

DOI: 10.13733/j.jcam.issn.2095-5553.2024.09.026

熊明明, 李晓娟. 基于图像与深度信息融合的西梅果实识别定位方法[J]. 中国农机化学报, 2024, 45(9): 172-177

Xiong Mingming, Li Xiaojuan. Recognition and positioning method of prune fruit based on fusion of image and depth information [J]. Journal of Chinese Agricultural Mechanization, 2024, 45(9): 172-177

基于图像与深度信息融合的西梅果实识别定位方法*

熊明明¹, 李晓娟^{1, 2}

(1. 新疆大学机械工程学院, 乌鲁木齐市, 830017; 2. 新疆工业经济和信息化研究院, 乌鲁木齐市, 830091)

摘要:为实现西梅采摘机器人在复杂自然生长环境下对西梅果实的快速、准确识别定位,提出一种基于图像与深度信息融合的西梅果实识别定位方法。首先采集西梅图像,并通过数据增强建立阴天、沙尘等环境和不同角度下的西梅图像数据集,然后利用彩色图像基于YOLOv7网络模型快速识别西梅果实,获取果实表面中心点,再通过匹配融合彩色图和深度信息,采用深度距离分割剔除背景干扰噪声,实现西梅果实表面中心点的三维空间定位。试验结果表明,西梅果实识别模型可在多种果实生长分布场景下实现西梅果实的识别,其识别 F_1 值最高为95.8%,最低为83.2%;融合图像与深度信息的定位方法具有良好的定位效果,当深度距离小于1 m时,算法在各轴向上的定位误差均在0.005 m内;当深度距离为1.5 m时,误差最高为0.013 m,可满足西梅果实的识别定位要求。

关键词:西梅;目标识别;三维定位;信息融合;YOLOv7

中图分类号:S662.4; TP391.4 **文献标识码:**A **文章编号:**2095-5553(2024)09-0172-06

Recognition and positioning method of prune fruit based on fusion of image and depth information

Xiong Mingming¹, Li Xiaojuan^{1, 2}

(1. College of Mechanical Engineering, Xinjiang University, Urumqi, 830017, China; 2. Postdoctoral Research Station, Xinjiang Institute of Industrial Economics and Information Technology, Urumqi, 830091, China)

Abstract: In order to realize the rapid and accurate identification and positioning of prune fruit by the prune picking robot in the complex natural growth environment, a method for prune fruit identification and positioning based on the fusion of image and depth information was proposed. Firstly, the image of prune was collected, and data enhancement was used to establish prune image data sets under cloudy days, dusty environment and different angles, and then the color image was used to quickly identify prune fruits based on the YOLOv7 network model, the center point of the fruit surface was obtained. By matching and blending the color image and depth information, the depth distance segmentation was adopted to eliminate the background interference noise, to realize the three-dimensional spatial positioning of the center point on the prune fruit surface. The experimental results show that the prune fruit recognition model can realize the recognition of prune fruits in a variety of fruit growth and distribution scenarios, and its recognition F_1 value is 95.8% at the highest and 83.2% at the lowest, the positioning method of fusion image and depth information has a good effect positioning effect, when the depth distance is less than 1 m, the positioning error of the algorithm in each axis is within 0.005 m, when the distance is 1.5 m, the maximum error is up to 0.013 m, which can meet the requirements of recognition and positioning of prune fruit.

Keywords: prune; object identification; three-dimensional positioning; information fusion; YOLOv7

0 引言

新疆喀什地区是我国西梅种质资源、果品的主要

分布和集中产区之一。西梅果实采收仍依赖人工,受产业高速发展影响,西梅生产需求同劳动力短缺矛盾日趋显现。因此,探索更为高效、自动化、智能化的采

收稿日期:2022年12月24日 修回日期:2023年2月16日

* 基金项目:国家自然科学基金资助项目(52265003);新疆维吾尔自治区科学技术协会科技重点咨询项目(xjkj-2021-019)

第一作者:熊明明,男,1997年生,河南新蔡人,硕士研究生;研究方向为机器视觉、采摘机器人。E-mail: 1348111069@qq.com

通讯作者:李晓娟,女,1987年生,新疆塔城人,博士,副教授;研究方向为农牧特种机器人。E-mail: lxj_xj903@163.com

摘方式是西梅产业发展的必然趋势,其中果实识别定位是西梅智能采摘的关键性支撑技术。

近年来,国内外学者根据不同的视觉处理算法开展了果实采摘点的识别定位研究,分为基于色值^[1, 2]、形态^[3, 4]的传统图像处理方法^[1-4]和基于神经网络与深度学习^[5-9]。Benavides等^[1]通过基于颜色特征的变换和形态学操作识别分割单个番茄果实区域,实现番茄果实中心和果梗的二维定位。宁政通等^[8]根据葡萄生长特点,采用改进Mask R-CNN结合区域生长算法分割果梗,计算果梗区域质心,进而确定采摘点。上述研究虽然实现了果实的识别定位,但仅是对平面图像信息进行处理,未能实现果实采摘点的空间三维定位。

随着深度相机技术发展,根据不同的定位原理,学者们开展了果实的定位研究,分为飞行时间法(ToF)^[10-13]、双目视觉法^[14, 15]和结构光法^[16-18]。麦春艳等^[10]通过基于飞行时间的深度相机获取目标点云信息,对苹果果树进行三维重构,并以色彩分割和立体形状提取实现果实的识别与定位。罗陆锋等^[14]基于双目立体视觉,通过图像分割、霍夫直线、圆检测和立体匹配法进行葡萄的识别与定位,构建葡萄空间包围体。Cupec等^[16]验证了RGB-D相机通过检测凸特性来检测水果的有用性,但仅使用深度图像来检测水果,会导致水果过度分割或分割不足,检测精度较低。上述研究的开展,验证了将机器视觉与不同定位方法结合实现果实识别定位的可行性,但适用性、实时性有限,难以适应生产需要。

西梅果实娇小、果梗细短,呈团簇状生长分布,存在果实交相重叠、枝叶遮蔽遮挡、自然光照环境变化多样等非可控因素,传统图像处理方法特征设计难度大、算法泛化和鲁棒性差,现行高精度深度学习方法存在部署能力弱、资源需求大等问题;基于飞行时间和双目视觉的定位方法,检测精度有限、计算复杂且耗时高,同时,目前未见针对西梅果实采摘识别定位的研究。为实现在复杂自然环境下西梅果实的快速、精准识别定位,结合结构光法深度相机技术和快速、精准且易于部署的YOLOv7网络,提出一种基于图像与深度信息融合的西梅果实识别定位方法。

1 算法构成与数据集构建

1.1 算法构成

基于图像与深度信息融合的西梅果实识别定位算法,由彩色、深度图获取、映射匹配、背景干扰噪声剔除、基于YOLOv7网络西梅果实检测识别和基于图像与深度信息融合西梅果实定位五部分组成,算法流程如图1所示。

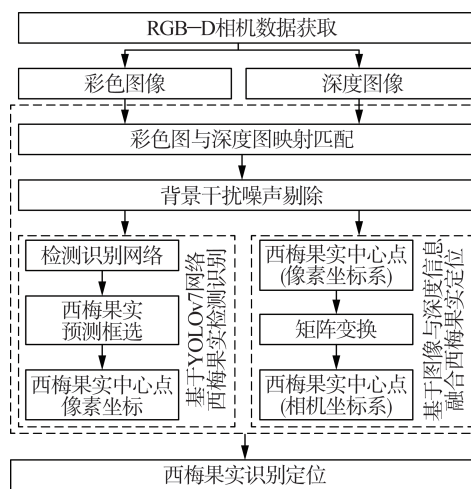


图1 西梅果实识别定位算法流程

Fig. 1 Prune fruit recognition and positioning algorithm process

1.2 数据集构建

1) 数据集采集与筛选。本文西梅图像数据采集自新疆喀什地区伽师县西梅种植基地。根据拍摄时景别的远近、果实的疏密程度和被遮挡果实的遮挡类别,选取2 870幅西梅图像组成西梅数据集。远近景别以拍摄时距果实远近为依据,1 m以内为近景;疏密程度以果实间距离、层叠数为基准,间距宽松、层叠2层以下为稀疏;轻重遮挡以被遮挡果实整体轮廓可见程度判断,整体轮廓可见为轻遮挡。

2) 数据集标注与增强。通过LabelImg对西梅数据图像进行标注,引入图像数据增强技术,通过进行图像亮度、色调、噪声、翻转、平移、裁剪、旋转等调整操作模拟西梅在不同天气环境和观察角度下的果实状态,同步扩充标注后的西梅数据集中图像和标签至8 610个,部分增强效果如图2所示,并将数据集按照7:1:2比例随机划分为训练集、验证集、测试集。

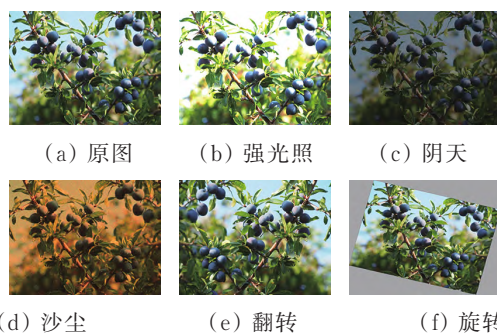


图2 部分增强效果图

Fig. 2 Partial enhancement effect diagram

2 西梅识别定位方法

2.1 基于YOLOv7西梅果实识别

2.1.1 西梅果实检测识别网络

西梅果实检测识别基于YOLOv7^[19]网络,其使用

了高效层聚合网络,采用模块级参数重参数化、辅助头检测与动态标签分配策略,具有较快的检测速度和较高的检测精度,其结构如图3所示。YOLOv7网络由Input(输入)、Backbone(骨干网络)、Head(头部)三部分组成,Input提供数据输入,Backbone用于提取特征,Head用于预测。西梅果实图像输入YOLOv7网络后,

经图片预处理统一缩放为640像素×640像素的分辨率大小,通过80×80、40×40、20×20三种不同尺度的特征图实现西梅大、中、小目标的检测,并获得在每种尺度特征图下3种不同尺寸的锚点,对各尺度特征图进行边界框预测,最终获取包含检测框坐标、检测类别和检测置信度的西梅目标预测信息。

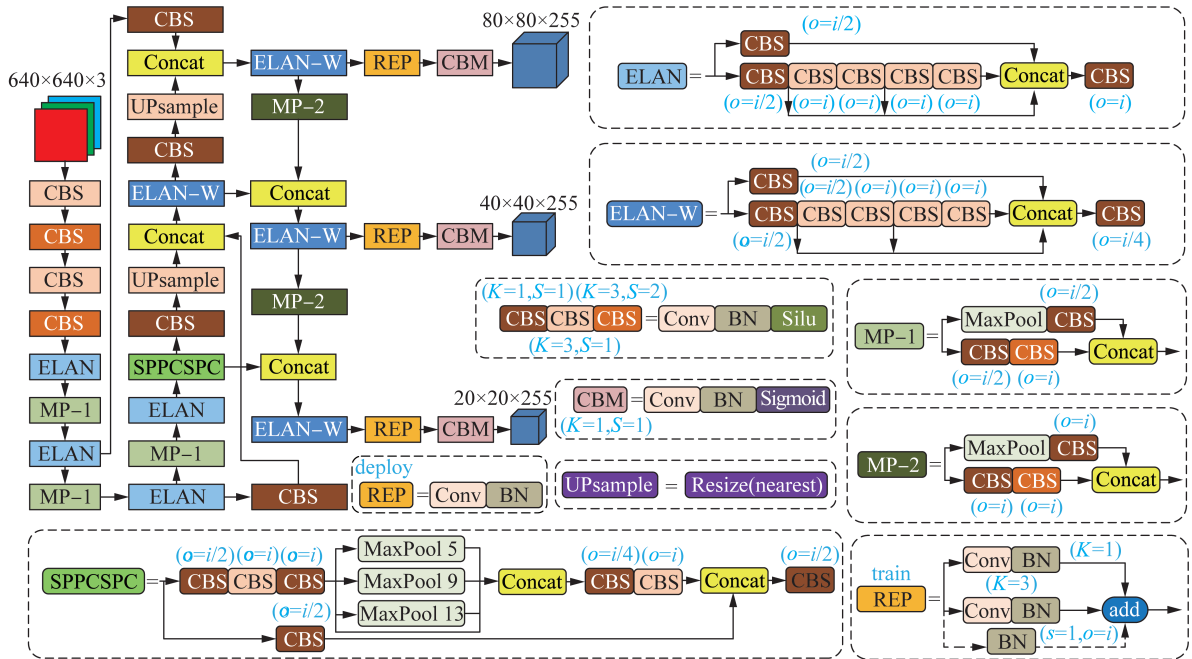


图3 西梅果实目标检测YOLOv7网络框架

Fig. 3 YOLOv7 network framework for prune fruit target detection

2.1.2 西梅果实表面中心点获取

相机采集的彩色图经过YOLOv7网络识别后,得到果实矩形检测框信息,其位置由像素坐标系下检测框左上角点 $P_{TL}(x_{TL}, y_{TL})$ 和右下角点 $P_{BR}(x_{BR}, y_{BR})$ 坐标确定,则检测框中心点为 $P_{CP}(\frac{x_{TL}+x_{BR}}{2}, \frac{y_{TL}+y_{BR}}{2})$,将西梅果实检测框的中心点作为西梅果实表面中心点,如图4所示。

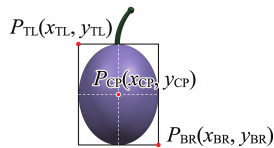


图4 西梅果实检测框参数示意图

Fig. 4 Prune fruit detection frame parameter diagram

2.2 基于图像与深度信息融合西梅果实定位

2.2.1 图像与深度信息融合

本文通过Intel RealSense™ Depth Camera D435i立体视觉深度相机采集西梅果实场景的彩色和深度图像,如图5所示。因深度相机中不同传感器参数和成像大小存在差异,采集的彩色图和深度图未一一对应,故需对彩色图和深度图配准,将深度图映射到彩色图上,实现图像与深度信息的融合,获取彩色图各像素

点对应的深度值。



(a) 彩色图 (b) 深度图

图5 RGB-D相机获取的彩色图和深度图

Fig. 5 Color and depth images obtained by RGB-D camera

2.2.2 背景干扰噪声剔除

西梅果实生长分布场景复杂,前后景物混合叠加且噪声多,导致西梅果实识别定位效果差。为解决以上问题,根据采摘机器人作业可达范围,将范围内的景物视为前景,其余视为背景,通过前、背景深度距离限制剔除复杂且无用的环境背景干扰噪声,提高果实识别精度,保障果实定位精度。深度距离分界定义如式(1)所示,剔除背景的色彩与深度融合图像如图6所示。

$$P_{RGB_depth}(x, y) = \begin{cases} (0, 0, 0) & P_{depth}(x, y) > threshold_{depth} \\ P_{RGB}(x, y) & otherwise \end{cases} \quad (1)$$

式中: $P_{depth}(x, y)$ ——点在深度图 (x, y) 处的深度值;
 $P_{RGB}(x, y)$ ——点在彩色图 (x, y) 处的RGB色彩值;

$P_{\text{RGB_depth}}(x, y)$ ——点在彩色图 (x, y) 处显示的 RGB 色彩值；

$threshold_{\text{depth}}$ ——深度阈值界限。



图 6 剔除背景的色彩与深度融合图像

Fig. 6 Color and depth fusion image without background

2.2.3 西梅果实中心点三维定位

经检测识别网络得到的基于像素坐标系的西梅果实表面中心点坐标 $P(u, v)$ ，需在彩色图和深度图配准映射后，经过坐标转换，得到基于相机坐标系下的西梅果实表面中心点三维坐标 $P_{\text{CP}}(x_{\text{CP}}, y_{\text{CP}}, z_{\text{CP}})$ ，实现空间三维定位，转换公式如式(2)所示。

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = z_c K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2)$$

式中： (x_c, y_c, z_c) ——相机坐标系下坐标值， z_c 经深度图与彩色图配准后获得；

(u, v) ——像素坐标系下坐标值；

K ——相机内部参数矩阵，由相机通过张式标定法标定。

3 结果与分析

3.1 模型训练与性能评估

3.1.1 训练环境及参数设置

模型训练采用 Windows11 64 位系统、PyTorch 等软件环境和 Intel i5-11400H 中央处理器，GeForce RTX 3050 显卡、16 GB 运行内存等硬件配置。采用 YOLOv7 网络模型进行训练，参数设置如表 1 所示。

表 1 训练参数设置

Tab. 1 Training parameter settings

参数	配置
预训练权重	YOLOv7_pt
训练轮次	500
批处理尺寸	8
学习率	0.01
权重衰减度	0.0005

3.1.2 模型评估和最优模型选取

本文选择衡量模型检测效果和评估模型整体性能的评价指标包含准确率(Precision, P)、召回率(Recall, R)、平均精度均值(mean Average Precision, mAP)，计算如式(3)~式(5)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n P(i)R(i) \quad (5)$$

式中： TP ——样本为正被检测为正样本数量；

FP ——样本为负被检测为正样本数量；

FN ——样本为正被检测为负样本数量；

n ——类别数量；

i ——类别编号。

经过 500 轮训练迭代后，绘制交并比阈值为 0.5 时的平均精度均值变化曲线，如图 7 所示。由图 7 可以看出，模型在前 11 次迭代时， mAP 值变化迅速，在迭代 117 次之后，趋于稳定，呈小幅度振荡。从 117~500 次迭代中选取第 426 次迭代生成的模型为最优模型，此时 mAP 值达到最大，为 0.974 6。

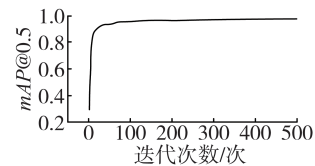


图 7 平均精度均值变化曲线

Fig. 7 Change curve of mean average precision

3.2 西梅果实识别测试

采用由精确率和召回率的调和均值表征的物体检测模型综合评价指标 F_1 值评估模型的识别性能，如式(6)所示。

$$F_1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (6)$$

为验证本文网络模型的识别性能，对不同场景下的西梅果实测试集图像进行识别测试，统计不同场景下西梅果实识别结果，如表 2 所示，部分识别效果如图 8 所示。

表 2 不同场景下西梅果实识别 F_1 值

Tab. 2 F_1 value of prune fruit recognition in different scenes

场景		F_1 值/%
近景	稀疏 轻遮挡	95.8
	稀疏 重遮挡	95.1
	密集 轻遮挡	93.6
	密集 重遮挡	91.4
远景	稀疏 轻遮挡	91.2
	稀疏 重遮挡	88.5
	密集 轻遮挡	84.9
	密集 重遮挡	83.2

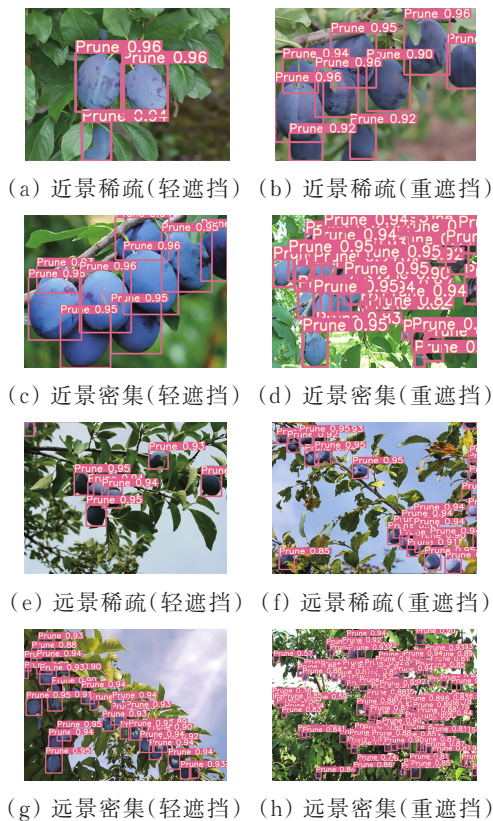


图8 不同场景下西梅果实识别效果

Fig. 8 Fruit recognition effect of prune in different scenes

由表2可知,本文模型在近景别下的识别效果优于远景别识别效果,近景别识别 F_1 值均在91%以上,远景别 F_1 值最高为91.2%,最低为83.2%;同时,在同景别下稀疏场景的识别效果总体要优于密集场景;在任意场景下,当果实存在遮挡时,果实在轻遮挡的识别 F_1 值均高于果实在重遮挡的识别的值。本文模型 F_1 值最高为95.8%,最低为83.2%。

3.3 果实定位测试

使用本文识别定位算法计算西梅果实表面中心点理论三维坐标 $P_{\text{Theoretical Value}}(x_{\text{TV}}, y_{\text{TV}}, z_{\text{TV}})$ 。试验时,固定相机于支架,在经过手眼标定的机械臂关节末端安装探针,移动西梅果实标靶至不同距离、高度处,将机械臂末端探针端点运动到西梅果实表面中心点,设在X、Y、Z方向上的定位误差分别为 Δx 、 Δy 、 Δz ,计算统计探针端点 $P_{\text{Measured Value}}(x_{\text{MV}}, y_{\text{MV}}, z_{\text{MV}})$ 与 $P_{\text{Theoretical Value}}(x_{\text{TV}}, y_{\text{TV}}, z_{\text{TV}})$ 的坐标差值,取同深度下误差均值为当前深度误差,并绘制不同深度距离下的定位误差曲线,如图9所示。果实表面中心点三维定位测试场景如图10所示。

由图9可以看出,定位误差随着距离的增大呈上升趋势,当深度距离小于1 m时,各轴向上的定位误差均小于0.005 m;当深度距离大于1 m时,其Z轴方向误差较其他轴误差增速明显;当深度距离为1.5 m时,最大误差来自Z轴方向,其定位误差为0.013 m。

深度误差增大的主要原因是果实目标在视野内的大小变化,受相机获取深度距离原理和精度限制,越小的果实获取的数据参数量越少,从而影响果实表面中心点识别定位精度,造成误差。在采摘机械臂可达工作范围内,其毫米级定位误差满足采摘精度要求,同时,通过末端采摘执行器的容差设计可消除误差的影响。

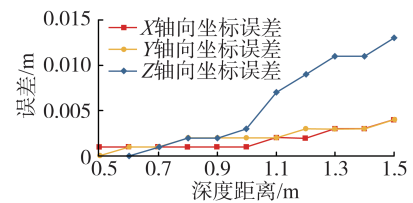
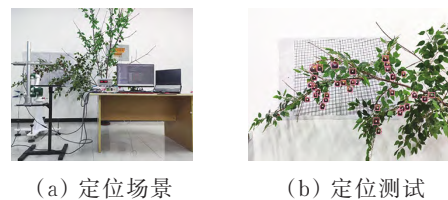


图9 深度距离定位误差曲线

Fig. 9 Depth distance positioning error curve



(a) 定位场景

(b) 定位测试

图10 果实三维定位测试场景

Fig. 10 Fruit three-dimensional positioning experiment scene

4 结论

1) 针对西梅采摘机器人在复杂自然生长环境下对西梅果实的快速、准确识别定位需求,本文研究基于YOLOv7网络模型和图像与深度信息融合的西梅果实识别定位方法。本文方法对不同场景下的西梅果实具有良好的识别定位效果,可以满足采摘机器人采摘工作对果实识别定位的需求。

2) 本文方法对不同生长场景下的西梅果实,其识别 F_1 值最高是近景稀疏轻遮挡果实为95.8%,最低是远景密集重遮挡果实为83.2%。定位误差随深度距离变化而变动,当深度距离小于1 m时,各轴向定位误差均小于0.005 m;当深度距离为1.5 m时,Z轴向误差最大,为0.013 m,X、Y轴向误差小于0.005 m。

参考文献

- [1] Benavides M, Cantón-Garbin M, Sánchez-Molina J A, et al. Automatic tomato and peduncle location system based on computer vision for use in robotized harvesting [J]. Applied Sciences, 2020, 10(17): 5887.
- [2] 王晓慧, 周昆鹏. 自然环境中的红色番茄图像识别方法研究[J]. 浙江大学学报(农业与生命科学版), 2021, 47(3): 395-403.
Wang Xiaohui, Zhou Kunpeng. Research on recognition methods for red tomato image in the natural environment [J]. Journal of Zhejiang University (Agriculture and Life

- Sciences), 2021, 47(3): 395–403.
- [3] 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.
- [4] 梁喜凤, 章艳. 串番茄采摘点的识别方法[J]. 中国农机化学报, 2016, 37(11): 131–134, 149.
Liang Xifeng, Zhang Yan. Recognition method of picking point for tomato cluster [J]. Journal of Chinese Agricultural Mechanization, 2016, 37(11): 131–134, 149.
- [5] Zheng C, Chen P, Pang J, et al. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard [J]. Biosystems Engineering, 2021, 206: 32–54.
- [6] 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.
- [7] 何斌, 张亦博, 龚健林, 等. 基于改进YOLOv5的夜间温室番茄果实快速识别[J]. 农业机械学报, 2022, 53(5): 201–208.
He Bin, Zhang Yibo, Gong Jianlin, et al. Fast recognition of tomato fruit in greenhouse at night based on improved YOLOv5 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 201–208.
- [8] 宁政通, 罗陆锋, 廖嘉欣, 等. 基于深度学习的葡萄果梗识别与最优采摘定位[J]. 农业工程学报, 2021, 37(9): 222–229.
Ning Zhengtong, Luo Lufeng, Liao Jiabin, et al. Recognition and the optimal picking point location of grape stems based on deep learning [J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(9): 222–229.
- [9] 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 Rendu, 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, 2019, 35(3): 164–173.
- [10] 麦春艳, 郑立华, 孙红, 等. 基于RGB-D相机的果树三维重构与果实识别定位[J]. 农业机械学报, 2015, 46(S1): 35–40.
Mai Chunyan, Zheng Lihua, Sun Hong, et al. Research on 3D reconstruction of fruit tree and fruit recognition and location method based on RGB-D camera [J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(S1): 35–40.
- [11] 刘德儿, 朱磊, 冀炜臻, 等. 基于RGB-D相机的脐橙实时识别定位与分级方法[J]. 农业工程学报, 2022, 38(14): 154–165.
Liu De'er, Zhu Lei, Ji Weizhen, et al. Real-time identification, localization, and grading method for navel oranges based on RGB-D camera [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(14): 154–165.
- [12] Malik M H, Qiu R, Gao Y, et al. Tomato segmentation and localization method based on RGB-D camera [J]. International Agricultural Engineering Journal, 2020, 28(4): 278–287.
- [13] 杨长辉, 刘艳平, 王毅, 等. 自然环境下柑橘采摘机器人识别定位系统研究[J]. 农业机械学报, 2019, 50(12): 14–22, 72.
Yang Changhui, Liu Yanping, Wang Yi, et al. Research and experiment on recognition and location system for citrus picking robot in natural environment [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 14–22, 72.
- [14] 罗陆锋, 邹湘军, 叶敏, 等. 基于双目立体视觉的葡萄采摘防撞空间包围体求解与定位[J]. 农业工程学报, 2016, 32(8): 41–47.
Luo Lufeng, Zou Xiangjun, Ye Min, et al. Calculation and localization of bounding volume of grape for undamaged fruit picking based on binocular stereo vision [J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(8): 41–47.
- [15] Wang C L, Tang Y C, Zou X J, et al. Recognition and matching of clustered mature litchi fruits using binocular charge-coupled device (CCD) color cameras [J]. Sensors, 2017, 17(11): 2564.
- [16] Cupec R, Filko D, Vidović I, et al. Point cloud segmentation to approximately convex surfaces for fruit recognition [C]. Proceedings of the Croatian Computer Vision Workshop, Zagreb, Croatia, 2014: 56–61.
- [17] Zhang Fan, Gao Jin, Zhou Hang, et al. Three-dimensional pose detection method based on key points detection network for tomato bunch [J]. Computers and Electronics in Agriculture, 2022, 195: 106824.
- [18] 冯青春, 程伟, 杨庆华, 等. 基于线结构光视觉的番茄重叠果实识别定位方法研究[J]. 中国农业大学学报, 2015, 20(4): 100–106.
Feng Qingchun, Cheng Wei, Yang Qinghua, et al. Identification and localization of overlapping tomatoes based on linear structured light vision system [J]. Journal of China Agricultural University, 2015, 20(4): 100–106.
- [19] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Train-able bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. arXiv e-prints, 2022.

