

doi:10.6041/j.issn.1000-1298.2023.11.017

基于改进 U-Net 的火龙果采摘图像分割和姿态估计方法

朱立学^{1,2} 赖颖杰¹ 张世昂¹ 伍荣达¹ 邓文乾¹ 郭晓耿¹

(1. 仲恺农业工程学院机电工程学院, 广州 510225;

2. 岭南现代农业科学与技术广东省实验室, 广州 510630)

摘要: 为了实现火龙果采收自动化作业,提出一种基于改进 U-Net 的火龙果图像分割和姿态估计方法。首先,在 U-Net 模型的跳跃连接(编码器与解码器部分特征图进行的连接操作)中引入通道和空间注意力机制模块(Concurrent spatial and channel squeeze and channel excitation, SCSE),同时将 SCSE 模块集成到残差模块(Double residual block, DRB)中,在增强网络提取有效特征能力的同时提高网络的收敛速度,得到一种基于注意力残差 U-Net 的火龙果图像分割网络。通过该网络分割出果实及其附生枝条的掩膜图像,利用图像处理技术和相机成像模型拟合出果实及其附生枝条的轮廓、果实质心、果实最小外接矩形框和三维边界框,进而结合果实及其附生枝条的位置关系进行火龙果三维姿态估计,并在火龙果种植园中获得一个测试集,以评价该算法的性能,最后在自然果园环境下进行实地采摘试验。试验结果表明,火龙果果实图像分割平均交并比(mIoU)和平均像素准确率(mPA)分别达到 86.69% 和 93.89%,三维姿态估计平均误差为 8.8°,火龙果采摘机器人在果园环境下的采摘成功率为 86.7%,平均采摘时间为 22.3 s。满足火龙果机械化作业要求。

关键词: 火龙果采摘机器人; 图像分割; 姿态估计; 注意力残差 U-Net; 三维边界框

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-1298(2023)11-0180-09

OSID:



Image Segmentation and Pose Estimation Method for Pitaya Picking Robot Based on Enhanced U-Net

ZHU Lixue^{1,2} LAI Yingjie¹ ZHANG Shiang¹ WU Rongda¹ DENG Wenqian¹ GUO Xiaogeng¹

(1. College of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

2. Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510630, China)

Abstract: In order to achieve automation of pitaya harvesting, an improved U-Net based method for pitaya image segmentation and pose estimation was proposed. Firstly, a concurrent spatial and channel squeeze and channel exception (SCSE) module was introduced into the skip connection (connection operation between the encoder and decoder feature maps) of the U-Net model. At the same time, the SCSE module was integrated into the residual module double residual block (DRB) to enhance the network's ability to extract effective features while improving its convergence speed, obtaining a pitaya image segmentation network based on attention residual U-Net. By using this network to segment mask images of fruits and their accompanying branches, image processing techniques and camera imaging models were used to fit the contours, centroids, minimum bounding rectangle boxes, and three-dimensional bounding boxes of fruits and their accompanying branches. Then based on the positional relationship of fruits and their accompanying branches, three-dimensional pose estimation of pitaya was performed. A test set was obtained in pitaya plantations to evaluate the performance of this algorithm. Finally, field picking experiments were conducted in a natural orchard environment. The experimental results showed that the average intersection and union ratio (mIoU) and the mean pixel accuracy (mPA) of image segmentation for pitaya fruit reached 86.69% and 93.89%, respectively. The average error of three-dimensional pose estimation was 8.8°. The success rate of pitaya fruit picking robot in orchard environment was 86.7%, and the average picking time was 22.3 s. The research results indicated that this method can provide technical support for developing an intelligent pitaya picking robot to achieve

收稿日期: 2023-08-03 修回日期: 2023-09-05

基金项目: 广东省农业科技创新“揭榜挂帅”项目(2022SDZG03-5)和岭南现代农业实验室科研项目(NZ2021038)

作者简介: 朱立学(1970—),男,教授,博士生导师,主要从事智能农业机械装备研究,E-mail: zhulixue@zhku.edu.cn

automated and precise picking.

Key words: pitaya picking robot; image segmentation; pose estimation; attention residual U-Net; 3D boundary frame

0 引言

火龙果因具有较高的经济价值、营养价值和加工利用价值,近年来在中国种植面积不断增加^[1]。目前火龙果的采摘方式主要依赖于人工,存在劳动强度大和劳动力成本高等问题。同时,火龙果枝条带刺,极易刺伤采摘者,因此采取智能化机械采摘是必然趋势。由于火龙果生长的位置、角度等无规律,加上受枝条遮挡的影响,采摘条件复杂多变,采摘机器人不易在复杂的果园环境中对火龙果进行精确识别和定位。

目前,机器视觉检测技术广泛应用于水果检测任务中^[2-4]。邓子青等^[5]结合最大类间方差(Otsu)和形态学方法,对火龙果果实图像进行分割,并计算出果实质心,从而实现了果实识别与定位,但未对火龙果生长位姿进行识别。LI 等^[6]对绿苹果 YUV 颜色空间中显著区域的 V 分量进行拟合并结合阈值分割方法,能够有效检测自然环境中的绿苹果,但这种方法在不规则光照条件下的分割效果容易受到较大影响。LIU 等^[7]提出一种基于 Cr-Cb 坐标的多椭圆边界模型,用于检测柑橘类水果和树干,其在自然光环境下识别正确率达 90.8%。付根平等^[8]提出一种基于背景饱和度压缩与差异阈值分割融合的方法,用于识别香蕉串,该方法对自然香蕉园环境中不同光照条件下香蕉果串的识别具有较好的鲁棒性。罗陆锋等^[9]基于改进聚类图像分割和形态学方法对葡萄进行检测定位,定位准确率达 88.33%。近年来,基于深度学习的神经网络在水果检测定位中发挥着重要作用^[10-12]。WANG 等^[13]提出一种改进轻量级卷积神经网络 YOLO v4-LITE 方法对火龙果进行检测,平均精度达 96.48%,但其研究只局限于正、侧两个视角方向火龙果,对火龙果的姿态估计适用性不强。LI 等^[14]提出一种改进 MobileNet-YOLO 网络,火龙果检测平均精度达 97%。商枫楠等^[15]提出一种嵌入注意力机制的改进 YOLOX 网络,模型满足在自然环境下对成熟火龙果进行快速检测的需求,但上述研究均未进行火龙果生长姿态的定位研究。HOU 等^[16]基于改进的 YOLO v5s 对成熟柑橘进行检测识别,再结合形态学处理和相机模型估计果实的三维姿态。JORDI 等^[17]基于实例分割 Mask R-CNN 和运动重建(SFM)算法实现在三维空间上对苹果的检测和定位。麦春艳等^[18]提

出使用 RGB-D 相机对苹果果树进行三维重构,并对果实的点云信息进行分割,实现对果实的三维定位。LIN 等^[19]基于改进卷积神经网络(FCN)分割番石榴果实和枝条,并提出一种三维线段检测方法重建枝条分支,利用果实和最近分支的位置信息估计果实三维姿态。朱立学等^[20]提出基于 U-Net 的轻量化分割网络,识别分割平均准确率达 97.32%,该算法在自然环境下对香蕉果串具有良好的检测效果。于雪莹等^[21]提出一种混合注意力机制的残差网络,有效提高了苹果病害的识别准确率。LEHNERT 等^[22]结合颜色分割和聚类方法对甜椒进行检测分割,然后采用非线性最小二乘法拟合甜椒点云,实现对其姿态的估计。

总的来看,目前针对自然环境中火龙果目标检测与位姿估计的研究较少,未见自然果园环境下成熟的智能化采摘技术报道。在非结构化果园环境中,可靠性和鲁棒性强的火龙果目标检测算法对于研制采摘机器人至关重要,同时,果实姿态估计是引导采摘机器人末端执行器接近果实进行精准采摘的重要因素。因此,本文针对自然果园种植环境下火龙果目标识别定位算法精度要求高、姿态估计困难的问题,提出一种基于注意力残差结构的改进 U-Net 火龙果检测分割和姿态估计算法,利用改进 U-Net 算法分割出果实及其附生枝条的图像,再结合相机成像模型和果实及其附生枝条的几何特征信息来估计火龙果的三维生长位姿,以期为火龙果采摘机器人的研制提供技术基础。

1 基于注意力残差 U-Net 的火龙果检测分割

1.1 数据集准备

本研究分别于 2022 年 8 月中旬和 9 月中旬在广州市白云区和从化区火龙果种植园拍摄采集火龙果(包含果实和枝条,下同)图像。拍摄设备为 RealSense D455 相机,图像分辨率为 640 像素 × 480 像素,成像距离 50 ~ 150 cm,RGB 图和深度图保存格式分别为 JPEG 和 PNG,并将图像尺寸调整为 512 像素 × 512 像素。共采集整株多果和任意单果原始火龙果图像 2 000 幅,分为晴天顺光、晴天逆光以及阴天 3 种类型图像,部分图像如图 1 所示。为提高模型训练的鲁棒性,对采集到的原始图像进行数据筛选,得到 1 844 幅火龙果原始图像,并使用 Labelme 软件对图像进行手动标注,再将图像数据

集按 8:1:1 随机划分为训练集、测试集和验证集。



图 1 部分图像数据

Fig. 1 Part of image data

1.2 基于注意力残差的 U-Net 网络模型构建

本研究须从输入图像背景中分割出火龙果果实及其附着枝条的信息,但在进行卷积运算时,图像背景的迭代积累会形成大量特征冗余,为高效抑制背景干扰和无关枝条的影响以获取准确的火龙果果实及其附着枝条的分割图像,本研究在原 U-Net 网络基础上引入通道和空间注意力机制(SCSE)^[23]以及残差模块(DRB)^[24],提出一种基于注意力残差(SCSE-DRB, SD)的 U-Net 网络模型(SD-UNet)。其中 SCSE 通过对特征图进行空间和通道压缩以及通道激励,得到一个权重向量,并对原始特征图进行加权处理,使网络能够更准确提取火龙果的特征信息。DRB 有利于网络获取更广泛的上下文信息,学习更复杂的特征,并提高网络的收敛速度,防止网络出现梯度消失和过拟合现象。因此 SD-UNet 通过尽可能提取目标更多的细节和边缘信息来增强目标特征的表达并抑制冗余信息,从而提高网络图像分割能力。其详细结构如图 2 所示。SD-UNet 由下采样和上采样两部分构成。左侧下采样用于提取图像局部的特征信息,右侧上采样根据上下文信息对图像的特征进行精准定位。基于原 U-Net 模型,本文将 SCSE 模块应用到网络的跳跃连接部分,以便为下采样中的每个特征映射赋予权重,更好地利用低级特征的细节信息和高级特征的语义信息;除此之外,为了缓解网络退化的问题,且更加关注特征通道之间的关系来增强重要特征,本文将 SCSE 注意力机制模块集成到 DRB 残差结构中,并将其命名为注意力残差结构 SD,如图 3 所示。由图 3 可以看出,注意力残差结构 SD 引入批规范化(BN)^[25]算法,可加速网络训练收敛速度,缓解梯度消失和梯度爆炸问题,增强了模型稳定性。除此之外还添加 Dropblock 层^[26],通过在训练中随机丢弃一部分神经元的特征图块,减少特征图的冗余信息,更好地保持特征图的空间结构和上下文信息,防止网络过拟合,有效提高网络泛化能力和鲁棒性。

网络模型下采样包含 4 个模块,每个模块由 SD 结构和规格为 2×2 的最大池化层组成。上采样同样包含 4 个模块,每个模块包括规格为 2×2 的转置

卷积,并与下采样 SCSE 模块加权的特征映射连接,通过 SD 模块处理后,经过一层 1×1 卷积进行降维处理,得到火龙果图像分割图。

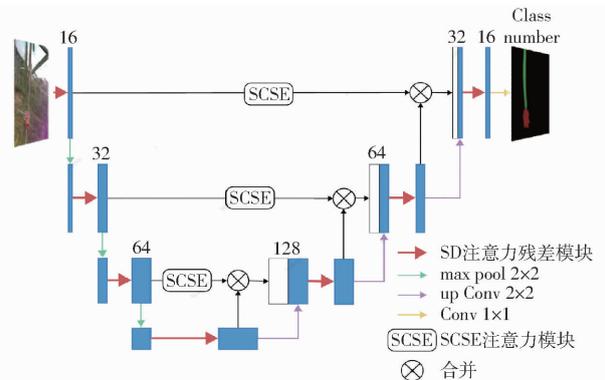


图 2 SD-UNet 结构图

Fig. 2 SD-UNet structure diagram

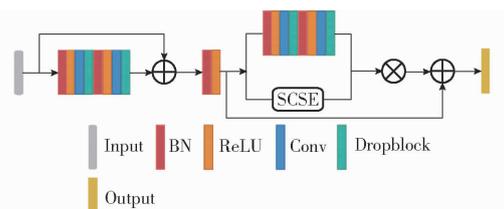


图 3 注意力残差模块 SD 结构图

Fig. 3 SD structure diagram of attention residual module

2 火龙果三维姿态估计方法

2.1 火龙果果实三维边界框提取

基于火龙果三维边界框的三维定位可确定果实空间位置和姿态信息,而果实三维框的获取需要将果实从三维空间映射到二维空间,通常通过小孔成像原理来计算实际三维空间点与实际成像平面间的关系。如图 4 所示,图像中每个像素是光圈与实际空间点连线与成像平面之间的交点,从而实现将三维空间点成像到二维图像。

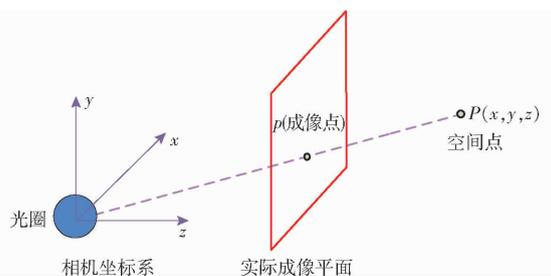


图 4 小孔成像模型

Fig. 4 Keyhole imaging model

在实际应用中,通过相机拍摄采集的图像,读取的数据是以像素点来计算的,通常以像素坐标系下的 $u'o'v$ 平面为基准,如图 5 所示。因此需要把像素坐标转换成对应二维空间的图像坐标,图像坐标系如图 6 所示。设每个像素尺寸为 dx, dy , 图像坐标系原点 o 对应的像素为 (u_o, v_o) , 通过给定任一具

体像素坐标 (u, v) , 可计算其图像坐标 (x, y) 。计算公式为

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} dx & 0 & -u_o dx \\ 0 & dy & -v_o dx \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (1)$$

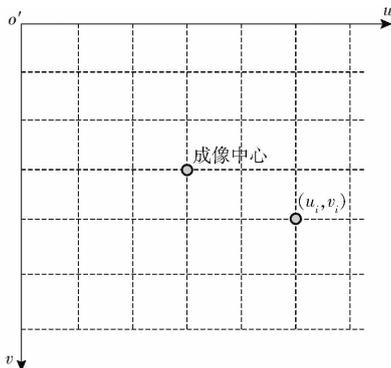


图 5 像素坐标系 $uo'v$

Fig. 5 $uo'v$ in pixel coordinate system

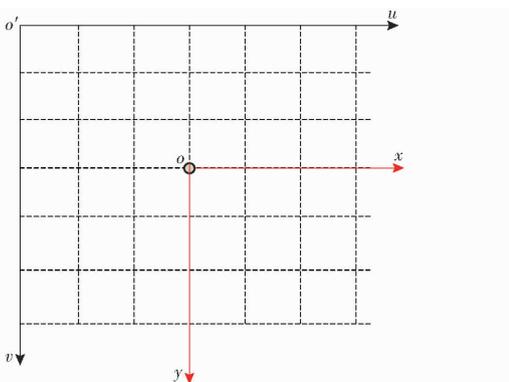


图 6 图像坐标系 xoy

Fig. 6 Image coordinate system xoy

进一步通过获取的空间深度信息可以实现二维图像到空间三维坐标的转换。根据相机的空间计算模型, 联立式(1)可得

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \begin{bmatrix} \frac{z_p}{f} & 0 & 0 \\ 0 & \frac{z_p}{f} & 0 \\ 0 & 0 & z_p \end{bmatrix} \begin{bmatrix} dx & 0 & -u_o dx \\ 0 & dy & -v_o dx \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2)$$

式中 f ——相机焦距

(x_p, y_p, z_p) ——点 P 的三维空间坐标

z_p ——实际空间点 P 在相机坐标系下的深度
从而计算出实际空间点在相机坐标系下的空间坐标。

将相机在 x, y 轴的焦距分别定义为 f_x, f_y , 化简式(2), 得像素坐标与空间三维坐标的关系为

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = z_p \begin{bmatrix} \frac{dx}{f_x} & 0 & \frac{-u_o dx}{f_x} \\ 0 & \frac{dy}{f_y} & \frac{-v_o dx}{f_y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (3)$$

实现图像中每个像素点与实际三维空间点的一一映射。

利用上述原理, 拟合出火龙果果实的三维边界框, 继而得到火龙果三维姿态信息。

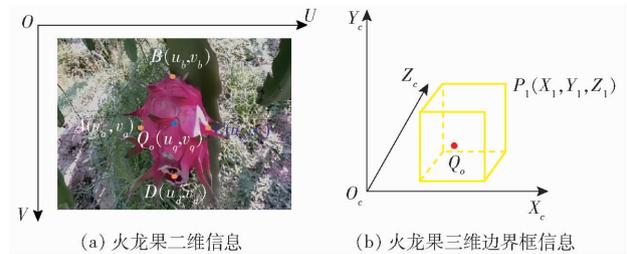


图 7 火龙果 2D 和 3D 位置信息示例

Fig. 7 Example of 2D and 3D location information of pitaya

如图 7a 所示, A, B, C 和 D 为火龙果果实最左边、最上面、最右边和最下面的端点, Q_o 为果实表面质心点, 其像素坐标为 (u_a, v_a) 、 (u_b, v_b) 、 (u_c, v_c) 、 (u_d, v_d) 和 (u_q, v_q) 。根据式(3)可得点 A, B, C, D 和 Q_o 的三维空间坐标为 (X_a, Y_a, Z_a) 、 (X_b, Y_b, Z_b) 、 (X_c, Y_c, Z_c) 、 (X_d, Y_d, Z_d) 和 (X_q, Y_q, Z_q) 。在 3 维透视投影中, 设 dx_c, dy_c 和 dz_c 分别为在 X_c, Y_c, Z_c 轴上的果实直径, 因火龙果果实横切面近似圆形, 将 dz_c 近似于 dx_c , 计算公式为

$$\begin{cases} dx_c = x_c - x_a \\ dy_c = y_d - y_b \\ dz_c = x_c - x_a \end{cases} \quad (4)$$

检测到的目标位置和姿态信息通常可以通过 3 维边界框来确定。设 P_1, P_2, \dots, P_8 是火龙果三维边界框的 8 个顶点, 坐标为 (X_i, Y_i, Z_i) ($i = 1, 2, \dots, 8$)。其坐标可以由从 P_i 到 Q_o 的相对几何位置获得, 如点 P_1 的三维坐标为

$$\begin{cases} X_1 = X_q + dx_c/2 \\ Y_1 = Y_q + dy_c/2 \\ Z_1 = Z_q + dz_c \end{cases} \quad (5)$$

2.2 火龙果果实生长姿态估计

为了获取火龙果生长的空间姿态, 在上述试验分割的图像中保留了火龙果及其附着生长的枝条, 剔除其他无果枝条和噪声的影响, 形成掩膜图像。本文基于连通组件标记算法 (Connected component labeling, CCL), 结合火龙果二值图像对掩膜图像进行连通区域分析, 即通过遍历二值图像中的像素, 将像素值相同的相邻像素标记为同一个连通区域。在遍历完所有像素后, 设定最小连通组件面积的阈值, 得到只有火龙果及其附着生长枝条的掩膜图像。将掩膜图像采取侵蚀、膨胀和孔洞填充的数学形态学操作来填充检测区域之间的间隙, 去除噪声并平滑区域边界, 使果实轮廓更加明显, 处理后得到的火龙

果掩膜图如图 8b 所示。

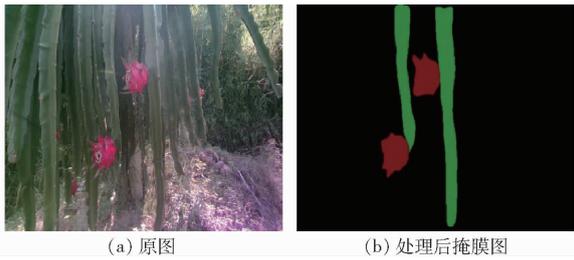


图 8 火龙果原图及掩膜图

Fig. 8 Original image and mask image of pitaya

在火龙果生长姿态估计时,先通过 RGB 颜色空间对上述掩膜图像分别进行绿色通道和红色通道图像的分离,得到火龙果枝条和果实的灰度图像,并对灰度图像进行高斯模糊,以减少噪声的影响,图像如图 9a、9b 所示;随后选取合适的阈值将灰度图转化成二值化图像,使目标区域像素值为 255,背景区域像素值为 0,如图 9c、9d 所示。

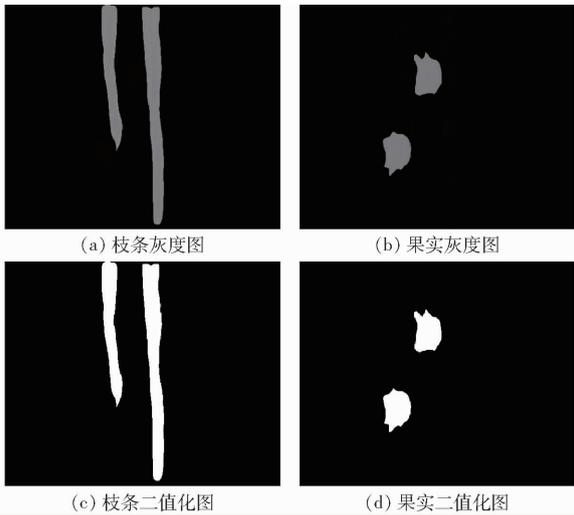


图 9 火龙果灰度图及二值化图

Fig. 9 Gray scale and binarization of pitaya

使用一种基于深度优先搜索的图像扫描算法来提取火龙果及其附着生长枝条轮廓。该算法从图像的左上角像素开始扫描,如果某个像素点是轮廓上的点,则将该点及其周围的像素点标记为已访问,直至扫描完整个轮廓,并将轮廓表示为一个向量,包含轮廓的所有像素点坐标。本文将火龙果及其附着生长枝条的交接点定义为果实生长点,则可利用果实质心与果实生长点的位置关系估计果实姿态,因此需要获取果实的质心坐标及果实生长点坐标。根据图像质心矩定义,火龙果质心坐标计算公式为

$$\begin{cases} x_c = \frac{\sum xf(x,y)}{\sum f(x,y)} \\ y_c = \frac{\sum yf(x,y)}{\sum f(x,y)} \end{cases} \quad (6)$$

式中 (x_c, y_c) ——质心坐标

(x, y) ——像素坐标

$f(x, y)$ ——二值化图像在点 (x, y) 处像素值

为了得到果实生长点坐标,遍历枝条轮廓上的所有点,并通过计算与果实质心点的距离,得到最小距离的轮廓点即为果实生长点。质心点距离 d 计算公式为

$$d = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad (7)$$

式中 (x_i, y_i) ——轮廓上的点坐标

火龙果质心与果实生长点构成的向量为

$$\mathbf{q} = (x_c - x_i, y_c - y_i) \quad (8)$$

即可获取火龙果果实的姿态 q_1 ,如图 10 所示。

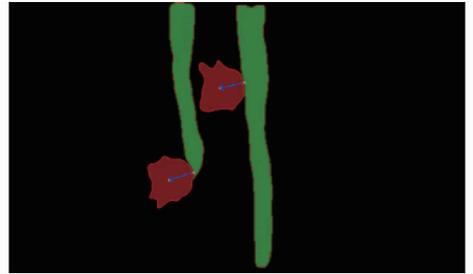


图 10 基于果实质心的火龙果估计姿态

Fig. 10 Pose estimation of pitaya based on fruit centroid

通过遍历轮廓上所有的点,并比较每个点的 x 和 y 坐标,找到最左、最上、最右和最下的点,即可得到轮廓的最小外接矩形,如图 11a 所示。将矩形各边的中点定义为火龙果果实轮廓的 4 个端点,结合相机成像原理,通过式(5)即可得到火龙果的三维边界框,如图 11b 所示。

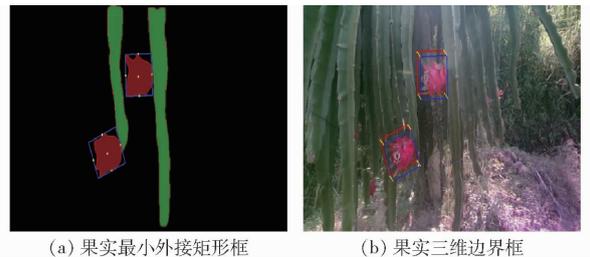


图 11 果实最小外接矩形框及三维边界框

Fig. 11 Fruit minimum external rectangular frame and 3D boundary frame

为了得到更准确的火龙果生长姿态,将果实三维边界框与上述果实估计姿态进行融合。由果实三维边界框的位置信息可得到三维边界框 6 个面的中心点坐标,计算果实生长点和三维边界框各个面中心点的距离,将距离最大的点与果实生长点构成的向量定义为果实姿态 q_2 。单个火龙果姿态估计结果如图 12a 所示。

再结合质心和三维边界框的姿态估计方法,即融合所得 2 个向量 q_1 、 q_2 的方向来定义果实最终估计姿态。本文计算三维边界框内 6 个面中点与果实



图 12 火龙果估计姿态

Fig. 12 Pitaya pose estimation

生长点之间的距离,求距离最大中心点与果实质心的中点坐标,并与果实生长点构成一个向量 q_s ,将该向量定义为火龙果最终估计姿态,如图 12b 所示。

3 结果分析

3.1 试验运行平台及评价指标

基于 PyTorch 的深度学习框架进行网络模型训练,主要依赖库为 OpenCV、Numpy、matplotlib 等。所提算法在 Ubuntu 18.04 操作系统运行,CPU 为 Intel Xeon E5 - 2620 V4 @ 2.10 GHz, RAM 为 64 GB, GPU 为 NVIDIA GeForce RTX 2080, CUDA 和 cuDNN 版本分别为 10.1 和 7.6.5。

试验使用平均交并比 (mIoU) 和平均像素准确率 (mPA) 作为衡量火龙果分割模型性能的评价指标,使用估计生长姿态与实际生长姿态(人为标注)之间的角度误差作为衡量火龙果姿态估计效果的评价指标。

3.2 火龙果图像分割试验

在自建测试集图像上进行验证和测试,计算各网络模型对火龙果图像识别结果的平均交并比和平均像素准确率,并作为各网络模型的评价指标。为了验证本文添加注意力机制 SCSE 和残差结构 DRB 的有效性,在测试集上进行消融试验,试验结果如表 1 所示。

表 1 消融试验结果

Tab. 1 Ablation experiment results %

模型	果实平均交并比	果实平均像素准确率	生长枝条平均交并比	生长枝条平均像素准确率
U-Net	85.66	92.51	60.56	69.35
U-Net + SCSE	85.78	92.76	62.75	69.96
U-Net + DRB	85.99	92.97	63.82	70.99
U-Net + DRB + SCSE	86.11	93.53	65.33	71.45
SD-UNet	86.69	93.89	66.56	72.39

从表 1 可以看出,在对 U-Net 改进后,各模型的性能均得到相应提升。其中将 SCSE 模块应用到 U-Net 的跳跃连接和 DRB 残差模块后,有效增强

了火龙果细节特征、抑制无关背景干扰,2 个指标 mIoU 和 mPA 均有明显提升,提高了模型对果实及其附着生长枝条的分割性能,也证实了所提算法的有效性。

将本文方法与 U-Net、FCN 和 PSPNet 网络在测试集图像上进行比较,由表 2 可以看出,对于火龙果及其附着生长枝条的图像检测分割,本文提出的 SD-UNet 模型性能明显优于其他 3 种网络模型,其中果实图像分割的平均交并比为 86.69%,平均像素准确率为 93.89%;附着生长枝条图像分割的平均交并比为 66.56%,平均像素准确率为 72.39%。从图 13 可以看出,虽然几种模型对火龙果果实图像都有较好的分割效果,但由于火龙果附着生长枝条的图像分割难度较大,采用 FCN 和 PSPNet 时容易出现误分割和图像掩膜缺失的问题,枝条轮廓不够平滑,而采用 SD-UNet 的图像分割效果均优于其他 3 种模型。

表 2 各网络模型评价指标

Tab. 2 Evaluation indexes of each network model %

网络模型	果实平均交并比	果实平均像素准确率	生长枝条平均交并比	生长枝条平均像素准确率
U-Net	85.66	92.51	60.56	69.35
FCN	84.15	90.36	58.15	68.36
PSPNet	82.12	89.99	54.12	62.99
SD-UNet	86.69	93.89	66.56	72.39

3.3 火龙果生长姿态估计试验

首先将本文算法在测试集上对火龙果进行生长姿态估计试验,以测试算法的稳定性,计算其估计生长姿态与实际生长姿态(人工标注)之间的角度来衡量火龙果姿态估计误差,角度越小表明误差越小,角度越大表明误差越大。估计生长姿态与实际生长姿态间的角度 θ_i 为

$$\theta_i = \arccos \frac{\tilde{q}_i q_i}{|\tilde{q}_i| |q_i|} \quad (9)$$

式中 \tilde{q}_i ——火龙果估计姿态

q_i ——火龙果实际生长姿态

现场实测部分火龙果的生长姿态估计结果如图 14 所示,图中黄点为果实生长点,红色箭头方向为算法估计的姿态,黑色箭头方向为真实姿态。可以看出,本研究算法的角度误差较小,有较好的姿态估计效果。同时,本研究也出现了部分姿态估计失败,如图 15 所示。姿态估计误差较大的原因主要有:①当火龙果枝条生长较为茂盛时,果实生长枝条的图像分割容易误分割和漏分割,导致果实与其附

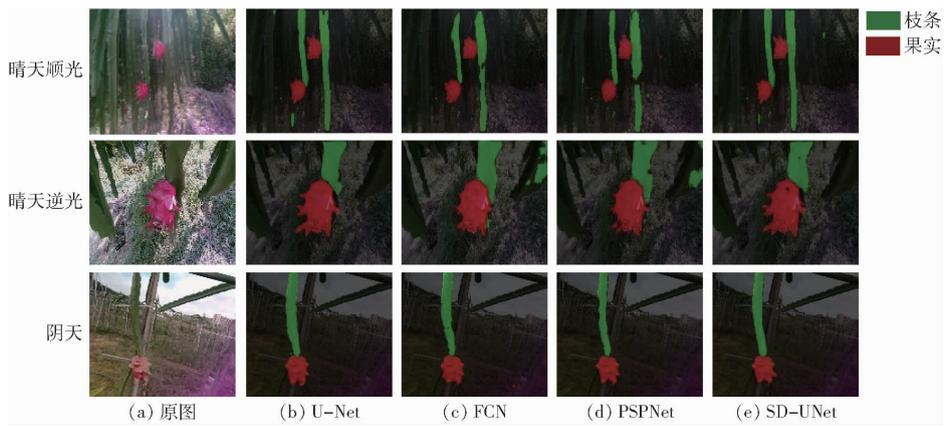


图 13 不同网络的图像分割效果

Fig. 13 Semantic segmentation results of different neural networks



图 14 火龙果姿态估计向量图

Fig. 14 Pose estimation vector diagram of pitaya



图 15 火龙果姿态估计失败示例

Fig. 15 Pitaya pose estimation failure samples

生枝条的位置关系不易确定。②本文算法将火龙果及其附着生长枝条的交接点定义为果实生长点,并将其定义为果实姿态估计向量的起始点,该点虽是生长枝条轮廓上离果实质心最近的点,却有可能不是最佳姿态估计向量的起始点。

角度误差累积分布比例如图 16 所示,角度误差在 10° 以内的比例为 66.6%,角度误差在 20° 以内的比例为 90%,所有角度误差平均值为 8.8° ,标准差为 7.2° 。

3.4 实地采摘试验

采摘试验在广州市白云区仲恺农业工程学院种植第 2 年的火龙果园进行,枝条生长较为稀疏。本文将基于改进 U-Net 的火龙果图像分割和姿态估

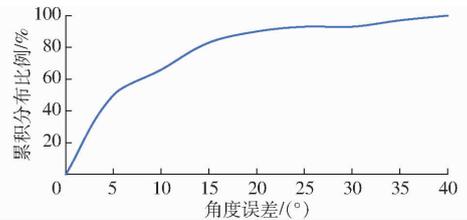


图 16 角度误差累积分布比例

Fig. 16 Cumulative distribution ratio of angular errors

计算法部署在边缘计算机 T600 上,并结合深度相机 RealSense D455、机械臂、末端执行器和履带式移动底盘构成火龙果采摘机器人,如图 17 所示。



图 17 火龙果采摘机器人样机试验

Fig. 17 Schematic of pitaya picking robot

火龙果采摘机器人工作流程如图 18 所示,机器人上电初始化后在火龙果园沿种植行开始移动,当检测到成熟红色的火龙果完整出现在机器人相机设定视野范围内时,机器人停止移动并开始解算果实位姿,再将位姿信息发送至机器人控制模块,经过路径规划后发送控制指令,控制机械臂的运动,完成火龙果采摘。若出现枝条严重遮挡果实情况,暂不设置机器人采摘,后续将做避障采摘研究。

火龙果采摘机器人实地采摘过程如图 19 所示,机械臂根据火龙果的位姿信息到达目标位置后,末端执行器夹住果实并旋转摘下。同时,本次试验也出现了部分果实采摘失败的例子,如图 20 所示。采摘失败的主要原因是火龙果姿态估计角度出现误差,末端机械爪发生弹性形变夹持不稳。采摘试验结果为:火龙果总数为 30,成功识别定位个数为 28

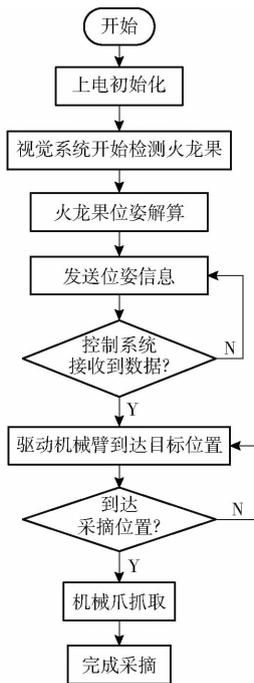


图 18 机器人工作流程图

Fig. 18 Robot workflow diagram

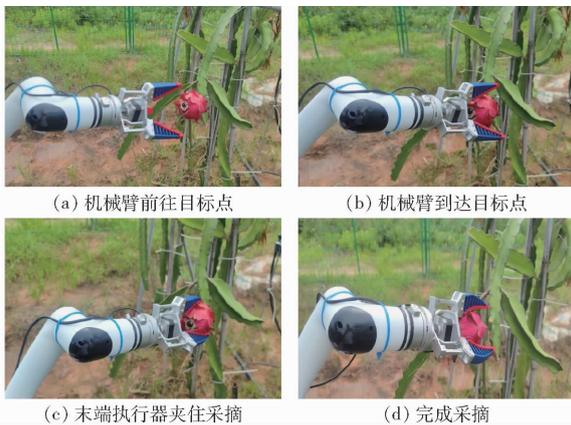


图 19 采摘试验过程示意图

Fig. 19 Process schematic of picking experiment

个, 识别定位成功率为 93.3%, 姿态估计角度误差平均值为 8.2° , 成功采摘 26 个, 采摘成功率为

86.7%, 平均采摘时间为 22.3 s, 表明所研制的火龙果采摘机器人基本能够完成智能化采摘作业。



图 20 采摘失败示意图

Fig. 20 Schematics of picking failure

4 结论

(1) 所建立的 SD-UNet 网络模型实现了火龙果及其附着生长枝条的目标检测和图像分割, 其中果实图像分割平均交并比为 86.69%, 平均像素准确率为 93.89%; 附着生长枝条图像分割平均交并比为 66.56%, 平均像素准确率为 72.39%。结果表明, 采用此模型对火龙果进行目标检测可以得到较好的图像分割效果。

(2) 基于 SD-UNet 网络模型的图像分割方法采用连通组件标记算法和形态学处理方法得到包含火龙果及其附着生长枝条的分割图, 再基于深度优先搜索的图像扫描算法以提取其轮廓, 由轮廓信息可得果实质心、果实生长点和果实最小外接矩形框, 再结合相机成像模型, 可以获得火龙果果实的三维边界框。基于果实三维边界框位置并结合果实质心与果实生长点的位置关系, 可以得到火龙果的估计姿态, 结果与其真实生长姿态角度平均误差为 8.8° , 标准差为 7.2° , 表明该方法可以用于火龙果果实姿态估计。

(3) 基于本文算法研制的火龙果采摘机器人可以根据控制指令准确到达目标位置进行采摘, 在自然果园环境下采摘成功率为 86.7%, 平均采摘时间为 22.3 s, 能够满足火龙果机械化采摘要求。

参 考 文 献

- [1] 王举兵. 广西火龙果产业高质量发展的思考[J]. 南方园艺, 2022, 33(4): 59-62.
WANG Jubing. Consideration on the high quality development of Guangxi Pitaya industry[J]. Southern Horticulture, 2022, 33(4): 59-62. (in Chinese)
- [2] 曾曦, 胡桂兵, 秦永华. 广东省火龙果产业发展现状及对策[J]. 中国果业信息, 2019, 36(9): 9-12.
ZENG Xi, HU Guibing, QIN Yonghua. Development status and countermeasures of dragon fruit industry in Guangdong Province [J]. China Fruit Industry Information, 2019, 36(9): 9-12. (in Chinese)
- [3] 赵玉山. 我国火龙果产量跃居全球第一[J]. 中国果业信息, 2023, 40(2): 52.
ZHAO Yushan. Chinese Pitaya production ranks the first in the world[J]. China Fruit Industry Information, 2023, 40(2): 52. (in Chinese)
- [4] 马聪, 陈学冬. 机器视觉技术在农业领域的应用及发展[J]. 宁夏农林科技, 2021, 62(7): 65-70.
MA Cong, CHEN Xuedong. Application and development of machine vision technology in agricultural field [J]. Ningxia Agriculture and Forestry Science and Technology, 2021, 62(7): 65-70. (in Chinese)
- [5] 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)

- [6] LI B, LONG Y, SONG H. Detection of green apples in natural scenes based on saliency theory and Gaussian curve fitting[J]. *International Journal of Agricultural and Biological Engineering*, 2018,11(1):192-198.
- [7] LIU Tianhu, REZA E, ARASH T, et al. Detection of citrus fruit and tree trunks in natural environments using a multi-elliptical boundary model[J]. *Computers in Industry*,2018,99:9-16.
- [8] 付根平,朱立学,张世昂,等. 基于背景饱和度压缩与差异阈值分割融合的香蕉识别[J]. *中国农机化学报*, 2021,42(6):151-158.
FU Genping, ZHU Lixue, ZHANG Shiang, et al. Bunch of bananas recognition based on background saturation compression and iterative fusion with different threshold segmentation[J]. *Journal of Chinese Agricultural Mechanization*, 2021, 42(6):151-158. (in Chinese)
- [9] 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 location of grape picking point by grape picking robot in natural environment[J]. *Transactions of the CSAE*, 2015, 31(2):14-21. (in Chinese)
- [10] 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)
- [11] 郑太雄,江明哲,冯明驰,等. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. *仪器仪表学报*, 2021,42(9):28-51.
ZHENG Taixiong, JIANG Mingzhe, FENG Mingchi, et al. A review of target recognition and localization methods for picking robots based on vision[J]. *Chinese Journal of Scientific Instrument*, 2021, 42(9):28-51. (in Chinese)
- [12] 景庄伟,管海燕,彭代峰,等. 基于深度神经网络的图像语义分割研究综述[J]. *计算机工程*, 2020,46(10):1-17.
JING Zhuangwei, GUAN Haiyan, PENG Daifeng, et al. A review of image semantic segmentation based on deep neural networks[J]. *Computer Engineering*, 2020, 46(10):1-17. (in Chinese)
- [13] WANG Jinpeng, GAO Kai, JIANG Hongzhe, et al. Method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. *Transactions of the CSAE*, 2020, 36(20):218-225.
- [14] LI X, QIN Y, WANG F, et al. Pitaya detection in or-chards using the MobileNet-YOLO model[C]//Technical Committee on Control Theory. Chinese Association of Automation, 2020.
- [15] 商枫楠,周学成,梁英凯,等. 基于改进YOLOX的自然环境中火龙果检测方法[J]. *智慧农业*,2022,4(3):120-131.
SHANG Fengnan, ZHOU Xuecheng, LIANG Yingkai, et al. Detection method of Pitaya fruit in natural environment based on improved YOLOX[J]. *Smart Agriculture*, 2022, 4(3):120-131. (in Chinese)
- [16] HOU C, ZHANG X, TANG Y, et al. Detection and localization of citrus fruit based on improved You Only Look Once v5s and binocular vision in the orchard[J]. *Frontiers in Plant Science*, 2022,13:972445.
- [17] JORDI G M, RICARDO S C, JOAN R P, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry[J]. *Computers and Electronics in Agriculture*, 2020, 169:105165.
- [18] 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(增刊):35-40.
MAI Chunyan, ZHENG Lihua, SUN Hong, et al. Fruit tree 3D reconstruction and fruit recognition and localization based on RGB-D camera[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015,46(Supp.):35-40. (in Chinese)
- [19] LIN G, TANG Y, ZOU X, et al. Guava detection and pose estimation using a low-cost RGB-D sensor in the field[J]. *Sensors*, 2019,19:428.
- [20] 朱立学,伍荣达,付根平,等. 基于多尺度串联空洞卷积的轻量化UNet香蕉图像分割[J]. *农业工程学报*, 2022,38(13):194-201
ZHU Lixue, WU Rongda, FU Genping, et al. Image segmentation of lightweight UNet banana based on multi-scale series cavity convolution[J]. *Transactions of the CSAE*, 2022, 38(13):194-201. (in Chinese)
- [21] 于雪莹,高继勇,王首程,等. 基于生成对抗网络和混合注意力机制残差网络的苹果病害识别[J]. *中国农机化学报*, 2022,43(6):166-174.
YU Xueying, GAO Jiyong, WANG Shoucheng, et al. Apple disease recognition based on generative adversarial network and hybrid attention mechanism residual network[J]. *Journal of Chinese Agricultural Mechanization*, 2022,43(6):166-174. (in Chinese)
- [22] LEHNERT C, SA I, MCCOOL C, et al. Sweet pepper pose detection and grasping for automated crop harvesting[C]//Proceedings of the IEEE International Conference on Robotics & Automation, 2016:2428-2434.
- [23] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel squeeze & excitation in fully convolutional networks[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2018. Cham:Springer International Publishing, 2018:421-429.
- [24] GUO C, SZEMENYEI M, YI Y, et al. Dense residual network for retinal vessel segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. Springer,Cham,2020:1374-1378.
- [25] KAUR R, GHOLAMHOSSEINI H, SINHA R, et al. Melanoma classification using a novel deep convolutional neural network with dermoscopic images[J]. *Sensors*, 2022, 22(3):1-15.
- [26] WANG H L, YANG J D. FBUNet: full convolutional network based on fusion block architecture for biomedical image segmentation[J]. *Journal of Medical and Biological Engineering*, 2021, 41(2):185-202.