

# 无规则扰动状态下柑橘果实在线目标检测与快速定位

娄欢欢, 李光林\*, 付兴兰, 李丽, 王旭, 黄伟东, 付泰戈

(西南大学工程技术学院, 重庆 400715)

**摘要:** 柑橘采摘机器人连续采摘过程中因各种因素会引起其他待采柑橘无规则扰动, 扰动状态下的柑橘无法在线快速准确检测与定位, 影响机器人采摘效率。针对此问题, 该研究提出一种基于改进 YOLOv5s+DeepSORT 算法的扰动柑橘在线目标检测与快速定位方法。首先在 YOLOv5 骨干网络中融入卷积注意力机制 (convolutional block attention module, CBAM), 提升模型对复杂目标的检测能力; 用 SIoU (scalable intersection over union) 损失函数增强预测框与标定框之间的方向匹配, 提升回归收敛速度。其次在 DeepSORT 算法中改进目标重识别网络 (re-identification, ReID), 增强网络特征提取能力, 提升目标跟踪准确度与精度; 在算法中融入 Count 计数机制, 实时反馈每个扰动柑橘跟踪帧数, 并改进算法实现对预测坐标值进行实时更新, 提升预测准确率。最后结合深度相机排除背景柑橘影响并限制每次跟踪目标数目为 3 个, 提升扰动柑橘预测定位速度。试验结果表明, 与原算法相比, 改进 YOLOv5s 算法的准确度、平均检测精度分别提升 3.9、1.1 个百分点, 检测速率 69.3 帧/s。改进 DeepSORT 算法的跟踪准确度、跟踪精度分别提升 9.2、5.4 个百分点, ID (identity) 切换次数减少 32 次。当预测定位时间为 3 s 时, 定位平均准确度为 81.9%, 在试验室进行模拟试验, 将盆栽柑橘果实沿不同方位随机摆动, 摆幅约 10 cm, 单个柑橘平均抓取时间为 12.8 s, 比未使用改进算法缩短 5.6 s, 效率提升 30.4%。该研究可为扰动状态下的柑橘快速采摘提供技术支持。

**关键词:** 扰动柑橘; 预测定位; 在线目标检测; 目标跟踪; YOLOv5; DeepSORT

doi: 10.11975/j.issn.1002-6819.202404120

中图分类号: S225; TP242; TP391.41

文献标志码: A

文章编号: 1002-6819(2024)-18-0155-12

娄欢欢, 李光林, 付兴兰, 等. 无规则扰动状态下柑橘果实在线目标检测与快速定位[J]. 农业工程学报, 2024, 40(18):

155-166. doi: 10.11975/j.issn.1002-6819.202404120 <http://www.tcsae.org>

LOU Huanhuan, LI Guanglin, FU Xinglan, et al. Online target detection and fast localisation of citrus fruits in irregularly disturbed state[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 155-166. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404120 <http://www.tcsae.org>

## 0 引言

中国是世界上最大的柑橘种植和消费国之一, 柑橘产量受需求影响持续增加<sup>[1]</sup>。目前, 中国柑橘采摘以人工采摘为主, 存在采摘劳动力不足、效率低等问题<sup>[2]</sup>。随着科技发展, 机器人采摘柑橘是重要的发展方向之一<sup>[3-4]</sup>。然而, 现有的柑橘采摘机器人连续采摘过程中, 由于受外界风力、机器人运动以及枝条承载重量变化等因素影响会引起柑橘的无规则扰动。在扰动状态下, 柑橘的检测和定位十分困难, 目前采用的方法多为等待扰动柑橘静止后再进行下一轮识别与采摘, 这不仅影响了机器采摘效率, 也限制了柑橘采摘机器人的应用与发展。因此, 为了提高柑橘采摘机器人连续采摘的效率, 研究柑橘在扰动状态下的在线目标检测与快速定位具有重要意义。

国内外学者对扰动状态下的水果目标检测与定位做了大量研究, 熊俊涛等<sup>[5]</sup>假设柑橘为微扰动, 利用

ANSYS 进行动力学分析并结合图像处理研究微扰动柑橘的运动规律, 确定扰动柑橘的有效采摘点。XIONG 等<sup>[6]</sup>对被干扰的动态荔枝簇进行研究, 使用带有两个相机的视觉系统来获取动态荔枝簇的图像, 对于静态或轻微扰动, 采用双目视觉立体匹配计算采摘点。LIU 等<sup>[7]</sup>利用 ORB-SLAM3(ORB-simultaneous localization and mapping) 算法估计相机在动态条件下的姿态并采用 YOLOv5 识别 RGB 图像中的苹果, 平均定位误差为 21.1 mm, 其中 90% 的误差小于 30 mm。叶敏等<sup>[8]</sup>针对荔枝采摘机器人的扰动视觉定位误差, 提出一种动态定位误差分析方法, 并采用统计方法进行了定量分析与评价。MEHTA 等<sup>[9]</sup>开发了一种基于图像的视觉伺服控制器, 用于柑橘目标存在未知运动的情况下将末端执行器调节到目标上。上述研究对扰动状态下水果的检测与定位虽具有一定的效果, 但实际采摘环境下柑橘是无规则扰动, 运动状态复杂, 且随着采摘智能化发展, 对柑橘检测与定位的实时性要求更高, 上述方法的速度与精度不能达到机器人在线实时需求, 无法满足采摘智能化的生产需要。

近年来, 采用机器视觉检测与跟踪目标的运动轨迹, 预测目标的未知运动已广泛应用于汽车的自动驾驶、机器人等领域<sup>[10-11]</sup>, 基于机器视觉和深度学习的目标检测与跟踪算法成为研究热点, 相对于传统的目标检测与跟踪算法, 基于深度学习的方法具有精度高、速度快、鲁

收稿日期: 2024-04-17 修订日期: 2024-06-06

基金项目: 重庆市科委产业化重点专项 (cstc2018jszx-cyzdX0051)

作者简介: 娄欢欢, 研究方向为深度学习、机器视觉。

Email: 1821978260@qq.com

\*通信作者: 李光林, 教授, 博士生导师, 研究方向为传感器与智能检测、农业智能装备。Email: liguanglin@swu.edu.cn

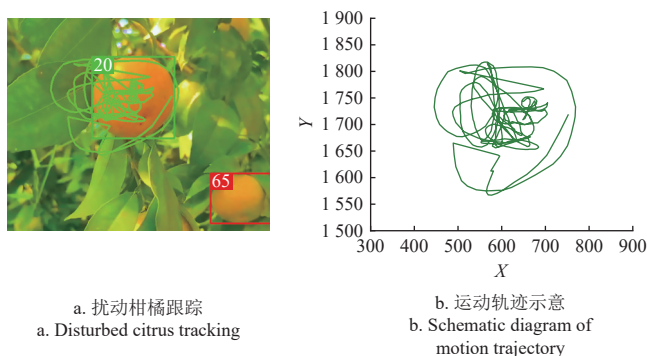
棒性好等特点<sup>[12-14]</sup>。常见的目标检测算法有 YOLO 系列<sup>[15]</sup>、Fast-RCNN 系列<sup>[16]</sup>、SSD<sup>[17]</sup>等。其中, YOLO 系列具有检测精度高、实时检测性能好等特点, 更适合用于动态目标检测<sup>[18]</sup>。常见的目标跟踪算法有 DeepSORT<sup>[19]</sup>、SiamRPN<sup>[20]</sup>系列、Bytetrack<sup>[21]</sup>等。其中, DeepSORT 算法可以提取目标外观特征对目标进行重识别, 拥有更好的跟踪性能, 能同时跟踪多个同类型目标并保持较少的 ID 切换<sup>[22]</sup>。因此, 本文拟采用改进的 YOLOv5s+DeepSORT 算法, 以 YOLOv5s 作目标检测器、DeepSORT 作目标跟踪器, 用于对扰动状态下的柑橘进行在线目标检测与定位。

## 1 材料与方法

通过在线检测与跟踪扰动柑橘的运动, 基于其运动轨迹对柑橘的停留位置预测定位。首先在 YOLOv5s 骨干网络中加入 CBAM<sup>[23]</sup> 注意力机制, 改善对复杂目标的检测能力; 并采用关注方向信息的 SIoU<sup>[24]</sup> 边界框回归损失函数替换原有的 CIoU<sup>[25]</sup> 损失函数, 提升回归收敛速度; 其次在 DeepSORT 算法中, 改进 ReID 特征提取网络, 提升模型检测精度与准确度, 同时融入 Count 计数机制反馈每个跟踪目标的跟踪帧数, 确定最优目标, 实时更新扰动柑橘预测位置, 提升算法预测定位的准确度。同时, 结合深度相机排除距离较远的柑橘, 且限制每次跟踪的目标不超过 3 个, 降低背景柑橘的影响, 提升跟踪速度。最后, 对上述改进模型进行逐一的试验验证, 评估模型的性能与效果, 以期提升采摘机器人连续采摘时的采摘效率。

### 1.1 扰动柑橘的定位方法

在果园环境下, 通过人工模拟机器人采摘时柑橘的扰动, 综合考虑果园环境下的各种复杂因素, 如风力影响、机械臂碰撞、枝条承载重力变化等, 经过多次反复试验, 对扰动状态下柑橘的运动轨迹进行跟踪, 绘制扰动轨迹如图 1 所示, 其中图 1a 是对视频中编号为 20 号的柑橘进行跟踪, 图 1b 是跟踪轨迹在坐标系中的运动轨迹示意。



注: X、Y 轴坐标单位为像素, 下同。

Note: X, Y axis coordinates in pixels. The same below

图 1 复杂扰动状态下柑橘的运动轨迹示意

Fig.1 Schematic diagram motion trajectory of citrus under complex disturbed state

从跟踪轨迹可以看出, 柑橘在复杂扰动条件下引起

的扰动几乎没有规律, 采用建立数学或物理模型的方法研究具有较大难度。本文的在线目标检测与快速定位方法主要分为目标检测、目标跟踪、快速定位三部分, 其流程如图 2 所示。

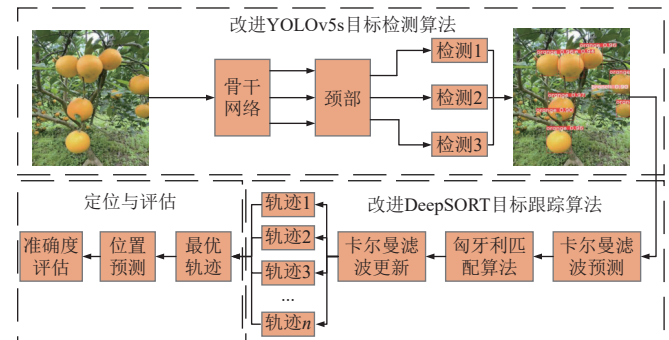


图 2 扰动柑橘检测与定位流程

Fig.2 Disturbance citrus detection and localization framework

1) 在线目标检测: 改进具有高精度和良好实时检测性能的 YOLOv5s 算法进行目标检测, 提升在自然环境下的检测性能。

2) 目标跟踪: 优化具有深度学习特征提取网络的 DeepSORT 多目标跟踪算法, 提升对扰动柑橘的跟踪性能, 主要由卡尔曼滤波算法<sup>[26]</sup>、匈牙利匹配算法<sup>[27]</sup>以及 ReID 网络组成。

3) 快速定位: 结合扰动柑橘的扰动范围随时间逐渐减小, 且扰动逐渐趋于范围内中心点的特点, 无论柑橘受到哪种形式的复杂扰动, 通过扰动柑橘的实时运动轨迹可对其静止时位置做出预测。具体如下: 对扰动柑橘进行跟踪和绘制轨迹的同时, 通过计数每个扰动柑橘的跟踪帧数, 以其中跟踪到帧总数最多的一个目标作为最优跟踪目标, 对其已有轨迹取上、下、左、右 4 个顶点值, 通过 4 个顶点值计算出一个中心点坐标, 即为此刻预测扰动柑橘的 XY 轴坐标, 因扰动柑橘运动范围随时间逐渐减小, 因此对各时刻运动轨迹的 4 个顶点值进行实时更新, 可以更精确预测定位 XY 坐标。同理, Z 轴坐标通过双目相机获取, 在扰动柑橘距离双目相机的最近点与最远点中, 实时取中间点作为 Z 轴坐标。

### 1.2 数据集制作

使用手机和相机分别在晴天、阴天的不同时间段内, 以不同角度采集柑橘树上果实图片共 721 张。为了增加柑橘目标检测的泛化能力, 从网络上获取柑橘图片 30 张, 并从视频数据集中截取图片 220 张, 数据集图片共 971 张, 按 8:2 划分训练集与验证集。为了对柑橘进行有效检测与跟踪, 将数据集分为有遮挡和无遮挡两类, 其中无遮挡的标记为“orange”、有遮挡的标记为“branch”, 通过 Labeling 工具进行标注保存。在数据集训练过程中, 采用数据增强的方式对训练集数据进行扩充, 分别对图片进行缩放、裁剪、翻转、Mosaic 处理等。

视频数据集采用手机和深度相机在固定位置进行拍摄, 位于扰动柑橘正前方约 30~60 cm。采用人工随机采摘模拟柑橘的复杂扰动, 共录制柑橘从扰动状态到完

全静止的视频 21 个，保存为 MP4 格式。随机选择 7 个视频作为 DeepSORT 跟踪算法的特征提取网络的训练集，3 个作为测试集，采用 DarkLabel 2.4 标注软件进行标注并保存。剩余 11 个视频作为预测定位准确率评估数据集。

## 2 算法改进

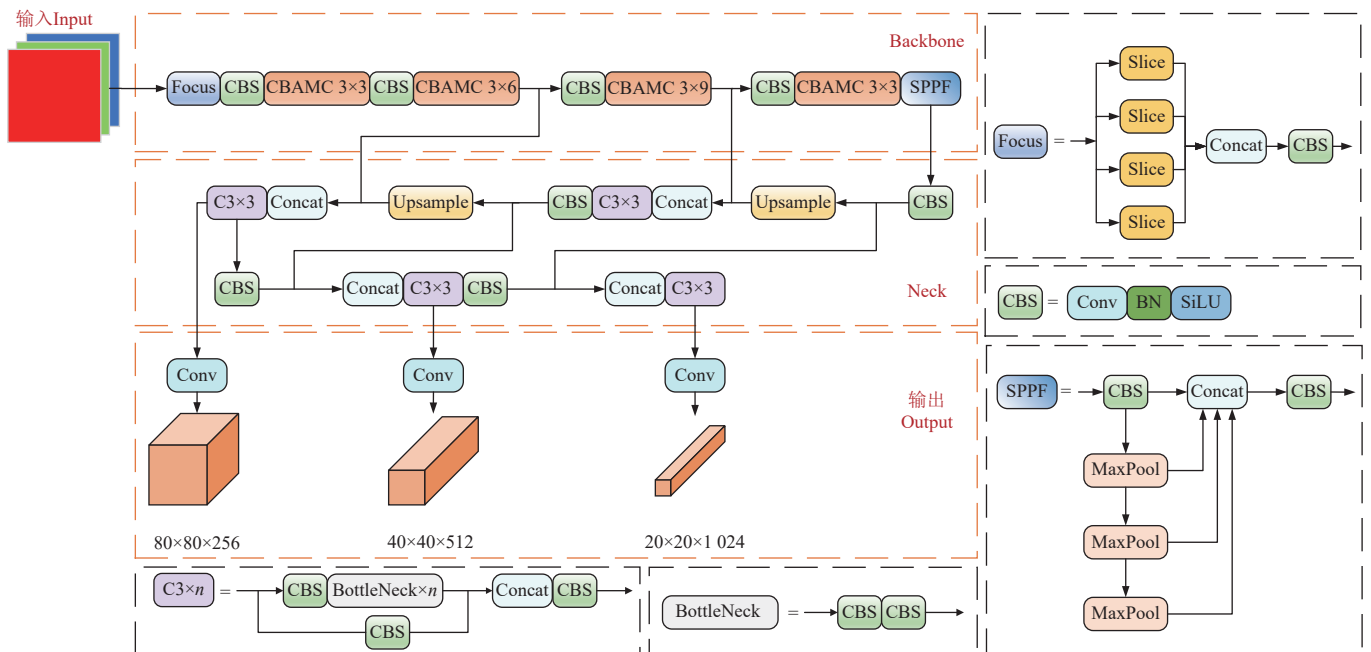
### 2.1 YOLOv5 检测算法改进

YOLO 系列算法是目前主流的目标检测算法，其中，YOLOv5 模型具有参数量少、精度高、速度快、鲁棒性好等优点，大量研究以其为基础将模型广泛应用于实际采摘场景中，且其在与 DeepSort 算法结合时具有稳定的检测性能<sup>[28-31]</sup>。因此，本文选用改进 YOLOv5s 作为基础目标检测模型。该模型主要由以下几部分构成：输入（Input）、

骨干网络（Backbone）、颈部（Neck）和输出（Output）。

#### 2.1.1 改进 YOLOv5s 网络模型

YOLOv5s 应用于自然环境下柑橘检测时，受光照影响较大，且在扰动状态下对柑橘实时检测会产生运动模糊，算法检测的精度和鲁棒性不足<sup>[32-34]</sup>。因此，本文对 YOLOv5s 模型进行如下改进：在 YOLOv5s 的骨干网络中使用融合 CBAM 注意力机制的 CBAMC3 模块取代原有的 C3 模块，使网络模型更加关注于重要的局部特征信息，在保证实时检测的基础上，提高柑橘检测精度，改善小目标和遮挡目标的检测效果。同时，考虑到边界框回归之间的向量角度，使用关注方向信息的 SIoU 损失函数作为边界框损失函数，重新定义惩罚指标，提升模型回归收敛速度，改进模型如图 3 所示。



注：Focus 为切片操作，将通道数扩展 4 倍；Slice 为切片后的特征图；Concat 为特征融合；CBS 为卷积模块，进行卷积操作并进行批归一化处理，由 Conv 卷积、BN 批归一化和 SiLU 激活函数组成；C3 为残差卷积模块，其中 BottleNeck $\times n$  表示进行  $n$  次残差操作；CBAMC3 为改进的 C3 模块；SPPF 为空间金字塔池化模块，对不同尺度的特征图进行池化操作，MaxPool 为最大池化；Upsample 为上采样。

Note: Focus is the slicing operation, which expands the number of channels by 4 times; Slice is the feature map after slicing; Concat is the feature fusion; CBS is the convolution module, which carries out the convolution operation and batch normalisation, and consists of the Conv convolution, the BN batch normalisation and the SiLU activation function; C3 is the residual convolution module, in which BottleNeck $\times n$  means to carry out the residual  $n$  times operations; CBAMC3 is the improved C3 module; SPPF is the spatial pyramid pooling module, which performs pooling operations on feature maps at different scales, with MaxPool being the maximum pooling; and Upsample is the upsampling.

图 3 改进的 YOLOv5s 网络模型

Fig.3 Improved YOLOv5s network model

#### 2.1.2 CBAMC3 模块

CBAM 注意力机制是一个简单有效的前馈卷积神经网络注意力模块，由通道注意力模块 (channel attention module, CAM) 和空间注意力模块 (spatial attention module, SAM) 组成。CBAM 注意力机制将通道和空间两个维度融合，使模块关注重要特征而抑制不必要的特征，提高卷积神经网络 (convolutional neural network, CNN) 的表征能力，有效改善柑橘小目标和遮挡目标难以检测的问题。CBAM 结构如图 4 所示。

通道注意力模块将输入特征  $F(h \times w \times c)$  分别经过宽度和高度的全局最大池化和全局平均池化得到两个  $1 \times 1 \times c$  的特征图，再将两个特征图送入共享多层感知机

(multilayer perceptron, MLP)，输出进行相加操作并经过 Sigmoid 激活函数，最终生成通道注意力特征  $M_c(F)$ 。计算如式 (1) 所示：

$$M_c(F) = \sigma(\text{MLP}(\text{MaxPool}(F)) + \text{MLP}(\text{AvgPool}(F))) \quad (1)$$

式中  $\sigma$  为 Sigmoid 激活函数。最后，将  $M_c(F)$  与输入特征图做乘法生成空间注意力模块的输入特征  $F'$ ，计算如式 (2) 所示：

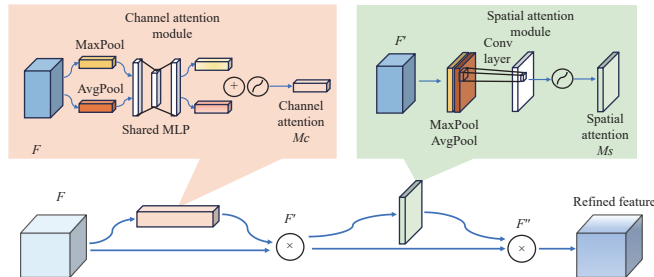
$$F' = M_c(F) \otimes F \quad (2)$$

空间注意力模块先沿着通道轴做全局最大池化与全局平均池化操作得到 2 个  $h \times w \times 1$  的特征图，再将这 2 个

特征图做特征融合并经过  $7 \times 7$  的卷积操作, 最后经过 Sigmoid 函数生成空间注意力特征  $M_s(F')$ 。计算如式 (3) 所示:

$$M_s(F') = \sigma(f^{7 \times 7}([\text{AugPool}(F'); \text{MaxPool}(F')])) \quad (3)$$

式中  $f$  为对特征图做  $7 \times 7$  的卷积操作。



注:  $F$  为通道注意力模块的输入特征、 $F'$  为空间注意力模块的输入特征、 $F''$  为输出特征。

Note:  $F$  is the input feature of the channel attention module,  $F'$  is the input feature of the spatial attention module, and  $F''$  is the output feature.

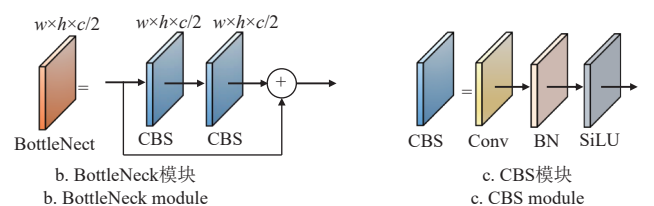
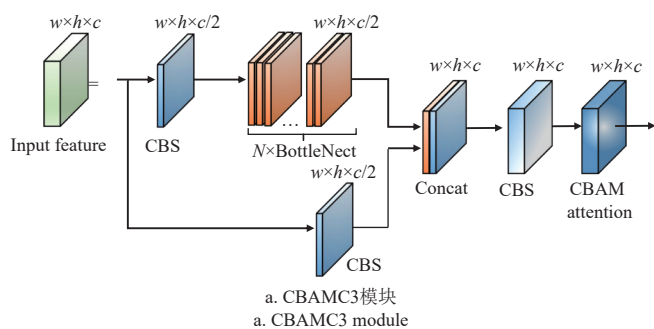
图4 CBAM 注意力机制  
Fig.4 CBAM attention mechanism

与输入特征做乘法生成  $F''$ , 计算如式 (4) 所示:

$$F'' = M_s(F') \otimes F' \quad (4)$$

C3 模块是 YOLOv5 模型骨干网络中的最重要组成模块之一, 结合了多个卷积核进行卷积操作, 由 3 个卷积模块 (CBS)、1 个残差网络模块 (BottleNeck) 组成, 另有 1 个特征融合 (Concat) 操作, 其主要作用是增加模型的感受野并减少计算参数量, 从而提高模型性能与效率。

在 C3 模块第 3 个卷积模块后加入 CBAM 注意力机制构成 CBAMC3 模块, 如图 5a 所示。使用 CBAMC3 模块替换骨干网络中所有 C3 模块, 可以有效增加网络的特征提取能力, 提升检测性能。其中, BottleNeck 模块和 CBS 模块结构如图 5b、5c 所示。



注:  $w$ 、 $h$  和  $c$  分别表示特征图的宽、高和通道数,  $N$  表示有  $N$  个 BottleNeck 模块。  
Note:  $w$ 、 $h$  和  $c$  denote the width, height and number of channels of the feature map, respectively.  $N$  denotes that there are  $N$  BottleNeck modules.

图5 CBAMC3 结构  
Fig.5 CBAMC3 structure

### 2.1.3 损失函数改进

边界框损失函数是计算预测框与标定框之间误差的函数, 使检测的目标定位更加接近真实值。为加快柑橘目标检测模型的回归速度以及提高模型检测准确率, 对边界框损失函数进行优化。在原 YOLOv5s 模型中, 边界框损失函数采用 CIoU 损失函数计算, 但 CIoU 只考虑预测框与标定框中心点之间的距离、宽高比以及重叠部分面积, 未考虑预测框与标定框之间的方向不匹配, 导致回归收敛速度慢。

针对以上问题, 本文采用 SIoU 损失函数作为边界框损失函数, 进一步考虑预测框与标定框之间的向量角度, 重新定义了相关的损失函数, 包含 4 个部分: 角度损失  $\Lambda$ 、距离损失  $\Delta$ 、形状损失  $\Omega$  以及 SIoU 损失。公式如下:

角度损失:

$$\Lambda = 1 - 2\sin^2\left(\arcsin\left(\frac{c_h}{\sigma'}\right) - \frac{\pi}{4}\right) \quad (5)$$

式中  $c_h$  为预测框与标定框中心点的高度差,  $\sigma'$  为预测框与标定框中心点的距离。

根据角度损失  $\Lambda$ , 距离损失  $\Delta$  公式如下:

$$\Delta = \sum_{t=x,y} (1 - e^{-(2-\Lambda)\rho_t}) \quad (6)$$

式中  $\rho_t$  表示  $\rho_x$  或  $\rho_y$ , 其中  $\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{C_w}\right)^2$ ,  $\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{C_h}\right)^2$ ,  $C_w$  和  $C_h$  分别是预测框与标定框最小外接矩形的宽和高,  $b_{c_x}^{gt}$  和  $b_{c_y}^{gt}$  分别是标定框中心横坐标与纵坐标,  $b_{c_x}$  和  $b_{c_y}$  分别是预测框中心横坐标与纵坐标。

形状损失  $\Omega$  如下:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (7)$$

式中  $\omega_t$  表示  $\omega_w$  或  $\omega_h$ , 其中  $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$ ,  $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$ ,  $w$  和  $h$  分别为预测框的宽与高,  $w^{gt}$  和  $h^{gt}$  分别为标定框的宽与高,  $\theta$  表示对形状损失的关注度, 其数值为 1。

IoU 损失如下:

$$\text{IoU} = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \quad (8)$$

式 (8) 表示预测框与标定框之间的交并比, 其中  $B \cap B^{GT}$  表示两框之间的交集,  $B \cup B^{GT}$  表示两框之间的并集。

因此, SIoU 损失函数的定义如下:

$$L_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \quad (9)$$

## 2.2 DeepSORT 跟踪算法改进

DeepSORT 算法采用了卡尔曼滤波算法与匈牙利匹配算法作为基础, 具有 ReID 特征提取网络以及采用级联匹配策略, 同时考虑了目标的运动特征与外观特征。由于柑橘果实外观极为相似, 在使用 DeepSORT 算法跟踪扰动柑橘时, 原始的 ReID 网络结构简单, 特征提取效果并不理想, 跟踪过程中仍会导致大量的 ID 切换。因

此本文对原始的 ReID 网络进行改进，增强 ReID 网络的特征提取能力，减少跟踪过程中的 ID 切换。同时，在算法每次进行卡尔曼滤波更新后，融入 Count 计数器累加跟踪的帧数值，在跟踪后进行最优轨迹判断。改进的 DeepSORT 算法及其结构如图 6 所示。

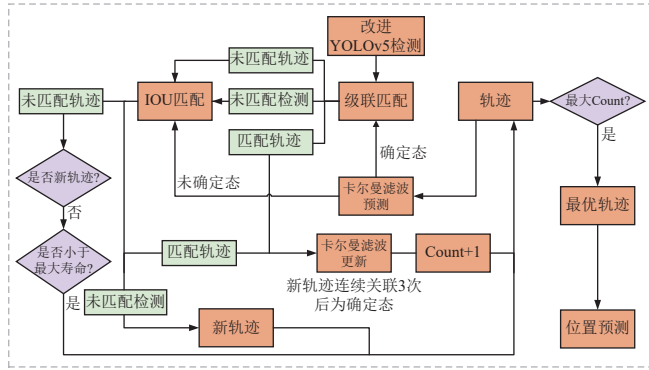


图 6 改进的 DeepSORT 算法与位置预测结构

Fig.6 Improved DeepSORT algorithm and localization structure

### 2.2.1 ReID 网络改进

DeepSORT 算法原始的 ReID 网络由 2 个卷积层、1 个最大池化层和 6 个残差网络模块组成，输出为 128 维向量表示目标的外观特征，并将特征投影到单位超球面上。针对柑橘果实具有极高相似性导致 ID 大量切换的问题，结合柑橘形状近似圆的特点，改进算法的图片输入尺寸为  $64 \times 64$  (宽  $\times$  高)，提升柑橘重识别准确率；同时，加深 ReID 网络模型深度，采用具有更好性能的特征提取网络提取特征，该网络在 ZQPei<sup>[35]</sup> 提出的 ReID 网络上进行改进，其结构如表 1 所示。

表 1 改进的 ReID 网络结构

Table 1 Improved ReID network structure

名称 Name	大小/步长 Size/stride	输出大小 Output size
Input	-	$3 \times 64 \times 64$
Conv	$3 \times 3 / 1$	$64 \times 64 \times 64$
MaxPool	$3 \times 3 / 2$	$64 \times 32 \times 32$
BasicBlock $1 \times 2$	$3 \times 3 / 1$	$64 \times 32 \times 32$
BasicBlock $2 \times 2$	$3 \times 3 / -$	$128 \times 16 \times 16$
BasicBlock $2 \times 2$	$3 \times 3 / -$	$256 \times 8 \times 8$
BasicBlock $2 \times 2$	$3 \times 3 / -$	$512 \times 4 \times 4$
AvgPool	$4 \times 4 / 1$	$512 \times 1 \times 1$
Dense	-	256

改进的特征提取网络结构主要包括 1 个卷积层 (Conv)、1 个最大池化层 (MaxPool)、8 个残差模块 (BasicBlock) 以及 1 个平均池化层 (AvgPool)，经过全链接层 (Dense) 后输出为 256 维向量表示目标的外观特征。

### 2.2.2 融入 Count 计数机制

进行柑橘采摘时，相机检测范围内常存在多个柑橘，且在扰动状态下柑橘位置时刻发生变化，在定位时，难以确定对哪一个柑橘进行定位准确率最高。针对此问题，本文在整个 DeepSORT 算法的基础上，融入 Count 计数机制，在每次卡尔曼滤波更新后生成新轨迹点时实现计数+1，用于统计每个跟踪目标的被跟踪的帧数。在跟踪时，通过 Count 的值，可以得到每个目标当前在时间段内跟踪的总帧数，最后比较 Count 值的大小可以确定一

个最优跟踪目标。其中，最大 Count 值表示跟踪到目标的轨迹最完整，对此目标预测会得到最好的预测结果。

### 2.3 实时更新预测坐标值

准确的定位是对目标精准采摘的关键，柑橘在受到外界因素产生扰动时，扰动幅度会随时间逐渐减小，直至幅度为 0 时保持静止状态。在复合扰动状态下，分别绘制在时间为 1、2、3、4、5、7、10 s 时以及柑橘最终停留位置时运动轨迹示意图，如图 7 所示。

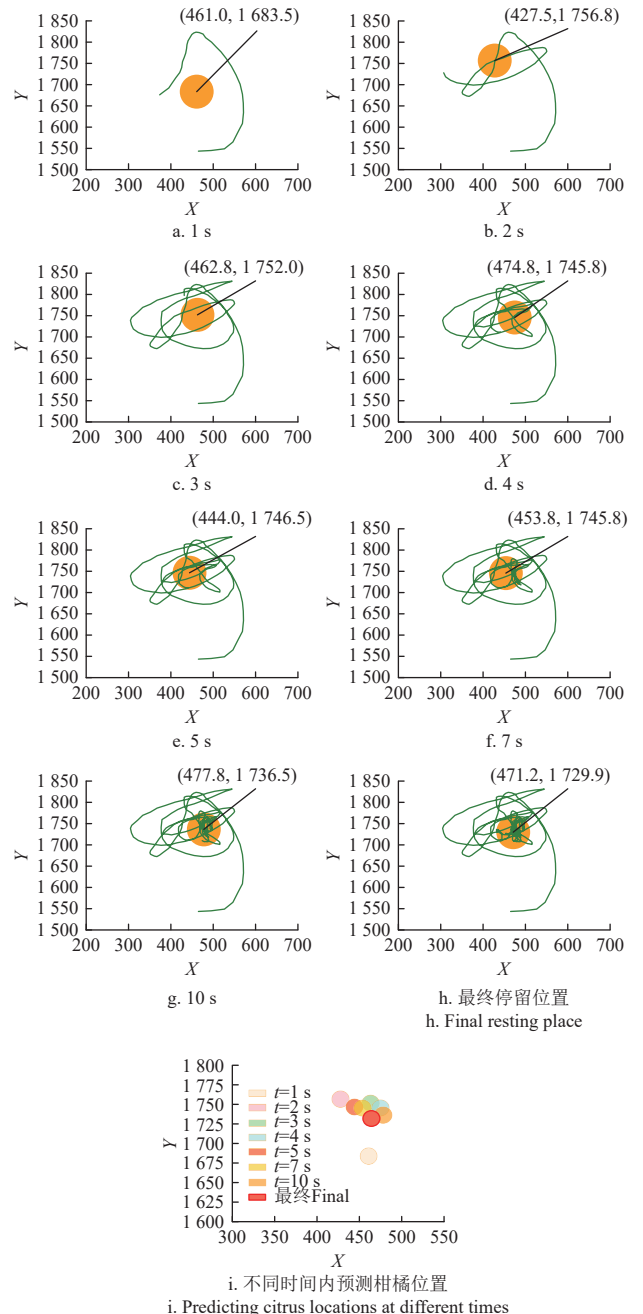


图 7 不同时刻预测中心点位置值变化

Fig.7 Changes in predicted center point position values at different times

图 7i 表示不同时间内柑橘预测位置分布，可以看出，1 s 时柑橘扰动范围较大，3 s 时扰动范围基本趋于稳定，当大于 3 s 时其预测定位点与最终停留位置只有微小差距。即时间  $t$  值越大，轨迹更新范围越小，其位置预测

的坐标值越接近最终停留位置的坐标值。因此,在算法中,采用对每个坐标值实时进行最大值最小值比较,对XY平面中运动轨迹的4个顶点坐标值进行实时更新,确保输出XY定位坐标是当前时刻最新的坐标值,即此刻是轨迹范围最小时的运动轨迹。同理,对双目相机测定的Z轴坐标值进行实时更新,提升扰动柑橘预测停留位置的三维坐标定位精度。

#### 2.4 结合深度相机优化跟踪目标

果园环境复杂,在采摘时存在多个柑橘目标,距离较远的目标在采摘过程中通常被视为背景目标,背景目标的存在会影响跟踪算法的跟踪速度。在采摘过程中,根据采摘机器人的采摘距离,限制深度相机到扰动柑橘的距离在某个区间内,可以排除背景目标的干扰,减少算法运行时的计算量,从而提升算法整体定位速度。同时,在限定的这个区间内仍可能存在多个被检测到的目标,若对这些目标都进行跟踪,跟踪速度可能受到影响。

随机选取3个视频,对不同数目的扰动柑橘进行跟踪,分别在柑橘数目为1、2、3、4、5、7、10个以及不限制柑橘数目时进行测试,同时对不同跟踪数目下的速度取平均值,统计结果如表2所示。

表2 不同柑橘跟踪数目对跟踪速度影响

数目 Number	1	2	3	4	5	7	10	不限制 Unlimited
帧率 FPS/(帧·s <sup>-1</sup> )	37.4	35.1	31.4	29.1	18.4	16.4	15.4	11.2

由表2可以看出,在跟踪柑橘时,不同数目的柑橘的跟踪速度具有一定的差异。由于需要保证跟踪算法的实时跟踪性能,本文中,限制柑橘目标跟踪数为3个,此时的跟踪速度比不限制柑橘数目时的跟踪速度提升20.2帧/s。因此,通过深度相机优化跟踪目标,保证跟踪精度的同时,确保跟踪的实时性。

### 3 试验与评估

#### 3.1 评估指标

##### 3.1.1 改进YOLOv5s算法评估指标

采用准确率(Precision,  $P$ )、召回率(Recall,  $R$ )、平均检测精度均值(mean average precision, mAP)、检测速率和模型权重大小作为改进的YOLOv5s目标检测算法的评估指标。其中 $P$ 、 $R$ 、mAP的计算如式(10)~(13)所示:

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (10)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (11)$$

$$A_p = \int_0^1 P dR \quad (12)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N A_p(i) \quad (13)$$

其中 $T_p$ 表示被正确地检测为正样本的数量; $F_p$ 表示被错误地检测为正样本的数量; $F_N$ 表示漏检的正样本数量;式(12)中的 $A_p$ 表示 $P$ - $R$ 曲线的面积; $N$ 表示检测类别数,本文中分为两类,故 $N=2$ 。

##### 3.1.2 改进DeepSORT算法评估指标

DeepSORT算法的性能评估采用TrackEval工具,主要的评估指标有:跟踪准确度MOTA(multi-object tracking accuracy)、跟踪精度MOTP(multi-object tracking precision)、目标ID切换次数IDSW(identity switches)以及跟踪速率。

MOTA是除去跟踪过程中所有漏检、误检以及错误匹配后的正确匹配占比,数值越大表明跟踪性能越好,计算式如(14)所示。

$$\text{MOTA} = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t} \quad (14)$$

其中 $t$ 为帧索引值; $GT_t$ 为第 $t$ 帧真实标定框数量; $FN_t$ 为第 $t$ 帧未跟踪到的目标数量; $FP_t$ 为第 $t$ 帧错误跟踪目标的数量; $IDSW_t$ 为第 $t$ 帧ID跳变的次数。

MOTP主要衡量标定框与预测框之间的位置误差,其计算式如(15)所示。

$$\text{MOTP} = \frac{\sum_t d_{t,i}}{\sum_t c_t} \quad (15)$$

其中 $c_t$ 为第 $t$ 帧中匹配成功的数目; $d_{t,i}$ 为第 $t$ 帧检测框与真实框的距离。

IDSW表示跟踪过程中目标ID切换的次数,其数值越小越好。

##### 3.1.3 扰动柑橘停留位置预测评估指标

采摘机器人进行柑橘采摘过程中,机械臂末端执行器通常适合采摘不同尺寸的柑橘,采摘中等或较小尺寸柑橘时,末端执行器有较大部分剩余空间。如图8a所示,末端执行器手爪上下、左右边缘与柑橘之间存在一定的间隙,手爪向内运动将柑橘置于果柄可剪切位置,由剪刀进行剪切,柑橘上下左右的间隙为剩余空间。因此,图像中的定位坐标(像素坐标系)与末端执行器坐标(末端执行器坐标系)具有一定的偏差,仍可对柑橘进行采摘。利用末端执行器在一定的定位误差范围内可以进行正常采摘的特点,将扰动柑橘在像素坐标下预测定位坐标与末端执行器坐标系联系起来,建立基于误差范围内的准确度评估指标。如图8b、8c所示。

柑橘与末端执行器的横向尺寸最大允许误差为 $(N+M)$  mm,纵向尺寸最大允许误差为 $(X+Y)$  mm,根据末端执行器坐标系向像素坐标系的变换,可以得到图像中指标圆的半径为 $R$ ,圆的半径 $R$ 计算公式为

$$R = \frac{(N+M+X+Y)}{4} \times \varnothing \quad (16)$$

其中 $\varnothing$ 表示从末端执行器坐标系转换为像素坐标系的坐标变换参数,其值根据深度相机与机器人的手眼标定获取。将指标圆均分为100份,从1%至100%作为评估准确度指标,如图9所示。

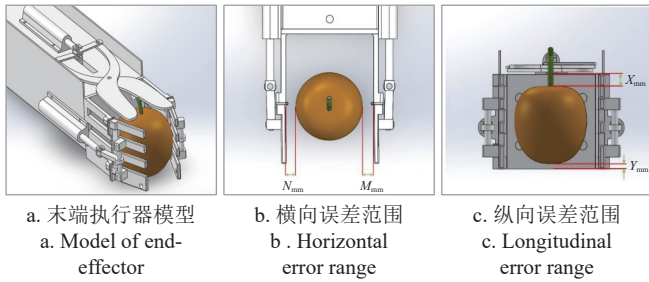
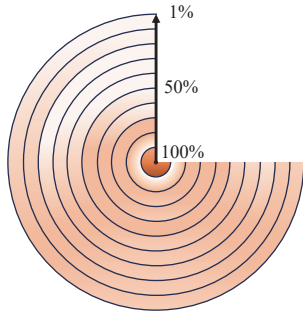


图 8 末端执行器与误差范围  
Fig.8 End-effector and error range



注：准确度数值由圆心向外逐渐减少。  
Note: Accuracy value decreases gradually from the centre of the circle outwards.

图 9 准确度指标圆示意图  
Fig.9 Schematic diagram of accuracy indicator circle

若预测坐标在圆外侧，此时两坐标中心点之间的欧式距离  $d > R$ ，则定位预测失败，如图 10a 所示，图例中“Orange”表示柑橘真实停留位置，“Predicted”表示柑橘的预测位置；当预测坐标与实际停留坐标完全重合时准确率  $p$  为 100%；当  $d < R$  时，此时预测准确率数值依次由圆心向外减小，直至圆的边缘为 1%，如图 10b 所示，其准确率计算公式为

$$p = \frac{d}{R} \times 100\% \quad (17)$$

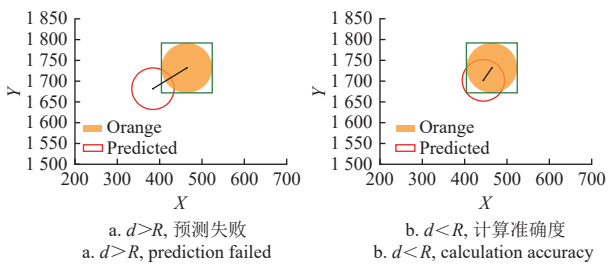


图 10 预测定位示意图  
Fig.10 Schematic diagram of predicted positioning

### 3.2 性能评估及分析

本文中数据集的训练采用的试验平台为远程服务器，Linux 操作系统，显卡型号为 Tesla V100-PCIE，运行内存为 32 GB，Python 版本为 3.7，基于 Pytorch 框架，torch 版本 1.13。

机器人型号为 AUBO-i10 轻型 6 自由度协作机器人，有效负载 10 kg，工作范围为半径 1350 mm 的球体空间区域。双目相机型号为 ZED 2i，并排输出分辨率  $2 \times (1280 \times 720)$ ，深度范围 0.3~20.0 m。

#### 3.2.1 改进 YOLOv5s 性能评估

YOLOv5s 模型输入图片尺寸为  $640 \times 640$  像素，优化器采用随机梯度下降法 SGD(stochastic gradient descent)，批量大小 (batch\_size) 为 32，进程 (works) 为 8，学习率为 0.01，训练次数 800 次。

为了验证 CBAM 注意力机制的有效性，分别在 YOLOv5s 主干网络中同一位置中加入不同的注意力机制进行对比试验，对比组分别有 YOLOv5s、YOLOv5s+CA、YOLOv5s+SENet、YOLOv5s+SimAM、YOLOv5s+CBAM，试验结果如表 3 所示。

表 3 不同注意力机制性能对比结果  
Table 3 Performance comparison results of different attention mechanisms

模型 Model	准确率 Precision /%	召回率 Recall /%	平均检测精度均值 Mean average precision mAP/%	帧率 Frame per second FPS/(帧·s <sup>-1</sup> )
YOLOv5s	92.6	87.1	92.2	69.4
+CA	93.7	85.1	91.7	68.1
+SENet	92.9	88.6	92.5	66.9
+SimAM	93.4	84.9	92.3	68.3
+CBAM	95.0	88.9	93.4	68.7

由表 3 可以看出，与 YOLOv5s、YOLOv5s+CA、YOLOv5s+SENet、YOLOv5s+SimAM 相比，YOLOv5s+CBAM 的准确率分别提高 2.4、1.3、2.1、1.6 个百分点，召回率分别提高 1.8、3.8、0.3、4.0 个百分点，平均检测精度分别提高 1.2、1.7、0.9、1.1 个百分点。因此，在模型检测帧率没有明显差异的情况下，YOLOv5s+CBAM 模型均比其他对比组注意力机制模型表现优异，验证了 YOLOv5s+CBAM 的具有较好的性能。

同时，为了验证 SIoU 损失函数对模型性能的提升，在 YOLOv5s 模型中使用不同的损失函数进行对比试验。试验包括 CIoU、DIoU、EIoU、GIoU、SIoU 共 5 组损失函数，试验结果如表 4 所示。

表 4 不同损失函数性能对比  
Table 4 Performance comparison of different loss functions

模型 Model	准确率 Precision/%	召回率 Recall/%	平均检测精度均值 Mean average precision mAP/%	帧率 Frame per second FPS/(帧·s <sup>-1</sup> )
CIoU	92.6	87.1	92.2	69.4
DIoU	94.1	85.3	91.8	68.9
EIoU	95.1	84.7	91.6	70.4
GIoU	95.6	84.6	91.6	69.9
SIoU	93.3	87.0	92.3	71.3

由表 4 可以看出，与 CIoU、DIoU、EIoU、GIoU 损失函数相比，SIoU 损失函数的 mAP 分别提升 0.1、0.5、0.7、0.7 个百分点，FPS 分别提升 1.9、2.4、0.9、1.4 帧/s。在模型准确率和召回率没有较大差异的情况下，SIoU 损失函数的 mAP 和检测帧率优于其他损失函数模型。因此，SIoU 损失函数可以有效提升模型的检测速率，同时提升检测性能。

为了验证改进的 YOLOv5s 的有效性，设计了消融试验，对加入的 CBAM 注意力机制和 SIoU 损失函数进行性能验证，试验结果如表 5 所示。

表 5 改进的 YOLOv5s 消融试验性能对比  
Table 5 Comparison of the performance of the improved YOLOv5s ablation experiments

序号 Index	YOLOv5s	SIOU	CBAMC3	准确率 Precision/%	召回率 Recall/%	平均检测精度均值 Mean average precision mAP/%	帧率 Frame per second FPS/(帧·s <sup>-1</sup> )	模型大小 Model size /MB
1	√			92.6	87.1	92.2	69.4	14.2
2	√	√		93.3	87.0	92.3	71.3	14.1
3	√		√	95.0	88.9	93.4	68.7	14.3
4	√	√	√	96.5	88.7	93.3	69.3	14.2

表 5 试验结果表明,改进的 YOLOv5s 在原 YOLOv5s 的基础上,准确率提升 3.9 个百分点、召回率提升 1.6 个百分点、平均检测精度均值提升 1.1 个百分点,检测速率和模型权重大小没有明显变化。

使用上述模型对不同数量柑橘以及不同复杂背景的柑橘图片进行检测,检测效果如图 11 所示,统计漏检和误检数量如表 6 所示。

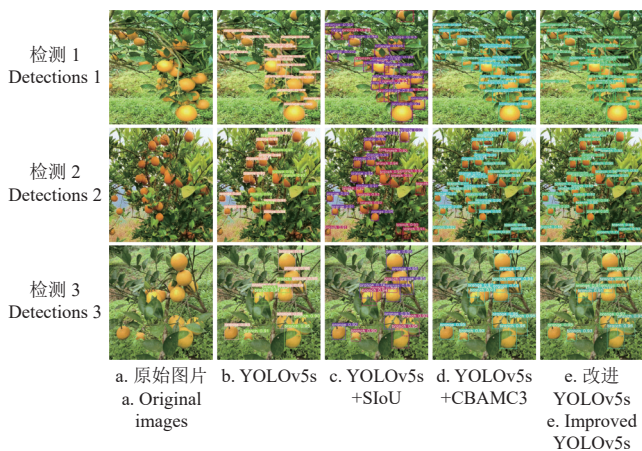


图 11 YOLOv5s 及其改进模型检测效果对比

Fig.11 Comparison of detection performance between YOLOv5s and its improved model

表 6 YOLOv5s 及其改进模型漏检与误检统计

Table 6 Missed and false detection statistics for YOLOv5s and its improved models

检测 Detection	漏检 与误检 Missed and false detections	YOLOv5s	YOLOv5s+ SIOU	YOLOv5s+ CBAMC3	改进 YOLOv5s Improved YOLOv5s
1	漏检	6	1	1	0
	误检	3	3	1	1
2	漏检	11	4	1	1
	误检	4	3	4	2
3	漏检	0	0	0	0
	误检	2	1	1	1

结合图 11 与表 6 统计结果可以看出:改进的 YOLOv5s 模型的漏检和误检数量均少于其他模型,该模型在检测柑橘小目标和遮挡目标时具有更好的检测效果,总体检测性能优于其他 3 个模型。

### 3.2.2 改进 DeepSORT 性能评估

将 7 个视频训练数据集分别用原始 DeepSORT 网络模型和改进 DeepSORT 网络模型进行训练,训练次数 300 轮并得到相应的权重文件。为了验证改进 ReID 网络跟踪性能,统一采用改进的 YOLOv5s 作为目标检测器,分别将网络模型用于 3 个视频评估数据集,并取均值得到结果如表 7 所示。

表 7 DeepSORT 中采用不同的 ReID 网络性能评估试验

Table 7 Performance evaluation experiments using different ReID networks in DeepSORT

模型 Model	视频 Video	跟踪 准确度 MOTA/%	跟踪 精度 MOTP/%	ID 切换次数 IDSW /次	FPS/(帧·s <sup>-1</sup> )
原始 ReID 网络 Original ReID network	1	60.4	81.4	67	34.3
	2	57.0	74.7	89	33.5
	3	60.7	79.3	78	33.7
均值 Means	/	59.4	78.5	78	33.8
改进 ReID 网络 Improved ReID network	1	70.3	88.4	39	32.1
	2	66.4	78.6	58	30.1
	3	69.1	84.9	42	30.9
均值 Means	/	68.6	83.9	46	31.0

由表 7 可以看出,改进 ReID 网络的帧率虽降低 2.8 帧/s,但跟踪准确度与跟踪精度分别提升 9.2、5.4 个百分点,ID 切换次数减少 32 次,验证了改进的 ReID 网络具有更好的跟踪性能。

表 8 为改进的 YOLOv5s+DeepSORT 算法与其他主流的跟踪算法的试验对比。每组采用 3 个相同的视频进行跟踪试验,并取每组各个评估指标的平均值。

表 8 主流跟踪算法性能对比试验

Table 8 Performance comparison test of different tracking algorithms

模型 Model	跟踪 准确度 MOTA/%	跟踪 精度 MOTP/%	ID 切换次数 IDSW/次	帧率 FPS /(帧·s <sup>-1</sup> )
YOLOv5s+SORT	40.2	48.0	127	46.3
YOLOv7s+DeepSORT	56.2	68.1	79	38.1
YOLOv8s+Bot-SORT	67.1	78.2	72	32.1
Byte-Tracker	62.3	77.1	65	34.6
本文	70.4	82.1	62	31.3

表 8 的对比试验结果表明：本文算法的跟踪准确度与 YOLOv5s+SORT、YOLOv7+DeepSORT、YOLOv8+Bot-SORT、Byte-Tracker 相比分别提升 30.2、14.2、3.3、8.1 个百分点；跟踪精度分别提升 34.1、14.0、3.9、5.0 个百分点；ID 切换次数分别减少 65、17、10、3 次；本文算法的检测速率是试验组中最低的，为 31.3 帧/s。综上，

试验表明本文算法具有最佳跟踪性能，帧率虽然低于其他算法，但仍可满足帧率大于 30 帧/s 的实时跟踪需求。

为了验证本文中所采用的检测算法与跟踪算法改进前后的有效性，采用逐次加入变量的方法对该算法进行对比试验，选取 3 个相同的视频分别输入模型进行试验评估，得到试验结果如表 9 所示。

表 9 算法改进前后对比试验  
Table 9 Comparison test before and after algorithm improvement

序号 Index	YOLOv5s	改进 YOLOv5s Improved YOLOv5s	DeepSORT	改进 DeepSORT Improved DeepSORT	MOTA/%	MOTP/%	IDSW/%	FPS/ (帧·s <sup>-1</sup> )
1	√		√		53.1	67.2	81	34.2
2	√			√	65.2	66.8	74	32.3
3		√	√		66.7	72.0	78	34.7
4		√		√	<b>70.4</b>	<b>82.1</b>	<b>62</b>	31.3

从表 9 中的试验结果可以看出，改进 YOLOv5s+改进 DeepSORT 算法的各项评估指标均高于改进前，检测速率有所下降，但也能达到实时检测的效果，说明了改进后的算法性能具有明显的提升。

### 3.2.3 扰动柑橘定位准确度评估

随机测量 100 个柑橘的横向及纵向直径，得到横向平均直径为 67.4 mm，纵向平均直径 69.2 mm，采用横向与纵向尺寸均为 100 mm 的末端执行器，计算得到位置预测评估指标圆的半径  $R=87$  像素。随机选取 5 个视频作为不同时间内预测定位准确度评估对象，预测时间分别为 1、3、5、7、10 s，计算得到不同时间内预测准确度结果如表 10 所示，对每个时间内的准确度取平均值，结果如表 11 所示。

表 10 不同时刻定位准确度评估  
Table 10 Assessment of positioning precision in different time periods

序号 Index	时间 Time/s	准确度 Precision/%
视频 1 Video 1	1	0
	2	42.1
	3	78.4
	5	80.3
	7	81.2
视频 2 Video 2	10	96.6
	1	12.1
	2	54.2
	3	80.4
	5	83.2
视频 3 Video 3	7	84.3
	10	96.2
	1	43.2
	2	67.9
	3	86.7
视频 4 Video 4	5	87.4
	7	88.4
	10	97.9
	1	0
	2	33.7
视频 5 Video 5	3	76.2
	5	77.8
	7	84.6
	10	90.1
	1	51.2
视频 5 Video 5	2	67.2
	3	87.9
	5	89.8
	7	93.1
	10	93.9

从表 10 以及表 11 可以看出，对于不同的视频，当时间 1、2、3、5、7、10 s 时，5 个视频预测定位的平均准确率分别为 21.3%、53.0%、81.9%、83.7%、86.1%、94.9%。当时间为 3 s 时，扰动柑橘位置预测效果如图 12 所示，其中绿色线条是柑橘的运动轨迹，红色圆圈表示预测准确度评估圆。

表 11 5 个视频在不同时刻定位的平均准确度评估  
Table 11 Mean precision assessment of positioning of 5 Video at different times

时间 Time/s	1	2	3	5	7	10
平均准确度 Means precision/%	21.3	53.0	81.9	83.7	86.1	94.9



a. 对 10 号柑橘预测定位  
a. Predicted positioning for Citrus 10  
b. 对 2 号柑橘预测定位  
b. Predicted positioning for Citrus 2

图 12 扰动柑橘位置预测效果

Fig.12 Predicted effects of disturbed citrus locations

其中图 12a 中对编号为 10 号和 2 号的柑橘进行轨迹绘制，经过 Count 计数器累加跟踪帧数后，10 号柑橘 Count 值大于 2 号柑橘，因此对 10 号柑橘进行预测定位，计算得到柑橘定位的准确率为 83.1%。同理，在图 12b 中分别对编号为 2 和 5 的柑橘进行跟踪，对 2 号柑橘进行预测定位，预测准确率为 88.3%。

### 3.2.4 扰动柑橘预测定位对抓取效率评估

为了验证本文提出的快速定位方法在采摘过程中应用的有效性，在试验室采用盆栽柑橘进行模拟抓取试验，试验场景如图 13 所示。分别模拟实际采摘过程中柑橘可能出现的各种扰动状态，将柑橘沿不同方位随机摆动(如上下摆动、左右摆动、前后摆动等，摆幅约为 10 cm)，如图 14 所示。扰动柑橘不仅需要满足实时快速定位，同时还具有较高的定位准确率，结合前文选择时间为 3 s 作为扰动柑橘最佳定位时间，通过对抓取时间进行统计，

比较扰动状态下的柑橘预测定位时间为 3 s 和未预测定位两种情况下的抓取时间。并取每个柑橘抓取的平均时间得到结果如表 12 所示。

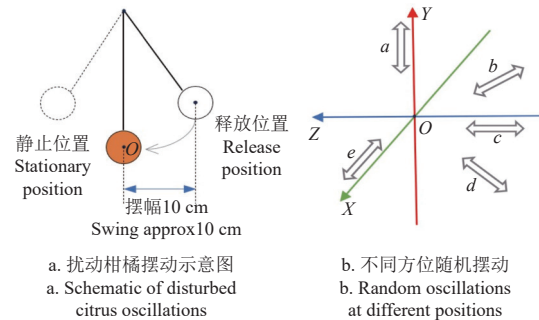


1. 控制箱 2. 电脑端 3. 机器人 4. 深度相机 5. 末端执行器 6. 柑橘  
1. Control box 2. Computer 3. Robot 4. Depth camera 5. End-effector 6. Orange

图 13 室内模拟抓取试验  
Fig.13 Indoor simulated gripping test

由表 12 可以看出, 柑橘在无预测定位下的平均抓取

时间为 18.4 s, 在时间为 3 s 时预测定位下的平均抓取时间为 12.8 s, 时间缩短 5.6 s, 平均抓取效率提升 30.4%, 验证了采用本文中的定位方法对扰动柑橘采摘效率提升的有效性。



注: 图 b 中的 a、b、c、d 和 e 分别表示随机上下摆动、左斜 45° 摆动、前后摆动、右斜 45° 摆动和左右摆动。

Note: a, b, c, d, and e in Fig. b indicate random up-and-down oscillations, left-oblique 45° oscillations, back-and-forth oscillations, right-oblique 45° oscillations, and left-right oscillations, respectively.

图 14 模拟扰动柑橘摆动示意图

Fig.14 Schematic of simulated disturbance for citrus oscillations

表 12 使用定位算法与未使用定位算法对扰动柑橘抓取效率影响对比试验

Table 12 Comparison test of picking without predicted positioning and with predicted positioning time 3 s

试验编号 Test No.	1	2	3	4	5	6	7	8	9	10	均值 Means
未预测定位 Unpredicted positioning	16.2	19.8	21.2	17.3	18.6	17.4	16.2	18.4	21.2	17.6	18.4
3 s 预测定位 3 s predicted positioning	12.4	11.9	13.4	12.7	12.3	13.0	12.5	12.1	13.8	14.1	12.8

## 4 结论

1) 扰动状态下柑橘在线目标检测与快速定位对机械化自动采摘有着重要的意义, 本文采用机器视觉和深度学习的方法具有时间短、准确度高的特点, 为扰动状态下的柑橘检测与快速定位提供一种新的途径。同时, 该方法可为其他扰动状态下水果快速定位提供参考。

2) 在 YOLOv5s 骨干网络的 C3 模块中加入 CBAM 注意力机制, 并使用 Siou 损失函数得到改进的 YOLOv5s 检测算法, 改善了对柑橘小目标和遮挡目标的检测性能, 提升了检测速度。改进的 YOLOv5s 算法准确率、召回率、平均检测精度均值比改进前分别提升了 3.9、1.6、1.1 个百分点, 平均检测速率 69.3 帧/s。

3) 改进 DeepSORT 中的 ReID 网络, 更适用于柑橘目标的特征提取, 同时采用 Count 计数机制反馈柑橘跟踪帧数以选择最优目标进行位置预测。改进 DeepSORT 算法的跟踪精确度, 跟踪准确度值分别提升了 9.2、5.4 个百分点, ID 切换次数减少了 32 次。

4) 对扰动柑橘的运动轨迹预测时, 采用对 XY 平面运动轨迹的上、下、左、右四个顶点值和 Z 轴深度坐标实时更新的方法, 可以更加精确对柑橘进行定位。并结合深度相机排除背景柑橘的干扰, 同时限制每一次跟踪柑橘数为 3 个, 跟踪速度比不限制柑橘跟踪数时提升 20.2 帧/s, 有效提升了扰动柑橘的定位速度。

5) 对扰动柑橘定位, 当时间分别为 1、2、3、5、7

和 10 s 时, 扰动柑橘定位的平均准确率分别为 21.3%、53.0%、81.9%、83.7%、86.1%、94.9%。实验室采用盆栽柑橘进行模拟抓取试验, 且将柑橘预测定位时间定为 3 s 时, 每个柑橘平均抓取时间为 12.8 s, 比无预测定位情况下其时间缩短 5.6 s, 平均采摘效率提升 30.4%。

## [参考文献]

- 毛海欧, 祁春节. 新冠肺炎疫情对中国柑橘贸易的影响研究[J]. 中国果树, 2023(8): 131-135.  
MAO Haiou, QI Chunjie. A study on the impact of the new crown pneumonia epidemic on China's citrus trade[J]. China Fruits, 2023(8): 131-135. (in Chinese with English abstract)
- 鲍秀兰, 马志涛, 马萧杰, 等. 丘陵果园自然环境下柑橘采摘机器人设计与试验[J/OL]. 农业机械学报: 1-13. [2024-03-24]. <http://dx.doi.org/10.6041/j.issn.1000-1298.2024.04.012>  
BAO Xiulan, MA Zhitao, MA Xiaojie. et al. Design and Experiment of Citrus Picking Robot in Hilly Orchard Natural Environment[J/OL]. Transactions of the Chinese Society for Agricultural Machinery: 1-13[2024-03-24]. <http://dx.doi.org/10.6041/j.issn.1000-1298.2024.04.012> (in Chinese with English abstract)
- 赵春江, 范贝贝, 李瑾, 等. 农业机器人技术进展、挑战与趋势[J]. 智慧农业(中英文), 2023, 5(4): 1-15.  
ZHAO Chunjiang, FAN Beibei, LI Jin, et al. Agricultural robots: Technology progress, challenges and trends[J]. Smart Agriculture, 2023, 5(4): 1-15. (in Chinese with English abstract)
- SHARMA V, TRIPATHI A K, MITTAL H. Technological

- revolutions in smart farming: Current trends, challenges & future directions[J]. *Computers and Electronics in Agriculture*, 2022, 201: 107217.
- [5] 熊俊涛, 邹湘军, 彭红星, 等. 扰动柑橘采摘的实时识别与采摘点确定技术[J]. *农业机械学报*, 2014, 45(8): 38-43. XIONG Juntao, ZHOU Xiangjun, PENG Hongxing, et al. Real-time identification and picking point localization of disturbance ditrus picking[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(8): 38-43. (in Chinese with English abstract)
- [6] XIONG J T, HE Z L, et al. Visual positioning technology of picking robots for dynamic litchi clusters with disturbance[J]. *Computers and Electronics in Agriculture*, 2018, 151: 226-237.
- [7] LIU T H, KANG H W, et al. ORB-Livox: A real-time dynamic system for fruit detection and localization[J]. *Computers and Electronics in Agriculture*, 2023, 209, 107834.
- [8] 叶敏, 邹湘军, 罗陆锋, 等. 荔枝采摘机器人双目视觉的动态定位误差分析[J]. *农业工程学报*, 2016, 32(5): 50-56. YE Min, ZHOU Xiangjun, LOU Lufeng, et al. Error analysis of dynamic localization tests based on binocular stereo vision on litchi harvesting manipulator[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016, 32(5): 50-56. (in Chinese with English abstract)
- [9] MEHTA S S, MACKUNIS W, BURKS T F. Robust visual servo control in the presence of fruit motion for robotic citrus harvesting[J]. *Computers and Electronics in Agriculture*, 2016, 123: 362-375.
- [10] MOZAFFARI S, AL-JARRAH O Y, DIANATI M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 23(1): 33-47.
- [11] 邓亚平, 李迎江. YOLO 算法及其在自动驾驶场景中目标检测研究综述[J/OL]. *计算机应用*, 1-12. [2024-03-21]. <https://link.cnki.net/urlid/51.1307.TP.20230904.1321.006>. DENG Yaping, LI Yingjiang. Review of YOLO algorithm and its application to object detection in autonomous driving scenesp[J/OL]. *Journal of ComputerApplications*, 1-12[2024-03-21]. <https://link.cnki.net/urlid/51.1307.TP.20230904.1321.006>. (in Chinese with English abstract)
- [12] ZHAO Z Q, ZHENG P, XU S, et al. Object detection with deep learning: A review[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(11): 3212-3232.
- [13] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: A review[J]. *Frontiers in Plant Science*, 2020, 11: 520170.
- [14] 李兴旭, 陈雯柏, 王一群, 等. 基于级联视觉检测的樱桃番茄自动采收系统设计与试验[J]. *农业工程学报*, 2023, 39(1): 136-145. LI Xingxu, CHEN Wenbai, WANG Yiqun, et al. Design and experiment of an automatic cherry tomato harvesting system based on cascade vision detection[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(1): 136-145. (in Chinese with English abstract)
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//*Computer Vision & Pattern Recognition*. IEEE, 2016.
- [16] GIRSHICK R, Fast R-CNN. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV '15)*[C]//*Computer Society, IEEE*, 2015: 1440-1448.
- [17] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//*Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016: 21-37*.
- [18] ZHANG Y, ZHANG M, LV X L. A review of research on small target detection based on deep learning[J]. *Computer Engineering and Applications*, 2022, 1684(1): 1-18.
- [19] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//*2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017: 3645-3649.
- [20] LI B, YAN J, WU W, et al. High performance visual tracking with siamese region proposal network[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 8971-8980.
- [21] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box[C]//*European Conference on Computer Vision, Switzerland, 2022: 1-21*.
- [22] DU Y, ZHAO Z, SONG Y, et al. Strongsort: Make deepsort great again[J]. *IEEE Transactions on Multimedia*, 2023,25: 8725-8737
- [23] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany, 2018: 3-19.
- [24] GEVORGYAN Z. Siou loss: More powerful learning for bounding box regression[EB/OL]. *arXiv preprint arXiv: 2205.12740*, 2022.[2024-05-21]<https://doi.org/10.48550/arXiv.2205.12740>
- [25] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34(7): 12993-13000.
- [26] WELCH G F. Kalman Filter[M]. *Computer Vision*. Springer, Cham. (2020-12-15). [https://doi.org/10.1007/978-3-030-03243-2\\_716-1](https://doi.org/10.1007/978-3-030-03243-2_716-1)
- [27] KUHN H W. The Hungarian method for the assignment problem[J]. *Naval research logistics quarterly*, 1955, 2(1 - 2): 83-97.
- [28] JIANG P, ERGU D, LIU F, et al. A review of yolo algorithm developments[J]. *Procedia Computer Science*, 2022, 199: 1066-1073.
- [29] 范天浩, 顾寄南, 王文波, 等. 基于改进 YOLOv5s 的轻量化金银花识别方法[J]. *农业工程学报*, 2023, 39(11): 192-200. FAN Tianhao, GU Jinan, WANG Wenbo, et al. Lightweight honeysuckle recognition method based on improved YOLOv5s[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(11): 192-200. (in Chinese with English abstract)
- [30] 范万鹏, 刘孟楠, 马婕, 等. 利用改进的 YOLOv5s 检测莲蓬成熟期[J]. *农业工程学报*, 2023, 39 (18): 183-191. FAN Wanpeng, LIU Mengnan, MA Jie, et al. Detecting lotus seedpod maturation using improved YOLOv5s[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*. 2023, 39(18): 183-191. (in

- Chinese with English abstract)
- [31] 彭炫, 周建平, 许燕, 等. 改进 YOLOv5 识别复杂环境下棉花顶芽[J]. 农业工程学报, 2023, 39(16): 191-197.  
PENG Xuan, ZHOU Jianping, XU Yan, et al. Cotton top bud recognition method based on YOLOv5-CPP in complex environment[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(16): 191-197. (in Chinese with English abstract)
- [32] 帖军, 赵捷, 郑禄, 等. 改进 YOLOv5 模型在自然环境下柑橘识别的应用[J/OL]. 中国农业科技导报: 1-10. [2024-05-14]. <https://www.nkdb.net/CN/10.13304/j.nykjdb.2022.0994>.  
TIE Jun, ZHAO Jie, ZHENG Lu, et al. Application of Improved YOLOv5 Model in Citrus Recognition in Natural Environment[J/OL]. Journal of Agricultural Science and Technology. 2024: 1-10. <https://www.nkdb.net/CN/10.13304/j.nykjdb.2022.0994>.
- [33] 孙月平, 孙杰, 袁必康, 等. 基于改进 YOLOv5s 的轻量化蟹塘障碍物检测与定位方法[J]. 农业工程学报, 2023, 39(23): 152-163.  
SUN Yueping, SUN Jie, YUAN Bikang, et al. Lightweight crab pond obstacle detection and location method based on improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(23): 152-163. (in Chinese with English abstract)
- [34] RAHAT S M S, AI PITOM M H, MAHZABUN M, et al. Lemon Fruit Detection and Instance Segmentation in an Orchard Environment Using Mask R-CNN and YOLOv5[M]// New York. Computer Vision and Image Analysis for Industry 4.0, Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: CRC Press, 2023: 28-40.
- [35] ZQPei, The modified wide residual network implementation [Z]. [2023-10-03] [https://github.com/ZQPei/deep\\_sort\\_pytorch](https://github.com/ZQPei/deep_sort_pytorch).

## Online target detection and fast localisation of citrus fruits in irregularly disturbed state

LOU Huanhuan, LI Guanglin<sup>\*</sup>, FU Xinglan, LI Li, WANG Xu, HUANG Weidong, FU Taige

(College of Engineering and Technology, Southwest University, Chongqing 400715, China)

**Abstract:** Picking robots have been widely used for citrus harvesting in recent years. However, the rest citrus during continuous picking can be irregularly disturbed by the wind, robot force, and the load weight of bearing branches under the natural environment. The citrus in the disturbed state cannot be rapidly and accurately detected, and then localized online, leading to the low efficiency of automatic robotic picking. In this study, online target detection and rapid localization were proposed using improved YOLOv5s+DeepSORT. The position of citrus at rest was predicted using the motion-tracking trajectory of disturbed citrus within a short period of time. The coordinates of the citrus were then obtained rapidly. Firstly, the CBAM (Convolutional Block Attention Module) attention mechanism was added to the YOLOv5s network, in order to detect the small and occluded targets. The SIOU loss function was used to enhance the direction matching between the prediction and the calibration frame, in order to improve the convergence speed of regression. Secondly, the target re-identification network was improved in the DeepSORT more suitable for the feature extraction of citrus targets. The feature extraction of the network was enhanced to improve the tracking performance on the disturbed citrus; The Count counter was used to accumulate the number of tracking frames in each citrus for an optimal target. Since the disturbance of the rest citrus was progressively propagated over time, the localization prediction and picking were only for targets with optimal tracking trajectories at a time. The real-time updating was realized in real time. Finally, the values of the depth camera were combined within the critical distance range, excluding the influence of background citrus on the detection speed. The number of tracking targets each time was limited to effectively improve the tracking speed of disturbed citrus. The experimental results show that the P (precision) and mAP (average detection accuracy) of improved YOLOv5s were improved by 3.9 and 1.1 percentage points, respectively, with a detection rate of 69.3 frames per second. The MOTA (Multi-Object Tracking Accuracy) and MOTP (Multi-Object Tracking Precision) of the improved DeepSORT were improved by 9.2 and 5.4 percentage points, respectively, whereas, the average number of ID (identity) switching times of targets was reduced by 32 times. Grasping experiments were conducted in the laboratory, in which the citrus was randomly swung along different orientations with an amplitude of about 10 cm. When the predicted localization time was 1, 2, 3, 5, 7, and 10 s, the average precision values of disturbed citrus localization were 21.3%, 53.0%, 81.9%, 83.7%, 86.1%, and 94.9%, respectively. The citrus picking test was conducted with the citrus localization time of 3 s. The average grabbing time for each citrus was 12.8 s, which was 5.6 s shorter than that without the optimization. The efficiency was improved by 30.4%. This finding can provide technical support and references for citrus picking in disturbed states.

**Keywords:** disturbed citrus; predictive localisation; online target detection; target tracking; YOLOv5; DeepSORT