

改进 Faster-RCNN 自然环境下识别刺梨果实

闫建伟^{1,2,3}, 赵 源¹, 张乐伟¹, 苏小东¹, 刘红芸¹,
张富贵^{1,3*}, 樊卫国², 何 林^{1,4}

(1. 贵州大学机械工程学院, 贵阳 550025; 2. 国家林业和草原局刺梨工程技术研究中心, 贵阳 550025;
3. 贵州省山地农业智能装备工程研究中心, 贵阳 550025; 4. 六盘水师范学院, 六盘水 553004)

摘 要: 为了实现自然环境下刺梨果实的快速准确识别, 根据刺梨果实的特点, 该文提出了一种基于改进的 Faster RCNN 刺梨果实识别方法。该文卷积神经网络采用双线性插值方法, 选用 Faster RCNN 的交替优化训练方式 (alternating optimization), 将卷积神经网络中的感兴趣区域池化 (ROI pooling) 改进为感兴趣区域校准 (ROI align) 的区域特征聚集方式, 使得检测结果中的目标矩形框更加精确。通过比较 Faster RCNN 框架下的 VGG16、VGG_CNN_M1024 以及 ZF 3 种网络模型训练的精度-召回率, 最终选择 VGG16 网络模型, 该网络模型对 11 类刺梨果实的识别精度分别为 94.00%、90.85%、83.74%、98.55%、96.42%、98.43%、89.18%、90.61%、100.00%、88.47%和 90.91%, 平均识别精度为 92.01%。通过对 300 幅自然环境下随机拍摄的未参与识别模型训练的刺梨果实图像进行检测, 并选择以召回率、准确率以及 F_1 值作为识别模型性能评价的 3 个指标。检测结果表明: 改进算法训练出来的识别模型对刺梨果实的 11 种形态的召回率最低为 81.40%, 最高达 96.93%; 准确率最低为 85.63%, 最高达 95.53%; F_1 值最低为 87.50%, 最高达 94.99%。检测的平均速度能够达到 0.2 s/幅。该文算法对自然条件下刺梨果实的识别具有较高的正确率和实时性。

关键词: 卷积神经网络; Faster RCNN; 机器视觉; 深度学习; 刺梨果实; 目标识别

doi: 10.11975/j.issn.1002-6819.2019.18.018

中图分类号: TP391.41

文献标志码: A

文章编号: 1002-6819(2019)-18-0143-08

闫建伟, 赵 源, 张乐伟, 苏小东, 刘红芸, 张富贵, 樊卫国, 何 林. 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. doi: 10.11975/j.issn.1002-6819.2019.18.018 http://www.tcsae.org
Yan Jianwei, Zhao Yuan, Zhang Lewei, Su Xiaodong, Liu hongyun, Zhang Fugui, Fan Weiguo, He Lin. Recognition of *Rosa roxbunghii* in natural environment based on improved Faster RCNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 143—150. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2019.18.018 http://www.tcsae.org

0 引 言

刺梨广泛分布于暖温带及亚热带地区, 在我国主要分布在贵州、云南、四川等地, 其中以贵州的盘县、龙里等刺梨资源最丰富、品种最多、产量最高^[1]。

近年来, 卷积神经网络 (convolutional neural network, CNN) 在目标识别与检测方面有广泛的应用。孙世鹏等^[2]采用机器视觉技术对冬枣黑斑病害和缩果病害进行无损检测, 分类正确率分别达到了 89.6%和 99.4%, 但是该方法过于依赖颜色分量, 在复杂背景下的冬枣识别效果有限; 傅隆生等^[3]提出一种基于 LeNet 卷积神经网络的深度学习模型进行多簇猕猴桃果实图像的识别方法, 该方法对田间猕猴桃图像具有较高的识别率和实时性, 但对于强反射光及重叠果实的识别效果不理想。孙云云等^[4]采用

AlexNet 经典网络模型对茶树病害进行图像识别, 平均测试准确率为 90%, 正确区分率分别为 85%、90%和 85%, 说明卷积神经网络在农作物的识别上具有高效性和可行性。Przybylo 等^[5]提出了利用卷积神经网络对橡胶种子切片颜色图像进行活性识别的方法, 该方法的准确度 (85%) 相当或略高于人工 (84%), 提高了作业效率。夏为为等^[6]提出了一种基于卷积神经网络的改进算法对宫颈癌细胞图像进行识别, 该改进算法降低了对宫颈癌细胞图像的识别错误率 (从 4.74%降到 4.38%左右), 说明神经网络在医学领域也有重要的应用。目标识别方法正在从传统的机器学习算法转向神经网络领域, 传统的机器学习算法由于对目标的颜色过度依赖, 使其对于复杂背景中的目标识别正确率较低。而卷积神经网络则通过对大量数据的训练, 学习目标具体特征, 以实现目标的精准识别与定位。卷积神经网络不仅在手写字识别^[7-9]、人脸识别^[10-14]、行为识别^[15-21]以及车辆检测^[22-23]等方面有较成熟的应用, 在苹果^[24-26]、猕猴桃^[3,27]和橘子^[28]等果实的识别方面也有广泛的应用, 但是, 还没有将神经网络用于刺梨果实识别方面的相关文献。

目前, 刺梨果实采摘是刺梨生产中最耗时、耗力的环节, 其投入的劳力约占生产过程 50%~70%。刺梨果实的采摘人工成本高、劳动强度大、采摘效率低^[29]。刺梨

收稿日期: 2019-03-26 修订日期: 2019-08-25

基金项目: 贵州大学培育项目 (黔科合平台人才[2017]5788); 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 (黔科合平台人才[2019]5616 号)

作者简介: 闫建伟, 博士, 副教授。主要从事农业智能化技术与装备研究。
Email: jwyang@gzu.edu.cn。

*通信作者: 张富贵, 博士, 教授, 主要从事农业机械化技术研究。

Email: zhufugui@vip.163.com

中国农业工程学会会员: 闫建伟 (E041201018S)

果实自身重力较小,且枝梗较硬,使得刺梨花苞朝向各个方向,且刺梨果实颜色特征与其枝条和叶片相近,这对实现自然环境下刺梨果实的识别和定位带来了困难。

本文结合自然环境下刺梨果实的生长特征,对 Faster RCNN 框架下的 VGG16 网络的结构和参数进行了调整、改进和优化,通过对刺梨数据集的训练,最终得到一个基于改进的卷积神经网络的刺梨果实识别模型,该模型能够高效快速地识别自然环境下的刺梨果实,以实现刺梨果实进行高精度、快速的识别。

1 数据采集与处理

1.1 数据采集

本文刺梨果实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,品种为贵龙 5 号。2018 年 8 月 8 日下午采集未成熟时期刺梨果实图像 1 500 幅,天气晴朗;2018 年 9 月 20 日下午采集成熟时期刺梨果实图像 1 600 幅,天气晴朗;共采集自然环境下刺梨果实原始图像 3 100 幅。本文所用图像采用尼康(Nikon)D750 单反相机多角度近距离(2 m 以内)进行拍摄,原始图像格式为 JPEG,分辨率为 6 016×4 016 像素。自然环境下的刺梨果实图像样本示例如图 1 所示。



图 1 自然环境下的刺梨果实图像样本示例

Fig.1 *Rosa roxburghii* fruit image under natural conditions

1.2 数据集样本及标签制作

本文从拍摄的 3 100 幅刺梨果实图像中选出 2 000 幅,将识别类别数设为 11 种。为避免参与训练的某类别数目过少而出现无法精确分类的欠拟合现象,以及某类别数目过多而出现在分类过程中过于注重某个特征的学习而导致分类错误的过拟合现象。本研究尽量保证各类别的刺梨果实图像样本数量均衡。

通过 Photoshop CS6 软件,将 2 100 幅大小为 6 016×4 016 像素的原图裁剪为多幅大小为 500×500 像素的完全包含刺梨果实的样本,借助 ACDSee 软件,对裁剪出的样本进行上下翻转以及旋转 45°、90°和 270°,增强图像数据集,同时统一批量重命名为 2018_000001.jpg 格式,最终处理完后得到 8 475 幅样本。再借助 labelImg 对其中 8 175 幅样本进行标签制作,本文使用 POSCAL VOC2007 数据集格式,制作样本标签。

1.3 刺梨果实分类方法

根据刺梨果实自然生长环境,按其成熟程度、独立与相邻、遮挡与否等情况,将刺梨果实图像分为 11 类: 1g0csnot、1g0csyes、1glcsnot、1glcsyes、2g0csnot、2g0csyes、2glcsnot、2glcsyes、ng0csnot、ng1csnot、

ng1csyes; 其中, 1g、2g、ng 分别表示相邻单元刺梨果实数为 1 个、2 个和 3 个以上(包含 3 个), 0cs 表示刺梨果实未成熟, 1cs 表示刺梨果实已成熟(颜色纯黄视为成熟,其余情况均视为未成熟), yes 表示有树叶或树枝等遮挡超过 1/4 而小于 3/4 的刺梨, not 表示没有遮挡或遮挡小于 1/4 或相互重叠而没有被枝叶遮挡的情况,对于 2 个和 3 个的情况,有 1 个未成熟则视为未成熟,有 1 个遮挡超过 1/2 或有枝叶横跨整个刺梨则视为遮挡。刺梨果实图像分类简图如图 2 所示。



注: 图为 1g0csnot 表示独立刺梨未成熟没有遮挡的情况, 其中: g 表示个数, 其前面的数字表示有几个相邻, 0cs 表示未成熟, 1cs 表示成熟, 颜色纯黄视为成熟, 其余情况均视为未成熟。yes 和 not 分别表示有遮挡和无遮挡。
Note: Figure 1g0csnot shows the immature condition of independent *Rosa roxburghii* without occlusion. Among them, g means number, the number before it represents several adjacent, 0cs means immature, 1cs means mature, pure yellow color is considered mature, the rest are considered immature. Yes and not denote occlusion and non-occlusion, respectively.

图 2 刺梨果实图像分类简图

Fig.2 Classification sketch of *Rosa roxburghii* fruit image

2 试验条件及方法

2.1 软件及硬件

本文试验条件为: Ubuntu 16.04、64 位操作系统, 采用 caffe 框架。相机: 尼康(Nikon D750), 常用变焦头: AF-S 尼克尔, 24-120mm f/4G ED VR 镜头。电脑配置: 台式电脑, GeForce GTX 1060 显卡, 6G 显存; Intel(R)Core(TM) i7-8700K 处理器, 主频 3.70 GHz, 磁盘内存 250 GB, 编程语言是 Python 编程语言。

2.2 试验方法

本文选择以 Faster RCNN 作为刺梨果实检测识别的基础网络框架, 根据刺梨果实的图像特征, 对该框架下的 VGG16、VGG_CNN_M1024 及 ZF 3 种训练模型的重要结构参数和模型训练策略进行了改进和优化, 使其能够更好地实现对刺梨果实图像的识别。

Faster RCNN 由 2 部分构成: 特征提取和 RPN+Fast RCNN。首先对特征进行提取, 再进入区域候选网络(region proposal network, RPN), 最后进入 Faster RCNN 的后半部分, 作为特征提取的网络, 再接入区域提取网络, 生成建议矩形框(Proposals)^[30]。各网络主要结构都由卷积层、激励层、池化层、RPN 层、ROI Align 层及全连接层构成, 网络结构如下。

2.2.1 卷积层 (Conv layer)

Faster RCNN 支持输入任意大小的图像, 卷积层输出图像的大小如公式 (1) 所示。

$$\text{output}_{\text{size}} = \frac{\text{input}_{\text{size}} - \text{kernel}_{\text{size}} + 2\text{pad}}{\text{stride}} + 1 \quad (1)$$

式中 $\text{output}_{\text{size}}$ 表示该卷积层输出到下一层的图像大小, $\text{input}_{\text{size}}$ 表示输入该层的图像大小, $\text{kernel}_{\text{size}}$ 表示卷积核大小, pad 表示填充的像素, stride 表示卷积核在图像上滑动的步长。

由于在卷积层图像的每一个像素点都有一个新值, 所以卷积层不会改变图像的大小。

2.2.2 激励层 (ReLU layer)

由于 ReLU 函数的收敛速度非常快, 所以选用修正线性单元函数 (the rectified linear unit, ReLU) 作为激励层函数。

2.2.3 池化层 (Pooling layer)

池化层选用了最大池化 (Max-pooling) 方法, 可以在一定的程度上降低卷积层参数误差造成的估计均值偏移所引起的特征提取的误差。通过卷积层、激励层和池化层完成了对输入图像特征图的提取。

2.2.4 RPN (Region proposal networks)

RPN 可以把一个任意尺度的图像作为输入, 输出一系列的建议矩形框 (object proposals), 每个矩形框都带一个目标框得分 (objectness score)。它是在特征图上提取候选框, 大幅加快了训练速度。

2.2.5 感兴趣区域校准 (ROI Align)

ROI Align 是 Mask-RCNN 框架中提出的一种区域特征聚集方式, 该方式能够解决 ROI Pooling 操作中 2 次量化造成的区域不匹配 (mis-alignment) 的问题^[31]。ROI Align 使用双线性内插的方法获得坐标为浮点数的像素点上的图像数值, 从而将整个特征聚集过程转化为一个连续

的操作, 解决了区域不匹配问题。

2.2.6 全连接层 (FC layer)

全连接层即两层之间的所有神经元都有权重连接, 它将 ROI Align 层输出的特征图 (feature map) 进行全连接操作。

3 网络模型训练

3.1 三种网络模型对比分析

本文使用 ImageNet 预训练的模型初始化权值, 选择交替优化训练方式 (alternating optimization), 在 VGG16、VGG_CNN_M1024 以及 ZF 3 种训练模型下进行了训练。

VGG16 训练模型具有 13 个卷积层, 13 个激励层, 4 个池化层, 是一种用于数据分别较多, 数据集较大的大型网络; VGG_CNN_M1024 有 5 个卷积层, 5 个激励层, 2 个池化层, 是一种中型网络; ZF 也有 5 个卷积层, 5 个激励层, 2 个池化层, 是一种用于分类较少数据较小的网络。

VGG16、VGG_GNN_M_1024 以及 ZF 3 种神经网络参数调整如下: 样本在 4 个阶段的训练迭代总次数为 280 000 次, 初始学习率设置为 0.001, 批处理图像为 128 幅, RPN 第一、二阶段 stepsize 均为 60 000, 最大迭代次数为 80 000, Faster RCNN 第一、二阶段 stepsize 均为 40 000, 最大迭代次数为 60 000, 动量因子均为 0.9, weight_decay 均为 0.000 5, 分别在 VGG16、VGG_GNN_M_1024 和 ZF 模型下训练出来的精确率-召回率曲线 (precision-recall, PR) 如图 3 所示。从图 3 中可以看出, 以 VGG16 网络训练出来的 P-R 曲线图效果最佳, 其召回率更接近于 1, 说明在这 3 种网络模型中, VGG16 网络下训练出来的模型所检测出的目标矩形框与制作标签时所画的矩形框重叠度最高。

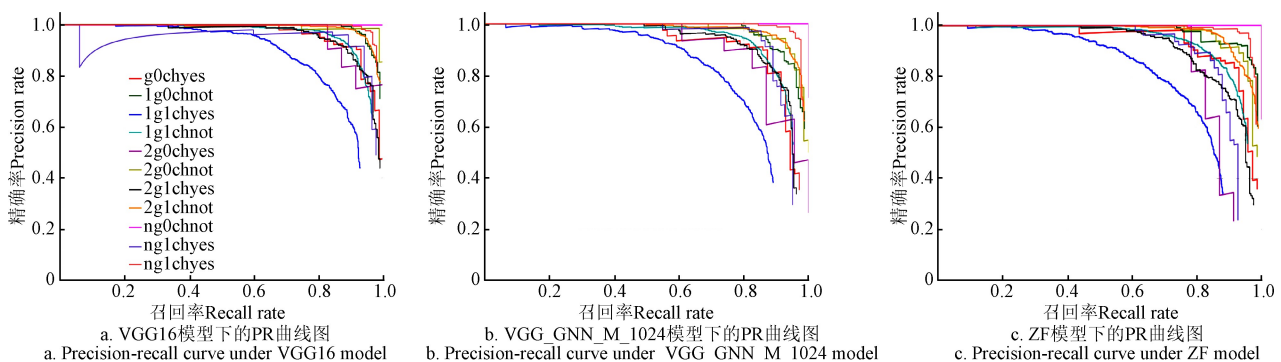


图 3 VGG16、VGG_GNN_M_1024 和 ZF 模型下的 PR 曲线图
Fig.3 Precision-recall curve under VGG16, VGG_GNN_M_1024 and ZF model

VGG16、VGG_GNN_M_1024 以及 ZF 3 种网络模型下各类别的训练精度如表 1 所示, 3 种网络模型下训练性能对比如表 2 所示。

从表 1 可以看出, 3 种网络训练模型中, VGG16 网络模型训练出的模型平均精度最高, 最低精度为 0.837 4, 最高精度为 1.000 0。所以本文采用 VGG16 网络模型进行训练。

3.2 模型训练

随机从 8 175 幅样本中选择 6 540 (80%) 幅作为训练验证集 (trainval), 作为训练集与验证集, 其余 20% 作为测试集 (test); trainval 的 80% 作为训练集, 其余 20% 作为验证集; 其余未参与训练的 300 幅用于对最终模型进行检验。

表 1 3 种网络模型下各类别的训练精度
Table 1 Accuracy under three network models

类别 Categories	VGG16	VGG_GNN_M_1024	ZF
g0chyes	0.940 0	0.875 1	0.882 3
1g0chnot	0.908 5	0.899 4	0.900 8
1g1chyes	0.837 4	0.758 4	0.743 9
1g1chnot	0.985 5	0.880 9	0.878 0
2g0chyes	0.964 2	0.904 3	0.833 7
2g0chnot	0.984 3	0.950 5	0.900 0
2g1chyes	0.891 8	0.877 9	0.861 3
2g1chnot	0.906 1	0.901 6	0.900 2
ng0chnot	1.000 0	1.000 0	1.000 0
ng1chyes	0.884 7	0.885 0	0.865 9
ng1chyes	0.909 1	0.908 5	0.907 4

表 2 3 种网络模型下训练性能对比

Table 2 Comparison of training performance under three network models

项目 Item	VGG16	VGG_GNN_M_1024	ZF
Max	1.000 0	1.000 0	1.000 0
Min	0.837 4	0.758 4	0.743 9
Mean AP	0.920 1	0.894 7	0.879 4

本文的网络可以直接将刺梨样本作为数据输入, 通过归一化将样本缩放为 500 像素×500 像素大小, 在卷积层中, 对归一化后的图像进行了扩边处理 ($pad=1$, 即图像四周填充一圈 0 像素点), 以保证不改变输入和输出矩阵的大小。使用 Relu 函数作为激励函数。经过一个大小为 2×2 的核进行池化 (下采样), 图像变为 250×250 像素大小, 网络的训练阶段采用批量随机梯度下降法, 经过 13 个卷积层, 13 个 Relu 层, 4 个池化层后, 生成大小为 31×31 像素大小的特征图, 该特征图的每一个特征点都与原图 16×16 像素大小的区域对应。RCNN 使用选择性搜索 (Selective Search) 方法^[8]生成检测框, 极大地提升了检测框生成速度。

本文选用损失函数 Softmax loss (对应 Softmax 回归分类器) 进行网络性能的对比分析。基于 VGG16 的 Faster RCNN 网络结构框架图如图 4 所示。

通过对 300 幅未参与训练的刺梨果实图像进行检验, 当图像大小为 500 像素×500 像素时检验速度最快, 平均速度可达到 0.2 s/幅。通过对 300 幅检验图像中实际存在的刺梨类别及其个数与检验出来的数据进行对比分析和对比, 即可得到识别准确率。

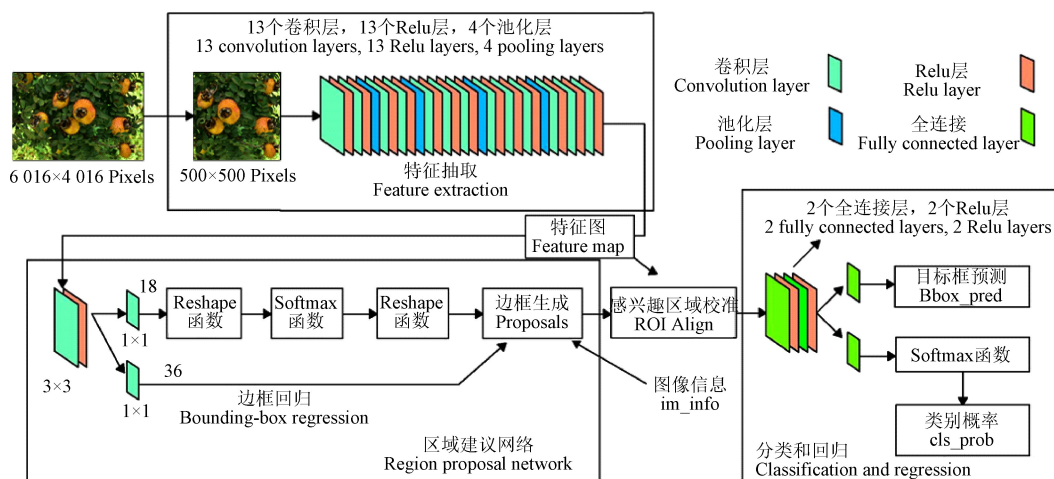


图 4 基于 VGG16 的 Faster RCNN 网络结构框架图

Fig.4 Framework diagram of Faster RCNN based on improved VGG16 network

改进的卷积神经网络进行刺梨果实识别模型训练的步骤如下:

1) 对刺梨果实原图进行预处理, 根据刺梨生长状况进行分类, 尽量使各类别刺梨果实图像数量相近;

2) 根据改进的卷积神经网络对输入数据大小的要求, 统一裁剪出 500×500 像素大小的样本, 同时通过旋转、镜像等扩展样本数量, 制作训练样本集;

3) 采用交叉验证的模型训练方法, 分别使用 VGG16、VGG_GNN_M_1024 以及 ZF 模型进行训练, 对比分析后, 决定使用模型训练精度较高的 VGG16 进行最终模型的训练。通过对输入网络的刺梨样本进行固定缩放至 500 像素×500 像素, 采用 ReLU 激励函数以及最大池化的下采样法, 经过 13 个卷积层+13 个 relu 层+4 个池化层提取特征图; 经过 RPN 网络的一个 3×3 的卷积, 生成前景锚点与边界框回归偏移量, 计算出

建议候选框;

4) 通过反向传播算法调节更新网络参数, 进行参数的调整与改进;

5) 通过区域特征聚集的方式, 从特征图中提取出建议候选框, 输入全连接层以及 softmax 网络进行分类, 得到各类的平均精度值 (average precision, AP) 以及所有类别的平均精度均值 (mean average precision, mAP)。最终得到矩形框几何中心位置, 即刺梨的近似质心坐标, 达到对刺梨的准确识别与定位。

4 模型训练结果对比分析

4.1 模型精度

改进的 VGG16 网络模型下训练的各阶段 (stage1_rpn、stage1_fast_rcnn、stage2_rpn、stage2_fast_rcnn) Loss 曲线图如图 5 所示。

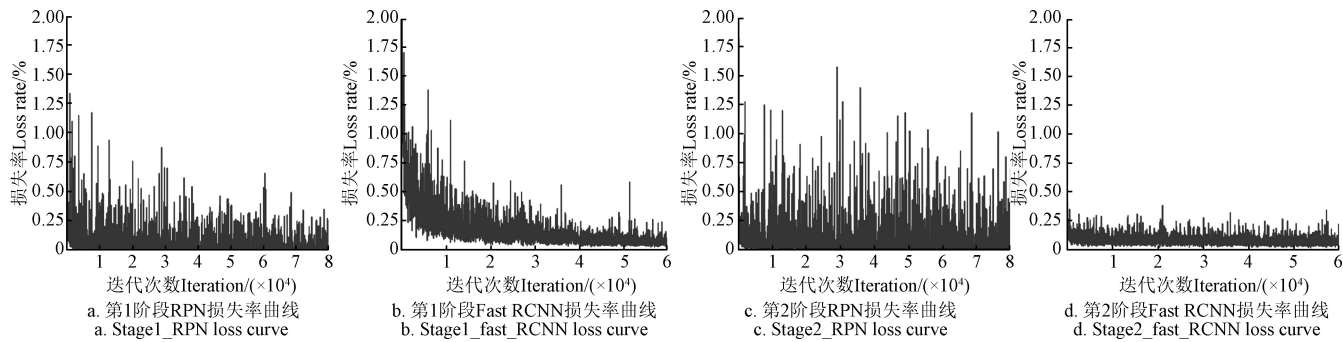


图 5 改进后的 VGG16 网络模型下训练的各阶段 Loss 曲线图
Fig.5 Loss curve of each stage of training under improved VGG16 network model

从以上训练过程的 4 幅损失率-迭代次数关系图可以看出，stage1_rpn 和 stage2_rpn 为区域生成网络（RPN）阶段，在这一训练阶段将生成大量的检测框，其损失率较低，说明生成的大多数检测框与标签中的目标框重叠度较高。stage1_fast_rcnn 和 stage2_fast_rcnn 为 Fast RCNN 网络训练过程的损失率。stage1_rpn 和 stage2_rpn 生成的检测框进行判别，区分目标是背景还是目标对象，并返回含有目标对象的矩形框与感兴趣区域的信息，stage2_rpn 再生成大量检测框，由 stage2_fast_rcnn 阶段接着上一阶段的数据进行判别。从 stage1_fast_rcnn 和 stage2_fast_rcnn 可以看出，当迭代次数达到 40 000 次时曲线收敛，损失率约为 0.05% 左右，训练结果比较理想。

4.2 模型的效果验证

本文的研究是为刺梨果实的智能化采摘提供刺梨果实的识别、类别及位置等提供依据，在采摘过程中，执行末端的允许误差半径为 10 mm，因此，需要识别

出刺梨果实的大部分区域即可实现刺梨果实的识别。当检测的红框区域与刺梨果实重叠部分超过约 3/4 且红框标记的类别与实际刺梨果实类别相同时，视为正确检测。

本文所用的模型性能评价指标有召回率（recall， R ）、准确率（precision， P ）以及 F_1 值， F_1 值计算公式如公式（2）所示，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）。

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (2)$$

式中 P 为准确率，%， R 为召回率，%。

根据对刺梨果实不同尺寸大小的图像进行检测，当图像尺寸大小为 500 像素×500 像素时检测用时最少，检测平均速度约为 0.2 s 每幅图像。本文将没参与训练的 300 幅大小为 500 像素×500 像素的图像对模型进行了验证，其中包含刺梨的 11 种类别。未参与训练的 300 幅刺梨图像在模型中检测的结果如表 3 所示。

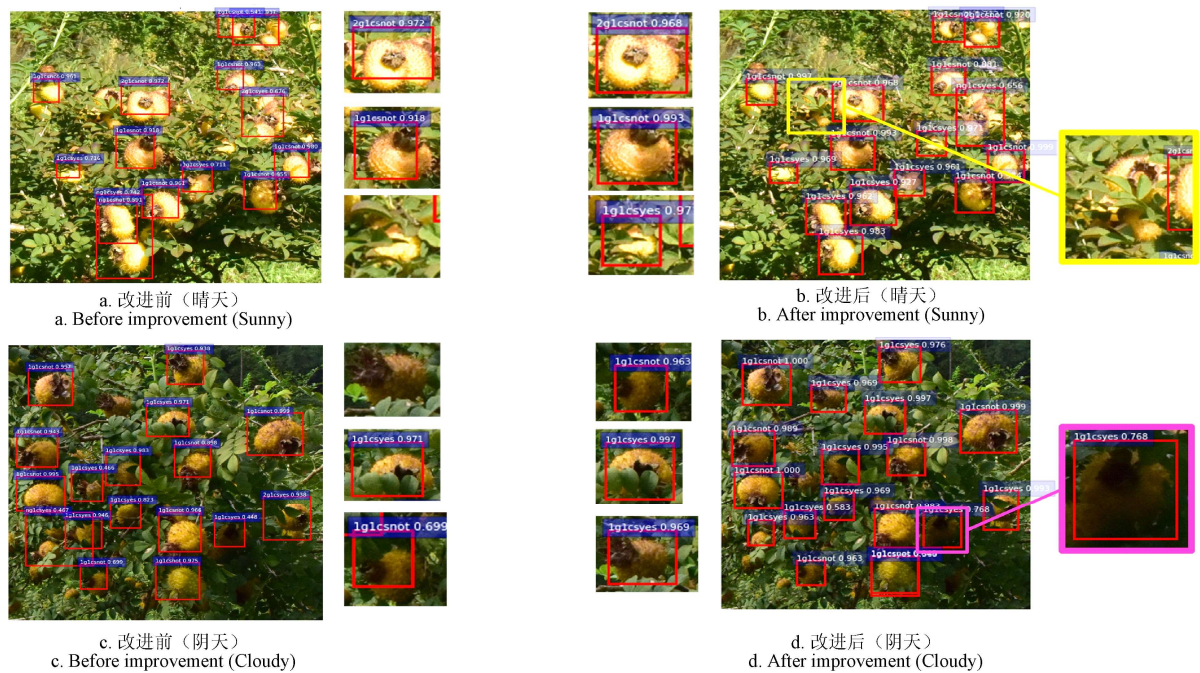
表 3 模型检测结果
Table 3 Test results of model detection

特征类别 Feature category	正确检测果实数 Recognized fruit number	误以为正确果实数 Mistakenly correct number	误以为错误果实数 Mistakenly error number	召回率 Recall /%	准确率 Precision /%	F_1 值 F_1 value/%
1g0chyes	235	11	26	90.00	95.53	92.68
1g0chnot	345	20	28	92.49	94.52	93.49
1g1chyes	344	30	20	94.51	91.98	93.23
1g1chnot	474	35	15	96.93	93.12	94.99
2g0chyes	70	4	16	81.40	94.59	87.50
2g0chnot	149	25	10	93.71	85.63	89.49
2g1chyes	147	10	20	88.02	93.63	90.24
2g1chnot	130	10	16	89.04	92.86	90.91
ng0chnot	140	7	15	90.32	95.24	92.71
ng1chyes	80	9	10	88.88	89.89	89.38
ng1chnot	76	5	8	90.48	93.83	92.12

由表 3 可知，本文选择 F_1 值来度量识别模型的优劣， F_1 值即准确率与召回率的一个平衡点，能同时考虑到准确率和召回率。 F_1 值均高于 87.50%，最高可达到 94.99%，说明本文训练出来的识别模型检测效果较好。改进前后训练的模型检测效果图如图 6 所示。

由图 6 可以看出，使用感兴趣区域池化（ROI pooling）训练出来的模型对刺梨果实的检测存在较大的偏差，改

进为感兴趣区域校准（ROI align）方法后，检测矩形框的精度有明显的改善。有极少数刺梨由于果实过小、遮挡面积较大或是模糊不清而不能被检测到。同时，也有小部分刺梨类别识别有误，一方面是由于在制作标签时带有人为判别误差导致，另一方面也有数据集样本不够多的因素在内。从效果图可以看出，该识别模型对光线较暗情况下的目标也能够检测到。



注：图中小图为具体果实识别情况。
Note: The small picture shows the specific situation of fruit recognition.

图 6 改进前后训练的模型检测效果图
Fig.6 Model inspection renderings of training before and after improvement

4.3 对比分析

目前，还没有刺梨果实图像识别相关研究，本文将应用卷积神经网络 Faster RCNN（ZFNet 网络）、LeNet 网络对猕猴桃、苹果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 Faster RCNN（VGG16 网络）的识别准确率及识别速率。具体对比分析见表 4 基于卷积神经网络的果实识别算法比较所示。

表 4 基于卷积神经网络的果实识别算法比较
Table 4 Comparison of fruit recognition algorithms based on convolutional neural networks

文献 References	类别 Category	拍摄方式 Shooting mode	图像特征 Image feature	作物特征 Crop characteristics	神经网络 Network	识别准确率 Recognition accuracy/%	单个果实识别时间 Single fruit recognition time/s
Fu 等 ^[32]	猕猴桃 Kiwi	远距离底部成像 Long distance and bottom imaging	多簇果实 Multiple clusters of fruits	遮挡、重叠、相邻、独立 Occlusion, overlap, adjacency, separated	Faster RCNN (ZFNet 网络)	92.30	0.27
傅隆生等 ^[3]	猕猴桃 Kiwi	远距离底部成像 Long distance and bottom imaging	多簇果实 Multiple clusters of fruits	遮挡、重叠、相邻、独立 Occlusion, overlap, adjacency, separated	LeNet 网络	89.29	0.27
程鸿芳等 ^[25]	苹果 Apple	远距离底部成像 Long distance and bottom imaging	单个果实 Single fruits	遮挡、重叠、相邻、独立 Occlusion, overlap, adjacency, separated	LeNet 网络	93.79	-
本文算法	刺梨 <i>Rosa roxbunghii</i>	近距离多角度成像 Close distance and multi angle imaging	单个果实 Single fruits	遮挡、重叠、相邻、独立 Occlusion, overlap, adjacency, separated	Faster RCNN (VGG16 网络)	95.16	0.20

从表 4 对猕猴桃、苹果、刺梨等果实识别的准确率和识别时间比较可以看出，本文的算法 Faster RCNN（VGG16 网络）对刺梨果实的识别精度高，达到 95.16%；在单个果实识别识别速度更快，平均每个刺梨果实的识别时间约为 0.20 s，平均时间上具有一定的优势，比 Fu 等^[32]的方法要快 0.07 s。同时，本文提出的算法对弱光照和强光照条件下的刺梨果实都有较好的识别效果，适合在复杂的田园环境中对刺梨果实进行有效地识别检测，能够达到刺梨果实自动化识别定位采摘的要求。

5 结 论

1) 为了实现刺梨果实的自动化采摘，本文建立了一种基于卷积神经网络的田间刺梨果实识别方法。通过对 Faster RCNN 架构下的 VGG16、VGG_GNN_M_1024 以及 ZF 网络模型的结构和参数进行调整优化，对比分析后选择了适用于刺梨果实识别模型训练的 VGG16 网络，本文训练出来的识别模型对自然条件下刺梨果实识别率较高，能够为刺梨果实的采摘奠定数据基础。

2) 本文中将 Faster RCNN 框架中的感兴趣区域池化

方法(ROI Pooling)改为 Mask RCNN 中提出的一种区域特征聚集方式-ROI Align, 提高了模型检测精度。本文算法对图像中刺梨果实的平均识别速度能够达到 0.2 s/个, F_1 值最低为 87.50%, 最高达 94.99%, 能够满足刺梨果实识别采摘的要求。

本文利用卷积神经网络对刺梨果实图像特征进行深度提取的文章, 为刺梨果实的智能化识别与采摘奠定了一定的基础, 为刺梨果实的自动化采摘技术的研究开启了新的征程。

[参 考 文 献]

- [1] 唐玲, 陈月玲, 王电, 等. 刺梨产品研究现状和发展前景[J]. 食品工业, 2013, 34(1): 175—178.
Tang Ling, Chen Yueling, Wang Dian, et al. The research status and the development prospect of *Rosa roxburghii* Tratt products[J]. Food Industry, 2013, 34(1): 175—178. (in Chinese with English abstract)
- [2] 孙世鹏, 李瑞, 谢洪起, 等. 基于机器视觉的冬枣病害检测[J]. 农机化研究, 2018(9): 183—188.
Sun Shipeng, Li Rui, Xie Hongqi, et al. Detection of winter jujube diseases based on machine vision[J]. Journal of Agricultural Mechanization Research, 2018(9): 183—188. (in Chinese with English abstract)
- [3] 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, Elkamil Tola, 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 (Transactions of the CSAE) 2018, 34(2): 205—211. (in Chinese with English abstract)
- [4] 孙云云, 江朝晖, 董伟, 等. 基于卷积神经网络和小样本的茶树病害图像识别[J]. 江苏农业学报, 2019, 35(1): 48—55.
Sun Yunyun, Jinag Zhaohui, Dong Wei, et al. Image recognition of tea plant disease based on convolution neural network and small samples[J]. Jiangsu Journal of Agricultural Sciences, 2019, 35(1): 48—55. (in Chinese with English abstract)
- [5] Przybyło J, Jabłoński M. Using deep convolutional neural network for oak acorn viability recognition based on color images of their sections[J]. Computers and Electronics in Agriculture, 2019, 156: 409—499.
- [6] 夏为为, 夏哲雷. 基于卷积神经网络的宫颈癌细胞图像识别的改进算法[J]. 中国计量大学学报, 2018, 29(4): 439—444.
Xia Weiwei, Xia Zhelei. An improved algorithm for cervical cancer cell image recognition based on convolution neural networks[J]. Journal of China University of Metrology, 2018, 29(4): 439—444. (in Chinese with English abstract)
- [7] Mane D T, Kulkarni U V. Visualizing and understanding customized convolutional neural network for recognition of handwritten marathi numerals[J]. Procedia Computer Science, 2018, 132: 1123—1137.
- [8] Rabby A S A, Haque S, Abujar S, et al. Using convolutional neural network for bangla handwritten recognition[J]. Procedia Computer Science, 2018, 143: 603—610.
- [9] Trivedi A, Srivastava S, Mishra A, et al. Hybrid evolutionary approach for devanagari handwritten numeral recognition using convolutional neural network[J]. Procedia Computer Science, 2018, 125: 525—532.
- [10] Li Ya, Wang Guangrun, Nie Lin, et al. Distance metric optimization driven convolutional neural network for age invariant face recognition[J]. Pattern Recognition, 2018, 75: 51—62.
- [11] O Toole A J, Castillo C D, Parde C J, et al. Face space representations in deep convolutional neural networks[J]. Trends in Cognitive Sciences, 2018, 22(9): 794—809.
- [12] Jiao Licheng, Zhang Sibo, Li Lingling, et al. A modified convolutional neural network for face sketch synthesis[J]. Pattern Recognition, 2018, 76: 125—136.
- [13] Banerjee S, Das S. Mutual variation of information on transfer-CNN for face recognition with degraded probe samples[J]. Neurocomputing, 2018, 310: 299—315.
- [14] Yang Meng, Wang Xing, Zeng Guohang, et al. Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person[J]. Pattern Recognition, 2017, 66: 117—128.
- [15] Aminikhanghahi S, Cook D J. Enhancing activity recognition using CPD-based activity segmentation[J]. Pervasive and Mobile Computing, 2019, 53: 75—89.
- [16] Hassan M M, Uddin M Z, Mohamed A, et al. A robust human activity recognition system using smartphone sensors and deep learning[J]. Future Generation Computer Systems, 2018, 81: 307—313.
- [17] Nweke H F, Teh Y W, Al-Garadi M A, et al. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges[J]. Expert Systems with Applications, 2018, 105: 233—261.
- [18] San-Segundo R, Blunck H, Moreno-Pimentel J, et al. Robust human activity recognition using smartwatches and smartphones[J]. Engineering Applications of Artificial Intelligence, 2018, 72: 190—202.
- [19] Ignatov A. Real-time human activity recognition from accelerometer data using convolutional neural networks[J]. Applied Soft Computing, 2018, 62: 915—922.
- [20] 张汇, 杜煜, 宁淑荣, 等. 基于 Faster RCNN 的行人检测方法[J]. 传感器与微系统, 2019, 38(2): 147—149.
Zhang Hui, Du Yu, Ning Shurong, et al. Pedestrian detection method based on Faster RCNN[J]. Transducer and Microsystem Technologies, 2019, 38(2): 147—149. (in Chinese with English abstract)
- [21] 李宗民, 邢敏敏, 刘玉杰, 等. 结合 Faster RCNN 和相似性度量的行人目标检测[J]. 图学学报, 2018, 39(5): 901—908.
Li Zongmin, Xing Minmin, Liu Yujie, et al. Pedestrian object detection based on Faster RCNN and similarity measurement[J]. Journal of graphics, 2018, 39(5): 901—908. (in Chinese with English abstract)
- [22] 张琦, 胡广地, 李雨生, 等. 改进 Fast-RCNN 的双目视觉车辆检测方法[J]. 应用光学, 2018, 39(6): 832—838.
Zhang Qi, Hu Guangdi, Li Yusheng, et al. Binocular vision vehicle detection method based on improved Fast-RCNN[J]. Journal of Applied Optics, 2018, 39(6): 832—838. (in Chinese with English abstract)
- [23] 史凯静, 鲍泓, 徐冰心, 等. 基于 Faster RCNN 的智能车道路前方车辆检测方法[J]. 计算机工程, 2018, 44(7): 36—41.
Shi Kaijing, Bao Hong, Xu Binxin, et al. Forward vehicle detection method of intelligent vehicle in road based on Faster RCNN[J]. Computer Engineering, 2018, 44(7): 36—41. (in Chinese with English abstract)
- [24] 车金庆, 王帆, 吕继东, 等. 重叠苹果果实的分离识别方法[J]. 江苏农业学报, 2019, 35(2): 469—475.
Che Jinqing, Wang Fan, Lv Jidong, et al. Separation and recognition method for overlapped apple fruits[J]. Jiangsu Journal of Agricultural Sciences, 2019, 35(2): 469—475. (in Chinese with English abstract)

- [25] 程鸿芳, 张春友. 自然场景下基于改进 LeNet 卷积神经网络的苹果图像识别技术研究[J]. 食品与机械, 2019, 35(3): 155—158.
Cheng Hongfang, Zhang Chunyou. Research on apple image recognition technology based on improved LeNet convolution neural network in natural scene[J]. Food and Machinery, 2019, 35(3): 155—158. (in Chinese with English abstract)
- [26] Park K, Hong Y K, Kim G H, et al. Classification of apple leaf conditions in hyper-spectral images for diagnosis of Marssonina blotch using mRMR and deep neural network[J]. Computers and Electronics in Agriculture, 2018(148): 179—187.
- [27] 詹文田, 何东健, 史世莲. 基于 Adaboost 算法的田间猕猴桃识别方法[J]. 农业工程学报, 2013, 29(23): 140—146.
Zhan Tianwen, He Dongjian, Shi Shilian. Recognition of kiwifruit in field based on adaboost algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(23): 140—146. (in Chinese with English abstract)
- [28] 毕松, 高峰, 陈俊文, 等. 基于深度卷积神经网络的柑橘目标识别方法[J]. 农业机械学报, 2019, 50(5): 182—186.
Bi Song, Gao Feng, Chen Junwen, et al. Detection method of citrus based on deep convolution neural network[J]. Transactions of The Chinese Society for Agricultural Machinery, 2019, 50(5): 182—186. (in Chinese with English abstract)
- [29] 雷倩, 杨永发. 便携式变径球形刺梨采摘器的设计[J]. 林业机械与木工设备, 2017, 45(3): 26—28.
Lei Qian, Yang Yongfa. Design of a portable variable-diameter spherical *Rosa roxburghii* Picker[J]. Forestry machinery and woodworking equipment, 2017, 45(3): 26—28. (in Chinese with English abstract)
- [30] Shaoqing Ren, Kaiming He, Ross Girshick, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137—1149.
- [31] Kaiming He, Georgia Gkioxari, Piotr Dollar, et al. Mask R-CNN[C]. in ICCV, 2017.
- [32] Fu Longsheng, Feng Yali, Majeed Yaqoob, et al. Kiwifruit detection in field images using Faster R-CNN with ZFNet[J]. IFAC-Papers OnLine, 2018, 51(17): 45—50.

Recognition of *Rosa roxburghii* in natural environment based on improved Faster RCNN

Yan Jianwei^{1,2,3}, Zhao Yuan¹, Zhang Lewei¹, Su Xiaodong¹, Liu Hongyun¹,
Zhang Fugui^{1,3*}, Fan Weiguo², He Lin^{1,4}

(1. College of Mechanical Engineering, Guizhou University, Guiyang 550025, China; 2. National Forestry and Prairie Bureau Rosa roxburghii Engineering Technology Research Center, Guiyang 550025, China; 3. Mountain Agriculture Intelligent Equipment Engineering Research Center of Guizhou Province, Guiyang 550025, China; 4. Liupanshui Normal University, Liupanshui, 553004, China)

Abstract: *Rosa roxburghii* is widely distributed in warm temperate zone and subtropical zone, mainly in Guizhou, Yunnan, Sichuan and other places in China. Panxian and Longli are the most abundant the most varieties and the highest yield *Rosa roxburghii* resources in Guizhou. The harvesting of *Rosa roxburghii* fruit is the most time-consuming and labor-consuming work in *Rosa roxburghii* production, and its labor input accounts for 50%-70% of the production process. Hand-picking of *Rosa roxburghii* fruit is of high cost, high labor intensity and low picking efficiency. In recent years, convolutional neural network has been widely used in target recognition and detection. However, there is no relevant literature on the application of neural network in *Rosa roxburghii* fruit recognition. In this paper, in order to realize rapid and accurate identification of *Rosa roxburghii* fruits in natural environment, according to the characteristics of *Rosa roxburghii* fruits, the structure and parameters of VGG16, VGG_CNN_M1024 and ZF network models under the framework of Faster RCNN were optimized by comparing them. The convolutional neural network adopted bilinear interpolation method and selected alternating optimization training method of Faster RCNN. ROI Pooling in convolutional neural network is improved to ROI Align regional feature aggregation. Finally, VGG16 network model is selected to make the target rectangular box in the detection result more accurate. 6 540 (80%) of 8 175 samples were selected randomly as training validation set (trainval), the remaining 20% as test set, 80% as training set, the remaining 20% as validation set, and the remaining 300 samples that were not trained were used to test the final model. The recognition accuracy of the network model for 11 *Rosa roxburghii* fruits was 94.00%, 90.85%, 83.74%, 98.55%, 96.42%, 98.43%, 89.18%, 90.61%, 100.00%, 88.47% and 90.91%, respectively. The average recognition accuracy was 92.01%. The results showed that the recognition model trained by the improved algorithm had the lowest recall rate of 81.40%, the highest recall rate of 96.93%, the lowest accuracy rate of 85.63%, the highest 95.53%, and the lowest F_1 value of 87.50%, the highest 94.99%. Faster RCNN (VGG16 network) has high recognition accuracy for *Rosa roxburghii* fruit, reaching 95.16%. The recognition speed of single fruit is faster, and the average recognition time of each *Rosa roxburghii* fruit is about 0.2 seconds. The average time has some advantages, which is 0.07 s faster than the methods of Fu Longsheng. In this paper, a Faster RCNN *Rosa roxburghii* fruit recognition network model based on improved VGG16 is proposed, which is suitable for *Rosa roxburghii* fruit recognition model training. The algorithm proposed in this paper has good recognition effect for *Rosa roxburghii* fruit under weak and strong illumination conditions, and is suitable for effective recognition and detection of *Rosa roxburghii* fruit in complex rural environment. This paper is the first study on the depth extraction of *Rosa roxburghii* fruit image features by using convolution neural network. This research has high recognition rate and good real-time performance under natural conditions, and can meet the requirements of automatic identification and positioning picking of *Rosa roxburghii* fruit. It lays a certain foundation for intelligent identification and picking of *Rosa roxburghii* fruit, and opens a new journey for the research of automatic picking technology of *Rosa roxburghii* fruit.

Keywords: convolutional neural network; Faster RCNN; machine vision; deep learning; *Rosa roxburghii*; target recognition