

吴烽云, 叶雅欣, 陈思宇, 等. 复杂环境下香蕉多目标特征快速识别研究 [J]. 华南农业大学学报, 2022, 43(2): 96-104.
WU Fengyun, YE Yaxin, CHEN Siyu, et al. Research on fast recognition of banana multi-target features by visual robot in complex environment[J]. Journal of South China Agricultural University, 2022, 43(2): 96-104.

复杂环境下香蕉多目标特征快速识别研究

吴烽云, 叶雅欣, 陈思宇, 艾璞晔, 邹湘军, 段洁利

(华南农业大学 工程学院, 广东 广州 510642)

摘要:【目的】针对野外环境下断蕾机器人对多特征的变量目标快速识别难题, 以及目标受到树叶、遮挡及光照影响精度的问题, 提出多特征目标的快速识别方法。【方法】提出对香蕉果实、果轴和花蕾这 3 个目标进行多尺度特征提取及模型分类, 融合聚类算法设计新的目标候选框参数, 提出改进 YOLOv3 模型及网络结构参数的 YOLO-Banana 模型; 为了平衡速度和准确度, 用 YOLO-Banana 和 Faster R-CNN 分别对变化尺寸的香蕉多目标进行试验, 研究算法对识别精度与速度的影响。【结果】YOLO-Banana 和 Faster R-CNN 这 2 种算法识别香蕉、花蕾和果轴的总平均精度分别为 91.03% 和 95.16%, 平均每张图像识别所需时间分别为 0.237 和 0.434 s。2 种算法精度均高于 90%, YOLO-Banana 的速度相对快 1.83 倍, 更符合实时作业的需求。【结论】野外蕉园环境下, 采用 YOLO-Banana 模型进行香蕉多目标识别, 可满足断蕾机器人视觉识别的速度及精度要求。

关键词: 深度学习; 花蕾; 果轴; 多特征; 多目标识别; 快速分类

中图分类号: TP391.41; S667.1; S225.93

文献标志码: A

文章编号: 1001-411X(2022)02-0096-

09

Research on fast recognition of banana multi-target features by visual robot in complex environment

WU Fengyun, YE Yaxin, CHEN Siyu, AI Puye, ZOU Xiangjun, DUAN Jieli

(College of Engineering, South China Agricultural University, Guangzhou 510642, China)

Abstract: 【Objective】Aiming at fast recognition of multi-feature variable target by robot vision in field environment, and considering the problem that the accuracy of target is affected by leaves, shade and light, a fast recognition method of multi-feature target was proposed. 【Method】A multi-scale feature extraction and classification model was proposed for three targets including banana fruit, fruit axis and flower bud. New target candidate box parameters were designed using clustering algorithm. The network structure parameters of YOLOv3 model were optimized and the YOLO-Banana model was proposed. In order to balance the speed and accuracy, YOLO-Banana and Faster R-CNN were used to conduct experiments on banana multi-targets with varying sizes. The effects of algorithms on recognition accuracy and speed were analyzed. 【Result】The average accuracies of YOLO-Banana and Faster R-CNN algorithms on banana fruit, fruit axis and flower bud were 91.03% and 95.16% respectively, average recognition time per photo was 0.237 and 0.434 s respectively.

收稿日期: 2021-08-10 网络首发时间: 2022-01-06 18:54:59

网络首发地址: <https://kns.cnki.net/kcms/detail/44.1110.S.20220106.1230.002.html>

作者简介: 吴烽云, 博士研究生, 主要从事农业智能采摘机器人及智能决策研究, E-mail: fyseagull@163.com; 通信作者: 段洁利, 副教授, 博士, 主要从事农业智能装备研究, E-mail: duanjieli@scau.edu.cn

基金项目: 国家重点研发计划 (2020YFD1000104); 广东省重点领域研发计划 (2019B020223002); 财政部和农业农村部: 现代农业产业技术体系建设专项 (CARS-31-10); 广东省基础与应用基础研究 (2020A1515010793)

Therefore the accuracies of two algorithms were both above 90%. YOLO-Banana had relatively 1.83 times faster speed than Faster R-CNN, better satisfying the requirement of real-time operation. 【Conclusion】 In the wild environment, utilization of YOLO-Banana model for banana multi-target recognition can meet the speed and accuracy requirements for visual recognition by the bud-breaking robots.

Key words: Deep learning; Flower bud; Fruit axis; Multi-feature; Multi-target recognition; Fast classification

野外环境下断蕾机器人需对多特征多目标进行识别, 特别是农业机器人在切断香蕉花蕾和估产时, 要同时识别香蕉果串、果轴和花蕾 3 种不同特征的目标, 其单一目标的大小是一种变量, 每次检测的目标尺寸和位置都不一样。另外, 目标还受到树叶、遮挡及光照影响, 研究复杂环境下的视觉检测方法意义重大。因此, 针对香蕉、花蕾及果轴的识别问题, 本研究提出多特征目标的快速识别方法。

断蕾机器人的关键视觉技术是对目标进行自动定位, 视觉检测一直被认为是最具挑战性的问题之一^[1-6]。Tang 等^[7]用多目视觉跟踪钢管的动态变形, 探索其破坏机理; Dai 等^[8]把深度学习算法用于交通管理; 研究还将视觉技术应用于水果采摘机器人识别^[9-13]、估产^[14]和病虫害检测^[15]等。

在植物目标的视觉定位检测中, 常用基于颜色、阈值分割、边缘检测等经典的图像处理算法^[16]。罗陆锋等^[17]结合颜色和聚类方法识别葡萄, 精确度达 87.63%。Hočevár 等^[18]基于 HSL(色调、饱和度、亮度)识别估算苹果花。然而, 颜色与阈值等识别方法受光照变化、枝叶或遮挡等因素影响。与经典模型和手工特征提取方法相比, 深度卷积神经网络在识别方面取得了较好的性能^[11], 在野外环境中检测目标具有优势。深度学习用于香蕉及果轴的多特征图像检测, 可让模型自我学习目标特征, 实现自动分类识别。

深度学习算法主要有 2 种类型的对象检测器, 第 1 类为两阶段目标检测算法, 它将检测分为 2 个阶段。以香蕉为例, 第 1 阶段使含有香蕉的候选区域 (Region proposal networks, RPN) 生成 1 个包含待检感兴趣的候选框, 选定后再作后续处理; 第 2 阶段是把所感兴趣的香蕉区域提议发送到通道中, 进行分类和边界框回归, 用卷积神经网络 (Convolutional neural network, CNN) 完成对候选框中候选香蕉位置与类别的预测, 例如 R-CNN、Fast R-CNN、Faster R-CNN 或 Mask R-CNN。CNN 是一种前馈神经网络, 其卷积计算具有很深的结构, Mask R-CNN 增加了图像分割功能, 准确率较高, 但速度较慢^[19]。第 2 类为单阶段检测器, 省去了第 1 阶段的 RPN, 直接用卷积神经网络模型预测目标

边界, 通过获取输入图像将对象检测视为回归问题, 同时学习边界框坐标和类标签概率。与 2 阶段的探测器相比, 该算法速度较快, 但准确率稍低。常见的单阶段目标检测算法有 SSD (Single shot multibox detector)、YOLO (You only look once) 及其改进的系列算法 YOLOv2、YOLOv3 等^[1, 20]。

对于第 1 类方法的研究, Alexe 等^[21]提出了一种线索综合性的方法来达到更好、更有效的预测效果。Lin 等^[22]提出一种具有重叠影响的草莓花识别系统, 试验表明 R-CNN、Fast R-CNN、Faster R-CNN 的目标检出率分别约为 63%、77% 和 86%。Tian 等^[23]提出改进 Mask-R-CNN 的图像分割算法, 用于苹果花的多特征检测, 准确率为 96%。对于第 2 类方法的研究, Koirala 等^[24]用 YOLO 和 R2CNN 检测芒果花穗, 认为 YOLOv3 模型用于计数时较优, R2CNN 模型用于分类时较准确。陈燕等^[25]改进了 YOLOv3 算法, 用于检测荔枝串, 平均检测精度为 94%, 结果可用于采摘机器人的预定位。

视觉检测及深度学习算法主要用于香蕉果实 (果串) 及树干 (假径)、荔枝果串等单一目标检测^[26-28]。但是机器人作业时需了解香蕉果实、花蕾及果轴情况, 为切下花蕾提供定位信息。不同香蕉果树的香蕉果实、花蕾和果轴大小不等, 且部分目标被遮挡, 增加了检测复杂度, 多特征目标的视觉精确识别方法仍然有待完善。

多种算法在检测精度和速度方面各有特点, 目标图像分类识别的结果最终应用于视觉机器人作业控制。因此, 研究者需要在目标检测的准确性和速度之间取得平衡^[29]。机器视觉对香蕉多目标的快速精确识别成为关键问题。为了能够实现对果轴的精确切割, 在图像采集设备将待采摘区域的图像数据输入视觉识别系统后, 视觉识别系统应迅速查找画面中的果实、果轴和花蕾, 使回归框的各边界准确地与目标边缘相切, 并将回归框区域中的画面提取出来, 以便进行进一步形态学处理和计算。对此, 本文提出一种探索多特征多目标检测方法, 优选模型参数, 融合聚类算法优化 YOLOv3 模型的网络结构参数, 并建立损失函数模型, 形成 YOLO-Banana

模型。分别选用 Faster R-CNN 和 YOLO-Banana 进行多目标分类,对香蕉、花蕾及果轴进行试验验证算法的有效性,为复杂场景多目标多特征精确快速识别提供支持。

1 试验设备和数据采集

1.1 试验设备

试验设备由硬件设备与软件组成。图像处理使用的计算机主要配置为:i7-7700K 处理器,内存为 16 G、2400 MHz;显卡为 GTX1080Ti、11G。样本采集时,用普通相机和高像素的手机,图像分辨率是 1600 万像素。

1.2 图像与数据采集

香蕉树通常只有香蕉果实串、花蕾及连接花蕾与香蕉的花序轴,简称果轴,花蕾及果轴朝下。为了提高产量,当香蕉果实串生长到一定时期,需要从果轴部位切下花蕾,称断蕾。本次采集样品包括香蕉果实、花蕾和果轴 (图 1),以便为机器人断蕾提供定位信息。



图 1 顺、逆光下的香蕉果实、花蕾和果轴
Fig. 1 Banana fruit, flower bud and fruit axis target detection under frontlight and backlight

野外机器人作业时无法通过增加标记或靶标等辅助手段来提高精度^[3,7],而是采用无标记的目标定位。为了适应实际作业场景,视觉检测时不在目标上做定位靶标。在采样时,我们的策略是,只要有香蕉和花蕾就采样,以便机器人学习判别香蕉开花,做出决策。

香蕉多目标样本试验图像采集拍摄分 2 批进行。2020 年 7 月 25 日,晴天,拍摄地点为广东省广州市水果世界岭南鲜果基地。2020 年 8 月 4、5 日,晴天转阴天,拍摄地点为广东江门果园。采集图像时,包括香蕉果实、果轴、花蕾。图像采集过程中,将相机与花蕾间的距离保持在 600 mm 左右,在不同背景和光照条件下共采集 1 285 幅图像。通过 Labelimg 标注软件对图片数据进行批量标注,通过

编写脚本按比例自动随机选取训练样本集合、验证样本集合以及测试样本集,其中训练集 1 029 张,占总样本 80%;验证集 64 张,占总样本 5%;测试集 192 张,占总样本 15%。

2 算法描述

2.1 图像的多目标分类

单个采样点会稍微变动拍摄角度来采集多幅图像,因此图片按照默认序列进行训练与检测时可能会出现测试集、验证集等均为少数采样点的若干相似图像的情况,使得训练、检测效果与样本本身有较强的关联性。为了避免这种关联性、减少随机性,在筛去无效图像以后,我们将剩余的图片随机打乱顺序,并按照 1 个标准格式进行命名。命名完成后,采用 Labelimg 软件对图像中的目标进行手工标注,并生成 1 个和图像同名的标签数据文档。图像数据与标签数据将按照 VOC2007 数据集格式进行分组与整理,按照 6:2:2 的比例划分为训练集、验证集和测试集。采集样图的像素为 721×960,在进行图像处理需要把它缩放至像素为 416×416。通过输入一张任意大小图片,缩放至像素为 416×416,作为网络的输入,即网络的输入是 1 张像素为 416×416、3 通道的 RGB 图。对于比例不为 1:1 的图片,通过在图片的上、下或者左、右分别添加 2 块区域,并将补上的区域以全黑 (0, 0, 0) 或全白 (255, 255, 255) 像素填充,将其补齐为一张 1:1 的 3 通道图片,由于补齐的区域不包含任何与待检测目标相似的内容,所以对检测结果没有影响。补齐以后通过升或降采样的方式将图片统一为 416×416×3 的图片,即可输入网络。

2.2 多尺度特征融合

采用多尺度对香蕉多目标进行检测。分辨率对预测的影响主要取决于分辨率信息,也就是像素的数量 (图 2)。在第 2 类方法的 YOLOv3 中,我们采用 (类似 FPN) 上采样 (Upsample) 和数据融合算法,融合了香蕉图像中的 3 个尺度 (13×13、26×26 和 52×52),在多尺度融合特征图上对目标进行独立检测,以优化效果。本文设计了一种新的分类器,通过设计多标签分类的 Logistic 分类器,用多个独立的 Logistic 分类器替代传统的 Softmax 层解决多标签分类问题,并保持准确性。传统 Softmax 分类器用于处理目标互斥的多分类问题,传统的 Logistic 分类器用于处理目标互斥的二分类问题。在采样得到的图片中,果轴、果实和花蕾并不存在明确的分界线,三者图像上均有一定的重叠区

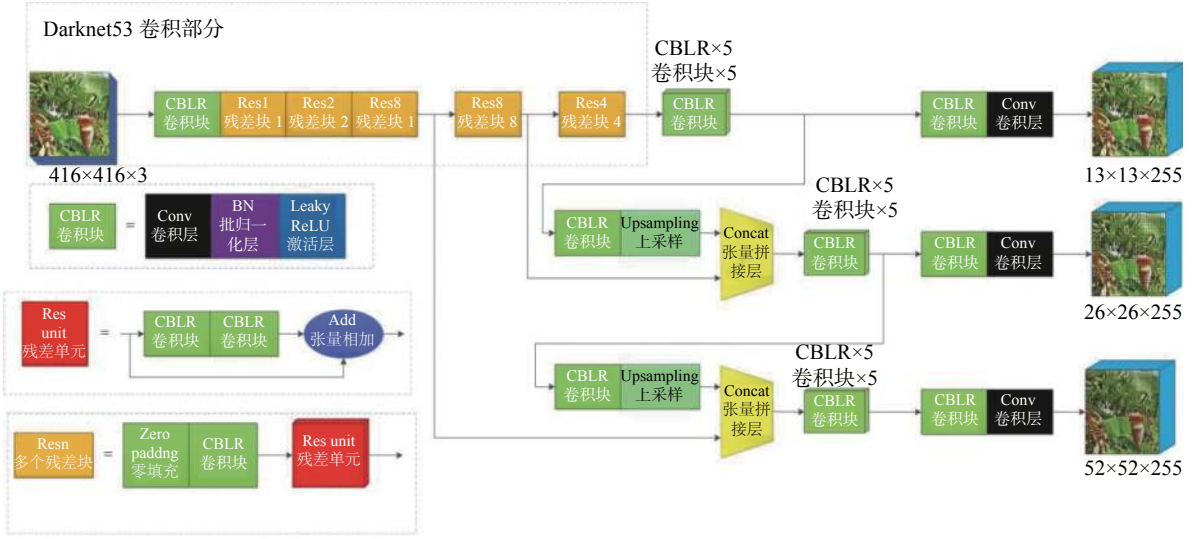


图 2 基于香蕉多特征目标的 YOLOv3 网络结构

Fig. 2 YOLOv3 network structure based on banana multi-feature target

域。若采用传统的 Softmax 分类器, 由于各标签是互斥关系, 所以这种区域重叠的情况会出现比较明显的分类错误或回归偏差; 若采用传统的 Logistic 分类器, 则无法处理这种多目标的检测问题。本文采用由多个独立的 Logistic 并联组成的分类器来解决这个问题, 既能实现多目标检测, 又能实现非互斥目标的准确分类。

2.3 基于香蕉多特征目标的 YOLO v3 网络结构

因算法采用多尺度融合特征的图像对多特征香蕉 (含轴和花蕾) 多目标进行单独检测, 有效地增强了对大小不同目标和被遮挡目标的检测效果, 并引入跃层连接强化收敛效果。因香蕉树生长于野外, 已有的 YOLOv3 结构仍需要改进 (图 2)。本文在损失函数的部分进行改进, 采用交叉熵损失函数代替方差损失函数, 二者均为损失函数, 损失函数可以衡量预测模型与真实模型的偏差, 其偏差越小 (即损失函数的值越小), 则预测模型与真实模型越接近。对于单样本的一次梯度下降过程中, 有如下关系:

$$\text{线性部分: } z = wx + b, \quad (1)$$

$$\text{非线性部分: } \hat{y} = a = \sigma(z), \quad (2)$$

对于均方差损失函数, 其损失函数定义为:

$$\text{Loss}_{\text{MSE}}(y, a) = \frac{1}{2}(y - a)^2, \quad (3)$$

对于交叉熵损失函数, 其损失函数定义为:

$$\text{Loss}_{\text{CE}}(y, a) = -[y \lg(a) + (1 - y) \lg(1 - a)], \quad (4)$$

对于一次梯度下降中权重(w)和偏置(b)的更新, 有

$$w = w - \alpha \frac{\partial \text{Loss}(y, a)}{\partial w}, \quad (5)$$

$$b = b - \alpha \frac{\partial \text{Loss}(y, a)}{\partial b}, \quad (6)$$

分别将均方差损失函数与交叉熵损失函数对参数求偏导:

$$\text{令: } z = wx + b, \quad (7)$$

则对于均方差损失函数, 有:

$$\frac{\partial \text{Loss}_{\text{MSE}}(y, a)}{\partial w} = -[y - \sigma(z)] \sigma'(z) x, \quad (8)$$

$$\frac{\partial \text{Loss}_{\text{MSE}}(y, a)}{\partial b} = -[y - \sigma(z)] \sigma'(z), \quad (9)$$

对于交叉熵损失函数, 有:

$$\frac{\partial \text{Loss}_{\text{CE}}(y, a)}{\partial w} = x [\sigma(z) - y], \quad (10)$$

$$\frac{\partial \text{Loss}_{\text{CE}}(y, a)}{\partial b} = \sigma(z) - y, \quad (11)$$

可见, 对于均方差损失函数, 其对参数的偏导中均含有 Sigmoid 函数的导数, 对于 Sigmoid 函数, 其表达式为:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (12)$$

可见, 当输入值很大或很小时, 其导数趋近于 0, 这将导致权重和偏置变化速度急剧下降, 即梯度消失。而交叉熵损失函数的偏导中不含有 Sigmoid 的导数项, 可以有效避免梯度消失的问题, 从而改善网络训练速度。

本文所采用的网络结构中该算法使用 Darknet53 作为骨架网络进行特征提取, 再对特征图进行坐标回归和分类, 结构中的最小组件由 Conv+BN+Leaky_relu 激活函数组成。

2.4 YOLO-Banana 模型与评估

在网络模型参数设计与训练中, 通常将 epoch 参数设置为 100 进行迭代, batch_size 为 16~32, 花费时间和内存消耗相对增加。

在建立的 YOLOv3 网络中, 优化了 Iteration、batch_size 和 epoch 这 3 个参数, 使损失函数 Loss 曲线趋于稳定, 再进一步地对 YOLOv3 模型的目标候选框维度进行聚类优化, 以提高分类识别的

精度。尽管 YOLOv3 自带默认的目标候选框数量、高宽比维度, 它具有一定的通用性, 但应用于野外复杂背景及变化光照环境下的香蕉多目标还需要优化改进。因此, 本文融合聚类算法来优化 YOLOv3 模型, 优化后的模型称为 YOLO-Banana 模型, 利用 k 均值聚类算法, 基于训练数据集, 重新获得适应于预测的目标候选框个数、新的高与宽, 并进行聚类分析, 优化参数见表 1。

表 1 香蕉多目标候选框参数
Table 1 Parameters of banana multi target candidate box for banana

候选框参数 Parameters of candidate box	1	2	3	4	5	6	7	8	9
新候选框高 Height of new candidate box	23	25	28	31	43	51	66	169	172
新候选框宽 Width of new candidate box	31	53	81	148	67	87	111	233	158
原候选框高 Height of original candidate box	10	16	33	30	62	59	116	156	373
原候选框宽 Width of original candidate box	13	30	23	61	45	119	90	198	326

为了评估模型的泛化能力并逐步优化模型, 评估参数包括精度 (Precision)、召回率 (Recall) 和二分类模型精确度的指标 F_1 。计算公式如下:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%, \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%, \tag{14}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%, \tag{15}$$

式中, TP 表示正类判定为正类; FP 表示负类判定为正类, “存伪”; FN 表示正类判定为负类, “去真”。

训练中, batch size 设计为 8, 即每次 Iteration 训练 8 张图片, 故 Iteration 为 128, 对每个香蕉目标的全数据集进行 1 次运行, 每次运行可得到 1 组 Precision 和 Recall。当对模型算法设定不同的阈值时, 会得到多组 Precision 和 Recall, 进而画出 1 条 PR 曲线, 该曲线的面积为平均精度 (Average precision, AP)。

3 结果与分析

用上文提出的方法, 在野外环境中完成香蕉多目标样本采集和训练模型。通过采样建立各类图像数据库, 训练模型。

用视觉平台在香蕉果园采集香蕉、花蕾和果轴样本。软件系统主要以 OpenCV 函数库与 YOLOv3 为基础编写而成, 硬件设备为有限通用水果采摘机器人及其视觉平台^[28], 为末端执行器切断花蕾提供定位 (图 3)。

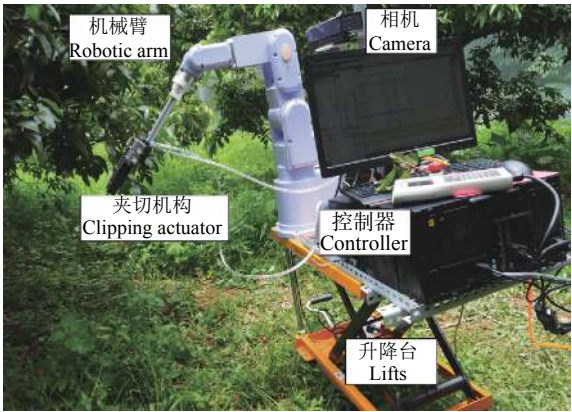


图 3 视觉机器人平台
Fig. 3 Visual robot platform

将本文数据集在 YOLO-Banana 模型的神经网络中进行训练和检测, 选择训练的最优模型在验证集中进行验证, YOLO-Banana 模型网络结构中的 Neck 部分采用的是 FPN 结构, 通过多尺度反复提取主干层和检测层的特征, 以提升网络检测微小目标效率。

用 YOLO-Banana 模型和 Faster R-CNN 模型进行比较。分别获得香蕉花蕾、香蕉果实和香蕉果轴的精度、召回率等评价参数。效果见图 4、图 5。

从图 4、图 5 可看出, 香蕉和花蕾的检测精度较高, YOLO-Banana 对较长的果轴检测精度较高, 与 Faster R-CNN 检测结果接近; 当果轴很短为小目标时, 部分被向上长的花瓣遮挡, 背景复杂且逆光, 用 YOLO-Banana 检测目标的置信度很低, 仅为 0.35, Faster R-CNN 则为 0.999 (图 5)。在这种极端条件下目标的检出置信度会有所下降, 但对回归位置没有显

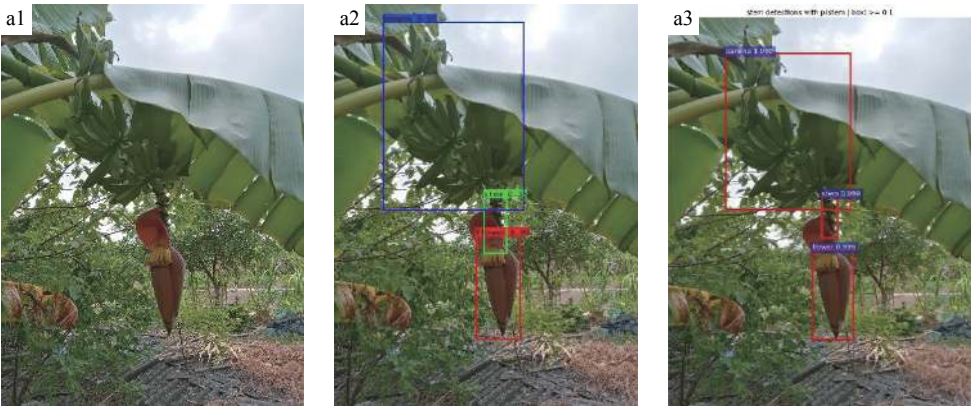


a1, b1, c1 表示香蕉原图; a2(1.00, 0.93, 0.99), b2(0.99, 0.92, 0.98), c2(1.00, 0.98, 1.00) 表示 YOLO-Banana 模型处理后的香蕉图像; a3(1.00, 1.00, 1.00), b3(1.00, 1.00, 0.999), c3(1.00, 1.00, 0.999) 表示 Faster R-CNN 模型处理后的香蕉图像; 括号内数字分别代表该算法对香蕉果实, 果轴和花蕾的置信度

a1, b1 and c1 are original images of banana; a2(1.00, 0.93, 0.99), b2(0.99, 0.92, 0.98), c2(1.00, 0.98, 1.00) are banana images treated by YOLO-Banana model; a3(1.00, 1.00, 1.00), b3(1.00, 1.00, 0.999), c3(1.00, 1.00, 0.999) are banana images treated by Faster R-CNN model; The numbers in parentheses represent the confidence of the algorithm for banana fruit, fruit axis and flower bud, respectively

图 4 顺光下香蕉测试结果对比图

Fig. 4 Comparison of banana test results under frontlight



a1 表示香蕉原图; a2(0.97, 0.35, 0.99) 表示 YOLO-Banana 模型处理后的香蕉图像; a3(1.00, 0.999, 0.999) 表示 Faster R-CNN 模型处理后的香蕉图像; 括号内数字分别代表该算法对香蕉果实, 果轴和花蕾的置信度

a1 is original image of banana, a2 (0.97, 0.35, 0.99) is banana image treated by YOLO-Banana model; a3 (1.00, 0.999, 0.999) is banana image treated by Faster R-CNN model; The numbers in parentheses represent the confidence of the algorithm for banana fruit, fruit axis and flower bud, respectively

图 5 逆光下香蕉测试结果对比

Fig. 5 Comparison of banana test results under backlight

著影响。此外实际应用中这种极端条件出现频率极低,此处仅展示在极端恶劣检测环境下的检出效果。

YOLO-Banana 与 Faster R-CNN 的香蕉多目标特征检测结果见表 2。表 2 表明, Faster R-CNN 对香蕉果实、花蕾、果轴的多特征目标的总平均识别精度为 95.16%, 平均每张图像处理的时间为

0.434 s。YOLO-Banana 对香蕉果实、花蕾、果轴的多特征目标的平均识别精度为 91.03%,平均每张图像处理的时间为 0.237 s。YOLO-Banana 对香蕉多目标精度精度比 Faster R-CNN 对香蕉多目标精度要低 4.13%, 但是其识别速度要比 Faster R-CNN 的快。

表 2 YOLO-Banana 与 Faster R-CNN 的香蕉多目标特征检测结果的对比

Table 2 Comparison of multi-target feature detection results for banana between YOLO-Banana and Faster R-CNN

模型 Model	花蕾 Flower bud		果实 Fruit		果轴 Fruit axis		总平均精度/%	
	平均精度/%	召回率/%	平均精度/%	召回率/%	平均精度/%	召回率/%	Mean of average	t/s
	Average precision	Recall	Average precision	Recall	Average precision	Recall	precision	
Faster R-CNN	99	99.48	98	98.05	89	89.51	95.16	0.434
YOLO-Banana	97	97.38	95	95.12	81	85.19	91.03	0.237

为了测试野外光照的影响,选用了 144 个图片组,每个图片组内图片均为对同一个样本进行采样,包含香蕉、花蕾、果轴这 3 个部位的顺光与逆光图片进行识别试验(图 4 和图 5),分别用本文提出的 YOLO-Banana 模型及 Faster R-CNN 进行试验,结果见表 3。表 3 表明,本文提出的 YOLO-Banana 模型在顺光时的平均识别精度与 Faster R-CNN 精

度相近,误差仅为 0.06%。

分析原因可知,因特征明显,且形状和色彩相似度高,花蕾识别精度最高;香蕉果实特征明显,但大小形状有一些差异,有时会被树叶遮挡,识别精度比花蕾低;果轴因其形状为近似小圆柱形,常被花瓣部分遮挡,识别率相对较小,特别是在逆光环境下,精度低。本文采用了多尺度检测算法,图像会

表 3 YOLO-Banana 与 Faster R-CNN 的顺光逆光下的香蕉多目标检测结果的对比

Table 3 Comparison of multi-target detection results for banana between YOLO-Banana and Faster R-CNN under frontlight and backlight conditions

模型 Model	花蕾精度		果实精度		果轴精度		总平均精度	
	Flower bud average precision		Fruit average precision		Fruit axis average precision		Mean of average precision	
	顺光	逆光	顺光	逆光	顺光	逆光	顺光	逆光
	Frontlight	Backlight	Frontlight	Backlight	Frontlight	Backlight	Frontlight	Backlight
Faster R-CNN	100	100	100	100	93	92	97.64	97.47
YOLO-Banana	100	98	100	93	93	83	97.58	91.24

以不同的分辨率进行检测识别,对不同大小的目标均有较好的检测效果。

断蕾主要考虑的技术指标是识别率,关键在于识别断蕾切割点所在的果轴,因此研究致力于提升果轴的识别率,比如通过聚类算法对各类别求出大小和比例不一的锚框进行网络训练。一般而言,果轴部分在整个生长周期中一直呈绿色,与未成熟的果实和香蕉叶片颜色相近,且四周均被香蕉果实包围,光照条件较差,边界纹理不清晰,且有些果轴短小易被遮挡,所以果轴的识别相对较难,效果不能达到果实和花蕾同时识别的要求。回归框位置的准确性会影响后续断蕾,针对特定目标框进行图像处

理,不同类别进行的图像处理方式不一致(纹理色彩等不一致),如对果轴而言,检测框越准确越能更好地进行后续边缘分割,从而获得果轴主方向,便于果轴切割。

由于深度学习无法确定断蕾位置,作者采用经典图像算法融合几何计算来确定断蕾点位置。在对多目标进行分类识别后,计算花蕾的边缘对称中心及其质心,建立花序轴求解方程,获取花序轴上的切断点。用边缘算法求得花蕾和花序轴图像边缘,利用花蕾的对称性求解花蕾质心与几何对称中心;然后,用几何方法计算花序轴上的切断点,该方法先求解检测到所有直线到质心之间的距离,再通过点线

距离最小约束选取距离最短的直线作为切断点所在花序轴的位置, 取线段中点坐标作为切断点。

在断蕾时, 对切断果轴的轴向尺寸的精度属于自由精度范围, 考虑到机器人在夹持切割机构及控制上具有一定的容错性, 末端执行器对香蕉果轴检测角度具有一定的鲁棒性, 当检测所得的香蕉果轴空间指向与实际香蕉果轴空间指向在一定锥角内, 末端执行器均能自适应地修剪, 并能获得较为一致的修剪效果。因此, 野外断蕾机器人作业时, 可以选择本文提出的 YOLO-Banana 算法。

4 结论

为了提高野外作业机器人综合工作效率, 考虑了检测精度和速度之间的平衡关系, 研究了果园环境下用 YOLOv3 对香蕉多目标的分类识别方法, 提出对香蕉目标进行多尺度特征提取及分类模型, 对 YOLOv3 模型的一些参数进行了修改, 并在此基础上, 提出基于 k 均值聚类算法改进的 YOLO-Banana 模型, 阐述了损失函数的误差原理, 在置信度和分类误差模型中采用交叉熵的损失函数。研究了光照对香蕉果实、花蕾及果轴识别精度的影响。试验结果表明, 用 YOLO-Banana 模型对香蕉果实、花蕾、果轴的检测精度分别为 97%、95%和 81%, 多特征的目标总平均精度为 91.03%, 平均每张所需时间为 0.237 s。Faster R-CNN 对香蕉果实、花蕾、果轴的检测精度分别为 99%、98% 和 89%, 多特征的目标总平均精度为 95.16%, 平均每张所需时间为 0.434 s。对比表明, 前者速度快 1.83 倍, 后者精度高。用 YOLO-Banana 模型与 Faster R-CNN 在顺光和逆光环境下进行了对比, 两者顺光识别精度的误差很小, 仅为 0.06%, YOLO-Banana 的顺光识别精度高于逆光识别精度, 误差为 8.26%。研究数据为机器人视觉识别与定位提供了参考。

参考文献:

[1] YIN Y H, LI H F, FU W. Faster-YOLO: An accurate and faster object detection method[J]. *Digital Signal Processing*, 2020, 102: 102756. doi: 10.1016/j.dsp.2020.102756.

[2] 何银水, 余卓骅, 李健, 等. 基于视觉特征的厚板机器人焊接焊缝轮廓的有效提取[J]. *机械工程学报*, 2019, 55(17): 56-60.

[3] 訾斌, 尹泽强, 李永昌, 等. 基于 YOLO 模型的柔索并联机器人移动构件快速定位方法[J]. *机械工程学报*, 2019, 55(3): 64-72.

[4] 孙国鹏, 郝向阳, 张振杰, 等. 多特征判断的合作目标识别方法[J]. *系统仿真学报*, 2018, 30(6): 2377-2383.

[5] 薛培林, 吴愿, 殷国栋, 等. 基于信息融合的城市自主车辆实时目标识别[J]. *机械工程学报*, 2020, 56(12): 165-173.

[6] 张青, 邹湘军, 林桂潮, 等. 草莓重量和形状图像特征提取与在线分级方法[J]. *系统仿真学报*, 2019, 31(1): 7-15.

[7] TANG Y C, LI L J, WANG C L, et al. Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision[J]. *Robotics and Computer-Integrated Manufacturing*, 2019, 59: 36-46.

[8] DAI Z, SONG H S, LIANG H X, et al. Traffic parameter estimation and control system based on machine vision[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020: 1-13. doi: 10.1007/s12652-020-02052-5.

[9] MIAN A S, BENNAMOUN M, OWENS R. Three-dimensional model-based object recognition and segmentation in cluttered scenes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(10): 1584-1601.

[10] WEI X Q, JIA K, LAN J H, et al. Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot[J]. *Optik*, 2014, 125(19): 5684-5689. doi: 10.1016/j.ijleo.2014.07.001.

[11] 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. doi: 10.1016/j.compag.2015.05.021.

[12] TANG Y C, CHEN M Y, WANG C L, et al. Recognition and localization methods for vision-based fruit picking robots: A review[J]. *Frontiers in Plant Science*, 2020, 11: 510. doi: 10.3389/fpls.2020.00510.

[13] FU L H, DUAN J L, ZOU X J, et al. Banana detection based on color and texture features in the natural environment[J]. *Computers and Electronics in Agriculture*, 2019, 167: 105057. doi: 10.1016/j.compag.2019.105057.

[14] ANDERSON N T, UNDERWOOD J P, RAHMAN M M, et al. Estimation of fruit load in mango orchards: Tree sampling considerations and use of machine vision and satellite imagery[J]. *Precision Agriculture*, 2018, 20(4): 823-839.

[15] ROUT R, PARIDA P. A review on leaf disease detection using computer vision approach[J]. *International Conference on Innovation in Modern Science and Technology*, 2019: 863-871. doi: 10.1007/978-3-030-42363-6_99.

[16] PATRICIO D I, RIEDER R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review[J]. *Computers and Electronics in Agriculture*, 2018, 153: 69-81. doi: 10.1016/j.compag.2018.08.001.

[17] 罗陆锋, 邹湘军, 王成琳, 等. 基于轮廓分析的双串叠贴葡萄目标识别方法[J]. *农业机械学报*, 2017, 48(6): 15-

22. doi: 10.6041/j.issn.1000-1298.2017.06.002.

[18] HOČEVAR M, ŠIROK B, GODEŠA T, et al. Flowering estimation in apple orchards by image analysis[J]. *Precision Agriculture*, 2014, 15(4): 466-478.

[19] GEORGESCU M I, IONESCU R T. Teacher-student training and triplet loss for facial expression recognition under occlusion[C]//2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021: 2288-2295. doi: 10.1109/ICPR48806.2021.9412493

[20] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. 2018 [2020-07-20].<http://predddie.commedia/files/YOLOv3.pdf>.

[21] ALEXE B, DESELAERS T, FERRARI V. Measuring the objectness of image windows[J]. *IEEE Transactions on Software Engineering*, 2012, 34(11): 2189-2202.

[22] LIN P, CHEN Y M. Detection of strawberry flowers in outdoor field by deep neural network[C]// 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). Chongqing: IEEE, 2018: 482-486. doi:10.1109/ICIVC.2018.8492793.

[23] TIAN Y N, YANG G D, WANG Z, et al. Instance segmentation of apple flowers using the improved mask R-CNN model[J]. *Biosystems Engineering*, 2020, 193: 264-278. doi: 10.1016/j.biosystemseng.2020.03.008.

[24] KOIRALA A, WALSH K B, WANG Z L, et al. Deep learning for mango (*Mangifera indica*) panicle stage classification[J]. *Agronomy*, 2020, 10(1): 143. doi: 10.3390/agronomy10010143.

[25] 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. doi: 10.11975/j.issn.1002-6819.2019.23.006.

[26] SONG S S, DUAN J L, YANG Z, et al. A three-dimensional reconstruction algorithm for extracting parameters of the banana pseudo-stem[J]. *Optik*, 2019, 185: 486-496. doi: 10.1016/j.ijleo.2019.03.125.

[27] CHEN M Y, TANG Y C, ZOU X, et al. Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology[J]. *Computers and Electronics in Agriculture*, 2020, 174: 105508.

[28] ZOU X, YE M, LUO C, et al. Fault-tolerant design of a limited universal fruit-picking end-effector based on vision-positioning error[J]. *Applied Engineering in Agriculture*, 2016, 32(1): 5-18. doi: 10.1016/j.compag.2020.105508.

[29] SOVIANY P, IONESCU R T. Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction[C]//2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). Timisoara, Romania: IEEE, 2018: 209-214. doi:10.1109/SYNASC.2018.00041.

【责任编辑 霍 欢】