

doi:10.6041/j.issn.1000-1298.2023.10.020

# 基于实例分割的番茄串视觉定位与采摘姿态估算方法

张勤<sup>1</sup> 庞月生<sup>1</sup> 李彬<sup>2</sup>

(1. 华南理工大学机械与汽车工程学院, 广州 510641;  
2. 华南理工大学自动化科学与工程学院, 广州 510641)

**摘要:** 准确识别定位采摘点,根据果梗方向,确定合适的采摘姿态,是机器人实现高效、无损采摘的关键。由于番茄串的采摘背景复杂,果实颜色、形状各异,果梗姿态多样,叶子藤枝干扰等因素,降低了采摘点识别准确率和采摘成功率。针对这个问题,考虑番茄串生长特性,提出基于实例分割的番茄串视觉定位与采摘姿态估算方法。首先基于 YOLACT 实例分割算法的实例特征标准化和掩膜评分机制,保证番茄串和果梗感兴趣区域 (Region of interest, ROI)、掩膜质量和可靠性,实现果梗粗分割;通过果梗掩膜信息和 ROI 位置关系匹配可采摘果梗,基于细化算法、膨胀操作和果梗形态特征实现果梗精细分割;再通过果梗深度信息填补法与深度信息融合,精确定位采摘点坐标。然后利用果梗几何特征、八邻域端点检测算法识别果梗关键点预测果梗姿态,并根据果梗姿态确定适合采摘的末端执行器姿态,引导机械臂完成采摘。研究和大量现场试验结果表明,提出的方法在复杂采摘环境中具有较高的定位精度和稳定性,对 4 个品种的番茄串采摘点平均识别成功率为 98.07%,图像分辨率为 1 280 像素 × 720 像素时算法处理速率达到 21 f/s,采摘点图像坐标最大定位误差为 3 像素,深度误差 ± 4 mm,成功定位采摘点后采摘成功率为 98.15%。与现有的同类方法相比,采摘点图像坐标定位精度提高 76.80 个百分点,采摘成功率提高 15.17 个百分点,采摘效率提高 31.18 个百分点,满足非结构化种植环境中番茄串采摘需求。

**关键词:** 番茄串; 机器人; 实例分割; 视觉定位; 采摘姿态预测

中图分类号: TP242.3      文献标识码: A      文章编号: 1000-1298(2023)10-0205-11      OSID: 

## Visual Positioning and Picking Pose Estimation of Tomato Clusters Based on Instance Segmentation

ZHANG Qin<sup>1</sup> PANG Yuesheng<sup>1</sup> LI Bin<sup>2</sup>

(1. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China  
2. School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China)

**Abstract:** Recognizing and positioning the picking points (spatial position and coordinate points), and determining the appropriate picking pose according to the direction of fruit stem, are the keys for the robot to achieve efficient and lossless picking during harvesting. However, the harvesting environment is complex and changeable, the color of fruit stem is similar to the branches and leaves, and the tomato clusters are always with different colors and shapes. Furthermore, tomato clusters grow in different directions, and the end effector frequently interferes with the leaves and vine during picking, there are often situations of “not picking when robot see it”, which reduces the recognition accuracy of picking points and picking rate. Aiming at this problem, considering the growth characteristics of tomato clusters, a method for visual positioning and picking pose estimation of tomato clusters based on instance segmentation was proposed. Firstly, based on the instance feature standardization, and the mask scoring mechanism of the YOLACT algorithm, the high quality and reliable region of interest (ROIs) and masks of tomato clusters were collected. Specifically, in order to efficiently achieve the coarse segmentation of fruit stems via the YOLACT. Then, according to the stem mask information and the neighbor relationship between tomato ROIs and stem ROIs, the ROIs of pickable stems were determined. Meanwhile, the

pickable stem edges were finely extracted from the stem ROI by using the thinning algorithm, together with expansion operation and shape characteristics of stem. Secondly, the picking point in image coordinate system was obtained, which was set as the center point of stem skeleton along the  $X$  (or  $Y$ ) axis. Subsequently, the depth map of pickable stem ROI was used for obtaining the original depth value of picking point. Specifically, due to the large depth value errors, or even a lack of depth values when capturing small objects by the economical RGB - D depth camera. By using only the depth map corresponding to the stem mask area, the average depth value of picking point was calculated. The accurate depth value of picking point was obtained by comparing the average with the original depth value. Thirdly, according to the geometric features of fruit stem, the tangent slope of fruit stem at the picking point was calculated, and the search algorithm was used for finding the endpoints of fruit stem. Correspondingly, fruit stem direction was estimated by the vector composed of two endpoints of fruit stem. Finally, the picking point was converted to the robot coordinate system. Simultaneously, according to the tangent slope and the direction of fruit stem, the picking pose of the end effector was determined. Eventually, the robot was guided to complete the picking task with an appropriate pose. A large number of field test verified that the average recognition rate of pickings point was 98.07%, while the image resolution was 1 280 pixel  $\times$  720 pixel, the processing rate of the algorithm was 21 f/s, the maximum positioning error of the image coordinates of picking points was 3 pixels, and the depth value error was  $\pm 4$  mm. After the picking points were successfully positioned, the picking rate was 98.15%. Compared with the existing similar methods, the positioning accuracy of the picking point was increased by 76.80 percentage points, the picking rate was increased by 15.17 percentage points, and picking efficiency was increased by 31.18 percentage points. Therefore, the proposed method fully met the requirements for robots in unstructured environment during harvesting.

**Key words:** tomato cluster; robot; instance segmentation; visual positioning; picking pose estimation

0 引言

机器人采摘可降低种植成本、提高生产率、减轻工人劳动强度、保证果实质量,且适时采摘是未来的发展方向<sup>[1-5]</sup>。由于番茄串采摘环境复杂、品种多样且个体差异较大,开发通用、稳定、精确的番茄串视觉定位方法,引导机械臂以合适姿态完成采摘,是实现智能采摘的重要环节,也是实现无损、安全、高效、连续采摘的重要保证。

动态变化的非结构化作业环境、果梗纤细且姿态多样、传感器精度不足等因素,增加了采摘点定位和果梗姿态预测的难度,所以现有的智能采摘设备在采摘效率、识别准确率和稳定性等方面远低于人工采摘,无法满足商业化应用需求。

近年来,数字图像处理和神经网络广泛应用于采摘点的识别与定位中,加速了采摘机器人的产业化进程<sup>[6-10]</sup>。现有的果梗采摘点定位方法主要分为两类,第1类基于图像处理技术,根据果实颜色、形状或纹理等基本特征定位采摘点。JI等<sup>[11]</sup>基于颜色特征  $2R - G - B$  识别番茄果实与辅助标记,将拟合的果梗曲线与辅助标记边缘的交点作为采摘点,采摘成功率为 88.6%。LUO等<sup>[12-13]</sup>基于颜色特征分割葡萄串,根据果梗与果实的约束关系定位采摘点,准确率 88.33%,平均耗时 0.346 7 s。冯青春等<sup>[14]</sup>基于黄瓜的形状特征和 HIS 空间阈值提取

黄瓜轮廓,并从黄瓜轮廓顶部提取采摘点,定位精度达到 2 mm。梁喜凤等<sup>[15]</sup>基于果梗骨骼线角点定位番茄串采摘点位置信息,成功率为 90%。熊俊涛等<sup>[16]</sup>基于 HIS 空间对荔枝果实和果梗进行分割,利用 Hough 直线拟合确定有效的果梗采摘区域,通过连续处理多帧图像定位采摘点,深度误差小于 6 cm。YOSHIDA等<sup>[17]</sup>利用颜色和点云特征识别番茄串采摘点,单帧图像耗时 1 s,难以满足实时采摘要求。多特征融合比单特征识别方法鲁棒性更强,但两者都容易受光照变化和障碍物遮挡的影响,对环境依赖性强、通用性差且检测性能不稳定。此类方法多数要求果实与果梗的颜色、形态单一,相对位置固定。

第2类基于神经网络识别果实或果梗,基于颜色或形状等特征对识别结果进一步处理定位采摘点,神经网络的应用极大地提升了采摘点识别定位的速度和精度。陈燕等<sup>[18]</sup>基于改进 YOLO v3 与双目立体视觉对荔枝串预定位,处理速度为 22 f/s,最大定位绝对误差为 36 mm。YU等<sup>[19]</sup>提出了 R - YOLO 用于识别并估计草莓生长姿态,根据边界框轴线旋转角度计算采摘点,平均识别率为 94.43%,采摘成功率为 84.35%。XU等<sup>[20]</sup>基于改进的 Mask R - CNN 识别距离相机最近的番茄串作为采摘目标,准确率为 93.76%,单帧图像处理时间为 0.04 s。宁政通等<sup>[21]</sup>基于改进 Mask R - CNN 与颜色特征分

割葡萄果梗并定位采摘点,单帧图像耗时 4.9 s。KALAMPOKAS 等<sup>[22]</sup> 基于回归卷积神经网络识别葡萄串果梗,结合识别结果的形状和边缘特征定位采摘点,深度定位误差 10 mm,算法速率为 8.3 f/s。张勤等<sup>[23]</sup> 基于 YOLO v4 与果梗颜色特征,融合深度信息分割果梗,番茄串采摘点定位成功率为 93.83%,单帧图像耗时 54 ms。番茄串果梗纤细且颜色与枝叶相似,背景噪声复杂,受光照变化影响大,难以精确分割;番茄串形态差异大,果实与果梗之间无固定位置关系,作业环境复杂,增加了预测果梗姿态的难度;采摘过程中植株晃动,经济型深度相机精度不足和深度缺失,难以精确识别定位采摘点;由于果梗姿态多变,即使成功定位采摘点,也会由于剪刀的剪切姿态和干涉等问题造成采摘失败。现有的研究主要侧重于采摘点的精确定位,忽略了果实的可采摘性和生长姿态多样化的问题,造成“看到采不到”的现象,影响了采摘成功率和效率,限制了番茄串采摘机器人在实际中的广泛应用。

针对上述问题,本文提出基于实例分割的串番茄视觉定位与采摘姿态估算方法。基于 YOLACT 实例分割算法<sup>[24]</sup> 的实例特征标准化和掩膜评分机制,输出番茄串的掩膜和 ROI,保证果梗掩膜的连续性,过滤大量背景噪声实现果梗粗分割,基于果梗掩膜信息和 ROI 位置关系匹配可采摘果梗;其次,通过细化算法和膨胀操作对可采摘果梗进行精细分割,提取果梗骨骼线、定位采摘点图像坐标,降低光照和背景噪声对分割果梗的影响;基于果梗深度信息填补法与坐标系转换,融合深度信息定位采摘点空间坐标;通过果梗几何特征,识别果梗关键点预测果梗姿态,估算采摘姿态,引导机械臂以合适的姿态完成采摘。

1 数据集构建与标注

为解决非结构化种植环境下,番茄串采摘点识别定位、果梗姿态预测的问题,提出基于实例分割的串番茄视觉定位与采摘姿态估算方法,方法流程如图 1 所示。

数据采集地点为广东省某农业技术推广中心数字设施馆水培区,数据集包含 4 个品种的番茄串图像,分别为以色列红、金玲珑、粤科达 202 和鸿海 HH10,每个品种的图像分别为 1 307、2 942、1 065、1 206 幅;上述图像使用不同设备、在不同时间段、从不同角度和不同光照条件下进行拍摄,相机与番茄串之间的距离为 0.4 ~ 1.2 m,图像尺寸均为 1 280 像素×720 像素;每个品种按 20:3 进行划分,然后汇总为训练集和测试集。测试集中的果梗数量为 949,可采摘果梗数量为 797,用于测



图 1 基于实例分割的番茄串视觉定位与采摘姿态估算方法

Fig. 1 Method for visual positioning and picking pose estimation of tomato clusters based on instance segmentation

试模型的识别准确率。

使用开源标注工具 LabelMe<sup>[25]</sup> 对数据集进行标注,共有 2 个标注类别,其中“stem”为果梗,“tomato”为成熟番茄串,其余部分均视为“背景”。图 2 为以色列红番茄串标注示例,分别使用多边形和矩形标注果梗和成熟番茄串,有效减少数据标注时间。为了提高模型的鲁棒性和防止训练过程中模型过拟合,标注完成后对训练集进行数据集增强,主要操作包括镜像、旋转、模糊、调节亮度和对比度等,增强后训练集图像总数为 15 311 幅。对 4 个品种的番茄串图像数据进行训练,能够有效增强番茄串和果梗识别定位的鲁棒性,满足该模型对不同番茄品种的适应性。



图 2 数据标注示例

Fig. 2 Example of data annotation

2 番茄串视觉定位与采摘姿态估算

2.1 基于实例分割的番茄串视觉定位方法

2.1.1 基于 YOLACT 的 ROI 识别与果梗粗分割

YOLACT 实例分割算法属于单阶段模型,在保

持高质量分割的同时有效地减少计算成本,具有速度快、精度高的优点,能够快速识别番茄串与果梗 ROI,有利于在复杂背景中分割果梗轮廓。采用残差网络(Residual network 101, ResNet101)和特征金字塔网络(Feature pyramid network, FPN)作为特征提取网络训练番茄串和果梗的检测模型,ResNet 基于残差模块(Residual blocks)允许信息在不同深度之间流动,使得网络能够更好地提取图像特征,有助于提高模型的分割精度,具有更快的模型训练速度。

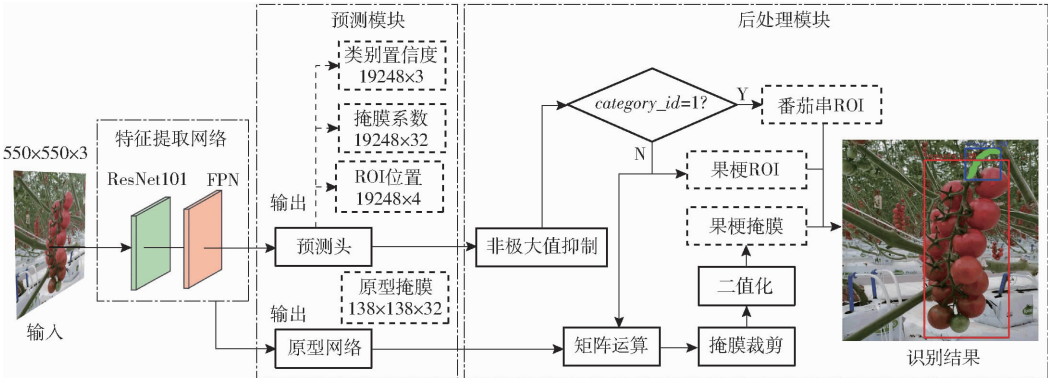


图3 基于实例分割的番茄串视觉定位方法

Fig. 3 Method for visual positioning of tomato clusters based on instance segmentation

由于番茄串果梗纤细且种植环境复杂,为了能够提取完整的果梗轮廓,在后处理模块通过降低掩膜精度对果梗进行粗分割,使果梗掩膜覆盖果梗轮廓,保证果梗的连续性;后续通过细化算法与形态学操作精确提取果梗轮廓,保证采摘点的识别定位精度和果梗姿态预测的可靠性。将识别结果存储于容器  $\text{vector}\langle \text{ROI} \rangle \text{Result}$ , ROI 包含下列参数:  $x$ 、 $y$ 、 $w$ 、 $h$ 、 $\text{confidence}$ 、 $\text{category\_id}$ ; 其中  $(x, y)$  为 ROI 左上角顶点坐标,  $w$  和  $h$  分别为 ROI 的宽和高;  $\text{confidence}$  为预测框置信度;  $\text{category\_id}$  为 0 时表示果梗 ROI, 为 1 时表示番茄串 ROI, 果梗 ROI 与果梗掩膜一一对应。

**2.1.2 基于 ROI 位置和掩膜信息匹配可采摘果梗**

由于识别结果中番茄串和果梗 ROI 数量不同,且位置具有随机性,需要进行匹配确定可采摘果梗,后续只处理可采摘果梗从而提高算法效率。主要依据为:①判断果梗掩膜是否连续,剔除部分可见区域较小、不利于提取果梗轮廓的番茄串。②果梗与果实的 ROI 位置关系。对识别结果进行分析,假设果梗和番茄串 ROI 的左上角顶点坐标分别为  $(x_i, y_i)$ 、 $(x_j, y_j)$ ,  $\text{Result}[i]$  和  $\text{Result}[j]$  分别表示当前处理的果梗 ROI 和番茄串 ROI(后文不再重述),具体步骤为:

(1) 基于轮廓检测判断果梗掩膜是否连续,若果梗掩膜不连续则剔除  $\text{Result}[i]$ 。

如图 3 所示,将图像调整为 550 像素  $\times$  550 像素后输入特征提取网络,提取番茄串和果梗的重要特征;预测模块对特征提取网络输出的重要特征进行整合与利用,由预测头分支输出各类候选框的类别置信度及位置信息、原型掩膜的掩膜系数;原型网络分支输出原型掩膜;原型掩膜和相应的掩膜系数进行组合得到番茄串和果梗的掩膜;后处理模块通过非极大值抑制、掩膜过滤、裁剪和二值化处理得到最终的掩膜结果与 ROI。

(2) 将  $\text{Result}$  内剩余的 ROI 按  $x$  坐标值由小到大进行排序。

(3) 判断 ROI 是否有交集:  $\text{if}(\text{Result}[i] \cap \text{Result}[j] \neq \emptyset \ \&\& \ y_i < y_j + 0.5h_j)$ ,  $\text{Result}[i]$  为可采摘果梗。

在步骤(3)中使果梗 ROI 下边框向下偏移  $0.15h_i$ , 番茄串 ROI 上边框向上偏移  $0.15h_j$ , 增强该方法对不同果梗长度的适应性。

2.2 采摘点定位方法

**2.2.1 基于细化算法和膨胀操作的果梗精细分割**

由于原始果梗掩膜存在番茄果实、植株主干或叶子等背景噪声,为了精确提取果梗轮廓,提高采摘点识别定位精度与稳定性,去除背景噪声的步骤为:

(1) 基于 ZHANG 等<sup>[26]</sup>快速并行细化算法提取原始果梗骨骼线。

(2) 基于果梗 ROI 长宽比确定 ROI 中心线  $L_{mid}$  的取值

$$L_{mid} = \begin{cases} 0.5h_i & (h_i > w_i) \\ 0.5w_i & (h_i \leq w_i) \end{cases} \quad (1)$$

(3) 求取原始果梗掩膜图在第  $L_{mid}$  行(或列)像素值非零点的总数,即果梗掩膜宽度  $L$ (单位:像素),如图 4 红色线段所示。

(4) 对原始果梗骨骼线进行膨胀操作,获得新的果梗掩膜:膨胀操作内核形状为 MORPH\_ELLIPS,以果梗掩膜宽度  $L$  为参数调整内核尺寸。



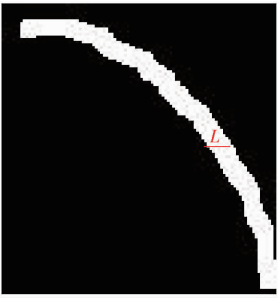


图 4 果梗掩膜宽度

Fig. 4 Mask width of stem

(5) 对新果梗掩膜进行细化处理提取果梗骨骼线。

### 2.2.2 基于果梗深度信息填补的采摘点定位

首先计算果梗骨骼线图第  $L_{mid}$  行(或列)中第 1 个像素非零的点  $(x_c, y_c)$ , 设定  $(x_i + x_c, y_i + y_c)$  为采摘点图像坐标  $(P_x, P_y)$ , 其中  $x_i$  和  $y_i$  分别为果梗 ROI 左上角顶点的坐标值, 然后融合配准后的深度图信息精确提取采摘点的深度  $P_z$ 。由于深度相机获取细小物体的深度信息时误差较大且不稳定, 导致某些像素点丢失深度, 为保证采摘点图像坐标  $(P_x, P_y)$  深度的可靠性, 提出果梗深度信息填补法。设图像坐标点  $(u, v)$  的深度为  $d(u, v)$  (单位: mm):

(1) 新果梗掩膜上所有满足  $0 < d(u, v) < 1\ 000\text{ mm}$  的像素点总数记为  $N$ , 计算这些像素点的深度之和, 果梗掩膜区域的平均深度  $d_{average\_1}$  计算式为

$$d_{average\_1} = \frac{\sum_{i=1}^N d(u, v)}{N} \quad (2)$$

(2) 提取点  $(P_x, P_y)$  的原始深度  $d_{original}$ 。

(3) 统计果梗掩膜宽度  $L$  上所有像素点的深度  $d(u, v)$ , 记所有满足  $0 < d(u, v) < d_{average\_1}$  的像素点总数为  $M$ , 统计这些点的深度之和, 采摘点平均深度  $d_{average\_2}$  计算式为

$$d_{average\_2} = \frac{\sum_{i=1}^M d(u, v)}{M} \quad (3)$$

(4) 选取最优采摘点深度  $P_z$

$$P_z = \begin{cases} d_{original} & (|d_{original} - d_{average\_2}| < 20\text{ mm}) \\ d_{average\_2} & (|d_{original} - d_{average\_2}| \geq 20\text{ mm}) \end{cases} \quad (4)$$

步骤(3)中通过计算线段的平均深度  $L$  作为采摘点的平均深度, 即使采摘点原始深度异常或丢失, 可以使用采摘点平均深度  $d_{average\_2}$  填补原始深度  $d_{original}$ , 以保证采摘点的定位精度和可靠性。

### 2.3 果梗姿态预测方法

由于番茄串果梗姿态各异, 机械臂以固定姿态采摘时存在采摘效率低、植株或果实损坏等问题。

为了避免采摘过程中末端执行器与果实或植株主干发生干涉, 导致植株晃动剧烈影响识别定位和采摘, 需要在实施采摘任务前对果梗生长姿态进行预测, 从而引导机械臂以适合的姿态靠近果梗, 提高采摘成功率和效率, 保证果实质量。

如图 5 所示, 以果梗骨骼线图上的采摘点(绿色圆点)图像坐标  $(x_c, y_c)$  为中心, 分别向骨骼线两端各拓展 10 个像素值非 0 点, 如图 5 中虚线框所示。记拓展后所得的上端点和下端点的图像坐标分别为  $(x_1, y_1)$  和  $(x_2, y_2)$ , 以两端点构成的直线斜率近似代替采摘点处的果梗切线斜率, 果梗的倾斜角  $\theta$  计算式为

$$\theta = \arctan \frac{y_2 - y_1}{x_2 - x_1} \quad \left( \theta \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right] \right) \quad (5)$$

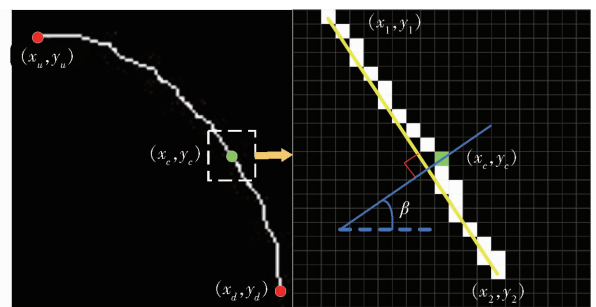


图 5 果梗骨骼线与采摘点处切线

Fig. 5 Skeleton of stem and tangent line at picking point

根据果梗的倾斜角  $\theta$  预测垂直生长的果梗, 经过现场试验表明, 当果梗倾斜角满足  $\theta \in ([4\pi/9, \pi/2] \cup [-\pi/2, -4\pi/9])$ , 可判断果梗为垂直生长, 否则按如下方法继续预测果梗姿态: 首先基于八邻域端点检测算法快速提取果梗的关键点, 即果梗骨骼线的上端点和下端点图像坐标  $(x_u, y_u)$  和  $(x_d, y_d)$ , 其中  $y_u < y_d$ , 如图 5 中红色圆点所示; 然后根据 2.2.2 节的方法提取  $(x_u, y_u)$  和  $(x_d, y_d)$  处的深度  $z_u$  和  $z_d$  并转换到机械臂坐标系(单位: m), 机械臂坐标系如图 6 所示。假设果梗骨骼线的上端点和下端点在机械臂坐标系下的坐标分别为  $P_u(x_{u1}, y_{u1}, z_{u1})$  和  $P_d(x_{d1}, y_{d1}, z_{d1})$ , 以方向向量  $\mathbf{l}_{P_u P_d}$  对番茄串果梗进行姿态预测。将向量  $\mathbf{l}_{P_u P_d}$  向  $XY$  平面投影, 得到  $P'_u(x_{u1}, y_{u1})$  和  $P'_d(x_{d1}, y_{d1})$ ,  $\mathbf{l}_{P'_u P'_d}$  与  $X$  轴正方向的夹角  $\gamma$  计算式为

$$\gamma = \text{atan2}(y_{d1} - y_{u1}, x_{d1} - x_{u1}) \quad (\gamma \in (-\pi, \pi]) \quad (6)$$

通过  $\gamma$  的取值预测果梗相对于植株主干的生长姿态。

### 2.4 采摘姿态与采摘方法

#### 2.4.1 采摘姿态估算

由于番茄串果梗木质化使其含有大量不易剪断

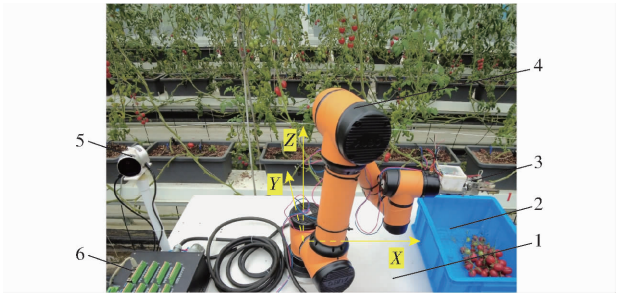


图 6 番茄串采摘机器人结构

Fig. 6 Structure of tomato cluster harvesting robot

1. 移动平台 2. 收纳筐 3. 末端执行器 4. 机械臂 5. 深度相机 6. 机械臂控制器

的纤维,若果梗未被剪断,机械臂拖拽果梗会损伤番茄植株;番茄串生长姿态各异,如果末端执行器以固定角度进行剪切,靠近目标时容易与果梗或果实发生干涉,同时干扰了视野中已识别的可采摘目标,导致植株晃动,严重影响番茄串的识别定位和采摘。因此选择适合采摘目标生长特点的剪切姿态是成功采摘的关键。

如图 5 所示,通过采摘点处的近似切线计算果梗的剪切角 $\beta$ (逆时针为正方向,单位:rad);机械臂初始状态末端执行器剪切平面与地面平行,剪切角 $\beta=0$ ,如图 5 中蓝色虚线所示。由于受果实重力的影响,大部分果梗的生长姿态为斜向下弯曲或垂直的状态,可以通过果梗倾斜角 $\theta$ 判断果梗被果实压弯的方向,剪切角计算式为

$$\beta = \begin{cases} \theta - \theta_s & (\theta \geq 0) \\ \theta + \theta_s & (\theta < 0) \end{cases} \quad (7)$$

式中 $\theta_s$ 为变量(单位:rad),用于调节末端执行器的剪切平面(图 5 蓝色线段)与果梗切线(图 5 黄色线段)的夹角,满足不同的农艺需求和应用场景。但对于不同品种的番茄串,由于果梗直径、硬度、强度、果实重量不同等因素,有部分果梗为斜向上生长,如按式(7)计算剪切角,此时剪切平面与果梗切线平行,无法成功剪切果梗,或剪切后无法夹持番茄串,导致采摘失败。此时需要引入果梗和番茄串 ROI 的左上角顶点坐标 $(x_i, y_i)$ 、 $(x_j, y_j)$ ,以及 ROI 对应的宽度 $w_i, w_j$ ,判断番茄串和果梗 ROI 的相对位置,剪切角计算式为:

当 $\theta \geq 0$ 时

$$\beta = \begin{cases} |\theta - \theta_s| & (x_i + w_i > x_j + w_j) \\ -|\theta - \theta_s| & (x_i + w_i \leq x_j + w_j) \end{cases} \quad (8)$$

当 $\theta < 0$ 时

$$\beta = \begin{cases} -|\theta + \theta_s| & (x_i < x_j) \\ |\theta + \theta_s| & (x_i \geq x_j) \end{cases} \quad (9)$$

为了保证夹剪一体的末端执行器夹持稳定性,

充分发挥末端执行器的容错性,选择切割强度最大的剪切姿态,如图 5 所示,即末端执行器的剪切平面与果梗切线垂直,此时 $\theta_s = \pi/2$ 。

2.4.2 采摘机器人系统

番茄串采摘机器人主要由工业控制计算机、可升降移动平台、L515 型深度相机(Intel RealSense)、Aubo-i5 机械臂、末端执行器和收纳筐组成,如图 6 所示。

2 个深度相机对称安装于机械臂坐标系 X 轴线上,分别用于识别轨道两侧的番茄串;机械臂的工作半径为 886.5 mm,末端执行器为夹持和剪切一体设计,最大开口宽度为 23 mm,可剪切直径 5 mm 以内的果梗。机器人通过视觉识别系统实时获取图像,输出采摘点位置和果梗姿态信息,控制各个组成模块的协调运行,机械臂通过果梗姿态自主调节剪切姿态完成采摘;移动平台通过识别磁条与 RFID 实现自动导航与上下轨道,能够自动切换采摘方向并调节高度,实现对双侧番茄串的智能化采摘。

2.4.3 手眼坐标转换

首先根据深度相机内参和采摘点深度 $P_z$ ,调用接口将采摘点图像坐标 $(P_x, P_y)$ 转换到深度相机坐标系得到点 $P_{camera}$ ,然后利用张氏标定法<sup>[27]</sup>获取深度相机坐标系到机械臂坐标系的齐次变换矩阵 $T_{camera}^{robot}$ ,将点 $P_{camera}$ 由深度相机坐标系转换到机械臂坐标系,即

$$\begin{bmatrix} P_{robot\_x} \\ P_{robot\_y} \\ P_{robot\_z} \\ 1 \end{bmatrix} = T_{camera}^{robot} \begin{bmatrix} P_{camera\_x} \\ P_{camera\_y} \\ P_{camera\_z} \\ 1 \end{bmatrix} \quad (10)$$

手眼标定测试中,重复标定误差控制在 1 mm 以内,为机械臂提供准确的采摘点位置信息。采摘过程中,机械臂根据果梗姿态预测结果 $\gamma$ 和果梗倾斜角 $\theta$ ,即根据果梗生长方向,选择合适的姿态接近果梗,避免与植株或果实发生干涉;将采摘任务进行分解并分段规划采摘路径,同时利用剪切角 $\beta$ 调整末端执行器的剪切姿态,引导机械臂以合适的姿态完成采摘<sup>[28]</sup>。

3 试验

3.1 模型训练与部署

Windows 10(64 位)系统下使用 Anaconda3 创建训练环境,Python 版本为 3.7.13,配备 AMD Ryzen 5 1400 处理器、16 GB 运行内存、GPU 为 GeForce GTX 1080Ti,使用构建的数据集训练模型;

训练结束后将模型文件转换成深度学习模型开放格式 (Open neural network exchange, ONNX), 基于 C++ 部署该模型至采摘机器人系统并测试, 开发环境为 AMD Ryzen 5 1400 处理器、8 GB 运行内存、GeForce RTX2060 GPU、Qt Creator 4.11.0、VS2017、CUDA 10.2、Opencv-Contrib-4.5.4。

3.2 番茄串视觉定位方法性能试验

3.2.1 模型识别准确率

为了测试模型对番茄串和果梗 ROI 的识别准确率,以召回率、预测精度作为主要评价指标。为了验证模型的分割精度,以平均交并比 (Mean intersection over union, mIOU)作为评价指标。试验过程中以重叠系数<sup>[29]</sup>评估模型的 ROI 识别准确率,重叠系数表示预测框与实际目标框之间的重叠比例,计算式为

$$O_c = \frac{A_T \cap A_D}{A_T \cup A_D} \tag{11}$$

式中  $A_T$ ——实际目标框区域  
 $A_D$ ——预测框区域

在试验中设定重叠系数大于 0.5 时,目标被正确识别。对第 1 节构建的测试集进行识别与统计,识别结果如表 1 所示,得到该模型对番茄串的预测精度和召回率分别为 97.67% 和 98.89%,对果梗的预测精度和召回率分别为 98.83% 和 98.00%;得到平均交并比为 69.04%。图像分辨率为 1 280 像素 × 720 像素时,算法处理速率达到 21 f/s。

表 1 模型识别结果  
Tab.1 Model recognition results

| 实际类   | 预测类   |     |    |
|-------|-------|-----|----|
|       | 成熟番茄串 | 果梗  | 背景 |
| 成熟番茄串 | 1 427 |     | 16 |
| 果梗    |       | 930 | 19 |
| 背景    | 34    | 11  |    |

3.2.2 采摘点识别成功率

使用同一模型对不同品种、不同姿态的番茄串进行采摘点识别定位,如图 7 所示。图 7 中红色箭头表示采摘点搜索方向,绿色圆点表示采摘点,若采摘点落入果梗轮廓内则采摘点识别定位成功。测试集中可采摘果梗实例测试结果如表 2 所示,以色列红、金玲珑、粤科达 202 和鸿海 HH10 番茄串的采摘点识别成功率分别为 98.20%、98.94%、97.25% 和 97.89%,不同品种的番茄串采摘点平均识别成功率为 98.07%。试验结果表明,提出方法对不同品种番茄串采摘点的识别定位成功率和速度能够满足采摘需求。

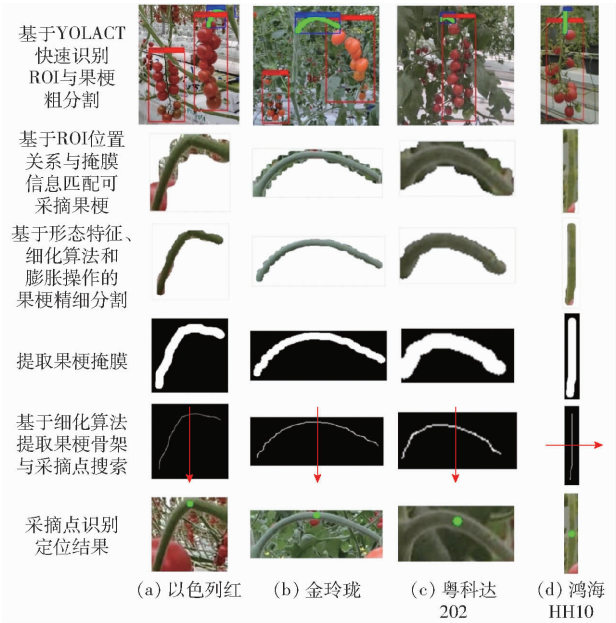


图 7 不同品种番茄串采摘点识别定位结果

Fig.7 Recognition and positioning of different varieties of tomato cluster picking points

表 2 测试集采摘点识别成功率

Tab.2 Picking point recognition accuracy of test set

| 品种      | 可采摘果梗数 | 成功定位采摘<br>点数 | 采摘点识别<br>成功率/% |
|---------|--------|--------------|----------------|
| 以色列红    | 167    | 164          | 98.20          |
| 金玲珑     | 379    | 375          | 98.94          |
| 粤科达 202 | 109    | 106          | 97.25          |
| 鸿海 HH10 | 142    | 139          | 97.89          |

3.2.3 采摘点图像坐标定位精度

试验地点位于广东省某农业技术推广中心数字设施馆水培区。为了确定采摘点的定位精度,从粤科达 202 品种中选取 20 串可采摘番茄,其中包含不同光照条件和不同姿态的果梗实例。机器人运动至指定位置后,对每串番茄连续拍摄 6 帧图像 (1 280 像素 × 720 像素) 并识别定位采摘点,同时保存拍摄的彩色图与深度图,使用文献[23]方法进行采摘点定位精度测试。

设最佳采摘点图像坐标为果梗 ROI 第  $L_{mid}$  行 (或列) 与果梗中轴线的交点,记为  $(O_x, O_y)$ 。算法识别的采摘点与最佳采摘点之间的误差  $\Delta_e$  计算式为

$$\Delta_e = \sqrt{e_x^2 + e_y^2} \tag{12}$$

其中  $e_x = |O_x - P_x| \tag{13}$

$e_y = |O_y - P_y| \tag{14}$

式中  $e_x$ ——采摘点在行方向上的偏差  
 $e_y$ ——采摘点在列方向上的偏差

按照 2.2.2 节的方法提取采摘点的原始深度  $d_{original}$  与平均深度  $d_{average\_2}$ ,当两者均为非零值时,采



摘点平均深度与原始深度的误差  $\varepsilon$  计算式为

$$\varepsilon = d_{original} - d_{average\_2} \tag{15}$$

分别记录每串番茄的 6 次采摘点识别定位结果,如图 8 所示,其中图 8a 记录了采摘点处平均深度误差的平均值和标准差,可知采摘点处平均深度误差  $\varepsilon$  为  $\pm 6\text{ mm}$ ,证明当采摘点原始深度异常或丢失时,可以采用平均深度代替原始深度。由于文献[23]方法的番茄串检测模型和数据集差异,有 4 串番茄未能成功识别,因此图 8b 仅统计了 16 组数据。由图 8 可知,提出方法和文献[23]方法的采摘点图像坐标最大定位误差分别为 3 像素和 10 像素,平均定位误差分别为 0.84 像素和 3.62 像素,其中输入图像中果梗直径为 7~15 像素,与文献[23]方法相比,采摘点平均定位精度提高 76.80 个百分点。此外,提出的方法对同一帧图像连续进行 10 次采摘点识别定位,每次输出的采摘点坐标( $P_x, P_y, P_z$ )均为定值,表明该方法的定位精度和稳定性能够满足采摘需求。

3.2.4 番茄串视觉定位与果梗分割性能对比试验

为验证提出方法的有效性,将该方法与文献[23]提出的方法进行对比试验,图 9 为两种方法对不同形态的番茄串处理结果,蓝色圆点为文献[23]方法识别结果,绿色圆点为提出方法的识别结果。当番茄串背景环境复杂或光照不稳定时,文献[23]方法通过深度分割算法,随机选取聚类中心进行 K-means 聚类分割果梗,鲁棒性较差,难以精确、稳定地定位采摘点图像坐标和深度。如图 10 所示,文献[23]方法在相同条件下对试验 2 中的番茄串重复识别,分割果梗时存在较多背景噪声且聚类结果不稳定,导致采摘点定位误差较大。

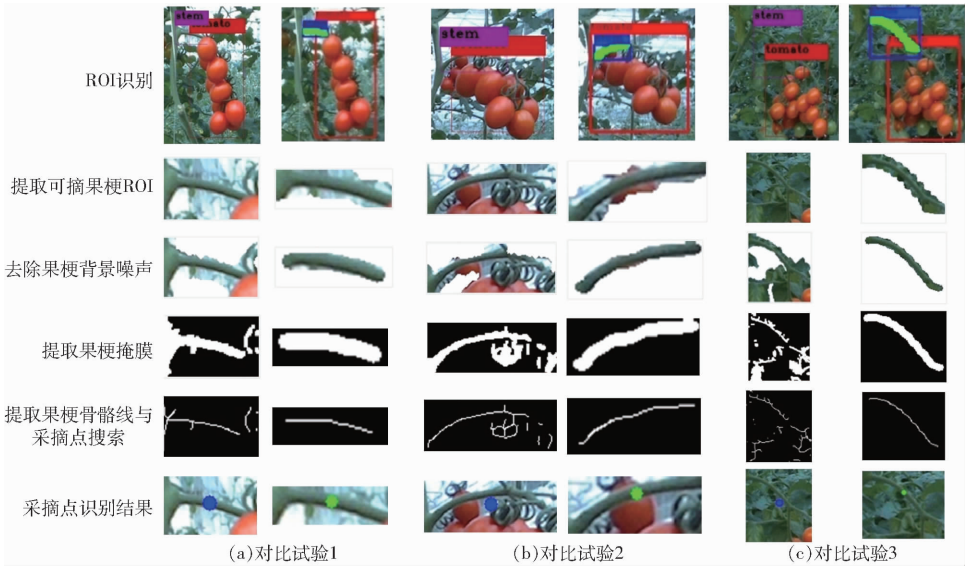


图 9 本文方法与文献[23]方法采摘点识别对比试验

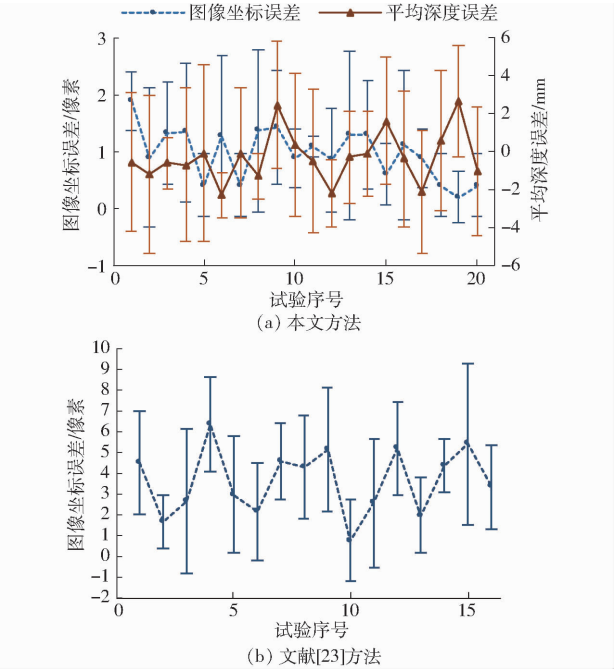


图 8 采摘点定位误差分布  
Fig. 8 Picking point error

试验结果表明,在相同场景下提出的方法具有更高的准确率和稳定性,能够准确分割果梗并提取果梗骨骼线,有利于精确定位采摘点和预测果梗姿态。

3.2.5 采摘方法对比试验与误差分析

以往的采摘方法,基于番茄串和果梗 ROI 位置关系,将番茄串生长姿态分为向左、向右和垂直生长 3 个类别,经常出现果梗姿态预测错误的情况,且末端执行器只有 3 个对应的采摘姿态,容易与果实、果梗或植株主干发生干涉,存在“看到采不到”的现象(方法 1)。提出的采摘方法不局限于 3 种采摘姿态,机械臂通过估算的果梗姿态自主调节剪切姿态,

Fig. 9 Comparative experiments of picking point recognition between proposed method and reference [23] method



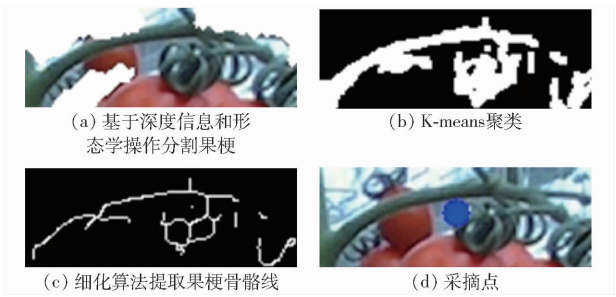


图 10 文献[23]方法重复识别试验

Fig. 10 Repeated recognition of reference[23] method

以合适的姿态靠近果梗,避免发生干涉(方法 2)。

基于同一个番茄串检测模型,使用方法 1 和方法 2,对不同果梗姿态的番茄串进行采摘试验,试验对象为 1 垄粤科达 202 番茄和 1 垄鸿海 HH10 番茄。采摘机器人在轨道上运行,通过视觉识别系统实时控制机器人移动,识别到可采摘目标后机器人触发制动指令,重新获取最新一帧图像并识别定位采摘点;机器人根据采摘点与果梗姿态信息调整采摘姿态,规划采摘路径,采摘完成后或连续 3 次采摘失败,机器人继续向前运动,重复上述步骤实现连续采摘,使用方法 2 采摘粤科达 202 番茄串过程如图 11 所示。

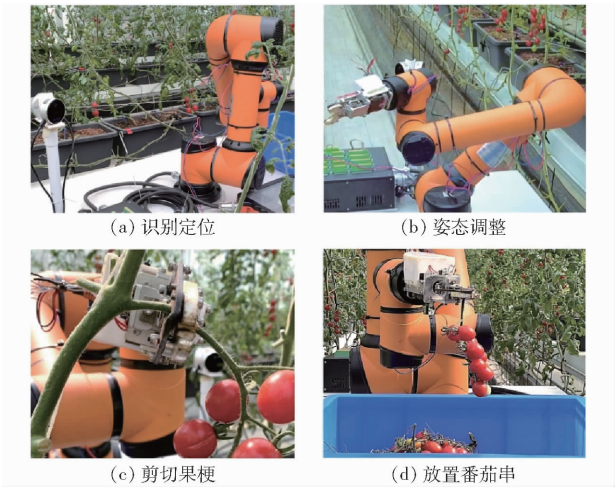


图 11 番茄串采摘过程

Fig. 11 Action sequence of tomato cluster harvesting

两种方法的采摘试验结果如表 3 所示,方法 1 的采摘成功率为 82.98%,平均采摘效率(成功采摘的番茄串总数量与采摘动作总次数的比值)为 55.71%。方法 2 的采摘成功率为 98.15%,平均采摘效率为 86.89%;相比于方法 1,采摘成功率提高 15.17 个百分点,平均采摘效率提高 31.18 个百分点。采摘试验过程中,采摘失败通常会导致番茄植株剧烈晃动,严重影响后续的番茄串识别定位和采摘。图 12 为方法 2 采摘失败案例,试验过程中为了限制机械臂的运动范围,对获取的图像进行裁剪。由于该番茄串处于图像的顶部被裁剪区域,因此番

茄串在距离相机较远时才进入识别范围,此时部分果梗被番茄植株主干遮挡,且相机成像平面与果梗平面不平行时,相机深度缺失、误差较大,使采摘点定位存在较大偏差,导致采摘失败。

表 3 采摘试验结果  
Tab. 3 Results of picking test

| 采摘方法 | 品种      | 成功识别数量 | 成功采摘数量 | 采摘动作次数 |
|------|---------|--------|--------|--------|
| 方法 1 | 粤科达 202 | 19     | 16     | 28     |
|      | 鸿海 HH10 | 28     | 23     | 42     |
| 方法 2 | 粤科达 202 | 21     | 21     | 23     |
|      | 鸿海 HH10 | 33     | 32     | 38     |



图 12 采摘失败案例

Fig. 12 Harvesting failure

图 13 为不同果梗姿态的剪切角,以末端执行器剪切刀刃中点为参考点,如图 13a 黄色标记所示,设果梗采摘点与参考点的相对距离为深度误差  $\Delta_d$ ;试验过程中选择切割强度最大的剪切姿态,即末端执行器的剪切平面与果梗切线垂直,如图 13b 所示,设末端执行器剪切平面与果梗切线正方向(指向植株主干的方向为正方向)的夹角为  $\alpha$ ;若剪切平面垂直于果梗切线,即  $\alpha = 90^\circ$ ,此时剪切角误差为  $0^\circ$ ,剪切角误差评价指标  $\Delta_\alpha$  计算式为

$$\Delta_\alpha = \alpha - 90^\circ \tag{16}$$

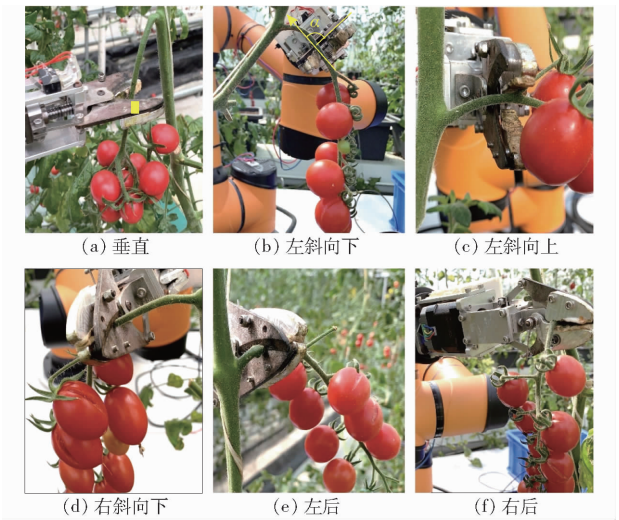


图 13 不同果梗姿态的剪切角

Fig. 13 Cutting angles of different stems

对方法 2 中成功采摘的番茄串进行深度误差与剪切角误差分析,误差分布结果如图 14、15 所示,深度误差  $\Delta_d$  为  $\pm 4$  mm,剪切角误差  $\Delta_a$  为  $\pm 6^\circ$ 。误差来源主要包括深度相机测量误差、手眼标定误差、采摘点图像坐标定位误差、移动平台振动、试验场地排气扇和采摘过程使植株产生晃动等。试验结果表明,使用方法 2 进行采摘试验,即使采摘点识别定位和果梗姿态预测存在一定误差,高容错率的末端执行器依然能够成功采摘番茄串。

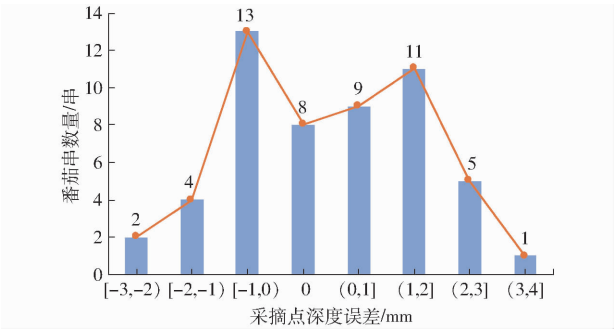


图 14 采摘点深度误差分布

Fig. 14 Distribution of picking point depth value error

4 结束语

考虑番茄串果实生长姿态的多样性和可采摘性,基于实例分割提出番茄串的视觉定位与采摘姿态估算方法。通过果梗的粗分割、精细分割以及深

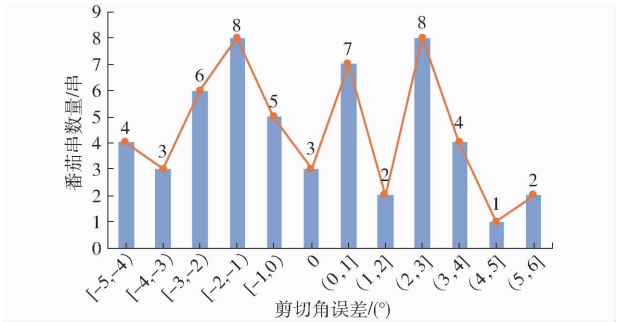


图 15 剪切角误差分布

Fig. 15 Distribution of shear angle error

度信息的填补、融合,精确定位采摘点;基于果梗关键点预测果梗姿态,并根据果梗姿态确定适合采摘的末端执行器采摘姿态,有效提高了采摘成功率。研究和试验结果表明,该方法对 4 个品种的番茄串采摘点平均识别成功率为 98.07%,采摘点图像坐标最大定位误差为 3 像素,深度误差为  $\pm 4$  mm,相对于果梗的剪切角误差为  $\pm 6^\circ$ ,成功识别后采摘成功率为 98.15%,图像分辨率为 1 280 像素  $\times$  720 像素时算法处理速率达到 21 f/s。与现有的同类方法相比,采摘点图像坐标定位精度提高 76.80 个百分点,采摘成功率提高 15.17 个百分点,采摘效率提高 31.18 个百分点。该研究以番茄串采摘为例提出的视觉定位与姿态预测方法,同样适用于其他串收果实的采摘。

参 考 文 献

[1] 刘继展. 温室采摘机器人技术研究进展分析[J]. 农业机械学报, 2017, 48(12): 1-18.  
LIU Jizhan. Research progress analysis of robotic harvesting technologies in greenhouse[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 1-18. (in Chinese)

[2] 刘成良, 贡亮, 苑进, 等. 农业机器人关键技术研究现状与发展趋势[J]. 农业机械学报, 2022, 53(7): 1-22, 55.  
LIU Chengliang, GONG Liang, YUAN Jin, et al. Current status and development trends of agricultural robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 1-22, 55. (in Chinese)

[3] HUGHES J, LIDA F, BIRRELL S A. Field robotics for harvesting: a review of field robotics approaches for harvesting[J]. Digital Agritechology, 2022, 11(4): 69-94.

[4] 宋健, 张铁中, 徐丽明, 等. 果蔬采摘机器人研究进展与展望[J]. 农业机械学报, 2006, 37(5): 158-162.  
SONG Jian, ZHANG Tiezhong, XU Liming, et al. Research actuality and prospect of picking robot for fruits and vegetables [J]. Transactions of the Chinese Society for Agricultural Machinery, 2006, 37(5): 158-162. (in Chinese)

[5] LIU J, LI Z, LI P. History and present situations of robotic harvesting technology: a review[J]. Rapid Damage-Free Robotic Harvesting of Tomatoes, 2021, 9(2): 107-125.

[6] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: a review[J]. Frontiers in Plant Science, 2020, 11: 510.

[7] ZHAO Y, GONG L, HUANG Y, et al. A review of key techniques of vision-based control for harvesting robot[J]. Computers and Electronics in Agriculture, 2016, 127: 311-323.

[8] GONGAL A, AMATYA S, KARKEE M, et al. Sensors and systems for fruit detection and localization: a review [J]. Computers and Electronics in Agriculture, 2015, 116: 8-19.

[9] 宋怀波, 尚钰莹, 何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报, 2023, 54(1): 1-19.  
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1): 1-19. (in Chinese)

[10] ZUJEVS A, OSADCUKS V, AHRENDT P. Trends in robotic sensor technologies for fruit harvesting: 2010-2015 [J]. Procedia Computer Science, 2015, 77: 227-233.

- [11] JI C, ZHANG J, YUAN T, et al. Research on key technology of truss tomato harvesting robot in greenhouse[C] // *Applied Mechanics and Materials*, 2014: 480 – 486.
- [12] LUO L, TANG Y, LU Q, et al. A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard[J]. *Computers in industry*, 2018, 99: 130 – 139.
- [13] 罗陆锋, 邹湘军, 熊俊涛, 等. 自然环境下葡萄采摘机器人采摘点的自动定位[J]. *农业工程学报*, 2015, 31(2): 14 – 21.  
LUO Lufeng, ZOU Xiangjun, XIONG Juntao, et al. Automatic positioning for picking point of grape picking robot in natural environment[J]. *Transactions of the CSAE*, 2015, 31(2): 14 – 21. (in Chinese)
- [14] 冯青春, 袁挺, 纪超, 等. 黄瓜采摘机器人远近景组合闭环定位方法[J]. *农业机械学报*, 2011, 42(2): 154 – 157.  
FENG Qingchun, YUAN Ting, JI Chao, et al. Feedback locating control based on close scene for cucumber harvesting robot[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2011, 42(2): 154 – 157. (in Chinese)
- [15] 梁喜凤, 金超杞, 倪梅娣, 等. 番茄果实串采摘点位置信息获取与试验[J]. *农业工程学报*, 2018, 34(16): 163 – 169.  
LIANG Xifeng, JIN Chaoqi, NI Meidi, et al. Acquisition and experiment on location information of picking point of tomato fruit cluster[J]. *Transactions of the CSAE*, 2018, 34(16): 163 – 169. (in Chinese)
- [16] 熊俊涛, 邹湘军, 陈丽娟, 等. 采摘机械手对扰动荔枝的视觉定位[J]. *农业工程学报*, 2012, 28(14): 36 – 41.  
XIONG Juntao, ZOU Xiangjun, CHEN Lijuan, et al. Visual position of picking manipulator for disturbed litchi[J]. *Transactions of the CSAE*, 2012, 28(14): 36 – 41. (in Chinese)
- [17] YOSHIDA T, FUKAO T, HASEGAWA T. Cutting point detection using a robot with point clouds for tomato harvesting[J]. *Journal of Robotics and Mechatronics*, 2020, 32(2): 437 – 444.
- [18] 陈燕, 王佳盛, 曾泽钦, 等. 大视场下荔枝采摘机器人的视觉预定位方法[J]. *农业工程学报*, 2019, 35(23): 48 – 54.  
CHEN Yan, WANG Jiasheng, ZENG Zeqin, et al. Vision pre-positioning method for litchi picking robot under large field of view[J]. *Transactions of the CSAE*, 2019, 35(23): 48 – 54. (in Chinese)
- [19] YU Y, ZHANG K, LIU H, et al. Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot[J]. *IEEE Access*, 2020, 8: 116556 – 116568.
- [20] XU P, FANG N, LIU N, et al. Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation[J]. *Computers and Electronics in Agriculture*, 2022, 197: 106991.
- [21] 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 Jiaxin, et al. Recognition and the optimal picking point location of grape stems based on deep learning[J]. *Transactions of the CSAE*, 2021, 37(9): 222 – 229. (in Chinese)
- [22] KALAMPOKAS T, VROCHIDOU E, PAPAKOSTAS G A, et al. Grape stem detection using regression convolutional neural networks[J]. *Computers and Electronics in Agriculture*, 2021, 186: 106220.
- [23] 张勤, 陈建敏, 李彬, 等. 基于 RGB – D 信息融合和目标检测的番茄串采摘点识别定位方法[J]. *农业工程学报*, 2021, 37(18): 143 – 152.  
ZHANG Qin, CHEN Jianmin, LI Bin, et al. Method for recognizing and locating tomato cluster picking points based on RGB – D information fusion and target detection[J]. *Transactions of the CSAE*, 2021, 37(18): 143 – 152. (in Chinese)
- [24] BOLYA D, ZHOU C, XIAO F, et al. YOLACT: real-time instance segmentation[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 9157 – 9166.
- [25] RUSSELL B C, TORRALBA A, MURPHY K P, et al. LabelMe: a database and web-based tool for image[J]. *Int. J. of Computer Vision*, 2008, 77(1): 157 – 173.
- [26] ZHANG T Y, SUEN C Y. A fast parallel algorithm for thinning digital patterns[J]. *Communications of the ACM*, 1984, 27(3): 236 – 239.
- [27] ZHANG Z. A flexible new technique for camera calibration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330 – 1334.
- [28] 张勤, 刘丰溥, 蒋先平, 等. 番茄串收机械臂运动规划方法与试验[J]. *农业工程学报*, 2021, 37(9): 149 – 156.  
ZHANG Qin, LIU Fengpu, JIANG Xianping, et al. Motion planning method and experiments of tomato bunch harvesting manipulator[J]. *Transactions of the CSAE*, 2021, 37(9): 149 – 156. (in Chinese)
- [29] YU Y, ZHANG K, YANG L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask – RCNN[J]. *Computers and Electronics in Agriculture*, 2019, 163: 104846.