

基于改进 YOLO v5 的复杂环境下柑橘目标精准检测与定位方法

李丽^{1,2} 梁继元^{1,2} 张云峰^{1,2} 张官明^{1,2} 淳长品³

(1. 西南大学工程技术学院, 重庆 400715; 2. 重庆市丘陵山区农业装备重点实验室, 重庆 400715;

3. 中国农业科学院西南大学柑橘研究所, 重庆 400700)

摘要: 针对自然环境下柑橘果实机械化采收作业环境复杂和果实状态多样等情况, 提出了一种多通道信息融合网络——YOLO v5-citrus, 以解决柑橘果实识别精准度低、果实分类模糊和定位精准度低等难题。将不同的柑橘目标通过不同遮挡条件分为“可采摘”和“难采摘”两类, 这种分类策略可指导机器人在真实果园中顺序摘取, 提高采摘效率并减少机器人本体和末端执行器损坏率。YOLO v5-citrus 中, 在颈部网络插入多通道信息融合模块, 对柑橘的深浅特征信息进行处理, 提高柑橘采摘状态识别精度, 同时修改颈部网络拼接方法, 针对目标柑橘大小进行识别, 训练后在识别部分嵌入聚类算法模块, 将训练部分识别模糊的柑橘目标进行最后区分。识别后进行深度图像和彩色图像的像素对齐, 并通过坐标系转换获取柑橘目标三维坐标。在使用多种增强技术处理的数据集中, YOLO v5-citrus 比原始 YOLO v5 在平均精度均值和精确率上分别提高 2.8 个百分点与 3.7 个百分点, 表现出更优异的泛化能力。与 YOLO v7 和 YOLO v8 等其他主流网络架构相比较, 保持了更高的检测精度和更快的检测速度。通过真实果园的检测与定位试验, 得到柑橘目标的三维坐标识别定位系统的定位误差为 (1.97 mm, 0.36 mm, 9.63 mm), 满足末端执行器的抓取条件。试验结果表明, 该模型具有较强的鲁棒性, 满足复杂环境下柑橘状态识别要求, 可为柑橘园机械采收设备提供技术支持。

关键词: 柑橘采摘机器人; 目标检测; 状态区分; 三维坐标获取; 复杂环境; YOLO v5

中图分类号: S225; TP391

文献标识码: A

文章编号: 1000-1298(2024)08-0280-11

OSID:



Accurate Detection and Localization Method of Citrus Targets in Complex Environments Based on Improved YOLO v5

LI Li^{1,2} LIANG Jiyuan^{1,2} ZHANG Yunfeng^{1,2} ZHANG Guanming^{1,2} CHUN Changpin³

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

2. Chongqing Key Laboratory of Agricultural Equipment for Hilly and Mountainous Regions, Chongqing 400715, China

3. Citrus Research Institute, Southwest University-Chinese Academy of Agricultural Sciences, Chongqing 400700, China)

Abstract: Aiming at the challenges of mechanized citrus fruit harvesting in natural environments, such as complex environments and diverse fruit states, a multi-channel information fusion network (YOLO v5-citrus) was developed, to solve the problems of low accuracy of citrus fruit recognition, fuzzy fruit classification and low accuracy of localization. Different citrus targets were categorized into “pickable” and “hard-to-pick” by different occlusion conditions, and this classification strategy guided the robot to pick them sequentially in a real orchard, which improved the picking rate and reduced the damage rate of the robot body and end-effector. In YOLO v5-citrus, a multi-channel information fusion module was inserted into the neck network to process the depth feature information of citrus to improve the recognition accuracy of the citrus picking state. At the same time, the splicing method of the neck network was modified to recognize the size of the target citrus. The clustering algorithm module was embedded in the recognition part after training to make the final distinction between the citrus targets blurred by the recognition in the training part. Pixel alignment of a depth image and a color image was performed after

收稿日期: 2024-04-08 修回日期: 2024-05-15

基金项目: 重庆市杰出青年科学基金项目(2022NSCQ-JQX0030)、宜宾市双城协议保障科研经费项目(XNDX2022020015)、中央高校基本科研业务费专项资金项目(XDJH202302)和重庆市研究生科研创新项目(CYB23125)

作者简介: 李丽(1982—), 女, 教授, 博士生导师, 主要从事智能农机装备研究, E-mail: cqulily@163.com

recognition and 3D coordinates of citrus targets were obtained by coordinate system transformation. In the dataset processed using multiple enhancement techniques, YOLO v5-citrus improved mAP and precision by 2.8 percentage points and 3.7 percentage points, respectively, compared with the original YOLO v5, respectively. It maintained higher detection accuracy and faster detection speed than other mainstream network architectures such as YOLO v7 and YOLO v8. Through the detection and localization test in the real orchard, the localization error of the 3D coordinate recognition localization system for the citrus target was obtained as (1.97 mm, 0.36 mm, 9.63 mm), which satisfied the grasping condition of the end-effector. The experimental results showed that the model had strong robustness, meeting the requirements of citrus state recognition in complex environments, and can provide technical support for mechanical harvesting equipment in large-field citrus orchards.

Key words: citrus picking robot; object detection; state differentiation; 3D coordinate acquisition; complex environments; YOLO v5

0 引言

柑橘类水果在消费类水果中占据重要地位,具有巨大的商业价值,自 2018 年以来,已有 120 多个国家和地区开始参与柑橘贸易,柑橘总种植面积已增加到 $1.139 \times 10^7 \text{ hm}^2$ ^[1]。在中国,柑橘种植基地已经扩展到 20 多个省市,柑橘总产量提升数倍。但机械化程度低,农业生产劳动力不足等问题限制了柑橘产业的提质增效^[2]。针对柑橘产业目前现状,国家有关部门出台了一系列政策文件鼓励因地制宜发展柑橘产业,反映了智能化柑橘采摘配套技术研发对于柑橘产业的重要性^[3]。

柑橘机械化采摘难点在于解决真实果园中复杂环境和柑橘果实自身多样的状态问题。针对实际情况,魏博等^[4]设计了一种欠驱动式柑橘采摘末端执行器,使机器人末端执行器可以采摘不同形状的柑橘。虞浪等^[5]设计了一种欠驱动关节型末端执行器,扩大了柑橘采收范围。王毅等^[6]设计了一种仿蛇嘴咬合式采摘末端执行器,可以采摘多角度果实倾角的柑橘。综上所述,针对采摘时复杂环境和柑橘自身状态的问题,很多学者对采摘末端执行器提出了改进方法。但全自动化柑橘采摘时,存在树叶大量覆盖和枝干遮挡,采摘机器人由于识别和定位不精准,末端执行器无法对柑橘目标完全包围,最终导致当前位置采摘失败^[7-8]。这说明精准识别与定位是目前实现柑橘采摘全自动化必须要解决的问题^[9-10]。

为了解决复杂环境对柑橘采摘设备的影响,推动柑橘产业智能化,需为柑橘采摘设备搭建高效准确的柑橘识别系统^[11],在识别可直接采摘柑橘的同时,分辨设备当前位置被遮挡严重的柑橘目标。本文中,前者被作为采摘中的第 1 目标,指引设备进行直接采摘,后者作为第 2 目标,在第 1 序列目标采摘完毕后,对被遮挡柑橘遮挡姿态进行分析后采摘,本文主要对柑橘采摘作业中柑橘目标的识别与定位展

开工作。在此之前,国内外水果识别领域,已经有很多创新型算法被提出,XU 等^[12]提出了一种高精度轻量化的 YOLO v4 检测方法,以轻量级特征提取网络 GhostNet 为骨干,增强柑橘特征提取能力,识别精度达到 93.45%。YAN 等^[2]提出了一种基于 YOLO v5 的遮挡检测算法,可以分辨复杂环境中被各种遮挡物覆盖的苹果目标,识别精度达到 86.75%。APPE 等^[13]提出了一种改进神经网络,对番茄的复杂环境与部分遮挡做出识别。WU 等^[14]结合聚类模型分割提取桃子点云数据,通过圆拟合来提高识别准确性,对遮挡状态的桃子识别精准度达到 88.68%。熊俊涛等^[15]改进了一种夜间环境的柑橘识别方法,提高了夜间柑橘识别精度。

综上所述,目前在多种水果上均有视觉识别研究,但针对柑橘,目前仅有对柑橘果实的单纯识别,采摘中柑橘目标状态的识别和定位研究较少,同时现有成果在识别率及定位精度上仍有提升空间。针对实际采摘时遇到的复杂环境和枝干遮挡、树叶遮挡等问题,本文提出一种基于改进 YOLO v5 的柑橘识别分类算法 YOLO v5-citrus。提出一种多通道信息融合网络 (Multi-channel information fusion network, MCIFNet),可对更深层的柑橘状态特征进行提取,其次改变网络连接方式检测更符合目标柑橘大小的特征图,最后更改损失函数并添加基于颜色空间变换和图像掩膜的柑橘识别算法,完成柑橘目标识别。基于 YOLO v5-citrus 识别算法搭建立体视觉定位系统,完成深度图与彩色图的像素对齐与坐标转化,获取柑橘目标的三维坐标。

1 柑橘图像数据获取与预处理

1.1 柑橘图像采集

本文以现代化标准柑橘园中的种植柑橘为研究对象,数据采集和试验地点位于西南大学柑橘研究所。数据集中的图像包括多种光照角度和不同遮挡情况的柑橘。图像采用 Intel Realsense D345i 型相

机拍摄,在不同拍摄距离(0.5 ~ 1.5 m)下选择多种角度进行图像采集,共获得 1 000 幅柑橘图像,具体拍摄柑橘类型为:自然光角度、背光角度和过曝角度柑橘等(图 1)。采集图像分辨率为 1 280 像素 × 720 像素,格式为 JPEG。



图 1 采集图像数据集示例

Fig. 1 Example of captured image dataset

从 1 000 幅图像中随机选取 150 幅图像作为测试集,700 幅图像作为训练集,其余 150 幅图像作为验证集。利用图像数据标注软件 Labelme,在柑橘图像中绘制出完全包裹柑橘目标的矩形框,实现柑橘目标的人工标注。其中,图像中未被遮挡的柑橘标记为“可采摘”类,仅被树叶遮挡的柑橘标记为“部分遮挡”类,枝干遮挡与树叶枝干混合遮挡的柑橘标记为“难采摘”类,数据集中各种柑橘目标数量如表 1 所示,训练集和测试集之间没有重叠。

表 1 不同状态柑橘数量分布

Tab. 1 Quantitative distribution of citrus in different states

数据集	可采摘	部分遮挡	难采摘	总和
训练集	2 907	2 741	903	6 551
验证集	410	377	131	918
测试集	393	422	105	920
总和	3 710	3 540	1 139	8 389

1.2 图像数据预处理

为增强数据多样性,提高模型泛化能力,对采集图像进行镜像旋转、镜像翻转、图像平移、亮度调节和添加噪声 5 种图像处理方式用以拓展数据集,处理后图像如图 2 所示。

通过 5 种数据扩充方式将训练集扩充至 4 200 幅图像,其中 3 500 幅增强图像和 700 幅原始图像作为训练柑橘状态识别模型的最终训练数据。

2 基于改进 YOLO v5 的柑橘目标检测算法

2.1 YOLO v5 网络

YOLO v5 属于 YOLO 架构系列,是 onestage 类

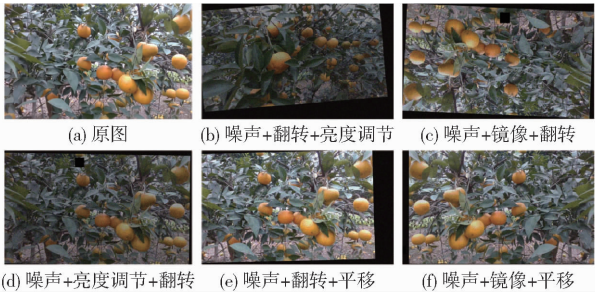


图 2 处理后图像数据集

Fig. 2 Processed image dataset

型识别算法。该网络模型检测精度高,推理速度快,同时权值文件较小,适用于柑橘采摘机器人部署^[16]。YOLO v5 网络拥有较小的模型和一定的精确度,但在试验中发现 YOLO v5 网络在一些遮挡柑橘目标的识别中,易出现误判,对于光照强烈和背光的柑橘,存在漏识别现象,同时精度和精准度还具有提升空间。而在实际采摘过程中,视觉识别模型的准确性、实时性对柑橘采摘机器人至关重要^[17],因此,本文基于 YOLO v5 架构,对柑橘采摘的深度识别网络进行改进。

YOLO v5 框架主要由 3 部分组成,包括:骨干网络、颈部网络和检测头(图 3)。骨干网络是指用来提取图像特征的网络,主要作用是将原始输入图像转化为多层特征图,以便后续的目标检测任务使用;颈部网络主要对骨干网络所提取到的深浅层次特征进行多尺度特征融合^[18],并将这些特征传递给最终检测网络;最终检测部分进行最后的回归预测,完成对图中所需特征的检测,得出图像的检测框及目标类别。

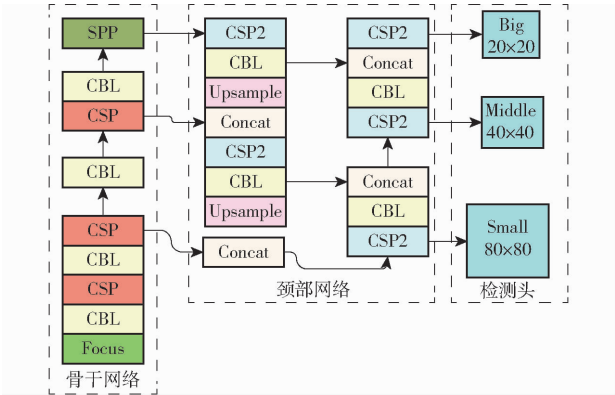


图 3 原始 YOLO v5 网络

Fig. 3 Original YOLO v5 network

2.2 YOLO v5 网络改进

本文使用的柑橘识别网络以原始 YOLO v5 作为基础网络,改进后整体框架如图 4 所示。将骨干网络中加入 CBAM (Convolutional block attention module)^[19],以提高神经网络对图像中柑橘目标特征的关注度,同时将本文提出的 MCIFNet 插入到网

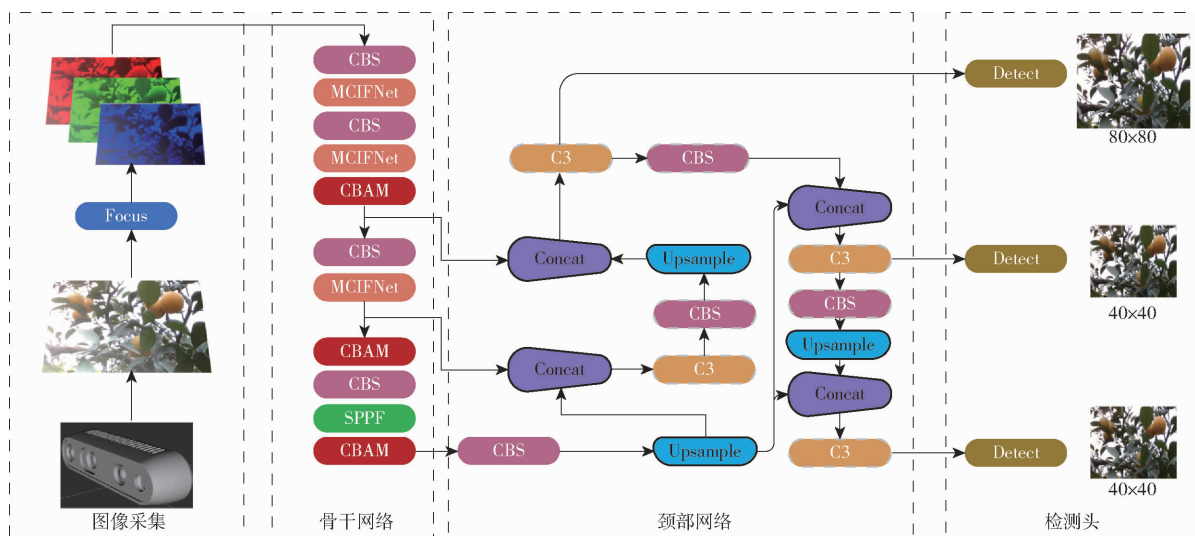


图 4 改进后 YOLO v5 网络结构图

Fig. 4 Structure of improved YOLO v5 network

络的骨干结构中,提高对柑橘目标深浅层特征的共同提取能力。由于相机识别区域有限,且经过统计,在就近采摘的前提下,相机中 40 像素 × 40 像素的目标更适合采摘^[20],因此在颈部网络改变检测部分的连接方式,在连接层增加适宜目标的检测权重。最后,将损失函数从 CIoU 变更为 SIoU^[21],在识别界面增加颜色掩膜检测算法,用以将部分遮挡的柑橘区分为可采摘柑橘与难采摘柑橘。

2.2.1 注意力模块添加

由于本文检测目标存在特殊外形特征及一定空间位置特征,提出 MCIFNet 模块,将连接方式改为三通道维度拼接,并在其中的一个通道中加入空间与通道注意力模块,结合网络层中的位置,可以更有效地提取浅层与深层特征,在 YOLO v5 骨干网络中插入 MCIFNet 模块可有效提高训练模型的网络特征提取能力和检测精度。骨干网络的最后几层隔层插入 CBAM 模块,由于其自适应特征优化,可提取更深层次的特征,同时拥有即插即用和轻量特质,因此可以在不影响模型部署权重的前提下插入网络中,MCIFNet 模块与 CBAM 的结合可以在检测网络进行拼接时提供包含更多图像重要信息的特征图,提升识别精度。

如图 5 所示,空间注意力模块 (Spatial attention module, SAM) 的输入是通道注意力模块上一步的输出特征图。SAM 对输入特征图 F 进行平均池化和最大池化后,得到 F_{avg}^S 和 F_{max}^S 聚合通道信息,两个特征图被缝合成一个特征图后用 7×7 的卷积生成一个二维空间特征图,空间注意力表达式为

$$M_s(F) = \sigma(f^{7 \times 7}(\text{AvgPool}(F); \text{MaxPool}(F))) = \sigma(f^{7 \times 7}[F_{avg}^S; F_{max}^S]) \quad (1)$$

式中 σ ——激活函数

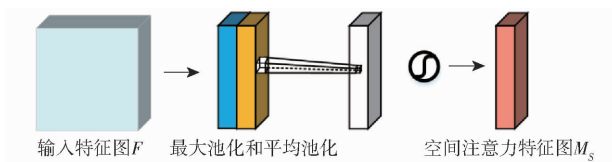


图 5 SAM 结构图

Fig. 5 Structure graph of SAM

如图 6 所示,通道注意力模块 (Channel attention module, CAM) 将特征图在空间维度上进行压缩,得到一个一维矢量后再进行后续计算。CAM 运行过程中会提取平均池化特征 F_{avg}^C 和最大池化特征 F_{max}^C 两个特征信息,输入后被传递到共享网络隐藏层 MLP 进行处理,通过共享隐藏层的激活函数 σ 后可得到通道注意力特征图 M_c 。多层感知模型中的两层参数用 α_1, α_2 表示,可获得注意力表达式为^[22]

$$M_c(F) = \sigma(\alpha_2(\alpha_1(F_{avg}^C)) + \alpha_2(\alpha_1(F_{max}^C))) \quad (2)$$

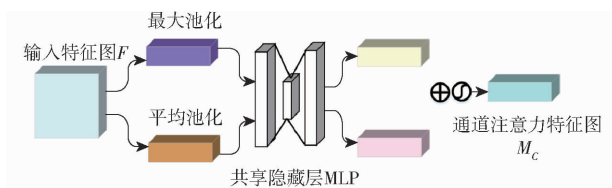


图 6 CAM 结构图

Fig. 6 Structure graph of CAM

MCIFNet 改进网络整体流程如图 7 所示,三通道连接方式在保留了原始特征的同时,通过通道注意力机制与空间注意力机制协同学习图像中的关键局部信息,为神经网络特征图中柑橘目标区域的典型特征分配较高的权重,将原始特征与两条分支的特征图拼接后即可提取到更详细的特征信息,有效地提高了神经网络对图像中柑橘各类状态的关注度,进而提升了网络特征学习和表达能力。

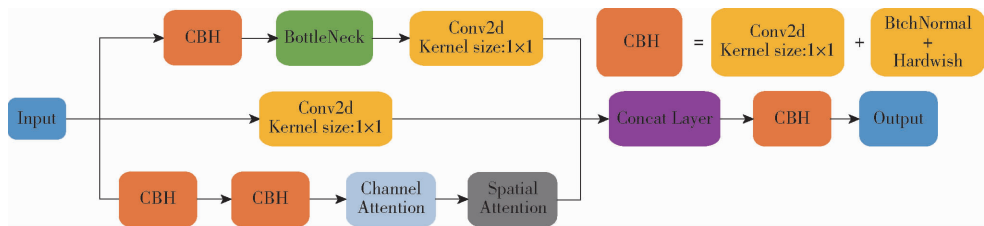


图7 MCIFNet改进网络整体流程图

Fig.7 MCIFNet improved overall network processes

2.2.2 连接方式改变

原始 YOLO v5 网络在颈部网络对骨干网络提取的特征层进行拼接,并向检测头输出 20 像素 × 20 像素、40 像素 × 40 像素和 80 像素 × 80 像素 3 种尺度的特征图用于检测不同尺寸目标。由文献[23]可知大特征图上检测小目标,由于实际采摘中机械手无法采摘远处柑橘(80 像素 × 80 像素),视觉系统也通常无法计算占据像素数太多的大目标质心(20 像素 × 20 像素),因此 40 像素 × 40 像素的特征图是实际采摘过程中的主要识别目标。经过添加 CBAM 和 MCIFNet,目前网络中 13 层输出 40 像素 × 40 像素特征图,蕴含目标丰富的特征信息,于颈部网络中将 13 层与 23 层和 13 层与 20 层进行拼接,可在 2 个 40 像素 × 40 像素检测头上进行目标柑橘检测与识别,加大可采摘柑橘目标的识别权重。改变连接方式后,采用 COCO 评价方法验证其性能,将原始 YOLO v5 模型与改变连接方式后的模型进行对比试验,试验结果见表 2,表中 AP₅₀ 表示 IoU 阈值取 0.5 时的识别平均精度,AP_S、AP_M、AP_L 分别表示小、中、大目标识别平均精度。

Tab.2 Comparison of target layer modification results				
%				
算法模型	AP ₅₀	AP _S	AP _M	AP _L
YOLO v5	88.5	21.8	64.1	64.5
YOLO v5-Middle	90.6	18.4	77.3	68.8

由表 2 可知,修改了连接方式的改进模型(YOLO v5-Middle)相比于原始模型,AP_M 提高 13.2 个百分点,AP_L 提高 4.3 个百分点,这表明修改连接方式后的网络结构能有效提高对目标对象的检测性能。

2.2.3 损失函数改进

计算机视觉中的一项基本任务是边界框回归,即使用矩形边界框预测目标对象在图像中的位置,旨在细化预测边界框位置。CIoU 损失函数使用矩形框损失、置信度损失和分类损失定义总损失。

YOLO v5 网络使用 CIoU 解决了损失因素考虑不足的问题,添加了重叠面积、中心点距离、长宽比

等多个因素在函数中。但对于柑橘目标,CIoU 差的回归样本对总损失产生了比较大的影响,而回归质量相对较好的样本则难以对总损失进一步优化。同时通过计算可以得到,柑橘的目标框长宽比在同一品种中不具有一致性,而在 CIoU Loss 中长宽损失一致,这导致经 CIoU 损失函数训练的目标框常常不能完全框住识别到的柑橘目标,为了解决这个问题,使真实框和预测框之间的方向同时匹配,本文选择新的损失函数 SIoU,其将目标框长宽固定损失变为长和宽的一致性损失,同时重新定义角度惩罚度量,使预测框快速漂移到最近的轴,可以进一步增加训练精度。SIoU 使用原始 IoU、距离成本 Δ、形状成本 Ω 来定义总损失,计算式为

$$SIoU_{Loss} = 1 - IoU + \frac{\Delta + \Omega}{2}$$

(3)

2.2.4 颜色图像分割分类算法

本文在训练时将待采摘柑橘分为可采摘、难采摘和部分遮挡柑橘,但在实地采摘中视觉系统需要具体识别一个柑橘是否可以采摘,需要解决中间态柑橘的识别问题,经过改进的神经网络训练后,视觉系统可以精准地框选中间态柑橘,由此可以得到一个部分遮挡柑橘的目标识别框及其在图中所处的位置。

本文使用图像颜色分割方法,在使用 OpenCV 库读取图像具体像素信息后,将图像格式从 BGR 转换到 RGB,通过图像切片的方式获取到被框选的中间态柑橘。由于在 RGB 颜色空间之内黄绿颜色的阈值区间十分宽泛,在视觉系统进行实时识别时,对硬件芯片的压力过大,所以将聚类中心点所表示颜色由 RGB 颜色空间转化到 HSV 颜色空间中,具体公式为

$$H = \begin{cases} 60^\circ \frac{G' - B'}{\Delta'} & (C_{\max} = R') \\ 60^\circ \left(\frac{B' - R'}{\Delta'} + 2 \right) & (C_{\max} = G') \\ 60^\circ \left(\frac{R' - G'}{\Delta'} + 4 \right) & (C_{\max} = B') \end{cases}$$

(4)

式中 R'、G'、B'——归一化后 R、G、B 值
C_{max}——R、G、B 归一化后最大值

$C_{\min}$ —— R 、 G 、 B 归一化后最小值
 Δ' —— $C_{\max}$ 与 $C_{\min}$ 差值
 H ——图像色调

在 HSV 颜色空间中,橙色 H 阈值在 $0 \sim 34$ 之间,绿色 H 阈值在 $35 \sim 99$ 之间,由于 S 和 V 表示饱和度与色调,在同一幅图像的不同颜色下饱和度与色调是相同的,所以通过对转换后图像的 H 值所在区间进行分析,即可判断颜色分类。设置生成掩膜的下界限为 $[0,30,30]$,上界限为 $[34,255,255]$,可以得到橙色区域的图像掩膜。

由于二值图像的局限性和 SV 的不稳定,得到的掩膜中会出现不稳定的白点,本文最终使用腐蚀与膨胀图像方法,具体方法为通过腐蚀去除掉周边的孤立白点,但同样的柑橘橙色特征的白色区域边界也会被腐蚀变小,可以用反向操作膨胀再填充回原尺寸得到稳定的掩膜区域。

而后可通过统计二值图像中掩膜区域或黑色区域大小在总像素中所占的比例来分析柑橘目标具体状态。经过试验验证,当叶子遮挡不超过 40% 柑橘总面积时,圆拟合可以有效推演出质心所在位置并进行抓取,所以本文选择当橙色特征总面积占总体面积比重大于 0.6 时,将当前识别目标归类入可采摘柑橘。经过验证本算法可以将部分遮挡柑橘划分到可采摘柑橘和难采摘柑橘中,效果如图 8 所示。

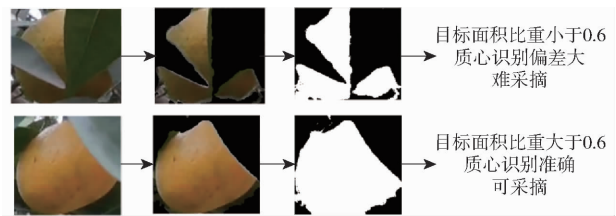


图 8 分类算法效果图

Fig. 8 Clustering algorithm effect diagram

2.3 基于双目立体视觉的可采摘柑橘空间定位

本文以自主研发的柑橘采摘机械容许定位误差作为衡量获取的三维坐标是否可行的标准,该设备进行柑橘采摘的流程为:首先启动相机识别节点并启动 YOLO v5 识别程序,对柑橘目标进行识别;随后深度相机获取可采摘柑橘的像素坐标,与相机的 K 矩阵计算可获得目标在相机坐标系的坐标;通过多次矩阵的转换可以最终将坐标映射到采摘机器人的机械臂坐标系中,最终实现机械臂对柑橘目标定位与抓取,相机与夹爪的坐标系位置示意图如图 9 所示。

经过统计计算,柑橘形状为宽高比 $0.75 \sim 0.85$ 之间的球形,柑橘直径为 $55 \sim 65$ mm。测量柔性夹爪张开的宽度约为 100 mm,夹爪坐标系在夹爪中

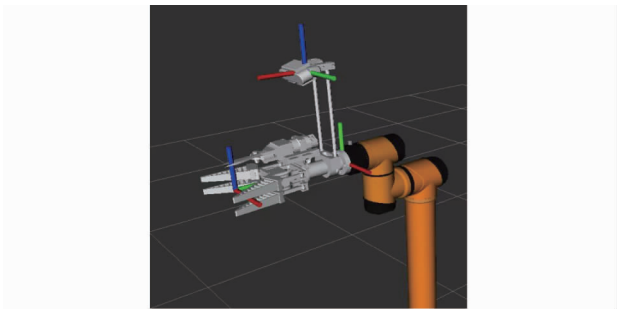


图 9 相机与夹爪坐标系

Fig. 9 Camera and gripper coordinate system

心,当 X 轴偏差小于 20 mm 时,夹爪可以包裹柑橘;测量柔性夹爪长度为 90 mm,大于一般柑橘直径 60 mm,故 Y 轴定位误差不大于 15 mm;当夹爪仅抓住柑橘上半部分或下半部分时,柑橘可能滑脱,而单柔性夹爪高度为 15 mm,故 Z 轴定位误差不大于 15 mm。综上,最大容许误差为 X 轴 20 mm、 Y 轴 15 mm、 Z 轴 15 mm。

2.3.1 相机标定及空间坐标系转换

选用 Intel 公司的 Realsense D435i 型系列双目相机开展研究,为获取柑橘目标的有效世界坐标并提高识别检测算法的准确度,需完成标定获取相机内部参数和畸变参数^[24],并将深度图对齐到彩色图,进而建立图像世界与三维世界的准确关系。

2.3.2 相机内部参数标定

对 D435i 型相机进行标定获取相机内部参数和畸变参数,从而在坐标系转换的矩阵运算中进行精度矫正,相机标定结果见表 3。

表 3 双目相机内部参数
Tab. 3 Binocular camera internal parameters

参数	相机标定结果
相机内部参数	$\begin{bmatrix} 909.087\ 4 & 0 & 640.991\ 8 & 0 & 907.876\ 4 \\ 360.328\ 5 & 0 & 0 & 1.0 & 0 \end{bmatrix}$
旋转矩阵	$\begin{bmatrix} 1.0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 1.0 \end{bmatrix}$
投影矩阵	$\begin{bmatrix} 909.087\ 4 & 0 & 640.991\ 8 & 0 & 0 & 907.876\ 4 \\ 360.328\ 5 & 0 & 0 & 0 & 1.0 & 0 \end{bmatrix}$

通过深度相机获取到的图像有彩色图像和深度图像,不经过矩阵对齐,会导致两个摄像头的位置偏差,从而像素点位置出现差异,最终造成识别和定位误差^[25]。因此,使用彩色像素进行图像识别并进行初步定位前,需对相机进行对齐处理,使彩色图像像素点与深度图像像素点对应,图像对齐公式为

$$T_{d2c} = \begin{bmatrix} R_{w2c}R_{w2d}^{-1} & t_{w2c} - R_{w2c}R_{w2d}^{-1}t_{w2d} \\ 0 & 1 \end{bmatrix} \tag{5}$$

式中 R_{w2c} ——世界坐标到彩色坐标旋转矩阵
 t_{w2c} ——世界坐标到彩色坐标偏移矩阵
 R_{w2d} ——世界坐标到深度坐标旋转矩阵

t_{w2d} ——世界坐标到深度坐标偏移矩阵
进行对齐后的深度图与彩色图如图 10 所示。

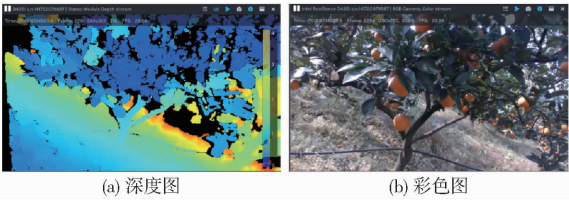


图 10 柑橘目标彩色图与深度图

Fig. 10 Color map and depth map of citrus

2.3.3 空间坐标系转换

得到双目相机的内部参数及畸变参数后,通过相机针孔成像原理和矩阵运算的方法对柑橘目标的世界坐标进行计算。在世界坐标获取中,需对像素坐标系、图像坐标系、相机坐标系和世界坐标系进行转换^[25]:通过识别网络获得柑橘目标在像素坐标系中的位置,将像素坐标系中的坐标转换处理后与相机的 K 矩阵相乘得到柑橘目标的相机坐标系,最终通过相机坐标系与基坐标系之间的相对关系获得最终的采摘坐标,转换示意如图 11 所示, $u-v$ 为像素坐标系,原点位于图像的左上角(单位为像素); $q(x,y)$ 为柑橘果实在图像坐标系中的投影点,相机焦距 f 为 O 与 O_c 间距离。柑橘目标在基坐标系中的坐标点为 $Q(X_w,Y_w,Z_w)$ (单位为 mm); $O_cX_cY_cZ_c$ 为双目相机坐标系(单位为 mm); O_{xy} 为图像坐标系(单位为 mm)。

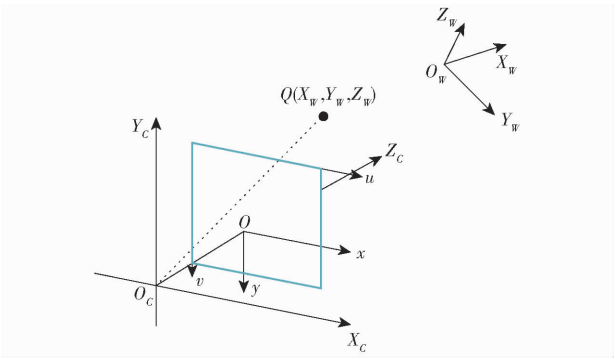


图 11 空间坐标系转换

Fig. 11 Spatial coordinates transform

读取图像像素坐标后,统计对应区域内中心的点云坐标获取柑橘表面中心点的三维坐标并建立 $O_cX_cY_cZ_c$ 的像素基础坐标系,沿深度方向伸展一个半径,即可获取柑橘目标中心坐标,并在柑橘中心处建立实际坐标系 $O_wX_wY_wZ_w$, d_z 为深度相机测算出柑橘目标的深度信息,具体过程如图 12 所示。

3 试验结果与分析

3.1 试验环境

本文采用的图像采集设备为 Intel Realsense

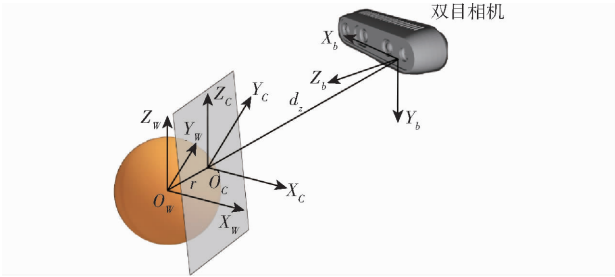


图 12 柑橘目标中心坐标示意图

Fig. 12 Acquisition of citrus target center coordinates

D435i 型相机,采集图像像素为 1 280 像素 $\times$ 720 像素。台式计算机 (Intel (R) Core (TM) i7 - 12600 CPU, 2.6 GHz, 16 GB 内存; NVIDIA Geforce RTX 3080GPU, 10 GB 显存), Windows 10 操作系统下构建 PyTorch 深度学习框架,利用 Python 语言编写程序代码。

3.2 模型性能评估指标

使用精确率 (Precision, P)、召回率 (Recall, R) 和平均精度均值 (Mean average precision, mAP) 作为验证模型性能的指标。设定两个边界框之间重叠区域和不重叠区域的交并比阈值为 0.5,若 IoU 超过该阈值为正样本,反之为负样本。

3.3 试验结果

3.3.1 注意力模块对比试验

为验证不同注意力模块对本文模型识别效果的影响,将 SE (Squeeze-and-ecitation module)、CA (Coordinate attention module)、CBAM 和 ECA (Efficient channel attention module) 4 种注意力机制分别按图 4 其他模块不做改动的方式插入网络,由表 4 可知,SE 与 CA 模块插入网络后精确率与平均精度均值远低于 CBAM 和 ECA 模块插入网络后的效果,而 CBAM 与 ECA 模块在精确率与平均精度均值相近的情况下,CBAM 模块对于网络的召回率提升远高于 ECA 模块。相比其他 3 种网络,CBAM 可以在考虑柑橘目标图像空间特征的同时,兼顾图像的通道特征,综上所述本文选择 CBAM 插入主干网络。

表 4 注意力机制对比试验				
Tab. 4 Comparison of attentional mechanisms				
注意力模型	精确率/ %	召回率/ %	平均精度 均值/%	模型内存占 用量/MB
CBAM	87.3	83.1	89.4	14.5
SE	84.1	81.9	86.3	14.5
CA	82.2	83.0	87.5	14.5
ECA	87.7	80.6	90.0	14.5

3.3.2 损失函数对比试验

对比 CIoU、GIoU、SIoU、EIoU 4 种损失函数,由

表 5 可知,针对复杂的柑橘特征,GIoU 仅考虑距离损失,精确率与召回率等指标全面下降;ElIoU 将纵横比的影响因子拆开分别计算,可小幅度提升模型平均精度;SIoU 对多种成本共同考虑,并在损失计算时将目标框长宽定义为一致性损失,模型平均精度均值可提升 2.1 个百分点。综合模型性能评估指标,本文选择 SIoU 作为网络损失函数。

表 5 损失函数对比试验
Tab.5 Comparison of loss function

损失函数	精确率/ %	召回率/ %	平均精度 均值/%	模型内存占 用量/MB
CIoU	88.5	82.6	88.3	14.5
GIoU	86.1	81.5	86.7	15.3
EIoU	87.2	83.3	89.1	16.1
SIoU	88.7	79.9	90.4	15.5

3.3.3 消融试验

为了进一步验证每种改进方式对最终算法的优化效果,分别将其添加进模型中进行消融试验。4 种改进方法分别为更改 SIoU 损失函数、中目标连接方式、MCIFNet 改进网络和 CBAM 模块,试验方法如下:在原始 YOLO v5 算法中分别加入 4 种改进方式,分别观察效果;加入全部改进模块,使用创新算法与原 YOLO v5 进行比对。从表 6 可以看出,对比改进前的 YOLO v5 模型,4 种改进方法分别在不同方面提高了总体网络性能,改进的 MCIFNet 在召回率和平均精度均值提升效果最为显著,精确率和平均精度均值分别提升 0.8 个百分点和 1.8 个百分点。而中目标改动对精准度的提升最大,达到 3.4 个百分点。在经过 4 项改进策略后的改进网络模型中,测试集平均精度从 88.3% 提高到 91.1%,精确率和召回率也分别提高 3.7 个百分点和 1.2 个百分点,说明改进后模型具有良好的收敛性。

表 6 消融试验结果对比
Tab.6 Comparison of ablation experiments

SIoU	MCIFNet	Middle	CBAM	精确 率/%	召回 率/%	平均精度 均值/%	模型内存占 用量/MB
×	×	×	×	88.5	82.6	88.3	14.3
√	×	×	×	88.7	79.9	90.4	15.5
×	√	×	×	89.1	83.4	90.1	13.5
×	×	√	×	91.9	81.7	90.6	14.7
×	×	×	√	87.3	83.1	89.4	14.5
√	√	√	√	92.2	83.8	91.1	13.9

注: × 表示未采用此改进方法,√ 表示采用此改进方法。

3.3.4 不同网络试验结果对比

为进一步评价改进网络对柑橘目标的检测效果,使用 YOLO v5s、YOLO v3、SSD、YOLO v7 和

YOLO v8 等算法模型在相同环境下进行训练和测试。由表 7 可以看出,YOLO v5-citrus 与其他模型相比,精确率和平均精度均值均有所提高,较改进前的 YOLO v5 网络精确率提升 3.7 个百分点,平均精度均值提升 2.8 个百分点,同时对比 YOLO v7 精确率和平均精度均值也提升 2.1 个百分点与 2.8 个百分点。检测速度达到 34.1 f/s,符合在柑橘采摘作业中,深度相机进行图像点云构建的要求,同时在添加多种改进因素的情况下降低了参数量,方便实际部署。

表 7 与主流模型对比试验结果
Tab.7 Comparison test results with mainstream models

算法模型	精确 率/%	召回 率/%	平均精度 均值/%	模型内存 占用量/ MB	检测 速度/ (f·s ⁻¹)
YOLO v5	88.5	82.6	88.3	14.3	45.3
YOLO v3	80.3	77.5	79.8	55.7	22.2
SSD	85.9	81.6	85.1	32.1	21.9
YOLO v7	90.1	82.2	88.3	74.8	31.8
YOLO v8	87.5	83.3	89.3	18.2	38.2
YOLO v5-citrus	92.2	83.8	91.1	13.9	34.1

综上所述,提出的 YOLO v5-citrus 是 6 种网络模型中最轻量级的,同时具有最高精确率和平均精度均值,因此 YOLO v5-citrus 在满足实用性能的前提下,保证了较高水平的召回率,综合性能最优。改进网络模型与原始 YOLO v5 网络识别结果如图 13 与图 14 所示。



图 13 密集与稀疏柑橘检测结果
Fig. 13 Detection results for dense and sparse citrus

图 13 为密集和稀疏条件下柑橘目标识别结果,图 14 为过曝和背光条件下柑橘目标识别结果。由图 13 与图 14 可知,本文提出的 YOLO v5-citrus 网络识别结果准确,结合原 YOLO v5 识别结果可以看出,在密集和稀疏条件下,YOLO v5 网络识别结果中存在未识别出难采摘柑橘的错误(图 13c 中由红色圆圈标记)。对于背光条件的识别,将可采摘和部分遮挡柑橘识别为难采摘柑橘的情况(图 14d 中用

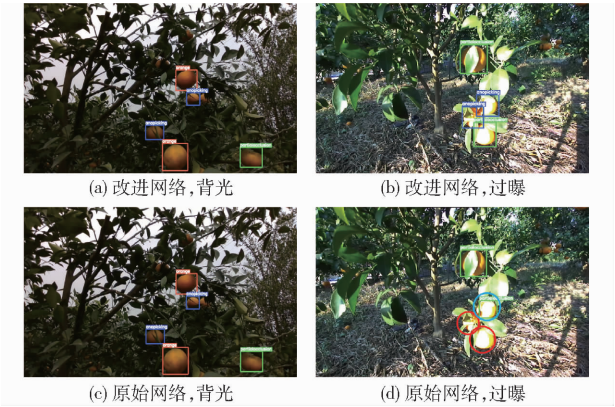


图 14 过曝与背光柑橘检测结果

Fig. 14 Detection results for exposure and backlight citrus
蓝色圆圈标记), 并且未识别出可采摘和难采摘柑橘(图 14d 中用红色圆圈标记)。

3.3.5 柑橘目标检测结果

柑橘目标三维定位使用 Vscode 作为开发软件, 安装 ROS、C++ 等插件, 实现了以下功能: 调用 YOLO v5-citrus 对柑橘目标进行识别; 读取深度相机数据获取目标像素坐标并得到目标的深度信息; 将柑橘目标的像素二维坐标转换为三维坐标, 并与机器人基坐标系进行运动学推理得到在基坐标系中的采摘坐标。

经过训练模型改进和识别模块添加, YOLO v5-citrus 对柑橘目标的漏识别减少, 识别精确率提高,

在密集、稀疏、过曝和背光等多种复杂环境下均有良好的识别表现, 最终识别结果如图 15 所示。

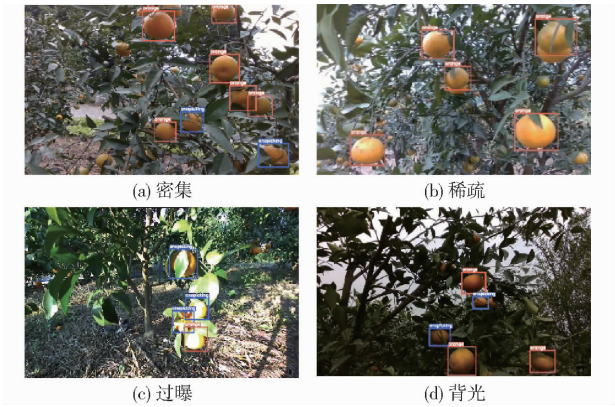


图 15 最终识别结果

Fig. 15 Final detection results

3.3.6 柑橘目标定位结果

通过柑橘定位验证定位结果, 试验使用双目相机与三脚架、激光雷达等仪器进行三维坐标测量。

将 Realsense D435i 型相机固定安装, 摄像头方向水平朝前, 测量镜头与地面竖直距离; 当相机镜头内含有帧率稳定的目标时, 测量柑橘目标离地竖直距离与距镜头的水平距离, 与相机 K 矩阵进行计算可以得到柑橘目标的实际三维坐标, 此值与算法识别的三维坐标进行对比并计算误差, 汇总测量结果见表 8, 算法识别界面见图 16。

表 8 柑橘果实三维坐标检测结果

Tab. 8 Results of citrus three-dimensional coordinate detection				cm
序号	真实值	检测值	误差	
1	(33.4, 56.9, 79.2)	(33.14, 57.43, 79.23)	(-0.26, 0.53, 0.03)	
2	(47.6, -14.4, 24.9)	(48.7, -14.86, 24.32)	(1.10, -0.46, -0.58)	
3	(79.4, 32.6, 63.9)	(79.86, 33.94, 64.73)	(0.46, 1.34, 0.83)	
4	(-3.0, 10.5, 63.7)	(-3.35, 9.44, 63.90)	(-0.35, -1.06, 0.20)	
5	(46.3, -9.8, 80.3)	(45.78, -9.30, 76.34)	(-0.52, 0.50, -3.96)	
6	(-32.1, -22.1, 100.5)	(-31.09, -23.25, 98.50)	(1.01, -1.15, -2.00)	
7	(-45.9, -29.4, 104.3)	(-46.72, -28.60, 101.70)	(-0.82, 0.80, -2.60)	
8	(-40.6, -49.5, 191.4)	(-39.46, -47.02, 193.90)	(1.14, 2.48, 2.50)	
9	(16.5, -22.8, 137.2)	(15.02, -21.02, 134.30)	(-1.48, 1.78, -2.90)	
10	(43.3, -16.4, 203.1)	(41.97, -16.15, 205.60)	(-1.33, 0.25, 2.50)	
11	(-12.3, -11.1, 61.3)	(-13.87, -11.69, 59.50)	(-1.57, -0.59, -1.80)	
12	(18.5, -8.5, 83.9)	(20.23, -9.83, 81.60)	(1.73, -1.33, -2.30)	
13	(40.1, -4.7, 74.9)	(38.83, -6.11, 73.70)	(-1.27, -1.41, -1.50)	
14	(-34.2, -19.5, 70.5)	(-33.93, -18.87, 69.00)	(0.27, 0.63, -1.50)	
15	(-11.2, -12.5, 43.1)	(-12.26, -14.27, 41.44)	(-1.06, -1.77, -1.66)	

由表 8 可得, X 轴最大误差为 17.3 mm, 平均误差为 1.97 mm; Y 轴最大误差为 24.8 mm, 平均误差为 0.36 mm; Z 轴最大误差为 39.6 mm, 平均误差为 9.63 mm, 误差在允许范围内, 符合使用要求。

最后通过柑橘采摘机器人进行实地验证, 对 30 个可采摘柑橘和 20 个难采摘柑橘进行了识别与

定位试验, 结果如表 9 所示。

最终总体识别成功率达到 88%, 准确定位率达到 82%。无遮挡的可采摘柑橘识别率与准确定位率达到 96% 与 94%, 可以验证视觉系统给定的识别框准确, 三维坐标转换无误, 柑橘目标世界坐标测算达到预期精准度, 识别与定位试验如图 17 所示。

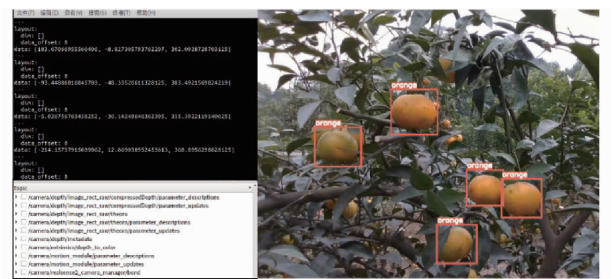


图 16 三维坐标算法识别界面

Fig. 16 3D coordinate algorithm detection interface

表 9 果园可采摘与难采摘柑橘成功识别与准确定位数

Tab.9 Successful identification and accurate localization of number of pickable and hard-to-pick citrus in orchards

果实类别	成功识别数	准确定位数
可采摘	28	27
难采摘	16	14
总和	44	41



图 17 果园识别与定位试验

Fig. 17 Orchard identification and localization

4 结论

- (1)通过分析果园复杂环境与果实自身状态来对柑橘目标分类,指引采摘机器人分序列对柑橘进行采摘,降低采摘机器人损耗风险,提高了末端执行器抓取成功率。
- (2)提出 YOLO v5-citrus 网络结构,在网络结构中插入 MCIFNet 模块,提高复杂柑橘特征的识别效果,修改颈部网络连接方式提高特定大小区间的识别权重,加入 CBAM 注意力机制,并修改损失函数为 SiIoU,识别模型精确率和平均精度均值达到 92.2% 与 91.1%,对比原 YOLO v5 分别提升 3.7 个百分点与 2.8 个百分点,对比 YOLO v7 与 YOLO v8 等均有提升,试验结果验证了 YOLO v5-citrus 可以有效解决各种复杂环境下柑橘目标的误识别和漏检问题。
- (3)将深度相机参数进行对齐,获取相机内部参数,通过像素坐标系到末端坐标系的换算提出柑橘目标中心定位算法,对识别到的柑橘目标做出准确定位。
- (4)在自然环境下分别进行了识别与定位试验。在真实果园中,对多种自然条件下的 50 个柑橘目标进行识别与定位,总体识别成功率和准确定位率达到 88% 与 84%。试验结果表明,本文提出的识别与定位算法能够实现果园中复杂状态柑橘的识别与定位。

参 考 文 献

[1] FAOSTAT. Food and Agricultural Organization of the United Nations[R]. 2024.

[2] YAN Bin, FAN Pan, LEI Xiaoyan, et al. A real-time apple targets detection method for picking robot based on improved YOLO v5[J]. Remote Sensing, 2021,13(9):1619.

[3] 何宇. 四川柑橘产业供给侧结构性改革绩效评价及推进路径研究[D]. 成都:四川农业大学, 2022.

HE Yu. Performance evaluation and of supply side structural reform of citrus industry in Sichuan research on promotion path [D]. Chengdu: Sichuan Agricultural University,2022. (in Chinese)

[4] 魏博,何金银,石阳,等. 欠驱动式柑橘采摘末端执行器设计与试验[J]. 农业机械学报, 2021,52(10):120-128.

WEI Bo, HE Jinyin, SHI Yang, et al. Design and experiment of underactuated end-effector for citrus picking[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(10):120-128. (in Chinese)

[5] 虞浪,俞高红,吴浩宇,等. 欠驱动关节型柑橘采摘末端执行器设计与试验[J]. 农业工程学报, 2023,39(17):29-38.

YU Lang, YU Gaohong, WU Haoyu, et al. Design and experiment of the end-effector with underactuated articulators for citrus picking[J]. Transactions of the CSAE, 2023,39(17):29-38. (in Chinese)

[6] 王毅,许洪斌,张茂,等. 仿蛇嘴咬合式柑橘采摘末端执行器设计与实验[J]. 农业机械学报, 2018,49(10):54-64.

WANG Yi, XU Hongbin, ZHANG Mao, et al. Design and experiment of bite-model end-effector for citrus harvesting by simulating with mouth of snake[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(10):54-64. (in Chinese)

[7] LIU Jizhan, YUAN Yan, ZHOU Yao, et al. Experiments and analysis of close-shot identification of on-branch citrus fruit with RealSense[J]. Sensors, 2018,18(5):1510.

[8] 蔡健荣,周小军,王锋,等. 柑橘采摘机器人障碍物识别技术[J]. 农业机械学报, 2009,40(11):171-175.

CAI Jianrong, ZHOU Xiaojun, WANG Feng, et al. Obstacle identification of citrus harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009,40(11):171-175. (in Chinese)

[9] 毕松,高峰,陈俊文,等. 基于深度卷积神经网络的柑橘目标识别方法[J]. 农业机械学报, 2019,50(5):181-186.

BI Song, GAO Feng, CHEN Junwen, et al. Detection method of citrus based on deep convolution neural network[J].

Transactions of the Chinese Society for Agricultural Machinery, 2019,50(5):181 – 186. (in Chinese)

[10] ZHANG Yunfeng, LI Li, CHUN Changpin, et al. Multi-scale feature adaptive fusion model for real-time detection in complex citrus orchard environments[J]. Computers and Electronics in Agriculture, 2024,219:108836.

[11] 王大明, 何逸霏, 李华英, 等. 柑橘采摘机器人图像识别算法研究[J]. 中国农机化学报, 2023,44(9):222 – 226, 264. WANG Daming, HE Yifei, LI Huaying, et al. Research on image recognition algorithm of citrus picking robot[J]. Journal of Chinese Agricultural Mechanization, 2023,44(9):222 – 226, 264. (in Chinese)

[12] XU Lijia, WANG Yihan, SHI Xiaoshi, et al. Real-time and accurate detection of citrus in complex scenes based on HPL – YOLOv4[J]. Computers and Electronics in Agriculture, 2023,205:107590.

[13] APPE S E, ARULSELVI G, BALAJI G N. CAM – YOLO: tomato detection and classification based on improved YOLO v5 using combining attention mechanism[J]. PeerJ Comput. Sci., 2023,9:e1463.

[14] WU Gang, ZHU Qibing, HUANG Min, et al. Automatic recognition of juicy peaches on trees based on 3D contour features and colour data[J]. Biosystems Engineering, 2019,188:1 – 13.

[15] 熊俊涛, 郑镇辉, 梁嘉恩, 等. 基于改进 YOLO v3 网络的夜间环境柑橘识别方法[J]. 农业机械学报, 2020,51(4):199 – 206. XIONG Juntao, ZHENG Zhenhui, LIANG Jiaen, et al. Citrus detection method in night environment based on improved YOLO v3 network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(4):199 – 206. (in Chinese)

[16] 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 Dean, 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 CASE, 2019,35(3):164 – 173. (in Chinese)

[17] 高新阳, 魏晟, 温志庆, 等. 改进 YOLO v5 轻量级网络的柑橘检测方法[J]. 计算机工程与应用, 2023,59(11):212 – 221. GAO Xinyang, WEI Sheng, WEN Zhiqing, et al. Citrus detection method based on improved YOLO v5 lightweight network [J]. Computer Engineering and Applications, 2023,59(11):212 – 221. (in Chinese)

[18] ZHU Linlin, GENG Xun, LI Zheng, et al. Improving YOLO v5 with attention mechanism for detecting boulders from planetary images[J]. Remote Sensing, 2021,13(18):3776.

[19] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[20] 刘建骐, 闫河, 王潇棠, 等. 改进金字塔和跳跃连接的 YOLO v5 目标检测网络[J]. 控制与决策, 2023,38(6):1730 – 1736. LIU Jianqi, YAN He, WANG Xiaotang, et al. YOLO v5 object detection network with improved pyramid and skip connection [J]. Control and Decision, 2023,38(6):1730 – 1736. (in Chinese)

[21] GEVORGYAN Z. Siou Loss: more powerful learning for bounding box regression Zhora Gevorgyan abstract[J]. arXiv Preprint, arXiv: 2205.12740, 2022.

[22] DAI Jian, ZHAO Xu, LI Lianpeng, et al. GCD – YOLO v5: an armored target recognition algorithm in complex environments based on array LiDAR[J]. IEEE Photonics Journal, 2022,14(4):1 – 11.

[23] LIU Hao, DING Meng, LI Shuai, et al. Small-target detection based on an attention mechanism for apron-monitoring systems [J]. Applied Sciences, 2023,13(9):5231.

[24] 刘嘉缘, 王国琿. 基于立体视觉的光度立体 LED 光源位置参数标定方法[J]. 激光与光电子学进展, 2022,59(16):299 – 306. LIU Jiayuan, WANG Guohui. Stereo vision-based calibration approach for position parameters of LED in photometric stereo [J]. Laser & Optoelectronics Progress, 2022,59(16):299 – 306. (in Chinese)

[25] 李丽, 卢世博, 任浩, 等. 基于改进 YOLO v5 的复杂环境下桑树枝干识别定位方法[J]. 农业机械学报, 2024,55(2):249 – 257. LI Li, LU Shibo, REN Hao, et al. Mulberry branch identification and location method based on improved YOLO v5 in complex environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(2):249 – 257. (in Chinese)