本文引用格式: 王红军,邹伟锐,谢启旋,等.基于机器视觉的香蕉果柄识别及采摘试验研究[J].自动化与信息工程,2023,44(5):14-21;51.

> WANG Hongjun, ZOU Weirui, XIE Qixuan, et al. Research on banana stem recognition and picking experiment ERING based on machine vision[J]. Automation & Information Engineering, 2023,44(5):14-21;51.

基于机器视觉的香蕉果柄识别及采摘试验研究*

王红军¹ 邹伟锐¹ 谢启旋² 郑文和³ (1.华南农业大学, 广东 广州 510642 2.比亚迪股份有限公司, 广东 深圳 518118 3.比亚迪电子(国际)有限公司,广东 深圳 518118)

摘要: 香蕉采摘是一个耗时费力的作业过程。为了实现香蕉的自动化采摘,首先,设计了一款香蕉智能采 摘装置,主要包括叉剪升降机构、3-RPS并联机构和 XY 直线移动机构等,通过控制这 3 个机构的运动变化,满 足其搭载的末端执行机构适应香蕉果柄的空间随机性要求: 然后, 搭建双目相机视觉识别系统, 采用 YOLOv5 算 法识别香蕉果柄,选择检测框的中心位置为采摘点;接着,根据采摘点位置反求 3-RPS 并联机构的位姿,驱动切 割机构对香蕉果柄定位,完成香蕉串的自动采摘;最后,研制了香蕉智能采摘装置样机,开展采摘试验。试验结 果表明,香蕉智能采摘装置的绝对位置误差小于 8.66 mm,相对位置误差小于 0.85%; y角度误差小于 1.10°,β 角度误差小于1.25°,满足香蕉采摘的定位要求。

关键词:机器视觉; YOLOv5;香蕉采摘; 3-RPS 并联机构;香蕉果柄识别

中图分类号: S225.93; TP391.41 文献标志码: A 文章编号: 1674-2605(2023)05-0003-09 DOI: 10.3969/j.issn.1674-2605.2023.05.003

Research on Banana Stem Recognition and Picking Experiment Based on Machine Vision

WANG Hongjun¹ ZOU Weirui¹ XIE Qixuan² ZHENG Wenhe³

(1.South China Agricultural University, Guangzhou 510642, China 2.BYD, Co., Ltd., Shenzhen 518118, China

3.BYD Electronics International Co., Ltd., Shenzhen 518118, China)

Abstract: Banana picking is a time-consuming and laborious process. In order to achieve automated banana picking, firstly, an intelligent banana picking device was designed, which mainly includes a fork shear lifting mechanism, a 3-RPS parallel mechanism, and an XY linear moving mechanism. By controlling the motion changes of these three mechanisms, the end effector mechanism it is equipped with meets the spatial randomness requirements of the banana handle; Then, build a binocular camera visual recognition system, use YOLOv5 algorithm to identify the banana stem, and select the center position of the detection box as the picking point; Next, reverse the pose of the 3-RPS parallel mechanism based on the picking point position, drive the cutting mechanism to position the banana stem, and complete the automatic picking of the banana string; Finally, a prototype of an intelligent banana picking device was developed and picking experiments were conducted. The experimental results show that the absolute position error of the intelligent banana picking device is less than 8.66 mm, and the relative position error is less than 0.85%; γ The angle error is less than 1.10°, β The angle error is less than 1.25° , meeting the positioning requirements for banana picking.

Keywords: machine vision; YOLOv5; banana picking; 3-RPS parallel mechanism; banana stem recognition

0 引言

近年来,我国香蕉的种植面积和产量不断提高, 香蕉产业具有较大的规模效益和发展潜力。根据海关 总署的数据显示,2021年我国香蕉进出口数量分别为 186.35万吨和2.03万吨,进出口金额分别为10.37亿 美元和0.18亿美元。我国香蕉产品在国际上的竞争力 较弱,香蕉采收的人工成本远高于东南亚国家^[1-2]。人 工采摘已成为阻碍我国香蕉产业发展的重要因素,而 农业机械化和自动化是解决该困境的有效途径。智能 化采摘机械可以提高采摘效率,减少人工成本,同时 也能减少采摘过程中对香蕉的伤害,提高香蕉品质^[3]。

围绕香蕉智能化采摘,国内外开展了相关的研究。 上世纪 70 年代,FRANKLIN 等件设计一种轻简的香 蕉采收辅助装置,具有同时砍断和夹持果柄的功能, 实现了香蕉采摘的部分机械化。澳大利亚农业部根据 微型农用拖拉机拖车、液压驱动的机械手臂、切割机 构和机械手末端的容器改装了一种香蕉采摘机^[5]。 MANUEL 等^[6]设计的香蕉采摘机利用末端夹持器夹 住香蕉茎,切割刀片将其切掉。唐之富等^[7]研制一种 模拟人工动作的香蕉采摘机械装置,具有升降,夹持 和切割香蕉串的功能。樊维^[8]设计一种能够自动采摘 香蕉的四自由度圆柱坐标型香蕉采摘机器人。刘远 力^[9]设计的香蕉采摘整机创新性地引入了前支式液压 伸缩臂与配重块相结合的采摘方式。

香蕉采摘装置的研究虽然取得了一定的进展,但 在智能化采摘方面还鲜有报导。本文利用双目相机视 觉识别系统获取香蕉串图像,采用 YOLOv5 算法识 别图像中的香蕉果柄,并选择检测框的中心位置为采 摘点;根据采摘点位置反求 3-RPS 并联机构的位姿, 驱动夹持切割机构对香蕉果柄定位,完成香蕉串的自 动采摘。

香蕉智能采摘装置

1.1 装置总体结构

香蕉智能采摘装置机械结构主要包括行走小车、 叉剪升降机构、3-RPS 并联机构、夹持切割机构、双

目相机、XY 直线移动机构等,结构图如图1所示。

1-行走小车; 2-叉剪升降机构; 3-3-RPS 并联机构; 4-夹持切割机构; 5-双目相机; 6-XY 直线移动机构。 图 1 香蕉智能采摘装置结构图

1.2 装置工作原理

香蕉智能采摘装置可模拟人工采摘行为实现移 动、升降、位置调整、角度调整、夹持和切割果柄等 动作。行走小车是承载其他机构的运动载体,实现香 蕉智能采摘装置在香蕉园移动;双目相机安装在 3-RPS 并联机构的前端,用于获取香蕉串图像;上位机 通过 USB 线与硬件系统的 Arduino 连接,完成双目 相机的图像采集及自动识别;通过反求定位切割点, 获得各控制电机的状态数据;系统控制各机械结构的 电机运动,实现夹持切割机构对香蕉果柄的准确定位; 由叉剪升降机构、3-RPS 并联机构和 XY 直线移动机 构组成的混联机构实现夹持切割机构以不同的空间 位姿靠近香蕉果柄。香蕉智能采摘的工作流程如图 2 所示。



图 2 香蕉智能采摘工作流程

1.3 装置运动坐标系

香蕉采摘作业过程中,智能采摘装置与香蕉串 之间的相对位置关系,即香蕉智能采摘坐标系如图3 所示。



图 3 香蕉智能采摘坐标系

在 3-RPS 并联机构上搭建的夹持切割机构通过 XY 直线移动机构完成定位,驱动 XY 直线移动机构 的滑块,使夹持切割点位置发生变化,完成夹持切割 点向香蕉果柄采摘点的移动定位。为建立夹持切割机 构与香蕉果柄切割点之间的运动关系,在行走小车的 四轮中间建立移动平台坐标系 *O*_D;在香蕉树假茎中 心与地面的交点建立世界坐标系 *O*_W,定位香蕉果柄 采摘点的位置;在双目相机安装位置建立相机坐标系 *O*_C;在夹持切割机构的采摘点处建立末端执行器坐标 系 *O*_E,其原点 *O*_E在移动平台坐标系 *O*_D下的位置为 *E*(*x*_e, *y*_e, *z*_e);在香蕉果柄采摘点处建立采摘坐标系 *O*_F, 其原点 *O*_F在世界坐标系 *O*_W下的位置为 *F*(*x*_f, *y*_f, *z*_f)。

2 香蕉果柄识别算法

香蕉智能采摘装置通过双目相机视觉识别系统 识别香蕉果柄并确定采摘点,并通过反求定位切割点 获得控制机构的运动参数。采用 YOLOv5 算法分别 检测左、右相机图像中的香蕉果柄,以检测框的中心 位置为采摘点,获得其在图像上的像素坐标。根据双 目测距原理和左、右相机得到香蕉果柄采摘点的二维 位置信息,获取香蕉果柄采摘点在相机坐标系下的三 维位置信息。

2.1 数据集

本文香蕉果柄数据集包括自采集和网络收集的 图像共 925 幅。其中,自采集野外图像 725 幅,分 别于早晨、中午和傍晚拍摄于广州市某香蕉园,如图 4 所示,图像采集设备为高像素智能手机,分辨率为 2 048×1 536 像素,摄像头与香蕉的距离为 500~ 1 500 mm,早上和中午各采集 300 幅图像,傍晚采集 125 幅图像;利用 Intel RealSense D435i 深度相机自采 集室内图像 100 幅,分辨率为 640×480 像素;网络收 集图像 100 幅。



(a) 早晨
(b) 中午
(c) 傍晚
图 4 早、中、晚采集的野外香蕉图像

2.2 图像预处理

采用图像标注软件 LabelImg 对 925 幅香蕉果柄 图像进行标注,框出香蕉果柄,得到数据集的标签文 件 XML,并转换成 YOLOv5 训练使用的 TXT 文件。 TXT 文件包含香蕉果柄在图像中的坐标位置、图像大 小和标签名称,组成香蕉果柄数据集。

2.3 YOLOv5 网络

YOLOv5 网络由主干网络(Backbone)、颈部(Neck)和头部(Head)组成,结构如图5所示。

2.3.1 输入端

2.3.1.1 数据增强

香蕉果柄数据集只有925幅图像,样本数量较少, 数据泛化能力不强。采用 Mosaic 图像预处理技术, 对香蕉果柄4幅图像进行随机裁剪、缩放和排布拼接 操作,合成1幅图像(如图6所示),既丰富了检测 物体的背景, 又增加了数据集样本的数量。



图 5 YOLOv5 网络模型



图 6 Mosaic 处理后的图像

2.3.1.2 自适应图像缩放

YOLOv5 采用自适应图像缩放的方式调整图像的宽和高,避免简单比例缩放导致目标变形,造成图像信息丢失。分别计算一幅图像的宽和高与标准特征图(*S*×*S*)的缩放比例,选择较小的缩放比例计算缩放后图像的尺寸。用图像的原始尺寸乘以缩放比例,得到新图像的长和宽。用标准特征图的边长*S*减去缩放后图像需要填充灰边的边,得出需要填充的像素值;需要填充的像素值除以2,得到左右或上下两边各自需要填充的像素值,缩放效果如图7所示。



图 7 自适应图像缩放效果图

2.3.1.3 自适应计算锚框

YOLOv5 在模型训练过程中计算锚框,即根据真 实框和预测框的将算法初始设定锚框的宽和高不断 地迭代更新。首先,YOLOv5 使用训练模型对目标进 行预测;然后,筛选预测框,保留宽和高都不小于 2 个像素的预测框;接着,根据预测框与真实框的偏移 量调整锚框位置,得到预测边界框;最后,筛选需要 输出的预测边界框。

2.3.2 主干网络

YOLOv5的主干网络包括 CBS 模块、CSP1_X 模 块和 SPPF 模块,主要通过相应的卷积操作在输入图 像中提取特征图。

2.3.2.1 CBS 模块

CBS 模块由卷积层 Conv, 批标准化 BN 和激活 函数 SiLU 组成, 如图 8 所示。



图 8 CBS 结构图

卷积层 Conv 是将输入图像上相近的像素值通过 加权叠加的方式生成另外一组数值的过程。批标准化 BN 通过规范手段使输入值处于标准正态分布范围内, 以避免输出值过大和梯度消失,解决了深层网络训练 困难、收敛速度慢等问题。激活函数 SiLU 是一种将 输入端数据映射到输出端的非线性函数,使训练模型 能更好地拟合非线性特征。SiLU 函数是 Sigmoid 函 数的加权线性组合,计算公式为

$$f(x) = x\sigma(x) \tag{1}$$

2023 年 第44 卷 第5 期 自动化与信息工程 17

其中 Sigmoid 函数 $\sigma(x)$ 的计算公式为

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

2.3.2.2 CSP1_X 模块

CSP1_X 模块包括 3 个 CBS 模块、X 个残差模块 ResUnit 和 1 个 Concat 模块,其中残差模块的数量由 训练网络中 X 的具体数值决定,如图 9 所示。



图 9 CSP1_X 结构图

CSP1_X 模块的输入和输出形状相同,它不改变 数据的宽、高和通道数量,可有效减少特征在传输过 程中的信息丢失和计算量。Concat 模块主要将 2 个或 多个数组拼接成 1 个数组。残差模块 ResUnit 结构如 图 10 所示,其中 CBS 卷积层中的卷积核大小不一, 可以有效减少训练参数。



2.3.2.3 SPPF 模块

SPPF 模块包括 2 个 **CBS** 模块、3 个最大池化层 MaxPool 和 1 个 Concat 模块,结构如图 11 所示。



图 11 SPPF 结构图

SPPF 模块将多尺度特征进行融合,首先,对输入数据进行1次卷积操作;然后,进行3次最大池化(核为5×5);最后,将每次操作得出的数据进行拼接后再进行1次卷积操作,得到输出结果。

2.3.3 颈部

颈部主要包括 CBS 模块、Upsample 模块、Concat 模块和 CSP2 模块,采用 FPN+PAN 结构。颈部特征 处理流程如图 12 所示。





FPN 结构利用 Upsample 模块向特征图中插值, 使特征图的尺度变大,并从上一层主干网络中获取相 对较浅的特征,通过 Concat 模块将浅层的图形特征 与浅层的语义特征结合在一起。PAN 结构利用 CBS 模块进行下采样,得到较深的特征,并通过 Concat 模 块将该较深的特征与 FPN 结构得到的特征结合。

颈部用于加强主干网络提取的特征,并把加强的特征融合后传递到头部。深层特征图携带更强的语义特征和较弱的定位信息;浅层特征图携带较强的定位信息和较弱的语义特征。FPN将多语义信息的深层特征和多细节信息的浅层特征利用下采样进行融合,从而增强多个尺度上的语义表达。PAN把浅层的定位信息传导到深层,增强多个尺度上的定位能力。

2.3.4 头部

头部只有 1 个卷积核为 1×1 的卷积层,将颈部 得到的特征图通道数扩展,是 YOLOv5 网络的输出 端,其借助颈部加强融合的特征图对图像进行预测和 分类,并输出目标的预测信息。

2.3.5 损失函数

YOLOv5的损失函数包括分类损失、定位损失和 置信度损失3类。其中,分类损失用于判断模型是否 能够准确地识别图像中的对象,并将其分类到正确的 类别中;定位损失为预测框与真实框之间的误差;置 信度损失为网络的置信度。YOLOv5利用二元交叉熵 损失函数计算分类损失和置信度损失,利用 CloU Loss 计算定位损失。

2.3.5.1 分类损失和置信度损失

YOLOv5 利用二元交叉熵对每个标签计算分类 损失及置信度损失,其计算公式为

$$L(w) = -\sum_{i=1}^{n} (y_i \lg(x_i) + (1 - y_i) \lg(1 - \sigma(x_i)))$$
(3)

式中:
$$\sigma(x) = \frac{1}{1+e^{-x}}$$
, x_i 为第 *i*个样本, y_i 为第

i个标签, n为样本总量。

2.3.5.2 定位损失

YOLOv5利用 CloU 值评价预测框和真实框的定 位损失。CloU 是 loU(预测框和真实框的交并比)的 一种推广, loU 计算公式为

$$IoU = \frac{\left|b \cap b^{gt}\right|}{\left|b \cup b^{gt}\right|} \tag{4}$$

式中: b、bst分别为预测框和真实框。

由公式(4)可知,如果2个物体不重叠,即 *loU* = 0,损失梯度为0无法对模型优化,因此不能使用 *loU* 作为定位损失的指标。CloU 考虑了真实框与预测框之间的距离、重叠率、边框尺度以及惩罚因子,使目标边框回归更加稳定,有效解决了 *loU* 在训练过程中发散的问题。CloU 的计算公式为

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v$$
 (5)

式中: ρ(b, bst)为预测框和真实框中心的距离,即 图 13 中的 d; c 是 2 个框的最小包围矩阵的对角线距 离; α 为权重系数,计算公式如式(6)所示; v 为衡量 长宽比一致性的参数,计算公式如式(7)所示。

$$\alpha = \frac{v}{1 - IoU + v} \tag{6}$$

$$v = \frac{4}{\pi^2} \left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{7}$$

定位损失的损失函数为

$$L_{\text{CIoU}} = 1 - CIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + av \quad (8)$$



2.3.5.3 总损失函数

总损失函数的计算公式为

$$L_{\text{total}} = \sum_{i}^{N} \left(\lambda_1 L_{\text{cls}} + \lambda_2 L_{\text{obj}} + \lambda_3 L_{\text{box}} \right)$$
$$= \sum_{i}^{N} \left(\lambda_1 \sum_{j}^{B_i} L_{\text{cls} j} + \lambda_2 \sum_{j}^{S_i \times S_i} L_{\text{obj}} + \lambda_3 \sum_{j}^{B_i} L_{\text{box}} \right) \quad (9)$$

式中: N 为检测层数; B 为标签分配到先验框的 目标数; S×S 为该尺度被分割成的网格数; L_{cls} 为分类 损失,对每个目标计算; L_{obj} 为目标置信度损失,对每 个网格计算; L_{box} 为边界框回归损失,对每个目标计 算; λ₁、λ₂、λ₃分别为这 3 种损失的权重。

2.3.6 训练结果

模型训练150批次后的检测效果和训练结果分别 如图14、15所示。



图 14 检测效果

2023年 第44卷 第5期 自动化与信息工程 19



由图 14、15 可知,随着训练轮数增加,训练集和 测试集的定位损失和置信度损失都逐渐下降并收敛, 而精确度和召回率逐渐提高。另外,香蕉果柄识别只 有1个类别,分类损失为0。

由图 15 可知,当 IoU 阈值为 0.5 时,平均准确率 AP 达到 98.034%,满足香蕉采摘识别的要求。

3 采摘定位试验

为验证香蕉果柄识别算法及夹持切割机构的合 理性,开发了香蕉智能采摘装置样机,在实验室开展 室内采摘定位试验,试验场景如图16所示。



图 16 采摘定位试验场景

采摘定位试验所用的仪器设备主要包括香蕉智 能采摘装置样机、香蕉串及果树、定位器、PE 线、电 子数显角度尺、钢卷尺、钢尺等。

香蕉智能采摘装置的运动参数范围如表 1 所示。

表1 香蕉智能采摘装置运动参数范围

化1 百点百能不同农业之初乡风花田							
试验 序号	运动参数名称	运动变量	变量运动范围/mm				
1	采摘切割点沿 X 轴的移动量	XA	270.0~570.0				
2	采摘切割点沿 Y轴的移动量	YA	-325.0~275.0				
3	采摘切割点沿 Z轴的移动量	ZA	690.0~840.0				
4	动平台绕X轴 的旋转量	γ	-15°~15°				
5	动平台绕 Y 轴 的旋转量	β	-15°~15°				

采摘定位试验过程如下:将定位器安装在夹持切 割机构末端执行器的抱掌上,保持定位器的尖点与香 蕉理想采摘点位置重合;将 PE 线一端捆绑在理想采 摘点位置,另一端垂钓一个配重铅锤,保证 PE 线沿 铅锤方向垂直接近地面;通过双目相机视觉识别系统 获取香蕉果柄的切割点位置参数;通过反求解算,机 构自动运动到香蕉果柄采摘位置;待到 PE 线不再左 右摆动,分别测量香蕉果柄采摘点在香蕉采摘装置中 的3个移动分量和2个旋转分量,测试结果分别如表 2、3 所示。

表 2 香蕉智能采摘装置位置精度数据

试验 序号	目	目标位置		实际位置		绝对位置	相对位置	
	x	у	z	x'	<i>y</i> ′	z'	庆左 <i>mb/</i> mm	庆左 rb/%
1	280	0	700	281	0	702	2.236 0	0.296 5
2	350	0	700	352	0	701	2.2360	0.285 71
3	420	0	700	419	1	703	3.316 6	0.406 28
4	490	0	700	494	1	700	4.123 1	0.482 5
5	560	0	700	565	2	700	5.385 1	0.600 7

表 3 香蕉智能采摘装置姿态精度数据

试验 序号	目标姿态		测量姿态		误差	误差
	γ	β	γ'	β	γb/(°)	βb/(°)
1	15	0	14.10	0.10	0.90	0.10
2	10	0	10.05	-0.95	0.05	0.95
3	5	0	4.10	-0.40	0.90	0.40
4	-5	0	-4.50	-0.10	0.50	0.10
5	-10	0	-9.20	-0.60	0.80	0.60
6	-15	0	-14.90	0.45	0.10	0.45

由表 2 可知:位置精度的最大绝对位置误差为 5.3851 mm,最大相对位置误差为 0.6007%。

由表 3 可知: 姿态测量数据的 γ 角度误差最大值 为 0.9°, β 角度误差最大值为 0.95°。在具有一定容 错的采摘装置中,夹持切割机构定位位置精度可满足 采摘香蕉串的需要。

结论

本文以香蕉串智能化采摘为研究目标,设计香蕉 智能采摘装置,提出基于 YOLOv5 网络模型的香蕉 果柄识别算法,建立香蕉智能采摘装置与香蕉果柄切 割点之间的坐标关系,通过反求获取运动机构参数。 经室内采摘试验验证了香蕉智能化采摘的可行性。

1)香蕉智能采摘装置包括行走小车、叉剪升降机 构、3-RPS 并联机构和夹持切割机构等。行走小车能平 稳在果园环境下行走;叉剪升降机构带动 3-RPS 并联 机构和夹持切割机构升降,使夹持切割机构达到香蕉 的采摘高度;3-RPS 并联机构控制夹持切割机构的位 置和姿态,使其达到采摘点位置,夹持并切断果柄。

2) 基于 YOLOv5 网络模型的香蕉果柄识别算法,采用 Mosaic 图像预处理技术,进行随机裁剪、缩放和排布拼接操作,合成图像,丰富了检测物体的背景,增加数据集样本。模型训练结果标明,该识别算法可有效识别香蕉果柄,确定切割点位置。该切割点位置用于香蕉智能采摘装置上末端执行器采摘点的位姿求解。通过上位机与硬件系统的 Arduino 连接,控制软件驱动机构运动电机,实现夹持切割机构对香蕉果柄的准确定位。

 3) 香蕉智能采摘装置样机在实验室环境下,验 证了基于 YOLOv5 网络模型的香蕉果柄识别算法及
香蕉智能采摘装置的可行性。

参考文献

- [1] 张卫明.一带一路经济植物[M].东南大学出版社,2017.
- [2] 陈清西,李冬香.香蕉周年管理关键技术[M].金盾出版社, 2012.
- [3] 彭舟.基于 3-RPS 并联机构的香蕉采摘机研究[D].广州:华南 农业大学,2022.
- [4] FRANKLIN J A. Robot hands and the mechanics of manipulation: editors: m.t. Mason and j.k. Salisbury, jr. Mit press isbn 0 262 13205 2[J]. 1988,3(1):51.
- [5] 晓明.香蕉采摘机[J].世界农业,1986(7):50.
- [6] MANUEL L P. Banana or pineapple harvester has clamp and cutting blade mounted on telescoping arm articulated to mast on tractor-mounted work surface or low frame: ES20040002089 [P]. ES2257170A1[2023-10-07].
- [7] 唐之富,王红军,邓建猛,等.面向大质量串型水果采摘的夹持 试验与仿真[J].机械设计,2018,35(11):47-52.
- [8] 樊维.香蕉采摘机器人机械手结构设计与分析[D].南宁:广西 大学,2018.
- [9] 刘远力.抗倾覆香蕉采摘整机的设计与研究[D].福州:福建农 林大学,2020.(下转第51页)

2023年 第44卷 第5期 自动化与信息工程 21