

基于优选 YOLOv7 模型的采摘机器人多姿态火龙果检测系统

王金鹏, 周佳良, 张跃跃, 胡皓若

(南京林业大学机械电子工程学院, 南京 210000)

摘要: 为了检测复杂自然环境下多种生长姿态的火龙果, 该研究基于优选 YOLOv7 模型提出一种多姿态火龙果检测方法, 构建了能区分不同姿态火龙果的视觉系统。首先比较了不同模型的检测效果, 并给出不同设备的建议模型。经测试, YOLOv7 系列模型优于 YOLOv4、YOLOv5 和 YOLOX 的同量级模型。适用于移动设备的 YOLOv7-tiny 模型的检测准确率为 83.6%, 召回率为 79.9%, 平均精度均值 (mean average precision, mAP) 为 88.3%, 正视角和侧视角火龙果的分类准确率为 80.4%, 推理一张图像仅需 1.8 ms, 与 YOLOv3-tiny、YOLOv4-tiny 和 YOLOX-tiny 相比准确率分别提高了 16.8、4.3 和 4.8 个百分点, mAP 分别提高了 7.3、21 和 3.9 个百分点, 与 EfficientDet、SSD、Faster-RCNN 和 CenterNet 相比 mAP 分别提高了 8.2、5.8、4.0 和 42.4 个百分点。然后, 该研究对不同光照条件下的火龙果进行检测, 结果表明在强光、弱光、人工补光条件下均保持着较高的精度。最后将基于 YOLOv7-tiny 的火龙果检测模型部署到 Jetson Xavier NX 上并针对正视角火龙果进行了验证性采摘试验, 结果表明检测系统的推理分类时间占完整采摘动作总时间的比例约为 22.6%, 正视角火龙果采摘成功率为 90%, 验证了基于优选 YOLOv7 的火龙果多姿态检测系统的性能。

关键词: 深度学习; 卷积神经网络; 采摘机器人; YOLOv7; 目标检测; 火龙果

doi: 10.11975/j.issn.1002-6819.202208031

中图分类号: TP301.6; TP181; S24

文献标志码: A

文章编号: 1002-6819(2023)-08-0276-08

王金鹏, 周佳良, 张跃跃, 等. 基于优选 YOLOv7 模型的采摘机器人多姿态火龙果检测系统[J]. 农业工程学报, 2023, 39(8): 276-283. doi: 10.11975/j.issn.1002-6819.202208031 http://www.tcsae.org

WANG Jinpeng, ZHOU Jialiang, ZHANG Yueyue, et al. Multi-pose dragon fruit detection system for picking robots based on the optimal YOLOv7 model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(8): 276-283. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202208031 http://www.tcsae.org

0 引言

近年来, 随着智慧农业的不断发展, 水果采摘机器人成为热门的研究课题^[1], 如何构建快速且准确检测果实的机器视觉系统是研究采摘机器人的关键问题之一^[2-3]。火龙果以其清爽的口感、显眼的外观和吉祥的名字而成为中国和东南亚最受欢迎的水果之一。目前火龙果的采摘方式多为人工采摘, 且火龙果的枝条较硬、带刺, 采摘劳动强度大、采摘成本高^[4], 因此, 实现火龙果智能化采摘是大势所趋。

果实识别与定位是采摘机器人智能化作业的重要环节, 目前对于果实识别与定位的研究较多^[5-7]。邓子青等^[8]基于 Otsu 算法与形态学对火龙果图像进行分割, 并求取处理后的质心, 实现了识别与定位, 与其他分割算法相比, 该方法对火龙果分割时的噪点大大减少, 边缘更加光滑。舒田等^[9]通过原始光谱、光谱不同形式变换和不同植被指数对火龙果植株冠层、果、枝、花等不同部位进行识别, 并构建 BP 模型来分析不同部位波段特征。结果表明绿红植被指数 (red green vegetation index) 对于

果实的识别能力最强, 准确率达 82.8%。肖冬娜等^[10]通过融合 4 种颜色指数, 构建融合颜色指数与点云数据空间结构的火龙果单株识别规则进行分割, 结果表明可见光波段差异颜色指数融合提取结果精度最高, 其单株面积值的均方根误差为 0.28 m²。

近年来, 随着深度学习技术的发展^[11], 越来越多的研究人员利用卷积神经网络来提取特征^[12-16], 规避了人为设计的特征提取过程, 因此有着更好的泛化能力^[17]。在农业领域, 单阶段目标检测算法 YOLO 系列以其高准确率和高效率而被广泛使用^[18-21]。WANG 等^[22]曾利用改进 YOLOv4 对火龙果进行识别, 将原本参数量较多的 CSPDarknet 主干网络替换为参数量较少的 MobileNetv3 以提高模型的检测速度, 将网络的第 39 层和 46 层直接进行上采样特征融合以提高对小目标的检测精度。改进后的 YOLOv4 平均检测精度达到 96.48%, 平均检测时间为 2.28 ms, 比原 YOLOv4 缩短了 160.32 ms。杨坚等^[23]利用改进 YOLOv4-tiny 对番茄果实进行识别, 增加了 76×76 的检测层以提高小番茄的识别准确率, 添加卷积注意力模块以提高遮挡番茄的识别准确率, 并使用密集连接的卷积网络以加强全局特征融合。该方法的平均精度达到 97.9%, 平均检测速度为每秒 111 帧。文斌等^[24]利用改进 YOLOv3 对三七叶片病害进行检测, 用注意力特征金字塔替代 YOLOv3 中原始特征金字塔以解决特征融合过程的干扰问题, 并使用双瓶颈层筛选注意力特征金字塔, 提取特征加强模型的泛化能力, 该方法相比原

收稿日期: 2022-08-03 修订日期: 2023-04-01

基金项目: 江苏省农业科技自主创新项目 (CX[22]3099); 国家林业和草原局应急科技项目 (202202-3); 江苏省现代农机装备与技术推广项目 (NJ2021-18); 江苏省重点研发计划项目 (BE2021016-2)

作者简介: 王金鹏, 博士, 副教授, 研究方向为智能化农业装备、智能采摘机器人。Email: jpwang@njfu.edu.cn

YOLOv3 模型, 平均精度提高了 1.47 个百分点, 且在复杂环境下抗干扰能力明显提高。刘天真等^[25]利用改进 YOLOv3 对自然场景下冬枣果实进行识别, 他们利用 SENet 中的 SE 结构将特征层的特征权重校准为特征权值, 提高了模型的识别精度, 该方法检测准确率达到 88.71%, 召回率为 83.80%, 平均精度为 82.01%, 与原模型相比平均精度提高了 4.78 个百分点。上述方法大多只考虑果实的位置, 很少考虑果实的姿态对采摘的影响。目前, 火龙果的识别定位研究有了一定的进展, 但仍存在不少问题: 1) 由于自然环境复杂多变, 目前存在的火龙果识别定位算法适应性普遍不足, 存在着较多误识别和漏识别的情况; 2) 采摘火龙果一般用剪切的方式, 由于火龙果的生长姿态比较复杂, 因此, 部分果实仅仅定位出果实中心点不足以引导机械手爪实现精准采摘。基于上述问题, 本文提出一种基于优选 YOLOv7 模型的多姿态火龙果检测方法, 在识别定位果实的基础上对不同生长姿态的火龙果进行分类, 通过对比 YOLOv7 系列模

型和其他先进的目标检测模型检测火龙果的性能, 最终选择检测速度快且精度较高的 YOLOv7-tiny, 并结合深度相机部署到移动设备上构建视觉检测系统, 最后, 对正视角火龙果进行了验证性采摘试验, 为火龙果的智能化精准采摘提供了一定的技术指导。

1 材料与方法

1.1 数据集的构建

本次试验使用的图像数据为果园实地拍摄所得, 拍摄地点为江苏省南京市溧水区溧乐农业火龙果园, 拍摄设备使用的是佳能 77D 相机, 图片保存格式为 JPEG, 分辨率为 2 400×1 600。受自然光照强度的影响火龙果的颜色略有不同, 向光果实呈现出亮红色, 背光果实呈现出暗红色。为了能够适应不同的工作环境, 在上午、下午以及傍晚进行图像数据采集, 共采集到 1 281 张火龙果的原始图像。包括强光、弱光及人工补光果实图像等, 图 1 为部分采集到的图像。



图 1 自然环境下的火龙果图像

Fig.1 Dragon fruit images in natural environment

使用 LabelImg 工具对原始图像进行标注, 并随机划分为训练集、验证集和测试集, 数量分别为 1 036、116 和 129 张。数据集按照三种光照条件划分结果如表 1 所示。

表 1 数据集划分表

Table 1 Table of dataset division

光照条件 Light conditions	训练集 Training set	验证集 Validation set	测试集 Test set	合计 Total
强光 Strong light	359	43	48	450
弱光 Weak light	436	48	51	535
人工补光 Artificial light	241	25	30	296
合计 Total	1 036	116	129	1 281

本研究依据火龙果相对于深度相机的位置不同将火龙果姿态分为侧视角火龙果和正视角火龙果。深度相机位于机器人腰关节正前方, 当视野中火龙果长在枝条的两侧时, 这类火龙果被检测为侧视角火龙果, 采用咬合式末端执行器的机器人需要调整末端执行器的接近角度, 使得末端执行器从火龙果的尾部方向靠近才能完成采摘; 另一方面, 当枝条上的火龙果尾部朝向机器人一侧时, 这类火龙果被检测为正视角火龙果, 咬合式末端执行器无需太多调整动作即可完成采摘, 图 2 为上述两类火龙果与机器人(末端执行器/视觉系统)的初始位置关系图。

如图 2 所示, 图 2a 中的果实能被现有的咬合式末端

执行器方便地剪切, 图 2b 中的果实不能很方便地被剪切采摘, 需要视觉系统更加精准的引导机器人对末端执行器进行精确的姿态调整。因此, 本研究对不同姿态的火龙果进行分类是有意义的。

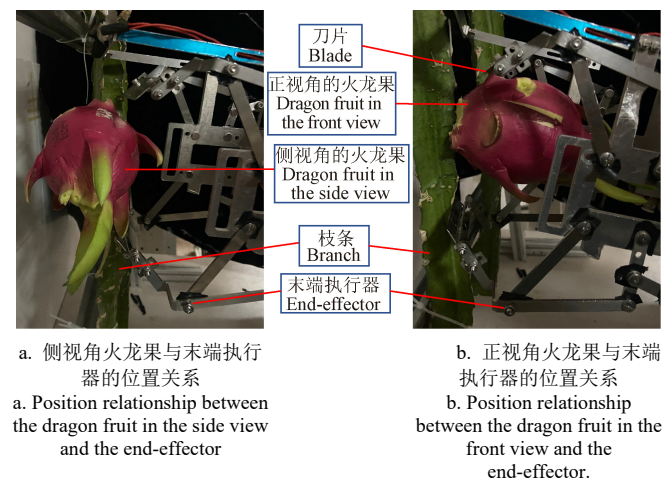


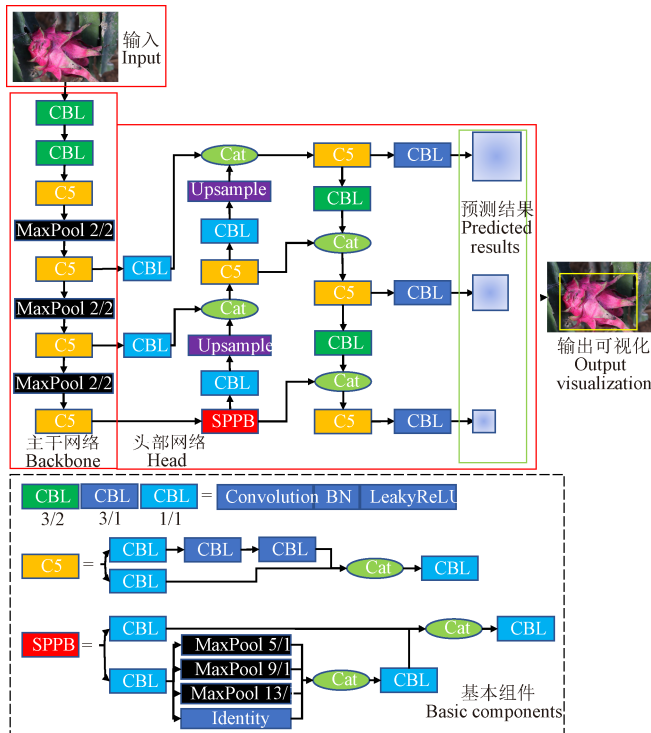
图 2 正视角火龙果和侧视角火龙果与末端执行器位置关系
Fig.2 Positional relationship between the dragon fruit in the front view and the dragon fruit in the side view and the end-effector

1.2 试验方法

1.2.1 YOLOv7 目标检测模型优选

YOLOv7 是 2022 年 7 月初推出的目标检测算法, 作

者在公共数据集上进行的对比试验结果表明, YOLOv7 有着目前目标检测领域最高的检测速度和检测精度。YOLOv7-tiny 是其中参数量最少的模型, 本研究拟优选 YOLOv7-tiny 目标检测模型构建视觉检测系统, 图 3 为 YOLOv7-tiny 模型结构示意图。输入图像经过 backbone 进行特征提取后获得 3 个不同尺度的特征图, 这 3 个尺度的特征图在 head 中进一步将不同尺度的特征进行整合, 最终输出 3 个尺度的预测结果分别用于检测大目标、中目标和小目标。



注: Convolution 指的是卷积层, BN 指的是批量归一化层, LeakyReLU 指的是激活函数。

Note: Convolution refers to the convolution layer, BN refers to the batch normalization layer, and LeakyReLU refers to the activation function.

图 3 YOLOv7-tiny 结构示意图

Fig.3 Structure of YOLOv7-tiny

1.2.2 试验环境

本文训练和测试使用的硬件环境为: CPU R7 5800X、GPU RTX 3070Ti、内存 8GB*2, 软件环境为 Win10 操作系统、Python 3.8、Pytorch 1.10、Torchvision 0.11、cuda 11.3、cudnn 8.2.0。

本文部署用于室外采摘试验的硬件主要为 Jetson Xavier NX 开发板、ZED 双目相机、S6H4D_Plus 六轴机械臂, 如图 4 所示。

1.2.3 模型训练

本次研究针对 YOLOv7 训练了 7 种网络, 分别为 YOLOv7、YOLOv7-d6、YOLOv7-e6、YOLOv7-e6e、YOLOv7-tiny、YOLOv7-w6 和 YOLOv7x。其中 YOLOv7-tiny、YOLOv7、YOLOv7-w6 分别是为边缘 GPU、普通 GPU 和云 GPU 设计的三种模型, YOLOv7x 则是对 YOLOv7 模型深度和宽度进行缩放获得, YOLOv7-e6 和 YOLOv7-d6 则是对 YOLO-w6 进行缩放所获得的, YOLOv7-e6e 则是由 YOLOv7-E6 使用 E-ELAN

后获得。

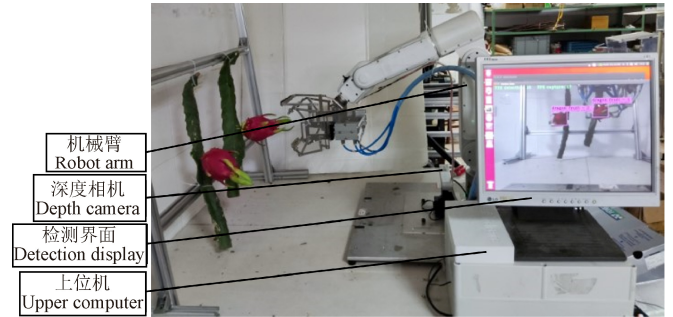


图 4 机器人采摘系统示意图

Fig.4 Schematic diagram of picking robot system

训练使用的超参数如下: 输入图像尺寸为 640×640 像素, 训练迭代次数为 200, 批量大小为 4, 参数优化器为 SGD (stochastic gradient descent), 动量 (momentum) 为 0.937, 权重衰减项 (weight decay) 为 0.000 5, 初始学习率为 0.01, 最小学习率为 0.001, 学习率下降为余弦退火算法, 训练的初始权重为随机初始化。

1.2.4 评价指标

本文选择准确率 (precision, P)、召回率 (recall, R)、准确率和召回率的调和平均 F1、平均精度 (average precision, AP)、平均精度均值 (mean average precision, mAP) 作为模型的评价指标^[12]。其中 P 可以反映模型的查准率, R 可以反映模型的查全率, $F1$ 可以反映准确率和召回率的综合性能, AP 可以反映单个类别的平均精度, mAP 可以反映所有类别的平均精度均值。

2 结果与分析

2.1 YOLOv7 系列模型的对比结果与分析

本节将对 YOLOv7 系列 7 种不同模型的检测性能, 为选择最适用于火龙果采摘机器人的模型提供依据。使用测试集中 129 张图片对 YOLOv7 系列的 7 种模型进行测试, 7 种模型的属性和检测性能如表 2 所示。

从表 2 各列中数值可以发现, YOLOv7-e6e 模型有着最高的准确率 85.0%, 这意味着使用该模型时误识别的概率最低, 采摘机器人错误抓取非目标果实的概率最低; YOLOv7x 模型有着最高的召回率 85.4%, 这意味着使用该模型时漏检率最低, 采摘机器人漏采摘的概率最低; YOLOv7x 模型也是检测侧视角的火龙果效果最好的模型, 使用该模型检测侧视角的火龙果平均精度为 92.3%; YOLOv7 模型检测的平均精度均值为 89.3%, 高于其他所有模型, 其也是检测正视角的火龙果效果最好的模型。当综合考虑准确率和召回率时, YOLOv7x 的 F1 值为 83.1%, 高于其他所有模型。YOLOv7-tiny 模型的参数量为 0.6×10^7 , 模型大小为 12 MB, 模型层数为 255, 推理时间为 1.8 ms, 其参数量、模型大小和模型层数最少, 这意味着使用该模型时, 单位时间内能够检测更多帧图像, 有利于提高采摘机器人的效率。根据表中 AP_0 和 AP_1 的结果, 可以发现, 正视角火龙果平均精度均低于侧视角的火龙果, 这意味着网络更难检测正视角的火龙果, 这

对机器人视觉系统和机械臂的空间布置提出了较高的要求。分析两类火龙果检测精度存在差异的原因后得知，

一些果实姿态介于正视角和侧视角之间，难以有定量的界定指标，尤其是正视角的果实。

表 2 7 种 YOLOv7 系列模型的检测性能和参数对比

Table 2 Comparison of detection performance and parameters of 7 YOLOv7 series models

模型 Models	<i>P</i> /%	<i>R</i> /%	<i>F1</i> /%	AP/%		平均精度均值 Mean average precision mAP/%	推理时间 Inference time/ms	参数量 Parameters	模型大小 Size of model/MB	层数 Number of layers
				AP ₀	AP ₁					
YOLOv7	82.3	83.7	83.0	91.9	86.7	89.3	6.3	3.72x10 ⁷	73	407
YOLOv7x	80.9	85.4	83.1	92.3	82.6	87.4	10.7	7.08x10 ⁷	139	459
YOLOv7-d6	82.9	74.5	78.5	87.8	69.8	78.8	10.5	13.29x10 ⁷	1015	702
YOLOv7-e6	84.3	79.8	82.0	90.8	86.3	88.6	8.7	9.65x10 ⁷	189	614
YOLOv7-e6e	85.0	80.7	82.8	89.7	84.4	87.1	13.7	15.10x10 ⁷	296	1032
YOLOv7-tiny	83.6	79.9	81.7	90.9	85.7	88.3	1.8	0.60x10 ⁷	12	255
YOLOv7-w6	82.7	82.4	82.5	90.4	85.9	88.2	5.8	6.98x10 ⁷	137	446

注：AP₀ 和 AP₁ 分别为对侧视角火龙果和正视角火龙果计算的平均精度。*P*、*R* 和 *F1* 分别为检测两类火龙果的平均准确率、平均召回率和调和级数。

Note: AP₀ and AP₁ are the average precision of the dragon fruits in the side view and dragon fruits in the front view, respectively. *P*, *R* and *F1* are the mean precision, mean recall, and harmonic progression for detecting two types of dragon fruits, respectively.

对于火龙果的采摘，经验表明，误识别相对于漏识别会造成更加严重的损害，当模型错误的将侧视角的火龙果识别为正视角的火龙果时，可能会损坏果实与枝条，甚至损坏末端执行器。选择模型的原则为保证召回率不会太低和保证检测的实时性的前提下选择准确率最高的模型。因此，对于配置较高的上位机，为了尽可能减少误识别率同时保证一定的检测速度，可以使用准确率最高的 YOLOv7-e6e 模型，该模型能最大程度降低误识别造成的影响，若上位机配置较高，也能保证检测的实时性。对于配置一般的上位机，选择参数量最小的 YOLOv7-tiny 模型，该模型最大的优势是轻量化。在一些配置较低的设备上也能实现实时检测，同时该模型的检测准确率为 83.6%，优于除了 YOLOv7-e6e 的所有模型。考虑到本次试验使用的上位机为 Jetson Xavier NX，YOLOv7-e6e 的检测时间过长，不利于机器人的实时性检测，因此优选 YOLOv7-tiny 作为后续机器人采摘试验的火龙果检测模型。

2.2 YOLOv7-tiny 与其他目标检测模型对比结果与分析

YOLO 系列是近些年热门的目标检测算法，为了充分说明 YOLOv7-tiny 的有效性。本文将 YOLOv7-tiny 与其他常用的目标检测模型进行对比，对比结果如表 3 所示。

从表 3 中可以看出，YOLOv7-tiny 与同量级的 YOLOv3-tiny、YOLOv4-tiny 和 YOLOX-tiny 相比准确率分别提高了 16.8、4.3 和 4.8 个百分点，mAP 分别提高了 7.3、21 和 3.9 个百分点；与 YOLOv5s、YOLOXs、YOLOv4G、YOLOv4M、YOLOv5x 和 YOLOXx 相比，准确率分别提高了 7.3、4.2、7.3、6.5、3.5 和 3.9 个百分点。与 EfficientDet、SSD、Faster-RCNN 和 CenterNet 相比 mAP 分别提高了 8.2、5.8、4.0 和 42.4 个百分点。

此外，CenterNet 模型检测火龙果的准确率最高（90.9%），但是该模型的召回率仅为 22.1%，这意味着存在大量的漏检测情况，不适用于采摘机器人。Faster-RCNN 模型检测火龙果的召回率最高（89.6%），但是该模型的准确率仅为 57.5%，这意味着存在大量误检测情况，也不适用于采摘机器人。YOLOv7-tiny 模型的平均精度均值最高为 88.3%，且准确率仅次于 CenterNet 模型，因此，YOLOv7-tiny 是最适合用于构建火龙果采摘机器人视觉检测系统的模型。

表 3 YOLOv7-tiny 与其他常用目标检测模型的检测性能对比

Table 3 Comparison of detection performance between YOLOv7-tiny and other commonly used object detection models %

模型 Models	评价指标 Evaluation metrics			
	<i>P</i>	<i>R</i>	<i>F1</i>	mAP
YOLOX-tiny	78.8	82.3	80.5	84.4
YOLOv4-tiny	79.3	48.8	60.4	67.3
YOLOv3-tiny	66.8	81.5	73.4	81.0
YOLOv5s	76.3	79.1	77.7	83.3
YOLOXs	79.4	85.3	82.8	86.5
YOLOv4G	76.3	66.2	70.9	63.1
YOLOv4M	77.1	54.4	63.8	60.5
YOLOv5x	80.1	78.6	79.3	83.3
YOLOXx	79.7	81.5	80.6	85.1
EfficientDet	83.3	68.5	75.2	80.1
CenterNet	90.9	22.1	35.6	45.9
SSD	73.1	86.0	79.0	82.5
Faster-RCNN	57.5	89.6	70.0	84.3
YOLOv7-tiny	83.6	79.9	81.7	88.3

注：YOLOv4G 和 YOLOv4M 分别代表主干网络为 GhostNet 和 MobileNetv3 的 YOLOv4 模型。

Note: YOLOv4G and YOLOv4M represent YOLOv4 models with GhostNet and mobilenetv3 backbone networks respectively.

2.3 YOLOv7-tiny 模型对自然光照环境下的火龙果检测结果

本研究根据火龙果生长姿态的不同将其分为正视角火龙果和侧视角火龙果，使用训练过的 YOLOv7-tiny 模型对 129 张测试集图像进行预测，其预测的混淆矩阵如表 4 所示。表中背景栏指的是除果实以外的物体以及模型在一些没有果实的位置检测出有果实的情况。

表 4 中可以看出侧视角火龙果和正视角火龙果的标签数量分别为 137 和 113，该 137 个侧视角火龙果有 116 正确分类，有 16 个错误的分类为正视角的火龙果，有 5 个漏识别；113 个正视角的火龙果有 85 个正确分类，有 18 个错误的分类为侧视角的火龙果，有 10 个漏识别；此外，有 4 个背景被误识别为侧视角的火龙果，1 个背景被误识别为正视角的火龙果。测试集中的 250 个目标果实，有 201 个被正确分类，分类准确率为 80.4%。原因分析：1) 极少部分火龙果的姿态处于正视角和侧视角的边缘，人工标注时可能存在少部分误差。2) 极少部分距离比较

远的果实在人工标注时被忽略,但是网络依然能识别它们,这就会导致将背景误识别为火龙果。

表 4 YOLOv7-tiny 预测的混淆矩阵
Table 4 Confusion matrix of YOLOv7-tiny

预测值 Predicted results	真实值 Ground truth		
	侧视角火龙果 Dragon fruits in the side view	正视角火龙果 Dragon fruits in the front view	背景 Background
侧视角火龙果的数量 Number of dragon fruits in the side view	116	18	4
正视角火龙果的数量 Number of dragon fruits in the front view	16	85	1
背景的数量 Number of background	5	10	

从在将模型部署到边缘计算设备上之前,应当验证模型有足够的泛化能力。本文使用的深度相机为 ZED 双目相机,其可以用于室外,但是该相机受光照影响较大。



a. 强光
a. Strong light



b. 弱光
b. Weak light



c. 人工补光
c. Artificial light

注: 图中矩形框表示检测出的果实的位置和边界; df_side 表示检测出的侧视角的火龙果; df_front 表示检测出的正视角的火龙果。

Note: The rectangular box in the figure represents the position and boundary of the detected fruit; df_side represents the detected dragon fruits in the side view; df_front represents the detected dragon fruits in the front view.

图 5 YOLOv7-tiny 在不同光照条件下检测结果

Fig.5 Detection results of YOLOv7-tiny under different light conditions

2.4 实地采摘试验

本次室外采摘试验地点为江苏省南京市溧水区溧乐农业,其种植方式为大棚式,当天天气为阴天,光照较弱。由于侧视角的火龙果不能直接根据识别到的中心点坐标直接采摘,本文只对识别到的正视角的火龙果进行采摘试验,对于识别到的侧视角需要进一步处理获取果实的生长方向等信息,作者在其他文献已有详细研究^[26-27]。

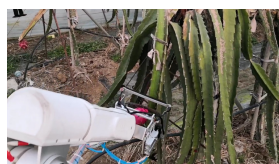
本文将 YOLOv7-tiny 部署到 Jetson Xavier NX 开发板



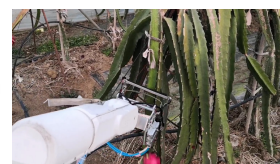
前往采摘点
Going to picking point



到达采摘点
Arriving at picking point



末端执行器剪切
Cutting action of end effector



成功采摘
Successful picking

图 6 机器人采摘火龙果的过程

Fig.6 Process of picking dragon fruit with a robot

图 6a 为软件系统处理完成后机械臂按一定的路径前往采摘点,6b 为机械臂到达采摘点,6c 为末端执行器执行剪切动作,6d 为成功采摘果实。

为了验证采摘系统的成功率以及 YOLOv7 模型的效率,进行了多组正视角的火龙果采摘试验,如果采摘的果实没有明显损伤,那么就认为采摘成功。表 6 中记录

因此,为了防止模型对某种光照条件下的火龙果检测性能太差,对表 1 测试集中 3 种不同光照条件分别进行测试,结果如表 5 所示。从表 5 中可以看出,模型在人工补光时的精度略低于强光和弱光时,分析其原因可能是测试集中人工补光的样本偏少,存在部分难以识别的样本。总体来说,在 3 种光照条件下 YOLOv7-tiny 模型均保持着较高的检测精度,这证明 YOLOv7-tiny 模型有着较好的泛化能力,能够用于自然环境下火龙果的检测。图 5 为 YOLOv7-tiny 模型对三种光照条件下的火龙果检测结果的可视化。

表 5 YOLOv7-tiny 模型对不同光照条件下火龙果的检测性能
Table 5 Detection performance under different lighting conditions of YOLOv7-tiny

光照条件 Light conditions	P/%	R/%	mAP/%
强光 Strong light	84.0	79.4	88.6
弱光 Weak light	84.7	81.3	89.2
人工补光 Artificial light	82.1	77.8	87.1

上,并结合深度相机与机械臂构建一套火龙果采摘系统,采摘系统的工作过程为:上位机处理深度相机采集到的火龙果图像得到火龙果在相机坐标系的三维坐标,再将相机坐标系内的坐标转换为机械臂坐标系的三维坐标,机械臂根据坐标规划路径到达采摘点。对于检测到的正视角的火龙果,会设置机械臂正常采摘,对于检测到的侧视角的火龙果,不设置机械臂的采摘。图 6 为机器人采摘火龙果的过程。

了 20 次采摘试验的数据。从表 6 中数据可以看出,上位机推理时间占整个采摘过程的比例较低,为 22.6%,说明本文部署的 YOLOv7-tiny 模型有较快的推理速度以及较高的定位准确率,能够满足实时性要求;正视角火龙果采摘成功率为 90%,说明本研究的采摘系统基本能够完成火龙果的自动化采摘。试验中两次采摘失败主要原

因为深度相机计算的距离存在一定误差，导致采摘时末端执行器剪切到果实。

表 6 采摘试验结果
Table 6 Results of picking experiments

试验序号 Experiment number	上位机推理时间 Upper computer inference time/s	总采摘时间 Total picking time /s	采摘结果 Picking results
01	4.37	17.82	采摘成功
02	4.56	19.26	采摘成功
03	4.26	20.36	采摘失败
04	4.59	18.86	采摘成功
05	4.13	19.78	采摘成功
06	4.66	20.12	采摘成功
07	4.17	18.85	采摘成功
08	4.63	18.95	采摘成功
09	4.21	19.53	采摘成功
10	3.97	19.62	采摘成功
11	4.36	19.43	采摘成功
12	4.51	18.84	采摘成功
13	4.62	20.88	采摘成功
14	4.17	20.46	采摘失败
15	4.38	18.64	采摘成功
16	4.11	17.95	采摘成功
17	4.42	19.23	采摘成功
18	4.32	20.15	采摘成功
19	4.37	19.46	采摘成功
20	4.62	18.65	采摘成功
平均值 Mean	4.37	19.34	90%

注：上位机推理时间指的是从整个程序开始运行到机械臂开始运动记录的时间，总采摘时间指的是从整个程序开始运行到机械臂采摘完成后复位记录的时间。

Note: The inference time of the upper computer refers to the time from the start of the entire program to the start of the motion recording of the robotic arm, and the total picking time refers to the time from the start of the entire program to the time when the robotic arm returns to the record after picking.

3 结 论

为了有效的实现火龙果自动化采摘，本研究基于 YOLOv7 模型，对不同姿态的火龙果分类，并对正视角的火龙果实现了采摘，其中主要结论如下：

1) YOLOv7 系列内的 7 种模型中，YOLOv7-e6e 模型的准确率最高，但是参数量太大，难以部署。YOLOv7-tiny 模型检测火龙果的准确率 P 为 83.6%，召回率 R 为 79.9%，平均精度均值 mAP 为 88.3%，火龙果姿态分类准确率为 80.4%，平均推理时间为 1.8ms，是 YOLOv7 系列中最适合部署在移动设备上的模型。与同参数量级 YOLO 系列中的 YOLOv3-tiny、YOLOv4-tiny 以及 YOLOX-tiny 相比较，准确率 P 分别提高了 16.8、4.3 和 4.8 个百分点，平均精度均值 mAP 分别提高了 7.3、21 和 3.9 个百分点。与 EfficientDet、CenterNet、SSD、Faster-RCNN 相比较，mAP 分别提高了 8.2，42.4，5.8 和 4.0 个百分点。在强光、弱光和人工补光的检测条件下，YOLOv7-tiny 均保持有较高水平的精度，说明 YOLOv7-tiny 模型有较强的泛化性。

2) 使用 YOLOv7-tiny 模型构建的采摘系统在自然环境下的采摘成功率为 90%，其中上位机计算时间占总采

摘时间的比例平均为 22.6%，验证了本文基于 YOLOv7-tiny 模型的采摘系统实现火龙果自动化采摘的可行性，为火龙果及类似水果在自然环境下的识别及定位提供了参考，为水果采摘机器人的研制提供了支撑。

【参 考 文 献】

- [1] XU Zhibo, HUANG Xiaopeng, HUANG Yuan, et al. A real-time zanthoxylum target detection method for an intelligent picking robot under a complex background, based on an improved YOLOv5s architecture[J]. Sensors, 2022, 22(2): 682.
- [2] 周裔扬, 邓三鹏, 祁宇明, 等. 基于 YOLOv5 的移动机器人目标检测算法的研究[J]. 装备制造技术, 2021, 8: 15-18. ZHOU Yiyang, DENG Sanpeng, QI Yuming, et al. Mobile robot target recognition algorithm based on YOLOv5[J]. Equipment Manufacturing Technology, 2021, 8: 15-18. (in Chinese with English abstract)
- [3] 王丹丹, 何东健. 基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别[J]. 农业工程学报, 2019, 35(3): 156-163. WANG Dandan, HE Dongjian. Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 156-163. (in Chinese with English abstract)
- [4] 黄凤珠, 陆贵锋, 武志江, 等. 火龙果种质资源果实品质性状多样性分析[J]. 中国南方果树, 2019, 48(6): 46-52, 58. HUANG Fengzhu, LU Guifeng, WU Zhijiang, et al. Diversity analysis of fruit quality traits in pitaya germplasm resources[J]. South China Fruits, 2019, 48(6): 46-52, 58. (in Chinese with English abstract)
- [5] 易诗, 李俊杰, 张鹏, 等. 基于特征递归融合 YOLOv4 网络模型的春见柑橘检测与计数[J]. 农业工程学报, 2021, 37(18): 161-169. YI Shi, LI Junjie, ZHANG Peng, et al. Detecting and counting of spring-see citrus using YOLOv4 network model and recursive fusion of features[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 161-169. (in Chinese with English abstract)
- [6] 杜文圣, 王春颖, 朱衍俊, 等. 采用改进 Mask R-CNN 算法定位鲜食葡萄疏花夹持点[J]. 农业工程学报, 2022, 38(1): 169-177. DU Wensheng, WANG Chunying, ZHU Yanjun, et al. Fruit stem clamping points location for table grape thinning using improved mask R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(1): 169-177. (in Chinese with English abstract)
- [7] WANG Peng, NIU Tong, HE Dongjian. Tomato young fruits detection method under near color background based on improved Faster R-CNN with attention mechanism[J]. Agriculture, 2021, 11(11): 1059.

- [8] 邓子青, 王阳, 张兵, 等. 基于 Otsu 算法与形态学的火龙果图像分割研究[J]. 智能计算机与应用, 2022, 12(6): 106-109.
DENG Ziqing, WANG Yang, ZHANG Bing, et al. Research on pitaya image segmentation based on Otsu algorithm and morphology[J]. Intelligent Computer and Applications, 2022, 12(6): 106-109. (in Chinese with English abstract)
- [9] 舒田, 陈智虎, 刘春艳, 等. 火龙果植株高光谱识别与特征波段提取[J]. 贵州农业科学, 2022, 50(3): 117-124.
SHU Tian, CHEN Zhihu, LIU Chunyan, et al. Hyperspectral identification and characteristic bands extraction of pitaya plant[J]. Guizhou Agricultural Sciences, 2022, 50(3): 117-124. (in Chinese with English abstract)
- [10] 肖冬娜, 周忠发, 尹林江, 等. 融合颜色指数与空间结构的喀斯特山地火龙果单株识别[J]. 激光与光电子学进展, 2022, 59(10): 489-503.
XIAO Dongna, ZHOU Zhongfa, YIN Linjiang, et al. Identification of single plant of karst mountain pitaya by fusion of color index and spatial structure[J]. Laser & Optoelectronics Progress, 2022, 59(10): 489-503. (in Chinese with English abstract)
- [11] 赵立新, 邢润哲, 白银光, 等. 深度学习在目标检测的研究综述[J]. 科学技术与工程, 2021, 21(30): 12787-12795.
ZHAO Lixin, XING Runzhe, BAI Yinguang, et al. Review on survey of deep learning in target detection[J]. Science Technology and Engineering, 2021, 21(30): 12787-12795. (in Chinese with English abstract)
- [12] 林森, 刘美怡, 陶志勇. 采用注意力机制与改进 YOLOv5 的水下珍品检测[J]. 农业工程学报, 2021, 37(18): 307-314.
LIN Sen, LIU Meiyi, TAO Zhiyong. Detection of underwater treasures using attention mechanism and improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 307-314. (in Chinese with English abstract)
- [13] PARICO A I B, AHAMED T. Real time pear fruit detection and counting using YOLOv4 models and deep sort[J]. Sensors, 2021, 21(14): 4803.
- [14] DING F, ZHUANG Z, LIU Y, et al. Detecting defects on solid wood panels based on an improved SSD algorithm[J]. Sensors, 2020, 20(18): 5315.
- [15] 龙洁花, 赵春江, 林森, 等. 改进 Mask R-CNN 的温室环境下不同成熟度番茄果实分割方法[J]. 农业工程学报, 2021, 37(18): 100-108.
LONG Jiehua, ZHAO Chunjiang, LIN Sen, et al. Segmentation method of the tomato fruits with different maturities under greenhouse environment based on improved Mask R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 100-108. (in Chinese with English abstract)
- [16] 蔡舒平, 孙仲鸣, 刘慧, 等. 基于改进型 YOLOv4 的果园障碍物实时检测方法[J]. 农业工程学报, 2021, 37(2): 36-43.
CAI Shuping, SUN Zhongming, LIU Hui, et al. Real-time detection methodology for obstacles in orchards using improved YOLOv4[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(2): 36-43. (in Chinese with English abstract)
- [17] GUO Q, CHEN Y, TANG Y, et al. Lychee fruit detection based on monocular machine vision in orchard environment[J]. Sensors, 2019, 19(19): 4091.
- [18] WANG Z, WALSH K, KOIRALA A. Mango fruit load estimation using a video based mangoYOLO-Kalman filter-Hungarian algorithm method[J]. Sensors (Basel, Switzerland), 2019, 19(12): 2742.
- [19] ZHENG Z, XIONG J, LIN H, et al. A method of green citrus detection in natural environments using a deep convolutional neural network[J]. Frontiers in Plant Science, 2021, 12: 705737.
- [20] 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)
- [21] 熊俊涛, 刘振, 汤林越, 等. 自然环境下绿色柑橘视觉检测技术研究[J]. 农业机械学报, 2018, 49(4): 45-52.
XIONG Juntao, LIU Zhen, TANG Linyue, et al. Visual detection technology of green citrus under natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery (Transactions of the CSAM), 2018, 49(4): 45-52. (in Chinese with English abstract)
- [22] WANG J, GAO K, JIANG H, et al. Method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(20): 218-225.
王金鹏, 高凯, 姜洪喆, 等. 基于改进的轻量化卷积神经网络火龙果检测方法(英文)[J]. 农业工程学报, 2020, 36(20): 218-225. (in English with Chinese abstract)
- [23] 杨坚, 钱振, 张燕军, 等. 采用改进 YOLOv4-tiny 的复杂环境下番茄实时识别[J]. 农业工程学报, 2022, 38(9): 215-221.
YANG Jian, QIAN Zhen, ZHANG Yanjun, et al. Real-time recognition of tomatoes in complex environments based on improved YOLOv4-tiny[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(9): 215-221. (in Chinese with English abstract)
- [24] 文斌, 曹仁轩, 杨启良, 等. 改进 YOLOv3 算法检测三七叶片病害[J]. 农业工程学报, 2022, 38(3): 164-172.
WEN Bin, CAO Renxuan, YANG Qiliang, et al. Detecting leaf disease for Panax notoginseng using an improved YOLOv3 algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(3): 164-172. (in Chinese with English abstract)
- [25] 刘天真, 滕桂法, 苑迎春, 等. 基于改进 YOLO v3 的自然场景下冬枣果实识别方法[J]. 农业机械学报, 2021, 52(5): 17-25.
LIU Tianzhen, TENG Guifa, YUAN Yingchun, et al. Winter

- jujube fruit recognition method based on improved YOLOv3 under natural scene[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5): 17-25. (in Chinese with English abstract)
- [26] ZHOU J, ZHANG Y, WANG J. RDE-YOLOv7: An improved model based on YOLOv7 for better performance in detecting dragon fruits[J]. Agronomy, 2023, 13: 1042.
- [27] ZHOU J, ZHANG Y, WANG J. A dragon fruit picking detection method Based on YOLOv7 and PSP-Ellipse[J]. Sensors, 2023, 23(8): 3803.

Multi-pose dragon fruit detection system for picking robots based on the optimal YOLOv7 model

WANG Jinpeng, ZHOU Jialiang, ZHANG Yueyue, HU Haoruo

(College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210000, China)

Abstract: Dragon fruit is one of the most popular fruits in Asia. The current manual picking cannot fully meet the requirement of large-scale production in recent years, due to the labor-intensive task. Alternatively, the automated picking of dragon fruit can be expected to greatly reduce labor intensity. Among them, the vision system can be one of the most important parts of the picking robot. The commonly-used recognition cannot consider the complex growth posture of dragon fruit. The hard branches and complex postures of dragon fruit can make it difficult to achieve automatic picking. It is a high demand to distinguish the dragon fruit with the different postures, and then guide the robotic arm to approach the fruit in an appropriate path. In this study, a multi-pose detection of dragon fruit was proposed for the automatic picking using optimal YOLOv7-tiny model. 1 281 images of dragon fruit were taken in the field, including 450, 535, and 296 images under strong, weak, and artificial light conditions. The image datasets were then divided into 1 036 images for training, 116 images for validation, and 129 images for testing, according to three light levels. Among them, the light conditions were the largest influencing factor on the detection performance. A series of experiments were conducted using the dataset. Firstly, the detection performance was compared with the seven models in the YOLOv7 series. The optimal models were given for the different devices, in terms of the number of model parameters and detection performance. Secondly, the detection performance of the YOLOv7 series models was compared with the other target detection models. Finally, the YOLOv7-tiny model was deployed into the mobile device. Specifically, the depth camera was combined with the robotic arm in the field picking. The results showed that the YOLOv7-e6e model in the YOLOv7 series presented the highest precision of 85.0%, while the YOLOv7x model was the highest recall of 85.4%, and the YOLOv7 model was the highest mean average precision (mAP) of 89.3%. The YOLOv7-tiny model shared the least parameters, weight files, layers, and inference time of 6×10^6 , 12MB, 255, and 1.8ms, respectively. It infers that the improved model was the most suitable for mobile devices, due to the fast inference speed. The detection precision of YOLOv7-tiny was 83.6%, the recall was 79.9%, the mAP was 88.3%, and the accuracy rate of classification for the multi-pose dragon fruits was 80.4%. Furthermore, the precision of YOLOv7-tiny increased by 16.8, 4.3, and 4.8 percentage points, respectively, whereas, the mAP increased by 7.3, 21, and 3.9 percentage points, compared with the YOLOv3-tiny, YOLOv4-tiny, and YOLOX-tiny. The precision of YOLOv7-tiny increased by 7.3, 4.2, 7.3, 6.5, 3.5, and 3.9 percentage points, respectively, compared with the YOLOv5s, YOLOXs, YOLOv4G, YOLOv4M, YOLOv5x, and YOLOXx. In addition, the mAP of YOLOv7-tiny increased by 8.2, 5.8, 4.0, and 42.4 percentage points, respectively, compared with the EfficientDet, SSD, Faster-RCNN, and CenterNet, indicating the high level of detection accuracy of YOLOv7-tiny model. The picking system of dragon fruit was constructed to verify by some picking experiments using the trained YOLOv7-tiny model. The experiment results show that the inference time of the vision system only accounted for 22.6% of the whole picking action time. The picking success rate of dragon fruits in the front view was 90%, indicating the higher performance of automatic picking than before. The conclusion can also provide technical support for fruit picking.

Keywords: deep learning; convolutional neural networks; picking robot; YOLOv7; object detection; dragon fruit