出了一种基于改进 YOLOv7 的苹果生长状态分类和果实姿态融合识别方法。首先改进多尺度特征融合网络, 在骨干网络中增加 160×160 的特征尺度层, 用于增强模型对微小局部特征的识别敏感度; 其次引入注意力机制 CBAM (convolutional block attention module), 改善网络对输入图片的感兴趣目标区域的关注度; 最后采用 Soft-NMS 算法, 能够有效避免高密度重叠目标被一次抑制从而发生漏检现象。此外, 结合 UNet 分割网络和最小外接圆及矩形特征获取未遮挡苹果姿态。试验结果表明, 改进 YOLOv7 的识别精确率、召回率和平均识别精度分别为 86.9%、80.5% 和 87.1%, 相比原始 YOLOv7 模型分别提高了 4.2、2.2 和 3.7 个百分点, 与 Faster RCNN、YOLOv5s、YOLOv5m 相比, 检测平均准确率分别提升了 18.9、7.2 和 5.9 个百分点, 另外苹果姿态检测方法的准确率为 94%。该文模型能够实现苹果生长状态分类及果实姿态识别, 可为末端执行器提供了抓取方向, 以期苹果无损高效的采摘奠定基础。

关键词: 图像处理; YOLOv7; 分类识别; 姿态识别; 深度学习; 苹果

doi: 10.11975/j.issn.1002-6819.202311080

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2024)-06-0258-09

陈青, 殷程凯, 郭自良, 等. 基于改进 YOLOv7 的苹果生长状态及姿态识别[J]. 农业工程学报, 2024, 40(6): 258-266.

doi: 10.11975/j.issn.1002-6819.202311080 http://www.tcsae.org

CHEN Qing, YIN Chengkai, GUO Ziliang, et al. Apple growth status and posture recognition using improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(6): 258-266. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202311080 http://www.tcsae.org

0 引 言

中国苹果产量居世界首位, 2022 年达到 4 700 万 t, 并呈现逐年上升的趋势。由于劳动力资源日趋紧张, 采用苹果采摘机器人替代人工进行果实采摘是农业发展的必然趋势。视觉系统是苹果采摘机器人的重要组成部分^[1], 然而果园实际作业环境较为复杂, 呈现非结构化和不确定性, 采集到的苹果图像中果实姿态多变, 严重制约着目标果实的识别效率和精度, 因此复杂环境下苹果的精准识别与定位是视觉系统研究的关键^[2]。

目前大多利用手工制作颜色、纹理以及形状等^[3-5]特征识别各特定场景下苹果, 但是传统图像处理方法在复杂自然环境中的性能仍然受到限制, 人工设计的特征难以应对多种场景混杂的情况^[6]。近年来深度卷积神经网络广泛应用于目标检测领域, 复杂环境下苹果目标识别已经有很多研究^[7-10]。然而目前对于复杂果园环境下的苹果识别算法, 大部分都将未遮挡、枝叶遮挡及果实遮挡等不同生长状态下的苹果识别为一类, 当苹果采摘机器人采摘果实时, 就无法获得苹果生长状态的具体分类信息, 不能对枝、果实遮挡等果实采取特定的采摘方式,

从而可能导致机械臂及末端执行器因碰到树枝而造成损伤。针对这一问题, LV 等^[11]将 YOLOv5 应用于多个苹果组成的生长状态整体分类识别, 平均识别准确率为 98.45%, 识别速度为 71 帧/s, 但该算法未对单一苹果生长状态进行识别。GAO 等^[12]利用 Faster R-CNN 实现密叶果树多类苹果识别, 4 类图像的平均识别精度为 87.9%, 平均识别速度为 241 ms。为提高识别速度, YAN 等^[13]基于改进 YOLOv5s 将苹果分为可抓性和不可抓性两类, 平均识别准确率为 86.75%, 平均每幅图像识别时间为 15 ms, 但存在枝遮挡苹果被漏采的问题。为解决上述问题, 闫彬等^[14]将改进 YOLOv5m 应用于不同枝干遮挡情形下的苹果识别, 识别准确率为 81%, 平均每幅图像识别时间为 25 ms, 为实现枝干遮挡下的苹果采摘提供了可能。虽然这类算法可以实现苹果生长状态分类识别, 但只能得到苹果的中心位置, 不能准确识别出苹果的姿, 无法使末端执行器以合适的方式采摘果实。对于苹果姿态的识别, 已经有了一定的研究, 王丹丹等^[15]通过转动惯量特性获取未遮挡苹果的对称轴, 平均角度误差为 4.542°, 识别时间为 7.5 ms。NIU 等^[16]利用凸壳理论和形状上下文算法提取出部分遮挡苹果的对称轴, 平均角度误差为 7.72°。然而这类方法只适用于单一生长状态的苹果, 无法适用于复杂环境下多生长状态的姿态识别, 且仅能获得苹果的对称轴, 不能为苹果采摘提供特定的矢量方向。

本文针对复杂自然环境下苹果识别面临的上述问题,

收稿日期: 2023-11-10 修订日期: 2024-01-18

基金项目: 江苏省重点研发计划项目 (BE2021016); 江苏省现代农机装备与技术示范推广项目 (NJ2021-18)

作者简介: 陈青, 博士, 副教授, 研究方向为智能农林装备。

E-mail: qchen@njfu.edu.cn

提出了一种苹果生长状态分类及果实姿态的融合识别方法。基于YOLOv7模型,引入多尺度检测层和注意力机制模块,增强模型对苹果不同生长状态间微小局部特征的提取能力,引入注意力机制,改善网络对输入图片的感兴趣目标区域的关注度,采用Soft-NMS算法,能够有效避免高密度重叠目标被一次抑制从而发生漏检现象,进而提高苹果分类识别的精度。在此基础上,采用Unet分割网络对未遮挡苹果进行轮廓特征提取,并根据轮廓最小外接圆及最小外接矩形特征获取果实姿态,从而实现苹果生长状态分类及果实姿态的融合识别。最后通过试验验证复杂环境下苹果生长状态分类及果实姿态识别效果,以期苹果无损高效的采摘奠定基础。

1 试验数据

1.1 数据样本采集

本文以红富士苹果为研究对象,为获得不同果园环境下苹果图像,分别在南京绿航果园和徐州大沙河集团果树试验站进行红富士苹果图像采集。采集设备是iPhone XR型手机,图像采集日期为2022年9—11月,时间是08:00-17:00,为丰富数据的多样性,提高模型泛化能力,所采集的图像类型包括顺光、逆光、无遮挡、枝叶遮挡、果实遮挡等,此外考虑到机械臂采摘的范围有限,在距离果树0.4~1.2 m处拍摄图像,共采集1580张分辨率为4032×3024像素的JPG格式苹果图像,在此基础上进行人工筛选,进一步剔除拍摄模糊、严重背光、背景中含有大量非目标小苹果的苹果图像,最终得到1300张用于构建数据集的苹果图像。

1.2 数据集构建

本文将苹果按照无遮挡、叶遮挡、果实遮挡及枝遮挡分为4类,具体生长形态分别用uo、lo、fo及bo标记,如图1所示,利用Labelimg标注软件对采集图像中不同生长状态果实进行人工标注,具体生长状态分别用uo、lo、fo及bo标记,标注信息文件的格式为txt文件,其中包含目标苹果的类别和边界框等信息。经过标注后统计,数据集中1300张图片总共包含8632个苹果果实,其中uo、lo、fo、bo四种样本数量分别为2103、2342、1772以及2415。

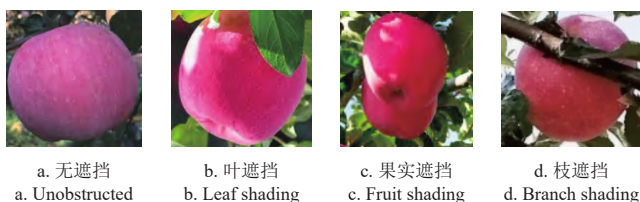


图1 苹果生长状态分类示意图

Fig.1 Schematic diagram of apple growth pattern classification

为训练出更加符合自然环境下苹果生长状态检测模型,制定了如下的标注规则:1)将bo>fo>lo>uo作为优

先级别,在高和低优先级并存时,标注高优先级的生长状态;2)边界框尽量选择目标果实的最小矩形框,如果在图片边缘出现的果实尺寸小于一半,则不进行标注;3)对于图像较远的非目标小苹果不进行标注。

由于数据集样本标签分为4类,为了达到丰富数据集信息,防止模型出现过拟合,提高模型识别精度和泛化能力的目的,采用添加椒盐噪声、水平翻转、亮度调整、随机平移、随机旋转等方式进行数据增强,扩展数据集,最后,经过数据增强的数据集达到5200张,将其按照8:1:1的比例分为训练集、验证集和测试集,其中部分经过数据增强的样本图像如图2所示。

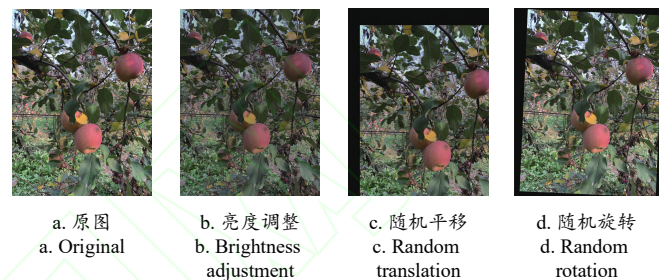


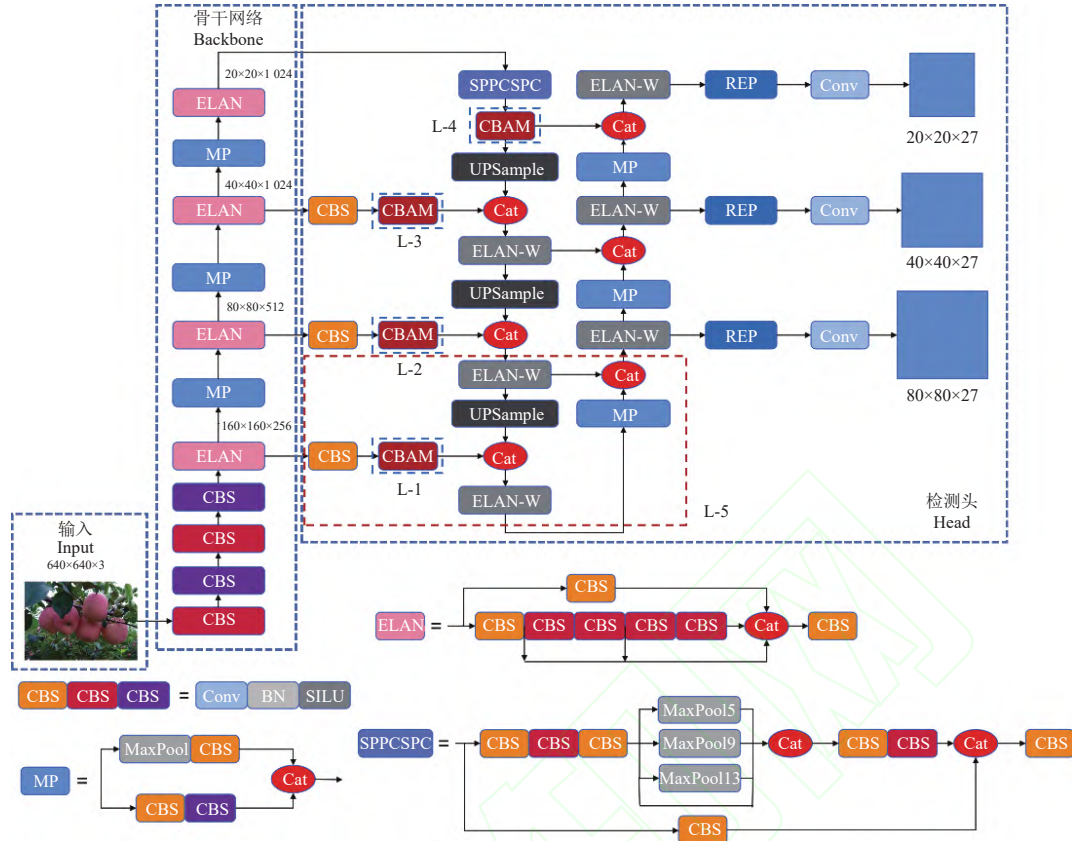
图2 苹果数据增强图像

Fig.2 Enhanced image of apple data

2 网络模型及改进

YOLOv7算法由输入端、骨干网络、预测头三部分构成^[17-21],其中骨干网络采用高效聚合网络结构,通过控制最短和最长的梯度路径,使网络能够学习到更多的特征,从而进行有效学习和收敛。此外,在预测头部分,采用模块重参化及辅助训练模块等方式,降低正样本分配的约束,避免丢失更多的信息,在降低模型参数的同时,提高模型的效率与性能。

由于果园环境复杂,且叶、枝及果实有时会遮挡目标果实较少部分,果实生长状态间的差异较小,原始图像归一化尺寸后目标果实被遮挡部分就会更小,由于卷积神经网络的浅层特征具有更多的局部信息,经多次卷积操作后易造成深层特征图丢失果实遮挡部分的关键信息,从而造成果实生长状态误识别。同时,自然环境下果实重叠的生长状态较多,检测网络在对重叠果实进行识别时容易出现将重叠程度较大的两个苹果识别为一个的情况,从而造成被遮挡果实的漏检。针对上述问题,本文对YOLOv7模型进行改进,首先,引入多尺度检测层,提高模型对微小遮挡局部特征的识别敏感度。为了提高模型对果实目标及生长形态特征的提取能力,在骨干网络中添加卷积注意力模块(convolutional block attention module, CBAM),从而改善网络对输入图像中感兴趣目标区域的关注度。另外,采用Soft-NMS的方法,以解决重叠遮挡苹果误检漏检的问题。最后,改进后的YOLOv7苹果生长状态检测模型如图3所示。

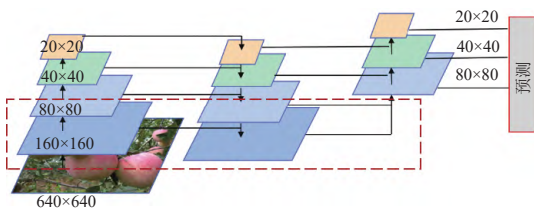


注: L-1~L-4 为 CBAM 模块的最终添加位置, L-5 为添加的 160×160 尺度特征。
Note: L-1 to L-4 are the final added locations of the CBAM modules, and L-5 is the added 160×160 scale feature.

图 3 改进 YOLOv7 网络模型
Fig.3 Improved YOLOv7 network model

2.1 多尺度特征融合改进

卷积神经网络的深层特征图上像素点数目较少且每个像素点感受野较大, 具有更多的语义信息, 容易检测到较大的目标。而原始图像中的小目标在多次下采样后保留的信息越来越少, 而为了能够获得更多的局部信息, 需要在尺寸较大的特征图提取特征。因此原有的网络结构在面对生长状态间差异较小的目标果实时, 并不能很好地提取微小遮挡的局部特征并进行融合, 从而导致果实遮挡部分的局部信息丢失, 造成果实生长状态误识别。YOLOv7 分别从骨干网络中 80×80、40×40 和 20×20 的特征图中提取关键特征, 并进行特征融合, 为了更好地对微小遮挡的苹果生长状态进行区分, 提高分类识别准确率, 在原有 3 个尺度的基础上, 增加了 160×160 的特征尺度, 该方法在骨干网络原有 160×160 的特征图上输出提取的特征, 并将该尺度特征与原有 3 个尺度特征进行融合, 具体结构如图 4 所示。



注: 虚线框为添加的 160×160 尺度特征。
Note: Dashed boxes are added 160 x 160 scale features.

图 4 多尺度特征融合改进
Fig.4 Improved multi-scale feature fusion

2.2 添加注意力机制

果园复杂环境中, 由于受到光照, 枝叶遮挡以及非结构化背景等因素的影响, 导致图片中出现诸多干扰信息。注意力机制是对图像输入重要信息的动态选择过程, 这个过程是通过特征自适应权重实现的, 聚焦图像的重要特征, 抑制不必要的区域响应。因此本文将注意力机制 CBAM^[22-23] 引入 YOLOv7, CBAM 将空间注意力模块和通道注意力模块进行结合, 在给定一张特征图, 它能够序列化地在通道和空间两个维度上产生注意力特征图信息, 然后两种特征图信息在与之前原输入特征图进行相乘进行自适应特征修正, 产生最后的特征图, 图 5 为 CBAM 的结构示意图。

CBAM 总体流程可用下式表示:

$$F' = M_C(F) \otimes F \quad (1)$$

$$F'' = M_S(F') \otimes F' \quad (2)$$

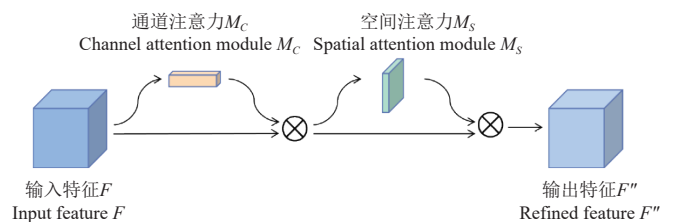


图 5 卷积注意力模块
Fig.5 Convolutional block attention module (CBAM)

式中 $F \in R^{C \times H \times W}$ 为输入特征图, $M_c \in R^{C \times 1 \times 1}$ 为通道注意力特征图, $M_s \in R^{1 \times H \times W}$ 为空间注意力特征图。

注意力机制的插入位置也会影响到网络的性能, 本文将CBAM分别嵌入主干网络的第11、24、37和51层, 将各层网络获得的特征图作为输入特征图, 经过CBAM模块后, 增加有关于苹果目标及生长形态特征的权重, 输出特征图, 从而进一步提升 160×160 、 80×80 、 40×40 以及 20×20 等不同特征尺度下的对于苹果关键特征的提取能力, 具体添加位置在图3分别用L-1~L-4表示。

2.3 柔性非极大值抑制 Soft-NMS 算法

非极大抑制是目标检测过程中的重要组成部分, 通过对所有检测框按其得分高低排序, 选择得分最大的检测框, 并将与之有显著重叠并超过一定阈值的其他检测框全部剔除。在自然环境条件下, 2个苹果果实高度重叠, 使其产生的检测框具有很高的重合度, 有可能出现高置信度得分的预测框抑制掉被遮挡苹果的检测框, 造成漏检的情况。

为解决上述重叠苹果检测框错误抑制而造成漏检的问题, 本文使用柔性非极大值抑制算法 (Soft-NMS) 替代非极大值抑制算法 (NMS)。Soft-NMS 与 NMS 最主要的区别是对与得分最大框的 IOU 大于阈值的检测框的处理方法^[24], 其为该类检测框的置信度得分设置一个衰减函数, 而不是直接将其设定为 0。通过这种方法, 可以避免出现某些高度重叠的果实检测框被彻底剔除的情况, 提升模型对于严重重叠遮挡苹果的检测能力。Soft-NMS 计算方法为

$$s_i = s_i e^{-\frac{iou(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (3)$$

式中 M 为当前得分最高框, b_i 为待处理检测框, iou 为交并比, s_i 为 b_i 框更新得分, D 为存放最终检测框的集合。

2.4 姿态识别网络

在对苹果生长状态进行分类识别后, 为使末端执行器以合适的方式采摘果实, 需要准确识别出苹果的姿态, 但在采摘过程中, 苹果采摘机器人一个视角采集的苹果图像通常包括被枝叶遮挡以及果实重叠, 当机器人移动时, 严重遮挡的苹果可能会从另一个视图中可见, 且重叠果实由于未遮挡果实被采摘后, 也变为未遮挡状态, 因此可以考虑仅识别未遮挡状态的苹果姿态。

在进行果实姿态识别时, 首先使用改进的YOLOv7对苹果的生长状态进行分类, 对未遮挡果实的区域进行裁剪, 获得独立的未遮挡苹果的图片, 将这些图片按照是否能够看见苹果花萼分为两类, 能看见苹果花萼的为生长类型1, 反之为生长类型2。在此基础上, 利用Unet算法实现对裁剪后图像中的苹果区域进行分割, 其中对于生长类型1的苹果, 还需对花萼区域进行分割。此后, 对分割后的图像进行图像二值化, 通过轮廓提取算法获得苹果及花萼轮廓特征。其中生长类型1利用果实及花萼轮廓的最小外接圆的两个圆心, 连接两个圆心, 并将花萼中心点到果实中心点的方向作为箭头方向, 从而获得生长类型1的苹果姿态方向。生长类型2的苹果利用果实轮廓最小外接矩形, 获取最小外接矩形的中心

点、宽高以及旋转角度等信息, 其中以 X_1 、 X_2 、 X_3 以及 X_4 为最小外接矩形4个点的坐标, 分别代表4个坐标中横坐标最小、纵坐标最小、横坐标最大以及纵坐标最大的点, 将 X_1 到 X_2 点的距离定义为高, X_2 到 X_3 点的距离定义为宽, 水平线顺时针旋转到第一条宽边的角度作为旋转角度, 最小外接矩形示意图如图6所示。

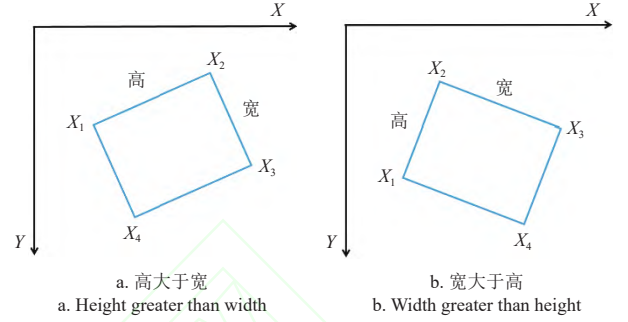
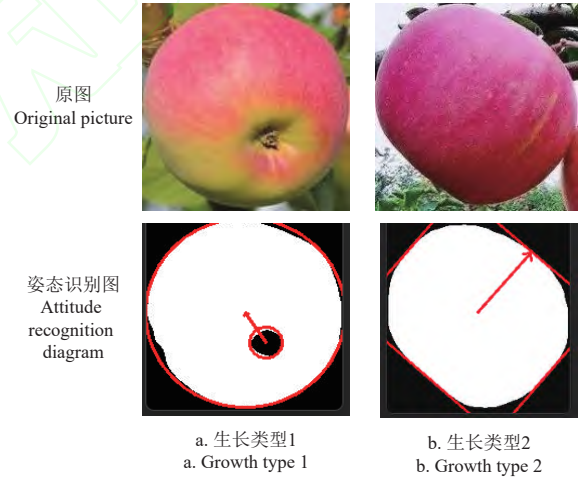


图6 苹果最小外接矩形示意图

Fig.6 Apple's smallest external rectangle schematic

当宽大于高时, 将最小外接矩形中心点到 X_2 和 X_3 的中点的方向作为箭头方向, 反之, 将最小外接矩形中心点到 X_1 和 X_2 的中点的方向作为箭头方向, 从而得到生长类型2的苹果姿态方向, 姿态识别结果如图7所示。



注: 箭头方向为苹果姿态, 生长类型1为能看见苹果花萼的类型, 生长类型2反之。

Note: The direction of the arrow is the apple stance, with growth type 1 being the type in which the apple calyx is visible and growth type 2 vice versa.

图7 苹果姿态识别示意图

Fig.7 Schematic diagram of apple's gesture recognition

2.5 试验平台配置与训练策略

本研究使用 Windows11 操作系统, 搭载 AMD Ryzen 75800H CPU @ 2.80 GHz 处理器、Nvidia GeForce GTX 3060 6GB 显卡和 16GB 运行内存。深度学习框架为 Pytorch 1.11.0, 采用 CUDA 和 CUDNN GPU 加速库, 编程平台为 PyCharm, 编程语言为 Python 3.9, 图像处理采用 OpenCV 库。

训练时选用 YOLOv7.pt 作为预训练权重, 数据集选用自己建立的苹果数据集, 输入图像大小为 640×640 像素, 经过多次调整和测试, 设置训练批次为 4, 迭代次数为 300, 初始学习率为 0.01, 衰减系数为 0.01, 动量因子为 0.9。

2.6 评价指标

选用精确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、平均检测时间、模型内存占用量作为苹果目标及生长状态识别评价指标, 选用平均像素准确率 (mean pixel accuracy, mPA)、平均交并比 (mean intersection over union, mIOU)、姿态检测准确率作为苹果姿态识别评价指标。

$$P = \frac{T_p}{T_p + F_p} \quad (4)$$

$$R = \frac{T_p}{T_p + F_N} \quad (5)$$

$$AP = \int_0^1 P \cdot (R) dR \quad (6)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (7)$$

$$mPA = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij}} \quad (8)$$

$$mIOU = \frac{1}{k} \sum_{i=1}^k \frac{P \cap G}{P \cup G} \quad (9)$$

式中 T_p 为真实的正样本数量, F_p 为虚假的正样本数量, F_N 为虚假的负样本数量, AP 代表 PR 曲线以下的面积, mAP 为所有类别的平均 AP 值, P 表示预测值, G 表示真实值, n 和 k 分别表示苹果生长形态检测和未遮挡苹果分割的类别, 在本文中, $n=4$, $k=3$ 。

3 结果与分析

3.1 改进 YOLOv7 模型消融试验分析

为了验证改进 YOLOv7 模型的性能, 本研究设置了 4 组消融试验, 在 YOLOv7 原始网络的基础上, 依次加入多尺度特征融合改进、注意力机制 CBAM 和 Soft-NMS 算法, 使用相同数据集进行训练和测试, 得到消融试验对比结果如表 1 所示。

从表 1 中可以看出, D-YOLOv7 和 DC-YOLOv7 相较于 YOLOv7 在 mAP 上分别提升了 2.4 和 1.3 个百分点, 这表明训练阶段引入改进多尺度特征提取层及注意力机制 CBAM 的方法可以有效提取果实目标及生长状态的特征, 提高模型的检测精度, 其中引入改进多尺度特征提取层相较加入注意力机制 CBAM 提升更为明显。DCS-YOLOv7 识别精确率、召回率和平均识别精度分别为 86.9%、80.5% 和 87.1%, 平均每幅图像识别时间为 14.8 ms, 相较于 DC-YOLOv7 模型, 在权重大小相同的情况下, mAP 提升了 0.6 个百分点, 这表明在测试阶段引入 Soft-NMS 算法, 能够有效避免高密度重叠目标被一次抑制从而发生漏检现象, 对于自然果园环境下密集遮挡苹果果实的目标检测具有积极影响。与原 YOLOv7 相比, DCS-YOLOv7 模型的精度、召回率以及平均检测精度提高了 4.2、2.2 和 3.7 个百分点, 虽然模型的识别时间略

有上升, 但仍可以满足机器人实时采摘的要求, 这表明本文提出的 DCS-YOLOv7 模型可以有效地用于检测苹果目标及生长状态分类。

表 1 改进 YOLOv7 模型的消融试验结果
Table 1 Improved YOLOv7 model ablation test results

模型 Model	精确率 Precision P/%	召回率 Recall R/%	平均精度均值 Mean average precision mAP/%	识别时间 Time t/ms	权重大小 Weight w/MB
YOLOv7	82.7	78.3	83.4	12.8	71.4
D-YOLOv7	85.2	79.7	85.8	13.7	84.6
DC-YOLOv7	86.1	79.6	86.5	14.6	95.7
DCS-YOLOv7	86.9	80.5	87.1	14.8	95.7

注: D-YOLOv7 表示在 YOLOv7 中修改多尺度特征提取层; DC-YOLOv7 表示在 YOLOv7 中同时修改多尺度特征提取层和引入 CBAM; DCS-YOLOv7 表示在 YOLOv7 中同时修改多尺度特征提取层和引入 CBAM 及 Soft-NMS 算法。
Note: D-YOLOv7 denotes the modification of the multi-scale feature extraction layer in YOLOv7; DC-YOLOv7 denotes the modification of the multi-scale feature extraction layer and the introduction of the CBAM in YOLOv7 at the same time; DCS-YOLOv7 denotes the modification of the multi-scale feature extraction layer and the introduction of the CBAM and the Soft-NMS algorithm in YOLOv7 at the same time.

3.2 不同目标检测模型的可视化对比

为验证 DCS-YOLOv7 在苹果生长状态数据集上检测效果, 利用训练得到的最优权重对测试集进行测试, 将 Faster RCNN^[25], YOLOv5s、YOLOv5 m、YOLOv7 及 DCS-YOLOv7 识别效果进行对比, 不同光照下识别结果如图 8 所示。

从图 8 可以看出, Faster RCNN 在不同光照条件下会产生重叠框现象, 即将同一苹果被区分成两种不同生长状态, 并且将两个重叠苹果识别成一个果实, 导致模型识别性能不佳。对于 YOLO 系列的识别算法, YOLOv5s 检测效果较好, 但由于网络模型参数量较少, 苹果特征提取不够充分, 对于枝叶遮挡较为严重的苹果, 存在漏检情况。YOLOv5m 对各类苹果生长状态均能较好识别, 但存在将枝遮挡下的单个苹果分成两个目标, 从而造成误识别。YOLOv7 模型检测置信度较高, 能够较好识别出不同光照下的苹果生长状态, 但也会出现漏检现象。总体来看, DCS-YOLOv7 具有较高的识别置信度, 能够识别出严重遮挡的苹果, 且对于较远种植行中的非目标小苹果不进行识别, 因此该模型对于不同光照条件下的苹果识别更加准确, 能够较好地满足苹果生长状态识别的需求。

模型的试验比较数据如表 2 所示, 其中 Faster RCNN 网络模型权重最大, 识别时间最长, 但 mAP 最低, 可以看出二阶段目标检测模型不适用于本文的苹果数据集。YOLOv5s 模型权重最小, mAP 远高于 Faster RCNN, 达到 79.9%, 说明一阶段目标检测模型在本文数据集能够达到更好的识别效果。在同样的迭代过程中, YOLOv7 网络的模型权重最大, 导致模型识别时间高于 YOLOv5s 和 YOLOv5m, 但 mAP 最高, 而平均每幅图片识别时间为 12.8 ms 的检测速度已能满足苹果采摘需求, 因此本文考虑在达到实时识别的基础上, 应当尽可能提高 mAP, 所以将 YOLOv7 选择作为苹果生长状态分类识别原模型。DCS-YOLOv7 在苹果数据集上的 mAP 比 YOLOv5s、

YOLOv5m、YOLOv7 分别高出 7.2、5.9、3.7 个百分点，因此，DCS-YOLOv7 综合性能最佳，对不同光照条件及存在遮挡的苹果目标均可准确识别，表明本文模型能够很好地检测出苹果生长状态，并且具有较好的鲁棒性。



注：蓝色圆圈框为误检和漏检的苹果。
Note: Blue circle boxes are misdected and missed apples.

图 8 不同模型对苹果的识别效果
Fig.8 Effect of different models for apple recognition

表 2 不同模型检测效果对比					
Table 2 Comparison of the detection effect of different models					
模型 Model	P/%	R/%	mAP /%	t/ms	w /MB
Faster RCNN	67.4	69.6	68.2	50.7	315
YOLOv5s	73.8	75.8	79.9	7.6	29.6
YOLOv5 m	73.4	80.1	81.2	12.7	40.5
YOLOv7	82.7	78.3	83.4	12.8	71.4
DCS-YOLOv7	86.9	80.5	87.1	14.8	95.7

3.3 姿态检测网络试验结果

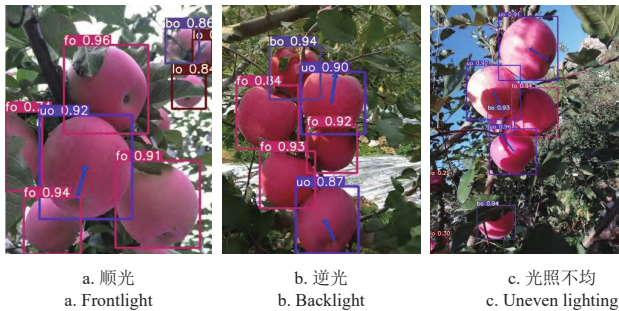
苹果姿态检测的第一步是使用语义分割模型对裁剪后的未遮挡苹果进行分割并消除背景干扰，优秀的分割模型具有良好的分割性能，有利于减少后续苹果姿态检测的误差。为了选择最适合本文的语义分割模型，将从原始图像裁剪出的未遮挡苹果图片进行苹果整体及花萼标注，构建一个苹果分割数据集，并采用近年来流行的语义分割模型 UNet^[26]、DeepLabv3+^[27] 和 PSPNet^[28] 进行分割试验，试验结果如表 3 所示。

表 3 不同模型分割效果对比			
Table 3 Comparison of segmentation effect of different models			
模型 Model	平均像素准确率 Mean pixel accuracy mPA/%	平均交并比 Mean intersection over union mIOU /%	时间 Time t/ms
UNet	95.3	85	15.5
DeepLabv3+	94.6	83.4	19.2
PSPNet	95.1	83.9	18.9

试验表明，UNet 模型在苹果图像分割中表现出最佳的性能，与 DeepLabv3+和 PSPNet 相比，平均像素准确率分别高出 0.7 和 0.2 个百分点，平均交并比分别高出了 1.7 和 1.1 个百分点，而且从分割速度可以看出，UNet

模型在分割效率上优于其他分割模型。由此可见，在处理小样本数据集时，采用基于编码器和解码器的 UNet 模型能够获得更好的性能。

在得到苹果分割图像后，利用苹果姿态识别算法，可以得到不同生长类型的苹果姿态。本文以 300 张苹果分割图像为样本，人工检验并计算准确度，试验证明，其中 282 幅图像的苹果姿态检测是正确的，准确率达到 94%，平均每幅图像处理时间为 15.7 ms，再加上 DCS-YOLOv7 识别时间，总的识别时间为 30.5 ms，比实时检测要求的 33 ms 少，采用该方法检测苹果姿态不会影响整体识别模型的实时检测，苹果生长状态分类及果实姿态融合识别算法的识别效果如图 9 所示。



注：uo，无遮挡；lo，叶遮挡；fo，果实遮挡；bo，枝遮挡；箭头方向表示未遮挡果实姿态。
Note: uo, no-shade; lo, leaf-shade; fo, fruit-shade; bo, branch-shade; and the direction of the arrows indicates the unobstructed fruit attitude.

图 9 苹果生长状态分类及果实姿态融合识别图
Fig.9 Classification of apple growth morphology and recognition of fruit posture fusion

3.4 与其他苹果识别算法对比

目前对于苹果生长状态的分类识别及姿态识别已有一些研究,为了验证本文所提出苹果检测算法的性能,将现有的相关研究与本文算法进行了比较,对比和分析结果见表4。

表4 与其他苹果识别算法的对比

模型 Model	检测对象 Detection objects	视觉信息 Visual information	准确率 Accuracy
Faster RCNN ^[12]	无遮挡、叶遮挡、枝遮挡和果实遮挡 4 类苹果	4 类苹果生长形态识别及分类	mAP: 87.9%
YOLOv5s ^[13]	可采摘与不可采摘 2 类苹果	2 类苹果生长形态识别及分类	mAP: 86.75%
文献 [15]	单个无遮挡苹果	苹果果实对称轴	角度误差: 4.542°
文献 [16]	单个部分遮挡的苹果	苹果果实对称轴	角度误差: 7.72°
本文算法 Our algorithm	无遮挡、叶遮挡、枝遮挡和果实遮挡 4 类苹果	4 类苹果生长形态分类及姿态识别	mAP: 87.1% 姿态准确率: 94%

从表4可以看出,现有苹果生长状态检测方法虽然可以实现苹果生长状态分类识别,但只能得到苹果的中心位置,不能识别出苹果的姿态,而通过获取果实轴对称进行果实姿态识别的这类算法只适用于单一生长状态的苹果,且不能为苹果采摘提供特定的矢量方向。本文提出的方法具有较高的检测精度,并将果实生长状态和姿态的识别进行融合,在对苹果生长状态进行分类的基础上,考虑了果实的姿态识别,使末端执行器以合适的方式采摘果实,为苹果无损高效的采摘奠定基础。

4 结 论

本文针对目前苹果在复杂环境下难以进行生长状态分类识别、姿态信息同步获取等问题,提出了一种基于改进 YOLOv7 的苹果生长状态分类和果实姿态融合识别方法,通过改进多尺度特征融合网络,用于增强模型对微小局部特征的识别敏感度,引入注意力机制 CBAM,改善网络对输入图片的感兴趣目标区域的关注度,采用 Soft-NMS 算法,能够有效避免高密度重叠目标被一次抑制从而发生漏检现象,从而提高模型对各类苹果生长状态的检测精度,并与其他典型算法进行对比,得出如下结论:

- 1) DCS-YOLOv7 可有效实现对苹果的生长形态识别,试验结果表明,识别精确率、召回率和平均识别精度分别为 86.9%、80.5% 和 87.1%,平均每幅图像识别时间为 14.8 ms。
- 2) 对比本文 DCS-YOLOv7 与 FasterRCNN, YOLOv5s、YOLOv5 m、YOLOv7 在苹果生长状态数据集上识别效果,结果表明,DCS-YOLOv7 的 mAP 比其他 3 种算法分别高出 7.2、5.9、3.7 个百分点,虽然改进模型的识别时间与原始 YOLOv7 相比略有上升,但仍可以满足机器人实时采摘的要求
- 3) 基于 UNet 和最小外接圆及矩形特征的苹果姿态检测方法的准确率为 94%,平均每幅图像处理时间为 15.7 ms,可以较好地实现苹果目标的姿态获取,验证了算法的有效性和正确性。

[参 考 文 献]

[1] JIA W, ZHANG Y, LIAN J, et al. Apple harvesting robot under information technology: A review[J]. International Journal of Advanced Robotic Systems, 2020, 17(3): 255688461.

[2] 王丹丹, 宋怀波, 何东健. 苹果采摘机器人视觉系统研究进展[J]. 农业工程学报, 2017, 33(10): 59-69.

WANG Dandan, SONG Huaibo, HE Dongjian, et al. Research advance on vision system of apple picking robot[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(10): 59-69. (in Chinese with English abstract)

[3] FAN P, LANG G, GUO P, et al. Multi-feature patch-based segmentation technique in the gray-centered RGB color space for improved apple target recognition[J]. Agriculture, 2021, 11(3): 273.

[4] 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)

[5] LI D, SHEN M, LI D, et al. Green apple recognition method based on the combination of texture and shape features[C]// 2017 IEEE International Conference on Mechatronics and Automation. Takamatsu, Japan, 2017.

[6] 陈青, 殷程凯, 郭自良, 等. 苹果采摘机器人关键技术研究现状与发展趋势[J]. 农业工程学报, 2023, 39(4): 1-15.

CHEN Qing, YIN Chengkai, GUO Ziliang, et al. Current status and future development of the key technologies for apple picking robots[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 1-15. (in Chinese with English abstract)

[7] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164-173.

ZHAO De'an, WU Rendi, LIU Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164-173. (in Chinese with English abstract)

[8] 赵辉, 乔艳军, 王红君, 等. 基于改进 YOLOv3 的果园复杂环境下苹果果实识别[J]. 农业工程学报, 2021, 37(16): 127-135.

ZHAO Hui, QIAO Yanjun, WANG Hongjun, et al. Apple fruit recognition in complex orchard environment based on improved YOLOv3[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(16): 127-135. (in Chinese with English abstract)

[9] JIA W, TIAN Y, LUO R, et al. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application

- in apple harvesting robot[J]. *Computers and Electronics in Agriculture*, 2020, 172: 105380.
- [10] 周桂红, 马帅, 梁芳芳. 基于改进 YOLOv4 模型的全景图像苹果识别[J]. *农业工程学报*, 2022, 38(21): 159-168.
Zhou Guihong, Ma Shuai, Liang Fangfang. Recognition of the apple in panoramic images based on improved YOLOv4 model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(21): 159-168. (in Chinese with English abstract)
- [11] LV J, XU H, HAN Y, et al. A visual identification method for the apple growth forms in the orchard[J]. *Computers and electronics in agriculture*, 2022, 197: 106954.
- [12] GAO F, FU L, ZHANG X, et al. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN[J]. *Computers and Electronics in Agriculture*, 2020, 176: 105634.
- [13] YAN B, FAN P, LEI X, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. *Remote sensing (Basel, Switzerland)*, 2021, 13(9): 1619.
- [14] 闫彬, 樊攀, 王美茸, 等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别[J]. *农业机械学报*, 2022, 53(9): 28-38, 59.
YAN Bin, FAN Pan, WANG Meirong, et al. Real-time apple picking pattern recognition for picking robot based on improved YOLOv5m[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(9): 28-38, 59. (in Chinese with English abstract)
- [15] 王丹丹, 徐越, 宋怀波, 等. 基于平滑轮廓对称轴法的苹果目标采摘点定位方法[J]. *农业工程学报*, 2015, 31(5): 167-174.
WANG Dandan, XU Yue, SONG Huaibo, et al. Localization method of picking point of apple target based on smoothing contour symmetry axis algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(5): 167-174. (in Chinese with English abstract)
- [16] NIU L, ZHOU W, WANG D, et al. Extracting the symmetry axes of partially occluded single apples in natural scene using convex hull theory and shape context algorithm[J]. *Multimedia tools and applications*, 2017, 76(12): 14075-14089.
- [17] 张万枝, 张弘毅, 刘树峰, 等. 基于改进 YOLOv7 模型的马铃薯薯芽眼检测[J]. *农业工程学报*, 2023, 39(20): 148-158.
ZHANG Wanzhi, ZHANG Hongyi, LIU Shufeng, et al. Detection of potato seed buds based on an improved YOLOv7 model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(20): 148-158. (in Chinese with English abstract)
- [18] 王小荣, 许燕, 周建平, 等. 基于改进 YOLOv7 的复杂环境下红花采摘识别[J]. *农业工程学报*, 2023, 39(6): 169-176.
WANG Xiaorong, XU Yan, ZHOU Jianping, et al. Safflower picking recognition in complex environments based on an improved YOLOv7[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(6): 169-176. (in Chinese with English abstract)
- [19] 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)
- [20] 祝志慧, 何昱廷, 李沃霖, 等. 基于改进 YOLOv7 模型的复杂环境下鸭蛋识别定位[J]. *农业工程学报*, 2023, 39(11): 274-285.
ZHU Zhihui, HE Yuting, LI Wolin, et al. Improved YOLOv7 model for duck egg recognition and localization in complex environments[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(11): 274-285. (in Chinese with English abstract)
- [21] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada, 2023.
- [22] WOO S H, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module [C]//15th European Conference on Computer Vision (ECCV). Munich, Germany, 2018.
- [23] 肖张娜, 罗陆锋, 陈明猷, 等. 基于改进 YOLO-v4 的果园环境下葡萄检测[J]. *智能化农业装备学报 (中英文)*, 2023, 4(2): 35-43.
XIAO Zhangna, LUO Lufeng, CHEN Mingyou, et al. Detection of grapes in orchard environment based on improved YOLO-v4[J]. *Journal of Intelligent Agricultural Mechanization*, 2023, 4(2): 35-43. (in Chinese with English abstract)
- [24] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS - Improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy, 2017.
- [25] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//29th Annual Conference on Neural Information Processing Systems (NIPS). Montreal, Canada, 2015.
- [26] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany, 2015.
- [27] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//15th European Conference on Computer Vision (ECCV). Munich, Germany, 2018.
- [28] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//30th Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, 2017.

Apple growth status and posture recognition using improved YOLOv7

CHEN Qing^{1,2}, YIN Chengkai², GUO Ziliang², WU Xuanbo², WANG Jinpeng^{1,2}, ZHOU Hongping^{1,2}

(1. Co-Innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University, Nanjing 210037, China;

2. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China)

Abstract: Manual picking cannot fully meet the large-scale production in China at present. Robotic picking has been an inevitable trend, particularly with the shortage of labor resources and the rapid development of mechanical automation. It is very necessary to accurately identify and position the apples in the complex environments. Fruit attitude fusion acquisition can be synchronously realized and then classified the apple information. Sometimes, only a small portion of target fruit is covered from the orchard environment, including the leaves, branches, and fruits. There are the small differences among the fruit growth patterns. The convolutional neural network is easy to cause the deep feature map, and then lose the key information of fruit covering parts after multiple convolution operations, resulting in the misrecognition of the fruit growth pattern. At the same time, the detection network can easily identify two apples as one for the overlapping fruits in the natural environment, thus causing the omission of the occluded fruits. In this study, an improved YOLOv7 model was proposed to recognize the apple posture from the growth morphologies. Firstly, the multi-scale feature fusion network was improved to add a 160×160 feature scale layer in the backbone network. The sensitivity of the model was enhanced to identify the tiny local features; Secondly, CBAM attention mechanism was introduced to improve the target region of interest; Finally, the Soft-NMS was used to effectively avoid the high-density overlapping targets being suppressed at one time, thus reducing the missed detection. The experimental results show that the recognition accuracy, recall and average recognition precision of DCS-YOLOv7 were 86.9%, 80.5% and 87.1%, respectively, which were 4.2%, 2.2% and 3.2% higher than the original YOLOv7 model. The average accuracy and speed were greatly improved to fully meet the requirements of picking robot. In addition, an apple gesture recognition was proposed using semantic segmentation and the minimum outer join features. Firstly, comparison tests showed that the Unet model exhibited the best performance in apple image segmentation. The average pixel accuracies were 0.7 and 0.2 percentage points higher than those of DeepLabv3+ and PSPNet. The average intersection and merger ratios were 1.7 and 1.1 percentage points higher as well. The average speed of segmentation also outperformed the rest. As such, the UNet instance segmentation network was chosen as the apple segmentation model. The apple image was segmented using UNet semantic segmentation network. The apple and calyx contour features were obtained by the contour extraction, and then the pose of unobstructed apple was obtained using the apple minimum external feature. The accuracy was 94% to detect the apple pose. The average processing time for each image was 15.7ms, indicating the better acquisition for the pose of apple target. The validity and correctness of recognition model were verified with the high detection accuracy to integrate the recognition of fruit growth pattern and posture. The recognition of fruit posture was considered to classify the growth pattern of apples. The end-effector can rapidly and accurately pick the fruits in a suitable way. The finding can lay the foundation for the non-destructive and efficient picking of apples.

Keywords: image processing; YOLOv7; classification recognition; pose recognition; deep learning; apple