

基于残差网络的自然环境中刺梨果实的识别*

闫建伟^{1, 2}, 赵源¹, 张乐伟¹, 张富贵¹

(1. 贵州大学机械工程学院/现代制造技术教育部重点实验室, 贵阳市, 550025;

2. 国家林业和草原局刺梨工程技术研究中心, 贵阳市, 550025)

摘要:针对传统的目标检测方法依赖人工提取特征,存在检测效率低、鲁棒性差和实时性差等缺陷。本文根据刺梨果实在自然环境中的生长特点,采用带有残差模块的 YOLO v3 神经网络进行刺梨果实识别模型的训练,该网络通过提取不同卷积层的特征图,将深层特征图进行上采样后与浅层特征图进行多次融合,以提取图像的更深层次的特征信息。通过对该网络的相关参数进行优化和改进,并对未参与模型训练的 70 幅刺梨图像进行检测,实验表明,本文算法能够有效地对自然环境下的 11 类刺梨果实进行识别,各类识别平均准确率为 88.5%,平均召回率为 91.5%, F_1 平均值为 89.9%,识别速率约为 20 f/s。本文算法在刺梨果实的识别上取得了理想的识别效果。

关键词:刺梨果实;深度学习;残差网络;YOLO v3;目标识别

中图分类号:TP391.41 **文献标识码:**A **文章编号:**2095-5553 (2020) 10-0191-06

闫建伟, 赵源, 张乐伟, 张富贵. 基于残差网络的自然环境中刺梨果实的识别[J]. 中国农机化学报, 2020, 41(10): 191-196

Yan Jianwei, Zhao Yuan, Zhang Lewei, Zhang Fugui. Recognition of *Rosa roxbunghii* fruit in natural environment based on residual network [J]. Journal of Chinese Agricultural Mechanization, 2020, 41(10): 191-196

0 引言

刺梨(*Rosa roxbunghii* Tratt)又称刺石榴,是蔷薇科植物缙丝花的果实,作为贵州省区域重点食药两用植物资源正在被大力开发。刺梨富含维生素 C、超氧化物歧化酶(superoxide dismutase, SOD)、黄酮及多种微量元素,具有极高的食药价值^[1]。

计算机视觉检测算法对图像的颜色特征过于依赖,且需要手工设计特征来实现,其检测效果受环境变化影响较大,检测准确率低,鲁棒性差^[2]。自然环境下的刺梨果实生长的背景复杂,传统的计算机视觉算法已无法满足识别要求,基于卷积神经网络的识别算法已成为智能识别研究的发展方向。目前,国内外基于深度学习的果实识别已有大量的研究。由于深度学习在目标识别与检测方面具有较强的鲁棒性,通过神经网络训练出来的识别模型能够有效地对目标进行检测,如对强光照、弱光照、阴影以及各种天气等复杂环境下的目标检测都具有较好的识别性能。深度学习算法能够无需人工设计特征而智能地对目标的特征进行提取^[3]。在深度学习领域,已有大量研究人员对基于

深度学习的目标检测算法进行了研究。2015 年, Ren 等^[4]提出了超快区域卷积神经网络(Faster R-CNN),检测平均准确率在 VOC 2007 数据集上达到了 73.2%,检测速度达到 5 f/s。Joseph Redmon 等^[5]提出了 YOLO 算法,在 VOC 2012 数据集上的检测平均准确率达到 57.9%,检测速度达到 45 f/s。2017 年, Redmon 等^[6-7]提出了 YOLO v2 算法,加入了批量归一化(batch normalization, BN),既防止了过拟合又加快了收敛速度,收敛速度提升了 2%,同时,采用了多尺度训练,检测准确率明显提升。2018 年, Redmon 等^[8]又提出了 YOLO v3 检测算法,检测性能更好,相比 SSD 算法,速度提高了 3 倍,相比 RetinaNet 速度提高 3.8 倍,能够达到对目标的实时检测。YOLO v3 在小目标检测中也有改进^[9],而且兼顾准确率和检测速度,达到了实时高效检测的要求。刺梨果实的检测、识别定位以及实时跟踪是刺梨采摘机器人视觉系统中的重要任务。闫建伟等^[10]提出采用 0.57R-0.18G-0.2B 色差分量和标记分水岭分割算法对刺梨图像进行二次分割,刺梨果实识别正确率高于 92%。闫建伟等^[11]又提出改进 Faster-RCNN 自然环境下识别刺梨

收稿日期:2020 年 2 月 27 日 修回日期:2020 年 3 月 26 日

* 基金项目:贵州省普通高等学校工程研究中心建设项目(黔教合 KY 字[2017]015);贵州省科技计划项目(黔科合重大专项字[2019]3014—3、黔科合成果[2019]4292 号、黔科合平台人才[2019]5616 号)

第一作者:闫建伟,男,1980 年生,河南鹿邑人,博士,副教授,硕导;研究方向为深度学习、智能装备等。E-mail: jwyan@gzu.edu.cn

果实的方法,通过比较 Faster RCNN 框架下的 VGG16、VGG_CNN_M1024 以及 ZF3 种网络模型训练的精度—召回率,最终选择 VGG16 网络模型,对 11 类刺梨果实的平均识别精度达到 92.01%。以上研究成果为刺梨果实识别提供了较好的借鉴。本文将对基于残差网络自然环境下刺梨果实识别进行研究,为实现刺梨果实的智能化采摘提供参考依据。

1 图像采集及预处理

1.1 图像采集

本文使用的刺梨图像采集于贵州省龙里县谷脚镇茶香村刺梨产业示范园区,分别采集了未成熟时期刺梨图像 1 500 幅,成熟时期刺梨图像 1 600 幅。图像分辨率均为 6 016 像素 \times 4 016 像素,图像格式为 JPEG。自然条件下的刺梨果实图像如图 1 所示。

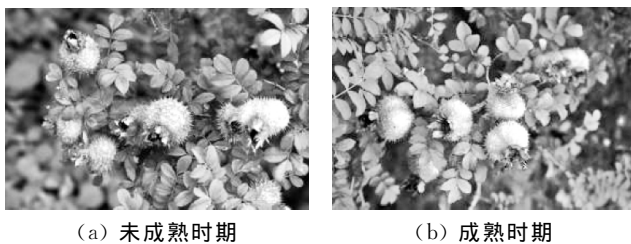


图 1 自然条件下的刺梨果实图像

Fig. 1 *Rosa roxbunghii* fruits under natural conditions

1.2 图像预处理

本文使用 Photoshop CS6 软件统一将采集的原始图像裁剪为 512 像素 \times 512 像素,以适应神经网络对输入图像尺寸大小的要求。为达到更好的模型训练效果,本文提出了一种增强刺梨图像数据集的方法,将裁剪的图像进行上下翻转以及分别旋转 45°、90°和 270°,最终得到 8 475 幅样本图像。本文使用 VOC2007 数据集格式,借助图像标签制作工具 labelImg 进行标签制作^[12]。本文根据刺梨的生长特点,从刺梨果实的成熟度、图像中刺梨果实相邻的个数以及刺梨果实是否存在遮挡三种情况考虑,将刺梨果实识别类别设置为 11 种类别。

2 试验条件及方法

2.1 硬件和软件

本文应用的硬件和软件包括:台式电脑,电脑配置: Intel(R) Core(TM) i7-8700K 处理器,主频为 3.70 GHz,磁盘内存为 250 GB,显卡采用 GeForce GTX1060 显卡,6G 显存。Ubuntu 16.04、64 位操作系统,采用 Darknet-53 网络进行刺梨数据集识别模型训练。相机型号及相关参数:尼康(Nikon D750),常用变焦头: AF-S 尼克尔,24~120 mm f/4G ED VR

镜头。

2.2 试验方法

本文根据刺梨果实在自然环境下的生长特征,对 YOLO v3 网络框架中的相关参数进行调整和改进,并将调整后的网络用于刺梨识别模型训练,该算法采用 YOLO v2 中的 K-means 聚类方式来做边框(anchor box)的初始化,这种先验知识对于边界框的预测具有促进作用。同时使用了 Darknet-53 网络结构对刺梨图像进行特征提取^[8]。几种典型网络在 ImageNet 数据集下的性能对比测试结果如表 1 所示。

表 1 性能测试结果
Tab. 1 Performance test result

Backbone	Top-1 /%	Top-5 /%	Bn Ops /10 ⁹	BFLOP /s	识别帧率 /(f \cdot s ⁻¹)
Darknet-19	74.1	91.8	7.29	1246	171
ResNet-101	77.1	93.7	19.70	1039	53
ResNet-152	77.6	93.8	29.40	1090	37
Darknet-53	77.2	93.8	18.70	1457	78

由表 1 可知,Darknet-53 检测速度较 ResNet-101 提高 2 倍,检测准确率较 Darknet-19 有较大提升,虽然速度上较之慢一些,但是已满足实时识别的要求。综合各方面性能,Darknet-53 具有较好检测效果。YOLO v3 的 Darknet-53 相对于 YOLO v2 的 Darknet-19 网络增加了残差模块^[13],保证了网络结构较深的情况下,仍能收敛,模型能继续训练下去,从而获得更好的特征提取效果。YOLO v3 的 Darknet-53 算法还采用多尺度特征融合、端到端及划分单元格等训练方式,其主要的内容包括以下几点。

2.2.1 多尺度特征提取

YOLO v3 框架的 Darknet-53 训练使用的最小图像尺寸为 320 像素 \times 320 像素,最大的图像尺寸为 608 像素 \times 608 像素,使得网络可以适应多种不同尺度的输入。多尺度特征融合示意图如图 2 所示。

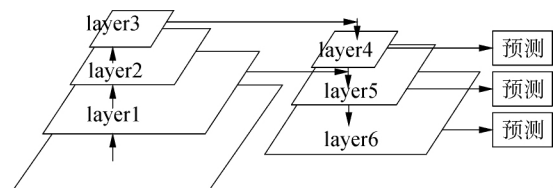


图 2 多尺度特征融合示意图

Fig. 2 Multi-scale feature fusion diagram

由图 2 可知 YOLO v3 使用上采样将高层特征与低层特征进行融合,最终获得 3 组特征图,同于对目标进行预测。当一幅样本图像输入网络后,Darknet-53

网络会用不同大小的刻度,将该图像平均分成 $S \times S$ 个网格(grid cell),如果某个待检测目标(object)的中心点(即标签中目标框的中心坐标)落在其中一个网格中,则这个网格就负责预测这个目标对象。每一个网络预测 B 个边框(bounding box),YOLO v3 网络中 B 为 3 个,这些边框中与标签中边框的交并比最大的将用来预测目标^[14]。同时,这些边框一方面要回归自身的位置,另一方面还要预测一个置信度。

2.2.2 k-means 维度聚类

YOLO v3 继续采用 YOLO v2 中使用的 K-means 维度聚类方法,其是在 Faster-RCNN 首选先验框方法上做的改进,将欧氏距离法改进为通过交并比(intersection over union, IOU)参与距离计算,交并比是一种测量在特定数据集中检测相应物体准确度的一个标准。YOLO v3 预测最终得到 9 个先验框,计算得到 9 组先验框维度,该维度也适用于本文的刺梨果实识别。

3 模型训练

本文采用的是 YOLO v3 框架进行刺梨果实识别模型的训练。YOLO v3 的网络结构图如图 3 所示。由图 3 YOLO v3 网络结构图可以看出,每一个 DBL 模块由卷积层、批归一化层和激励层构成, resn 表示该

残缺块(res_block)中有 n 个残缺单元(res_unit),张量拼接(concat)的功能主要是将 Darknet-53 的中间某一层与其后的某一层的上采样进行拼接。

YOLO v3 网络结构中 add 层有 23 层,主要用于 res_block 的构成,每个 res_unit 需要一个 add 层。批归一化层和激励层各 72 层,激励层紧跟在批归一化层后面。卷积层有 75 层,上采样和张量拼接各有 2 次,零填充有 5 个,总共有 252 层。与多数网络不同,该网络结构中没有池化层和全连接层,其通过改变卷积核的步长进行张量尺寸的变换。YOLO v3 将 YOLO v2 中的单标签多分类的 softmax 层换成用于多标签多分类的逻辑回归层,分类更加准确。

本文采用的网络参数如下:输入样本的宽度为 512 像素,高度为 512 像素。学习率(learning rate)为 0.001,一批训练样本的样本数量(epoch)为 16 个,一次性送入训练器的样本数量为 4 个,动量(momentum)为 0.9,权重衰减系数(weight_decay)为 0.000 5,训练的最大迭代次数为 20 000,学习率变化比例在迭代 10 000 次和 15 000 次时分别为 0.1 和 0.5,学习率调整策略为 steps。非极大值抑制阈值(nms)为 0.6,交并比阈值为 0.8。通过对训练参数的调整与改进,使改调整和进后的网络能够训练出适用于刺梨果实识别的网络模型。

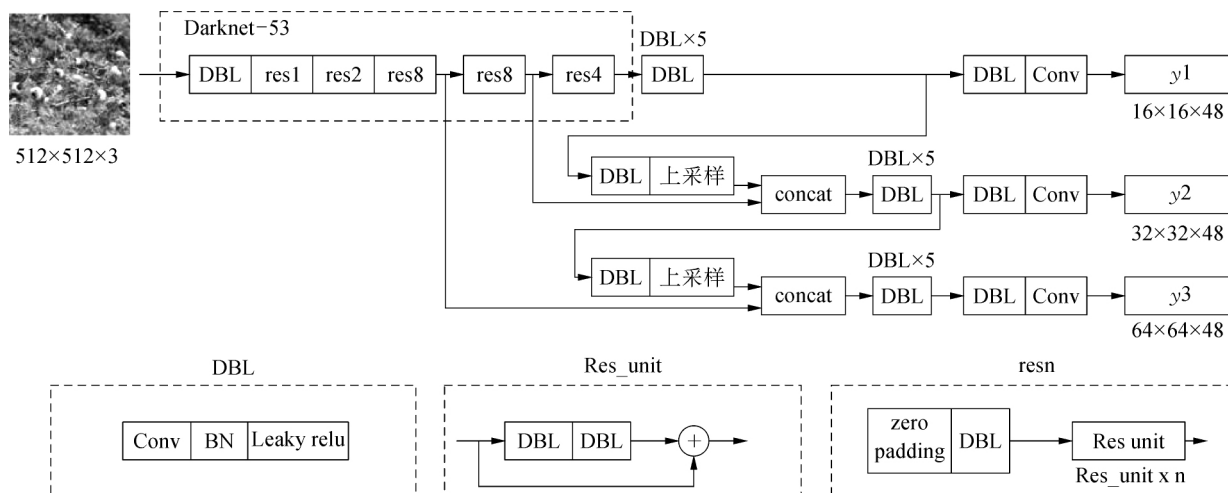


图 3 YOLO v3 网络结构图

Fig. 3 Network structure diagram of YOLO v3

4 试验结果及分析

4.1 模型验证

通过对刺梨数据集进行模型训练,得到适用于刺梨果实检测的识别模型。通过对训练日志进行分析,可获得平均损失(Avg loss)变化曲线图如图 4 所示,平均交并比(Avg IOU)变化曲线图如图 5 所示。

由图 4 平均损失变化曲线图可以看出,迭代次数

约达到 15 000 次后,平均损失曲线已收敛,损失率最低达到了 0.08%。由图 5 的平均交并比变化曲线图可以看出,平均交并比随着批次的增高,不断趋于 100%。说明模型训练效果比较好,模型的检测性能不断提高。

本试验选择了 70 幅未参与训练的刺梨图像进行模型性能的检测,本文选择以召回率(recall, R)、准确率(precision, P)以及 F_1 值作为识别模型的性能评价

指标, F_1 值是准确率和召回率的一个平稳点, 能同时考虑到准确率和召回率两个指标, 所以识别模型性能的最终评价指标是 F_1 值, 其计算公式如式(1)所示。

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (1)$$

其中: P = 检测正确 / (检测正确 + 检测误以为正确); R = 检测正确 / (检测正确 + 检测误以为错误)。

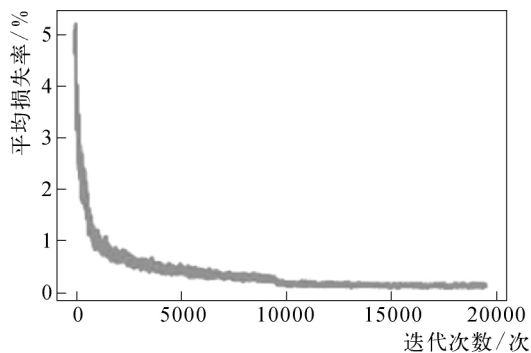


图 4 平均损失变化曲线图

Fig. 4 Average loss curve

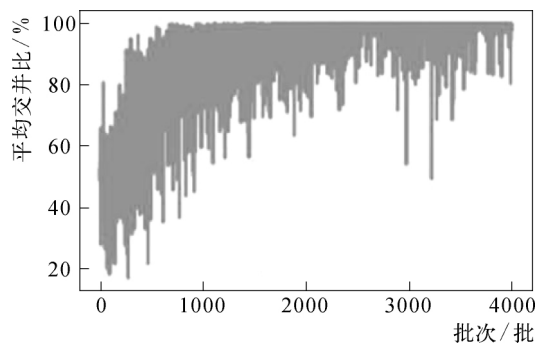


图 5 平均交并比变化曲线图

Fig. 5 Average IOU curve

该模型对 11 类不同生长特征的刺梨果实模型检测结果如表 2 所示。从表 2 可以看出, 11 类刺梨检测准确率均高于 81.5%, 召回率均高于 85.3%, F_1 值均高于 84.1%, F_1 值最高可达 94.5%。通过计算, 该模型对 11 类刺梨果实的检测平均准确率达到 88.5%, 平均召回率 R 为 91.5%, F_1 平均值为 89.9%。最终检测的效果图如图 6 所示。

表 2 模型检测结果

Tab. 2 Test results of model detection

果实类别	正确检测 果实/个	误以为正确 果实/个	误以为错误的 果实/个	召回率 /%	准确率 /%	F_1 值 /%
单个成熟无遮挡	175	12	14	92.6	93.6	93.1
单个成熟有遮挡	132	16	7	95.0	89.2	92.0
单个未成熟无遮挡	143	14	12	92.3	91.1	91.7
单个未成熟有遮挡	97	10	15	86.6	90.7	88.6
两个成熟无遮挡	85	3	7	92.4	96.6	94.5
两个成熟有遮挡	53	12	5	91.4	81.5	86.2
两个未成熟无遮挡	38	4	2	95.0	90.5	92.7
两个未成熟有遮挡	24	5	3	88.8	82.8	85.7
多个成熟无遮挡	20	3	1	95.2	93.0	94.1
多个成熟有遮挡	29	6	5	85.3	82.9	84.1
多个未成熟无遮挡	32	7	3	91.4	82.1	86.5

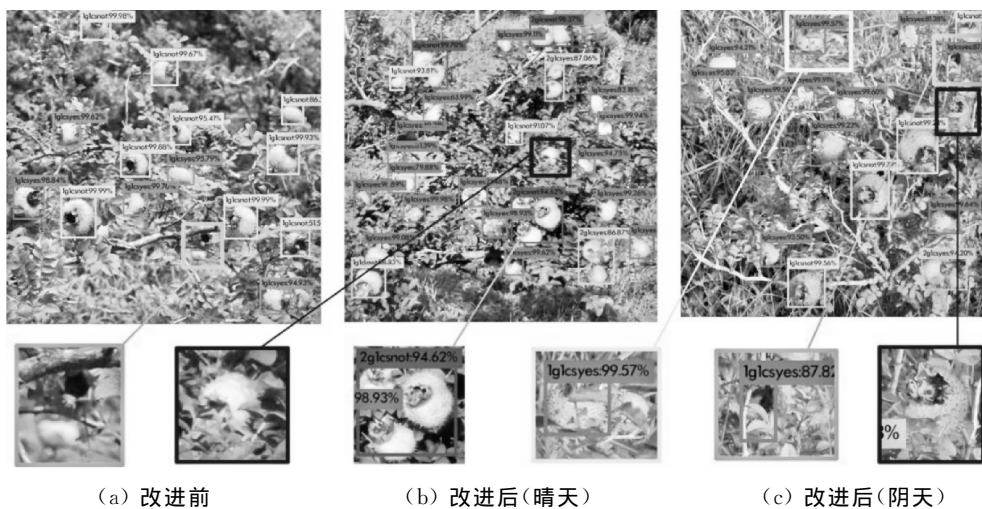


图 6 检测效果图

Fig. 6 Chart of detection effect

由图 6 的检测效果图可以看出,通过本文的方法训练出来的识别模型对自然环境下的刺梨果实具有较好的识别效果,对小目标的识别率也较高。图 6 中带类别名和准确率的边框为通过模型检测出的结果。在检测中,极大部分刺梨果实都被正确识别,并具有较高的准确率。但是,仍然有极少数几个目标由于其所处的位置极其特殊以及遮挡较为严重而没有被检测到,同时,也有极少数黄色树叶由于与刺梨果实极其相似而被误判为目标对象。虽然个别果实的类别识别有误,但成熟度和边框仍然正确。以上几点并不影响刺梨果实的采摘识别,总体来说,本文的模型检测性能能够达到刺梨果实智能化采摘的识别要求。

表 3 基于卷积神经网络的果实识别算法比较
Tab. 3 Comparison of fruit recognition algorithms based on convolutional neural networks

算法	类别	拍摄方式	图像特征	作物特征	神经网络	识别准确率/%
Fu 等 ^[16]	猕猴桃	远距离底部成像	多簇果实	遮挡、重叠、相邻、独立	Faster RCNN (ZFNet 网络)	92.30
傅隆生等 ^[15]	猕猴桃	远距离底部成像	多簇果实	遮挡、重叠、相邻、独立	LeNet 网络	89.29
闫建伟等 ^[11]	刺梨	近距离多角度成像	单个果实	遮挡、重叠、相邻、独立	Faster RCNN (VGG16 网络)	95.16
本文算法	刺梨	近距离多角度成像	两个果实	无遮挡、相邻、独立	Darknet-53 (YOLO v3)	96.60

5 结论

- 1) 本文在刺梨图像数据集制作的过程中采用了多角度旋转和镜像操作,对数据集进行了增加,首次建立了一个适用于刺梨果实实时识别的刺梨图像数据集。
- 2) 本文根据刺梨果实在自然环境下的生长特点,采用带有残差模块的 Darknet-53 网络结构对刺梨果实特征进行提取,对小目标的识别性能得到了增强。
- 3) 本文在 YOLO v3 框架下对刺梨图像数据集进行了模型训练,通过本文的方法所训练出来的识别模型对两个独立果实识别准确率达 96.60%,各类刺梨果实识别平均准确率达到 88.5%,平均召回率为 91.5%, F_1 平均值为 89.9%,识别速率约为 20 f/s,完全能够达到刺梨果实智能化采摘识别的要求。
- 4) 本文的研究主要是为实现自然环境下刺梨果实的智能化采摘,为刺梨果实智能化采摘机器人的研究提供最主要的位置数据,在一定程度上促进了农业从手工时代向智能化时代发展。

参 考 文 献

[1] 夏仕青, 张爱华. 刺梨的营养保健功能及其开发利用研究进展[J]. 贵州医科大学学报, 2018, 43(10): 1129—1132.

- 4.2 对比分析
- 目前,刺梨果实图像识别已有相关研究,同时,将本文算法与相关卷积神经网络 Faster RCNN(ZFNet 网络)、LeNet 网络对猕猴桃等其他果实图像进行识别及检测的结果进行了对比,以验证本文的算法 Darknet-53 (YOLO v3)的识别准确率。具体对比分析见表 3 基于卷积神经网络的果实识别算法比较所示。
- 从表 3 可以看出,本文的算法 Darknet-53 (YOLO v3 网络)对刺梨果实的识别精度高,达到 96.60%,比傅隆生^[15]等方法识别准确率高 7.31%。同时,本文提出的算法对刺梨果实进行有效地识别,能够达到刺梨果实自动化识别定位采摘的要求。

Xia Shiqing, Zhang Aihua. Research progress on nutrition and health care function of *Rosa roxburghii* Tratt and its development and utilization [J]. Journal of Guizhou Medical University, 2018, 43 (10): 1129—1132.

[2] 施辉, 陈先桥, 杨英. 改进 YOLO v3 的安全帽佩戴检测方法[J]. 计算机工程与应用, 2019, 55(11): 213—220.

Shi Hui, Chen Xianqiao, Yang Ying. Detection method of safety helmet wearing on improved Yolo V3 [J]. Computer Engineering and Application, 2019, 55 (11): 213—220.

[3] Le Cun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436—444.

[4] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real time object detection with region proposal networks [C]. International Conference on Neural Information Processing Systems, 2015: 91—99.

[5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 779—788.

[6] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517—6525.

[7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7263—7271.

[8] Redmon J, Farhadi A. YOLO v3: an incremental improvement [J]. IEEE Conference on Computer Vision

- and Pattern Recognition, 2018; 89–95.
- [9] 鞠默然, 罗海波, 王仲博, 等. 一种改进的 YOLO V3 及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 253–260.
- Ju Muran, Luo Haibo, Wang Zhongbo, et al. An improved YOLO V3 and its application in small target detection [J]. Acta Optica Sinica, 2019, 39(7): 253–260.
- [10] 闫建伟, 赵源, 苏小东, 等. 基于深度学习的自然环境下刺梨果实识别[J]. 农机化研究, 2020, 42(11): 23–28.
- Yan Jianwei, Zhao Yuan, Su Xiaodong et al. *Rosa roxbunghii* fruits recognition based on machine vision in natural environment [J]. Journal of Agricultural Mechanization Research, 2020, 42(11): 23–28.
- [11] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster-RCNN 自然环境下识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143–150.
- Yan Jianwei, Zhao Yuan, Zhang Lewei, et al. Recognition of *Rosa roxbunghii* in natural environment based on improved Faster RCNN [J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(18): 143–150.
- [12] 殷帅, 胡越黎, 刘思齐, 等. 基于 YOLO 网络的数据采集与标注[J]. 仪表技术, 2018(12): 22–25.
- Yin Shuai, Hu Yueli, Liu Siqi, et al. Data acquisition and annotation based on the YOLO network [J]. Instrument Technology, 2018, 12: 22–25.
- [13] 郑志强, 刘妍妍, 潘长城, 等. 改进 YOLO V3 遥感图像飞机识别应用[J]. 电光与控制, 2019, 26(4): 28–32.
- Zheng Zhiqiang, Liu Yanyan, Pan Changcheng, et al. Application of improved YOLO V3 remote sensing image in aircraft recognition [J]. Electro Optic and Control, 2019, 26(4): 28–32.
- [14] Bodla N, Singh B, Chellappa R, et al. Soft-NMS-improving object detection with one line of code [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561–5569.
- [15] 傅隆生, 冯亚利, Tola Elkamil, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205–211.
- Fu Longsheng, Feng Yali, Tola Elkamil, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(2): 205–211.
- [16] Fu Longsheng, Feng Yali, Yaqoob M, et al. Kiwifruit detection in field images using Faster R – CNN with ZFNet [J]. IFAC – PapersOnLine, 2018, 51(17): 45–50.

Recognition of *Rosa roxbunghii* fruit in natural environment based on residual network

Yan Jianwei^{1, 2}, Zhao Yuan¹, Zhang Lewei¹, Zhang Fugui¹

(1. College of Mechanical Engineering of Guizhou University/Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, Guiyang, 550025, China;

2. Research Center of *Rosa roxbunghii* Engineering Technology of National Forestry and Prairie Bureau, Guiyang, 550025, China)

Abstract: The traditional target detection method relies on manual to extract features, which has the defects of low detection efficiency, poor robustness and poor real-time performance. In this paper, the YOLO v3 network with residual module is used according to the growth characteristics of the *Rosa roxbunghii* fruit in the natural environment. The training of the *Rosa roxbunghii* fruit recognition model is carried out by extracting the feature maps of different convolutional layers, and then extracting the deep feature maps and superimposing them with the shallow feature maps to extract deeper feature information in the images. Through the adjustment and improvement of the relevant parameters of the network, and testing 70 images of *Rosa roxbunghii* that did not participate in the model training, the experiment shows that the algorithm can effectively identify the 11 types of *Rosa roxbunghii* fruits in the natural environment. The average accuracy rate was 88.5%, the average recall rate was 91.5%, and the average value of F_1 was 89.9%, and the recognition rate is about 20 f/s. The algorithm has achieved an ideal recognition effect on the identification of the *Rosa roxbunghii* fruit.

Keywords: *Rosa roxbunghii* fruit; residual network; deep learning; YOLO v3; target detection