

基于 OWD – YOLO 的火龙果成熟度识别与三维姿态估计方法

周 成¹ 蔡浩洋¹ 付 威² 姚立立¹ 殷承亮¹

(1. 湖州师范学院信息工程学院, 湖州 313000; 2. 海南大学机电工程学院, 海口 570228)

摘要: 针对复杂果园环境下火龙果成熟度识别与姿态估计困难导致采摘精度不高的问题,以 YOLO 11n – pose 为基础,提出一种轻量高效的 OWD – YOLO 实时检测模型,实现火龙果精准、高效采摘。在基础模型中引入重参数化卷积,并与 C3K2 模块结合,增强对火龙果多尺度特征和细节姿态特征的提取能力;在 SPPF 模块中引入小波池化与大核卷积注意力机制,减少光照变化、背景遮挡等环境因素干扰,提高模型检测精度;在主干网络部分引入 DGECA 注意力机制提升模型对火龙果果皮颜色和纹理等重要特征的识别能力,改善成熟度判别准确性;在复杂果园环境中部署基于 OWD – YOLO 模型的六自由度机械臂采摘机器人平台,并通过深度相机实现火龙果三维姿态估计。田间试验结果表明,OWD – YOLO 模型目标检测精确率、平均精度均值与关键点检测平均精度均值分别达到 88.0%、92.7%、93.3%,较基础模型分别提高 5.0、2.4、2.0 个百分点,平均帧率为 58.7 f/s,单果采摘成功率达 86.0%,平均采摘时间为 29.4 s,能够满足复杂果园环境下精准机械化采摘需求。

关键词: 火龙果采摘机器人; 姿态估计; YOLO 11n – pose; 成熟度判别; 图像识别; 深度相机

中图分类号: S24 文献标识码: A 文章编号: 1000-1298(2026)08-0203-11

OSID:



Dragon Fruit Maturity Classification and 3D Pose Estimation Method Based on OWD – YOLO

ZHOU Cheng¹ CAI Haoyang¹ FU Wei² YAO Lili¹ YIN Chengliang¹

(1. School of Information Engineering, Huzhou University, Huzhou 313000, China

2. School of Mechanical and Electrical Engineering, Hainan University, Haikou 570228, China)

Abstract: In response to the challenges of low harvesting accuracy due to difficulties in the maturity recognition and pose estimation of dragon fruits in complex orchard environments, a lightweight and efficient OWD – YOLO real-time detection model was proposed based on YOLO 11n – pose to achieve precise and efficient harvesting of dragon fruits. Firstly, reparameterized convolution was introduced into the base model, combined with the C3K2 module, to enhance the model's ability to extract multi-scale features and fine-grained pose details of the dragon fruit. Secondly, by incorporating wavelet pooling and large kernel convolution attention mechanisms into the SPPF module, the model reduced interference from environmental factors such as lighting variations and background occlusion, thereby improving detection accuracy. Additionally, a DGECA attention mechanism was introduced into the backbone network to enhance the model's ability to recognize key features such as the fruit skin color and texture, improving the accuracy of maturity classification. Finally, a six-degree-of-freedom robotic arm harvesting platform based on the OWD – YOLO model was deployed in a complex orchard environment, with three-dimensional pose estimation of the dragon fruit achieved via a depth camera. Field experiments demonstrated that the OWD – YOLO achieved an object detection precision of 88.0%, a mean average precision of 92.7%, and a keypoint mean average precision of 93.3%, with absolute improvements of 5.0, 2.4, and 2.0 percentage points over the baseline, respectively. The average frame rate was 58.7 f/s, with a single fruit harvesting success rate of 86.0%, and an average harvesting time of 29.4 s. These results met the requirements for precise mechanized harvesting in complex orchard environments.

Key words: dragon fruit harvesting robot; pose estimation; YOLO 11n – pose; maturity classification; image recognition; depth camera

0 引言

火龙果是我国南方重要经济作物,但目前采摘仍主要依靠人工,劳动力短缺及成本上涨制约了产业效益的提升^[1]。近年来,以精准控制和自主感知为代表的农业机器人技术逐渐应用于果蔬采摘,不仅降低了劳动成本,还显著提高了采摘效率^[2-4]。

在果蔬采摘机器人中,视觉系统是实现高效、精准采摘的关键^[5-8]。机器人利用视觉图像分析技术,准确识别水果特征,实现果实空间定位^[9]。现时,目标检测算法迅速发展,深度学习已成为提升识别精度的重要手段^[10-11]。

目前,研究人员利用深度学习技术实现了火龙果目标检测,并在复杂环境下取得了较高精度的识别效果。文献[12-14]主要集中在精准的目标检测上,但火龙果生长姿态和成熟度检测同样对自动化采摘至关重要。

研究人员目前已认识到准确识别火龙果果实生长特征、成熟度对自动化采摘的重要性。文献[15-17]分别实现了火龙果果实姿态估计和成熟度检测,其中,成熟度判断决定果实是否达到采摘标准,生长姿态对执行器路径规划起着关键作用。

以上火龙果在自然环境中的目标检测精度以及果实的单一生长特征(如果实成熟度、姿态)研究均取得显著进展。然而,以上研究仅聚焦于单一特征识别,未充分考虑火龙果采摘过程果实姿态与成熟度叠加影响,实际采摘应用仍存在一定局限。

以YOLO 11n-pose为基础,本文提出一种轻量高效的OWD-YOLO火龙果成熟度与姿态联合检测模型。将轻量化的多尺度跨通道注意力(Dual global-pooling efficient channel attention, DGECA)用于主干网络,以提高模型对不同成熟度火龙果识别的准确性;改进特征提取模块C3K2以提升多尺度和细节姿态特征提取能力;融合小波池化与大核可分离卷积注意力机制的WLA_SPPF模块(Wavelet and LSKA-augmented spatial pyramid pooling fast),以提升对光照变化、枝条遮挡等环境因素的抗干扰能力;设计融合关键点与深度相机的三维姿态估计方法,搭建六自由度机械臂抓取平台开展田间试验,以期为实现火龙果精准自动采摘、缓解果园劳动力不足提供新技术路径与实践支撑。

1 材料与方法

1.1 果实成熟度与生长姿态划分

1.1.1 果实成熟度分类

在火龙果实际采摘中,由于植株个体差异及环

境适应性的影响,同一植株上常同时存在成熟与未成熟果实,需分批采摘。因此,火龙果自动采摘过程中,准确识别果实成熟度至关重要^[18]。依据NY/T 3911—2021《火龙果采收贮运技术规范》^[19]第4.2节中火龙果采摘成熟度判别标准,即当果皮开始转呈粉红色,鳞片青绿色时,判断为基本成熟(七成熟至九成熟),可以进行采收;同时,参照文献[20],将火龙果成熟度划分为2类:①可采摘成熟果实,即果皮呈粉红色至深红色,鳞片青绿色至褐色。②不可采摘未成熟果实,即果皮呈浅绿色至浅粉色,鳞片呈浅绿色。

1.1.2 果实生长姿态分类

除同一时期果实成熟度差异外,自然生长状态下火龙果空间姿态具有明显多样性。因此,为提高采摘平台成功率,根据火龙果果实与视觉采集装置位置关系,参照文献[21],将火龙果划分为正视角、侧视角和背视角3种类型,表示不同果实生长姿态。不同成熟度与姿态火龙果图像如图1所示。



图1 不同成熟度与姿态的火龙果图像

Fig. 1 Images of dragon fruit at different maturity levels and respective postures

1.2 数据集获取

数据采集地点为浙江省湖州市吴兴区上营村,时间为2024年7—10月,涵盖火龙果渐熟至采收的全部阶段。为确保数据多样性和全面性,在晴天、阴天、早晨及傍晚等不同光照条件下进行采集。同时,为提升模型适应性,采用近景、远景以及仰视、俯视和平视等不同拍摄视角组合多角度拍摄策略,特别关注果实在拍摄过程出现的遮挡情况,涵盖果间遮挡和枝条遮挡等复杂情景。共采集3 019幅火龙果图像,经数据清洗后保留2 559幅有效图像,图像分辨率分别为4 032像素×3 024像素和8 192像素×6 144像素。部分样本图像如图2所示。

采用Labelme标注工具对采集的火龙果图像进行目标检测框与关键点标注。应用最小外接矩形方式对火龙果定位,标注类别包括未成熟果实、成熟侧视角果实和成熟正视角果实3种类型。其中,对成熟正视角火龙果在其头部中心处标记关键点“Head”,对成熟侧视角火龙果则分别在果实头部、



图 2 火龙果图像数据集示例

Fig. 2 Dragon fruit image dataset examples

中心和尾部标记关键点“Head”、“Centre”和“Rear”，对未成熟火龙果不作关键点标注，并以比例 8:1:1 将图像划分为训练集、验证集和测试集。为提高模型在光照变化、位置差异及噪声干扰等条件下鲁棒性，对训练集图像进行数据增强处理，对图像注入高斯噪声，标准差设为灰度范围的 10%，同时随机实施 180°旋转变换，以 0.8 ~ 1.2 倍随机调整亮度与对比度扩充样本。获得样本图像 6 681 幅，生成标签数量 13 819 个，如表 1 所示。

表 1 不同类别火龙果标签数量

Tab. 1 Label counts of different dragon fruit categories				
火龙果类别	训练集	验证集	测试集	总计
未成熟	2 71	92	134	2 497
成熟侧视角	5 745	251	233	6 229
成熟正视角	4 725	189	179	5 093
总计	12 741	532	546	13 819

2 检测模型构建

2.1 OWD - YOLO 网络

YOLO 11 系列目标检测算法以较高的检测精度和实时性，在农业目标检测、果实识别与分级等领域已被广泛应用^[22]。考虑应用平台硬件算力限制，选取其系列中浮点运算量最小的 YOLO 11n - pose 模型为基础进行改进。

YOLO 11n - pose 模型最初用于人体姿态关键点预测任务。为解决复杂果园环境中火龙果成熟度分类困难、环境噪声干扰及边缘计算设备算力受限等问题，提出轻量高效的 OWD - YOLO 模型。模型以 YOLO 11n - pose 为基础，首先在主干网络中设计多尺度融合的 OM_C3K2 (Online reparameterized convolution and mixed local channel attention enhanced C3K2 module) 模块，以替代原始的 C3K2 结构；其次，引入融合小波池化与大核可分离卷积注

意力机制 (Large separable kernel attention, LSKA)^[23] 的 WLA_SPPF 结构，取代传统 SPPF (Spatial pyramid pooling fast)；最后，在模型主干网络中添加 DGECA 注意力。OWD - YOLO 结构如图 3 所示，其中 k 表示卷积核尺寸， s 表示步长， p 表示填充大小。

2.2 OM_C3K2 模块

YOLO 11 的 C3K2 模块融合 CSP 与 Bottleneck 结构，平衡了模型效率与检测性能。然而在火龙果实景拍摄中，由于拍摄距离不同，果实尺度差异显著。且在正视角姿态和侧视角姿态下，火龙果果实的轮廓、尺寸及纹理等主要特征，表现出较高的相似性，从而导致这 2 种姿态下果实难以有效区分。这些问题限制了 C3K2 模块在特征提取中的有效性。为此，在 C3K2 基础上融入在线重参数化卷积 (Online convolutional re-parameterization, OREPA)^[24] 和 MLCA 注意力机制 (Mixed local channel attention)^[25] 构建了 OM_C3K2 模块。

OM_C3K2 结构如图 4 所示，其中 c 表示卷积层输出通道数， c_{out} 表示模块输出通道数。首先以 1×1 卷积扩张通道并分为 2 路：主路直接保留原始特征，旁路经若干 OM_Bottleneck 与 OM_C3K 串联以提取更深层表征；随后将 2 路特征拼接，并通过 1×1 卷积压缩输出。OM_Bottleneck 由 2 层 OREPA 卷积与 MLCA 注意力构成，并以残差连接稳定训练。OM_C3K 先用 1×1 卷积扩展通道，再分为直通与深分支；深分支堆叠 OM_Bottleneck 后与直通分支拼接，最后以 1×1 卷积融合压缩。

相较于 C3K2，OM_C3K2 模块通过多分支卷积与通道注意力的协同作用强化了在复杂果园场景中火龙果特征表达。首先，并行多尺度卷积分支针对不同拍摄距离自适应提取果实表面的纹理细节，从近景细粒度纹理到远景低分辨纹理均能有效捕获。MLCA 机制利用局部池化保留果实表面纹理和轮廓的细节特征，通过全局池化对果实整体形态建模，能够提升果实形状和位置特征的表达，局部与全局信息的结合使得模型通过细微的形态学差异有效区分果实姿态。最后，模块在训练阶段将多分支卷积融合为等效单核卷积执行，在降低模型浮点运算量的同时提高检测性能。

2.3 WLA_SPPF 模块

火龙果采摘需在复杂果园环境下完成，采集火龙果图像易受光照变化、枝条遮挡、果间遮挡、多源成像噪声等因素影响，导致目标难以准确预测。为了提升 YOLO 算法在复杂果园环境下目标识别精度，设计 WLA_SPPF 模块，融合小波池化、多尺度空间金字塔池化和大核可分离卷积注意力机制，旨在

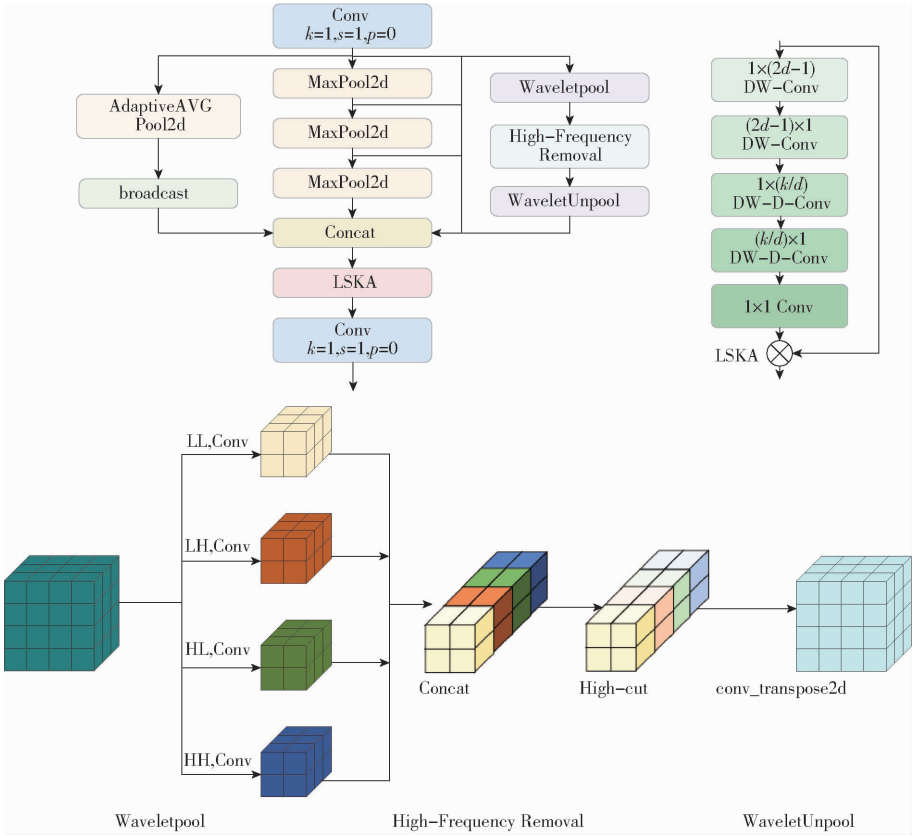


图 5 WLA_SPPF 模块结构

Fig. 5 Structure of WLA_SPPF module

周围枝条在颜色与纹理上差别微弱,严重影响了目标检测模型的检测效果;而转色期果实与成熟果实在颜色特征上相似,进一步增大了检测模型在成熟度判别上的难度。针对上述问题,设计一种轻量化的DGECA 通道注意力机制,以提高模型对火龙果成熟度检测的准确性。DGECA 注意力模块的设计基于ECA 注意力的机制(Efficient channel attention)^[26]。

DGECA 模块结构如图 6 所示,其中 h 表示特征图高度, w 表示特征图宽度, N 表示批量大小。首先,对尺寸为 $c \times h \times w$ 的特征图分别应用全局平均池化和全局最大池化操作,获得 2 个 $1 \times 1 \times c$ 的特征图,并将这 2 个特征图通过元素级加法融合。随后,使用动态卷积执行一维卷积操作,从而实现全局范围内对不同通道重要性的学习。其中,卷积核尺寸采用 ECA 注意力机制提出的自适应映射计算,由

通道数确定,并采用统一超参数设定,公式为

$$k = \left\lfloor \frac{\text{lb}c}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \tag{1}$$

式中 γ ——超参数,取 2

b ——超参数,取 1

最后,使用 Sigmoid 函数得到各个通道权重,将这些权重与原始特征图对元素相乘,得到最终输出特征图。

DGECA 模块通过全局平均池化和全局最大池化能够提取每个通道的语义信息和显著特征,再通过一维卷积动态调整各通道的权重,从而强化模型对果实颜色与表面纹理等关键信息通道的关注。这一机制使得模块能够有效捕捉果实与背景的视觉差异,从而提升火龙果成熟度检测的准确性和鲁棒性。

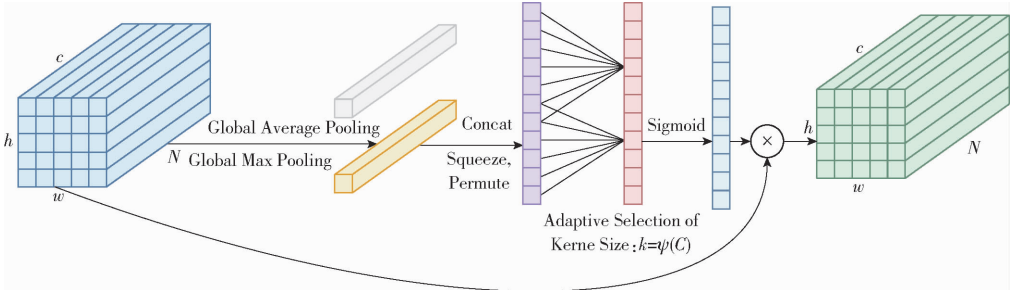


图 6 DGECA 模块结构

Fig. 6 Structure of DGECA module

3 结果与分析

3.1 试验环境及参数配置

在 Windows 10(64 位)操作系统环境下搭建并训练神经网络模型,并在边缘设备上对模型进行验证与测试。计算机配置为运行内存 16 GB、处理器 Intel(R)Core(TM)i5-12400F,以及显存 12 GB 的 NVIDIA GeForce RTX 4070 显卡。深度学习环境配置为 Python 3.9.19、PyTorch 2.4.0、TorchVision 0.19.0、CUDA 版本为 12.4。模型训练主要参数设置如表 2 所示。

表 2 主要训练参数
Tab.2 Primary training parameters

参数	数值/形式
训练图像尺寸/(像素×像素)	640×640
批量大小	8
迭代次数	120
初始学习率	0.01
周期学习率	0.01
学习率动量	0.937
权重衰减系数	0.000 5
Mosaic 数据增强	关闭

表 3 消融试验结果

Tab.3 Results of ablation test

模型	OM_C3K2	WLA_SPPF	DGECA	精确率/ %	召回率/ %	mAP@ 0.5 / mAP - kp@ 0.5/ %	浮点运算量	边缘设备帧率/ (f·s ⁻¹)	
YOLO 11n - pose	×	×	×	83.0	83.3	90.3	91.3	6.7 × 10 ⁹	70.9
模型 1	√	×	×	86.5	84.0	91.8	92.3	6.1 × 10 ⁹	65.9
模型 2	×	√	×	86.7	84.4	91.9	92.2	7.1 × 10 ⁹	59.2
模型 3	×	×	√	84.7	85.7	92.3	92.5	6.7 × 10 ⁹	69.2
OWD - YOLO	√	√	√	88.0	83.8	92.7	93.3	6.7 × 10 ⁹	58.7

注:√表示采用此模块,×表示不采用此模块。

将其与改进模型进行对比,并对每个改进模块进行定量评价。模型 1 将主干网络特征提取模块替换为 OM_C3K2 后,先通过不同的卷积分支学习特征,增强模型对不同尺度和姿态果实特征的捕获能力,随后将卷积分支重参数化为单一卷积,从而显著降低了模型复杂度。因此,火龙果识别精确率提高 3.5 个百分点,mAP@0.5 和 mAP-kp@0.5 分别提升 1.5、1.0 个百分点,且浮点运算量减少 6×10⁸。模型 2 引入 WLA_SPPF 模块后,减少光照变化、遮挡等环境因素的干扰,因此精确率、召回率、mAP@0.5 和 mAP-kp@0.5 分别提升 3.7、1.1、1.6、0.9 个百分点。模型 3 加入轻量化的 DGECA 注意力机制后,增强对颜色、纹理等重要通道特征的响应,火龙果识别精确率和召回率分别提高 1.7、2.4 个百分点,mAP@0.5 和 mAP-kp@0.5 分别提升 1.9、1.2

边缘设备配置为运行内存 16 GB、处理器 Intel(R)Core(TM)i7-8750H,以及显存 8 GB 的 NVIDIA GeForce GTX 1060 显卡。深度学习环境配置为 Python 3.9.23、PyTorch 2.7.1、TorchVision 0.22.1、CUDA 版本为 12.6。

3.2 评价指标

为全面评估模型对火龙果成熟度与姿态关键点识别任务的性能与效率,使用精确率(Precision,P),召回率(Recall,R),IOU 阈值为 0.5 时的平均精度均值(mean Average Precision,mAP@0.5)衡量模型目标检测性能,采用 IOU 阈值为 0.5 时的关键点平均精度均值(mean Average Precision for keypoints,mAP-kp@0.5)衡量模型关键点检测性能。采用浮点运算量和模型在边缘设备帧率作为效率评价指标。

3.3 消融试验

为评估模型有效性,验证本文 OM_C3K2、WLA_SPPF 和 DGECA 模块的效果,遵循控制变量原则,设计 4 组消融试验,模型在验证集上的结果如表 3 所示。

以未经改进的 YOLO 11n-pose 作为基准模型,

个百分点,浮点运算量与边缘设备检测帧率基本保持不变。融合上述 3 个模块后,模型对火龙果识别精确率和召回率分别提高 5.0、0.5 个百分点,达到 88.0%和 83.8%,mAP@0.5 和 mAP-kp@0.5 分别提升 2.4、2.0 个百分点,达到 92.7%和 93.3%,浮点运算量保持不变。结果表明,模型可以在不增加计算量的条件下更准确地实现果实检测。模型在算力受限的边缘计算设备上检测速率虽较基准模型降低 12.2 f/s,但仍能达到 58.7 f/s,能够满足采摘平台实时检测需求。

图 7 展示了 YOLO 11n-pose 基准模型与 OWD-YOLO 模型以 0.5 的置信度在密集遮挡的复杂果园环境下,火龙果不同成熟度及姿态检测效果,其中红色圆圈表示漏检,红色三角表示误检。在图 7a、7b 中,YOLO 11n-pose 因图像中果实较小与姿态特征不明

显,出现 2 次漏检与 1 次误检,OWD - YOLO 通过引入 OM_C3K2 模块,更好地捕捉不同尺度及姿态细节的果实特征,解决了遮挡引起的漏检和误检问题;在图 7c 中,YOLO 11n - pose 误将未成熟果实识别为背景枝条,导致 2 次漏检,模型引入 DGECA 模块后,强化了模型对未成熟果实与枝条间的区分度,降低了漏检率;图 7d 中由于强烈光照影响,图 7e 中果实被图

像边缘遮挡,YOLO 11n - pose 均有漏检现象产生,而 OWD - YOLO 通过引入 WLA_SPPF 模块,有效减轻了环境因素对果实特征的干扰,从而解决了漏检现象。结果表明,本文提出的 OM_C3K2、WLA_SPPF 以及 DGECA 模块有效提升了模型对果实特征的识别能力,充分验证了改进策略在复杂果园环境下火龙果成熟度与姿态识别任务中的有效性与实用性。

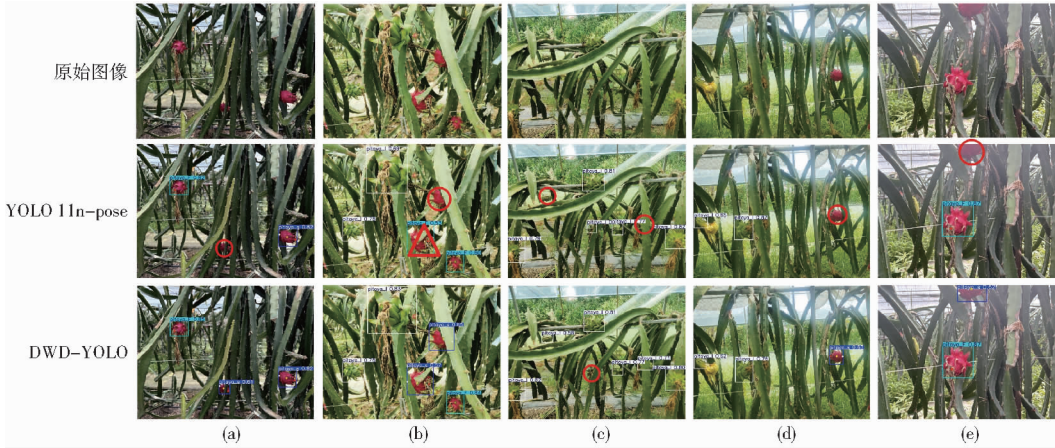


图 7 YOLO 11n - pose 与 OWD - YOLO 检测结果对比

Fig. 7 Comparison of detection results between YOLO 11n - pose and OWD - YOLO

3.4 不同检测模型性能试验

为进一步验证 OWD - YOLO 模型在火龙果检测中的有效性,基于自制火龙果数据集,与当前主流

的 YOLO 检测模型及轻量化主干网络实施性能对比分析。试验结果如表 4 和表 5 所示。

由表 4 可知,OWD - YOLO 模型的精确率、召回

表 4 不同 YOLO 检测模型试验结果

Tab. 4 Experimental results among different YOLO detection models

模型	精确率/%	召回率/%	mAP@0.5/%	mAP - kp@0.5/%	浮点运算量	边缘设备帧率/(f·s ⁻¹)
YOLO v5n - pose	80.5	83.4	90.0	90.5	6.1 × 10 ⁹	72.2
YOLO v6n - pose	83.5	81.4	89.7	90.6	1.2 × 10 ¹⁰	66.9
YOLO v8n - pose	83.6	82.3	90.2	91.0	7.1 × 10 ⁹	67.0
YOLO 11n - pose	83.0	83.3	90.3	91.3	6.7 × 10 ⁹	70.9
YOLO 11s - pose	87.5	84.2	91.4	92.2	2.2 × 10 ¹⁰	38.8
YOLO 11m - pose	83.6	89.6	93.2	93.2	7.1 × 10 ¹⁰	23.8
YOLO 12n - pose	82.0	83.6	90.2	90.9	6.1 × 10 ⁹	61.5
OWD - YOLO	88.0	83.8	92.7	93.3	6.7 × 10 ⁹	58.7

表 5 不同轻量化主干网络试验结果

Tab. 5 Experimental results among different lightweight backbone networks

模型	精确率/%	召回率/%	mAP@0.5/%	mAP - kp@0.5/%	浮点运算量	边缘设备帧率/(f·s ⁻¹)
EfficientNet2	80.2	82.0	88.4	89.0	4.2 × 10 ⁹	66.4
MobileNetV3	81.8	80.6	89.5	91.2	4.3 × 10 ⁹	60.6
GhostNetV2	82.3	80.9	89.2	89.8	6.4 × 10 ⁹	32.4
OWD - YOLO	88.0	83.8	92.7	93.3	6.7 × 10 ⁹	58.7

率、mAP@0.5 和 mAP - kp@0.5 均优于轻量级 n 系列的 YOLO 模型,分别达到 88%、83.8%、92.7% 和 93.3%。其中,精确率较 YOLO v5n - pose、YOLO v6n - pose、YOLO v8n - pose、YOLO 11n - pose 和 YOLO 12n - pose 分别提高 7.5、4.5、4.4、5.0、6.0

个百分点;召回率分别提升 0.4、2.4、1.5、0.5、0.2 个百分点,mAP@0.5 分别提升 2.7、3.0、2.5、2.4、2.5 个百分点;mAP - kp@0.5 分别提升 2.7、2.6、2.3、2.0、2.4 个百分点。结果表明 OWD - YOLO 模型与同等轻量级的 YOLO 模型相比,表现出更好的

检测性能。相较于 YOLO 11s - pose, OWD - YOLO 模型精确率、mAP@ 0.5 和 mAP - kp@ 0.5 分别提高 0.5、1.3、1.1 个百分点,浮点运算量减少 1.53×10^{10} ,帧率提高 19.9 f/s。与参数量较大的 YOLO 11m - pose 相比,OWD - YOLO 在精确率和 mAP - kp@ 0.5 上表现更优,召回率和 mAP@ 0.5 虽稍有欠缺,但浮点运算量减少 6.43×10^{10} ,帧率提高 34.9 f/s,更适合在算力受限的边缘设备上部署。

由表 5 可知,OWD - YOLO 模型在浮点运算量上略高于其他轻量化的主干网络,但在精确率、召回率、mAP@ 0.5 和 mAP - kp@ 0.5 上均优于其他模型。其中,精确率比 EfficientNet2、MobileNetV3 和 GhostNetV2 分别提升 7.8、6.2、5.7 个百分点;召回率分别提升 1.8、3.2、1.8 个百分点;mAP@ 0.5 分别提升 4.3、3.2、3.5 个百分点;mAP - kp@ 0.5 分别提升 4.3、2.1、3.5 个百分点。OWD - YOLO 帧率为 58.7 f/s,较 GhostNetV2 高 26.3 f/s。相较于主流的轻量化主干网络,本文模型提高了检测精度,减少了误检和漏检,更适用于火龙果检测任务。

3.5 复杂果园环境下模型性能检测

为验证 OWD - YOLO 模型在复杂果园环境下实际检测性能,将模型部署于自研的搭载深度相机和边缘计算设备的火龙果自动采摘机器人平台,并在田间开展火龙果实时检测试验。

图 8 展示了实地检测结果。从图 8 可以看出,本文模型在真实果园环境中能够准确识别火龙果的成熟度和果实姿态,即使在果实被枝条和果实相互遮挡的复杂场景下,模型仍能较为精确地检测到目标果实,并预测被遮挡果实的关键点位置,表明了 OWD - YOLO 模型具有较强的准确性和鲁棒性。



图 8 火龙果实地检测结果

Fig. 8 Field test results of dragon fruit

3.6 火龙果姿态检测方法

火龙果末端柄强度较高,因此通过旋转执行器

实现果实与枝条分离。为提高末端执行器抓取果实的精度,提出一种基于 OWD - YOLO 模型的火龙果三维姿态检测方法。

第 2.1 节提出的 OWD - YOLO 关键点检测模型已经实现对不同姿态果实端点的检测与分类,结合深度相机对检测的关键点进行空间定位,可以精确获取火龙果头部和根部三维空间位置。为准确表达火龙果果实相对于机械臂坐标系的空间姿态,以各姿态的关键点为基准,采用四元数表示方法。与传统方向余弦矩阵和欧拉角姿态表示方法相比,四元数不仅避免了“万向节锁死”问题,还提高了计算效率和姿态插值平滑度,满足火龙果自动化采摘任务对实时性和精度的要求。以正视角果实的头部关键点为基准,并以深度相机视轴为方向,构建正视角果实姿态四元数。针对侧视角火龙果的四元数姿态表达方法,定义果实中第 i 个未被遮挡的火龙果三维坐标为

$$P_i = (x_i, y_i, z_i) \tag{2}$$

式中 (x_i, y_i, z_i) ——火龙果第 i 个关键点在世界坐标系下 x 、 y 、 z 轴方向位置坐标

四元数法进行姿态表达,先计算 2 个关键点之间的单位方向向量,公式为

$$(v_x, v_y, v_z) = \frac{(d_x, d_y, d_z)}{\sqrt{d_x^2 + d_y^2 + d_z^2}} \tag{3}$$

其中 $(d_x, d_y, d_z) = P_1 - P_2$
式中 P_1, P_2 ——火龙果关键点世界坐标
 d_x, d_y, d_z ——关键点间连线向量分量
 v_x, v_y, v_z ——关键点单位连线向量分量
接着构造四元数旋转轴向量分量,公式为

$$(r_x, r_y, r_z) = u(v_x, v_y, v_z) \tag{4}$$

其中 $u = (0, 0, 1)$
式中 u ——初始方向向量
 r_x, r_y, r_z ——四元数旋转轴向量分量
再构造四元数旋转轴向量,公式为

$$\omega = 1 + u(v_x, v_y, v_z) \tag{5}$$

式中 ω ——四元数旋转轴向量
最后将姿态四元数单位归一化,得到侧视角火龙果空间姿态,公式为

$$q = \frac{(r_x, r_y, r_z, \omega)}{\sqrt{r_x^2 + r_y^2 + r_z^2 + \omega^2}} \tag{6}$$

式中 q ——侧视角火龙果姿态四元数

火龙果姿态的验证实质上是对关键点坐标在三维空间中准确性进行评估,关键点定位精度直接影响姿态评估的准确性。为此,构建火龙果姿态验证试验平台进行角度误差试验。在 50 次试验中,火龙

果以随机姿态和随机位置摆放,通过计算估计的生长姿态与实际生长姿态(测量获取)之间的角度差异,量化姿态估计误差,结果如表 6 所示。角度误差在 5°以内的火龙果数量为 27 个,占 54%;角度误差在 5°~10°之间的火龙果为 13 个,占 26%;角度误差在 10°~15°之间的火龙果为 6 个,占 12%;角度误差大于 15°的火龙果有 4 个,平均角度误差为 50.3°,该异常主要归因于果实关键点定位失败:原本应落于果实两端的关键点被模型错误预测于果实中心,致使用于构建主轴的两端点在三维空间与真实轴线产生显著偏离,从而导致果实姿态预测失败并造成大幅角度误差。从统计结果看,整体平均角度误差为 8.6°,且误差主要集中于 0°~10°。结合上述误差分布,本文所提姿态检测方法误差较为稳定,能够满足机器人精准采摘需求。

表 6 姿态估计误差结果

Tab. 6 Pose estimation error results

角度误差/(°)	火龙果数量/个	角度误差平均值/(°)
0~5	27	2.2
5~10	13	7.7
10~15	6	11.9
>15	4	50.3
总计	50	8.6

3.7 火龙果采摘试验

火龙果采摘试验地点为浙江省湖州市吴兴区妙西镇五星村,参照 NY/T 3911—2021《火龙果采收贮运技术规范》第 4.3 节提出的采收要求:露地种植火龙果宜在晴天或阴天采收,高温天、下雨天或鳞片、果实表面露水未干时,不宜采收。选择在 2025 年 7 月 23 日和 8 月 5 日 09:00—16:30 时间段进行火龙果采摘试验。试验期间,天气状况分别为阴及晴,符合采摘相关要求。采摘平台采用 ZED 2i 型深度相机实时获取 RGB-D 图像。运动执行单元采用遨博 AUBO-i5H 型六自由度协作机械臂,工作半径为 886.5 mm,有效负载为 5 kg,重复定位精度达 ±0.02 mm,能够满足火龙果的采摘作业需求。末端执行器选用 WHEELTEC 电动三指爪,适用于抓取尺寸为 10~120 mm、质量不超过 1 kg 的果实。移动底盘采用煜禾森 FR_mid 型阿克曼线控底盘,续航里程达 20 km,负载能力为 100 kg,尺寸为 1 320 mm×785 mm×490 mm,能够满足采摘机器人在复杂果园环境中移动和作业负载需求。

火龙果采摘机器人的工作过程为:边缘计算设备通过 OWD-YOLO 模型处理深度相机采集的图像,获得火龙果关键点的二维像素坐标,并进一步将其转换为世界坐标系下三维空间位置,利用姿态检

测方法估计火龙果位置与姿态。根据实际果园环境中果实间存在遮挡情况,设计对应的采摘策略,即根据识别确定果实位置,由近及远优先采摘无遮挡果实,再对被遮挡果实实施采摘。机械臂依据目标果实的坐标进行路径规划并调整至相应的采摘姿态,完成采摘后,将果实放入果篮,再进行下一目标的采摘任务。

火龙果采摘过程如图 9 所示,图 9a 为边缘设备利用深度相机实时获取 RGB-D 图像,并识别计算可采摘火龙果的三维空间位置与姿态;图 9b 为机械臂根据火龙果的姿态信息快速移动至预采摘位置;图 9c 为机械臂慢速接近目标果实并抓取;图 9d 为机械爪执行旋转动作实现果实与枝条分离;图 9e 为机械臂运送已采摘的果实至果篮;图 9f 为机械臂释放机械爪,将果实放置果篮内,完成采摘过程。



图 9 火龙果采摘平台作业过程

Fig. 9 Operation procedure of dragon fruit harvesting platform

为全面评估火龙果采摘平台工作性能,设计了针对正视角和侧视角 2 种典型生长姿态的火龙果采摘试验,并记录了平均采摘时间及采摘成功率,采摘试验结果见表 7。2 种视角分别进行 25 次采摘试验,共 50 次。结果表明,总体采摘成功率为 86.0%,平均单次采摘耗时 29.4 s。在正视角采摘中,成功率为 88.0%,平均采摘时间为 27.5 s。3 次失败采摘中,2 次由于视觉识别模型对果实采摘关键点坐标定位误差,导致机械爪未能稳定抓取果实,从而无法完成旋转分离动作;1 次为发生在果实成功采摘后,机械臂在转运过程由于枝条阻挡,未能顺

利到达果篮位置,最终导致果实放置失败。侧视角采摘试验成功率为 84.0%,平均采摘时间为 31.3 s。试验中有 4 次失败,2 次由于枝条阻挡机械臂运动,导致机械臂无法到达目标采摘位置;1 次由于视觉识别模型对果实采摘关键点坐标定位误差,导致机械爪未能稳定抓取果实而无法完成旋转分离动作;1 次由于机械臂与末端执行器的工作范围限制,未能成功到达采摘点。综合分析表明,正视角火龙果的采摘成功率和作业效率明显优于侧视角。主要原因在于,侧视角采摘姿态要求机械爪与果实轴线平行,末端执行器需深入到枝条密集区域,导致机械臂运动过程易受阻,影响采摘成功率,最终使成功率下降至 84.0%,与正视角的火龙果采摘相比平均采摘时间增加 3.8 s。

表 7 不同姿态的火龙果采摘性能

Tab.7 Harvesting performance for dragon fruits with different postures

火龙果 姿态	采摘 总数	失败 次数	单果平均 采摘时间/s	采摘成功 率/%
正视角	25	3	27.5	88.0
侧视角	25	4	31.3	84.0
总计	50	7	29.4	86.0

4 结论

(1)针对复杂果园环境下火龙果识别与定位中存在的问题,提出了一种轻量化模型 OWD-YOLO。在特征提取模块中引入结构重参数化卷积和 MLCA 注意力机制,降低计算繁琐度的同时提升了模型在

复杂背景下对火龙果多尺度和细节姿态特征的识别能力。设计了融合小波池化和大核可分离卷积注意力机制的 WLA_SPPF 模块,有效抑制了光照变化、枝条遮挡等环境因素影响。针对果实颜色和纹理差异微弱的问题,设计了 DGECA 通道注意力机制,通过动态调整通道权重,强化了对不同成熟度果实的分类表现。

(2)消融试验结果表明,OWD-YOLO 模型目标检测精确率、平均精度均值与关键点检测平均精度均值分别达到 88.0%、92.7%、93.3%,较基础模型分别提高 5.0、2.4、2.0 个百分点,同时平均帧率为 58.7 f/s,说明本文方法能够满足复杂果园环境下火龙果成熟度与姿态的精准实时识别需求。同时,在 OWD-YOLO 与 YOLO 其他版本的轻量级关键点检测模型及轻量化主干网络对比试验中,OWD-YOLO 以较小的检测速度损失换取了显著的精度提升,在火龙果检测任务中实现了与规模较大的 YOLO 11m-pose 相近的性能。

(3)提出了一种基于关键点的三维姿态识别方法,针对不同观察视角下的成熟火龙果分别制定了对应的姿态估计策略,并构建了基于 OWD-YOLO 模型的火龙果自动采摘平台。在复杂果园环境下开展了 50 次田间采摘试验。结果表明,采摘成功率达 86.0%,平均采摘时间 29.4 s。其中,对于正视角的火龙果,采摘成功率高达 88.0%;对于侧视角的火龙果,采摘成功率达到 84.0%。试验验证了 OWD-YOLO 模型与基于关键点的姿态检测方法在火龙果自动采摘任务中具备良好的性能表现。

参 考 文 献

[1] 赵文锋,黄袁爵,钟敏悦,等.基于 OrchardYOLOP 的火龙果园多任务视觉感知方法[J].农业机械学报,2024,55(11):160-170.
ZHAO Wenfeng, HUANG Yuanjue, ZHONG Minyue, et al. Multi-task visual perception method in dragon orchards based on OrchardYOLOP[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(11):160-170. (in Chinese)

[2] 陈天赐,张世昂,付根平,等.热带水果采摘机器人研究进展综述[J].农业机械学报,2025,56(5):184-201.
CHEN Tianci, ZHANG Shiang, FU Genping, et al. Review of research progress on tropical fruit harvesting robots[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(5):184-201. (in Chinese)

[3] WAKCHAURE M, PATLE B K, MAHINDRAKAR A K, et al. Application of AI techniques and robotics in agriculture: a review [J]. Artificial Intelligence in the Life Sciences, 2023, 3: 100057.

[4] SPAGNUOLO M, TODDE G, CARIA M, et al. Agricultural robotics: a technical review addressing challenges in sustainable crop production [J]. Robotics, 2025, 14(2): 9.

[5] YANG Y, HAN Y X, LI S, et al. Vision based fruit recognition and positioning technology for harvesting robots [J]. Computers and Electronics in Agriculture, 2023, 213: 108258.

[6] XU X, JIANG Y M, WANG Y N. Key technologies for machine vision for picking robots: review and benchmarking [J]. Machine Intelligence Research, 2025, 22(1): 2-16.

[7] MAGALHÃES S A, MOREIRA A P, SANTOS F N D, et al. Active perception fruit harvesting robots: a systematic review [J]. Journal of Intelligent & Robotic Systems, 2022, 105(1): 14.

[8] LIU J, LIU Z. The vision-based target recognition, localization, and control for harvesting robots: a review [J]. International Journal of Precision Engineering and Manufacturing, 2023, 25: 127.

[9] KALEEM A, HUSSAIN S, AQIB M, et al. Development challenges of fruit-harvesting robotic arms: a critical review [J].

AgriEngineering, 2023, 5(4): 2216 - 2237.

[10] DAZA A, RAMOS K Z, PAZ A A, et al. Deep learning and machine learning for plant and fruit recognition: a systematic review [J]. Journal of System and Management Sciences, 2024, 14(3): 133 - 154.

[11] 赵博,柳苏纯,张巍朋,等. 面向樱桃番茄采摘识别的轻量化 Transformer 架构优化研究[J]. 农业机械学报,2024, 55(10):62 - 71,105.
ZHAO Bo, LIU Suchun, ZHANG Weipeng, et al. Performance optimization of lightweight Transformer architecture for cherry tomato picking[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(10):62 - 71,105. (in Chinese)

[12] SHANG F, ZHOU X, LIANG Y, et al. Detection method for dragon fruit in natural environment based on improved YOLOX [J]. Smart Agriculture, 2022, 4(3): 120 - 131.

[13] 罗陈迪,李文涛,商枫楠,等. 基于 YOLOv7 和 YCrCb 的火龙果识别与分割方法[J]. 农机化研究,2025,47(7):59 - 64.
LUO Chendi, LI Wentao, SHANG Fengnan, et al. Identification and segmentation method of dragon fruit based on YOLOv7 and YCrCb[J]. Journal of Agricultural Mechanization Research, 2025,47(7):59 - 64. (in Chinese)

[14] 刘进一,晏伏山,董赫,等. 基于改进 MBS - YOLO v8 的火龙果目标检测[J]. 农业机械学报,2025,56(5):425 - 432.
LIU Jinyi, YAN Fushan, DONG He, et al. Pitaya fruit target detection and localization method based on improved MBS - YOLO v8[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(5):425 - 432. (in Chinese)

[15] 朱立学,赖颖杰,张世昂,等. 基于改进 U - Net 的火龙果采摘图像分割和姿态估计方法[J]. 农业机械学报,2023, 54(11):180 - 188.
ZHU Lixue, LAI Yingjie, ZHANG Shiang, et al. Image segmentation and pose estimation method for pitaya picking robot based on enhanced U - Net[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(11):180 - 188. (in Chinese)

[16] 周佳良. 基于深度学习的复杂采摘环境下火龙果位姿识别与控制系统研究[D]. 南京:南京林业大学,2023.
ZHOU Jialiang. Research of dragon fruit position recognition and control system in complex picking environment based on deep learning[D]. Nanjing: Nanjing Forestry University, 2023. (in Chinese)

[17] 马瑞峻,何浣冬,陈瑜,等. 基于改进 YOLOv5 的火龙果成熟度识别方法[J]. 沈阳农业大学学报,2023,54(2):196 - 206.
MA Ruijun, HE Huandong, CHEN Yu, et al. Recognition method for maturity of pitaya based on improved YOLOv5 [J]. Journal of Shenyang Agricultural University, 2023,54(2):196 - 206. (in Chinese)

[18] QIU Z, HUANG Z, MO D, et al. GSE - YOLO: a lightweight and high-precision model for identifying the ripeness of pitaya (Dragon Fruit) based on the YOLOv8n improvement[J]. Horticulturae, 2024, 10(8): 852.

[19] 农业农村部. 火龙果采收贮运技术规范:NY/T 3911—2021[S]. 北京:中国标准出版社,2021.

[20] 徐婷婷,宋亮,卢学鹤,等. 基于 YOLO v7 - RA 的火龙果品质与成熟度双指标检测方法[J]. 农业机械学报,2024, 55(7):405 - 414.
XU Tingting, SONG Liang, LU Xuehe, et al. Dual-index detection method of pitaya quality and maturity based on YOLO v7 - RA[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(7):405 - 414. (in Chinese)

[21] 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.
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 CSAE, 2023, 39(8): 276 - 283. (in Chinese)

[22] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [J]. arXiv Preprint, arXiv:2410. 17725, 2024.

[23] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.

[24] HU M, FENG J, HUA J, et al. Online convolutional re-parameterization [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, 2022: 568 - 577.

[25] WAN D, LU R S, SHEN S Y, et al. Mixed local channel attention for object detection [J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.

[26] WANG Q, WU B, ZHU P, et al. ECA - Net: efficient channel attention for deep convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, 2020: 11534 - 11542.