

基于改进 SSD 的柑橘实时分类检测

李善军^{1,2,3,4,5}, 胡定一^{1,2}, 高淑敏^{1,2}, 林家豪^{1,2}, 安小松^{1,2}, 朱明^{1,2*}

(1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070;
3. 国家现代农业(柑橘)产业技术体系, 武汉 430070; 4. 国家柑橘保鲜技术研发专业中心, 武汉 430070;
5. 农业农村部柑橘全程机械化科研基地, 武汉 430070)

摘要:针对人工分拣柑橘过程中, 检测表面缺陷费时费力的问题, 该文提出了一种基于改进 SSD 深度学习模型的柑橘实时分类检测方法。在经改装的自制打蜡机试验台架下采集单幅图像含有多类多个柑橘的样本 2 500 张, 随机选取其中 2 000 张为训练集, 500 张为测试集, 在数据集中共有正常柑橘 19 507 个, 表皮病变柑橘 9 097 个, 机械损伤柑橘 4 327 个。该方法通过单阶段检测模型 SSD-ResNet18 对图片进行计算和预测, 并返回图中柑橘的位置与类别, 以此实现柑橘的分类检测。以平均精度 AP (average precision) 的均值 mAP (mean average precision) 作为精度指标, 平均检测时间作为速度指标, 在使用不同特征图、不同分辨率和 ResNet18、MobileNetV3、ESPNetV2、VoVNet39 等 4 种不同特征提取网络时, 进行模型分类检测效果对比试验研究。研究表明, 该模型使用 C4、C5 特征图, 768×768 像素的分辨率较为合适, 特征提取网络 ResNet18 在检测速度上存在明显优势, 最终该模型的 mAP 达到 87.89%, 比原 SSD 的 87.55% 高出 0.34 个百分点, 平均检测时间为 20.27 ms, 相较于原 SSD 的 108.83 ms, 检测耗时降低了 436.90%。该模型可以同时多类多个柑橘进行实时分类检测, 可为自动化生产线上分拣表面缺陷柑橘的识别方面提供技术借鉴。

关键词: 目标识别; 模型; 无损检测; 柑橘; 表面缺陷; 深度学习; SSD; ResNet18

doi: 10.11975/j.issn.1002-6819.2019.24.036

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2019)-24-0307-07

李善军, 胡定一, 高淑敏, 林家豪, 安小松, 朱明. 基于改进 SSD 的柑橘实时分类检测[J]. 农业工程学报, 2019, 35(24): 307—313. doi: 10.11975/j.issn.1002-6819.2019.24.036 http://www.tcsae.org

Li Shanjun, Hu Dingyi, Gao Shumin, Lin Jiahao, An Xiaosong, Zhu Ming. Real-time classification and detection of citrus based on improved single short multibox detector[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(24): 307—313. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2019.24.036 http://www.tcsae.org

0 引言

水果品质分级是水果加工生产线的重要环节, 分级得当可以创造更多的经济价值, 而水果的表面缺陷情况是影响水果品质的重要指标之一^[1]。但目前筛除表面有缺陷水果的工作多以人工为主, 工作量大且耗费人力、财力。因此, 实现水果的自动化分类检测具有重要的意义。

目前, 国内外学者运用多种方法对水果进行表面缺陷识别。李江波等^[2]通过建立照度-反射模型对脐橙表面缺陷进行检测, 总体正确率超过 99%。赵杰文等^[3]利用支持向量机识别缺陷红枣, 识别准确率达到 96.2%。章海亮等^[4]应用高光谱成像技术对柑橘缺陷进行无损检测, 识别率达到 94%。Dong 等^[5]提出了一种基于高光谱成像技术结合主成分分析法和 B-样条光照校正技术的方法对柑橘缺陷进行检测, 准确率达 96.5%。Zou 等^[6]建立了 1 个由

3 台彩色相机组成的系统, 从不同角度拍摄了 9 张苹果的照片, 通过判断是否有 1 张照片中的感兴趣区域超过 1 个来识别表面缺陷的苹果, 误分率为 4.2%。Sharif 等^[7]通过优化加权分割和特征选择, 选出最优特征, 然后将选中的特征输入多类支持向量机进行最终的柑橘病变类型的分类, 该方法在其综合数据集上准确率为 89%。

上述水果缺陷识别方法存在样本较小问题, 且通常只能一次识别单个水果, 识别效率不高。而近年来快速发展的深度学习技术可以较好的解决这些问题。国内外学者也对基于深度学习的农业检测方向做了相关研究, 其方法主要分为语义分割^[8-10]、目标检测^[11-17]、实例分割^[18-20] 3 种。赵德安等^[21]使用 YOLO 模型对复杂背景下的苹果进行定位, 其平均精度 (AP, average precision) 的均值 (mAP, mean average precision) 为 87.71%, 检测视频的帧率达到 60 帧/s。王丹丹等^[22]利用 R-FCN 深度学习模型对疏果前期的苹果进行目标识别, 误识率为 4.9%, 处理一张图像的平均速度为 0.187 s。Dias 等^[23]以苹果花朵为研究对象, 构建了一种深度学习模型, 实现了对苹果花朵的实例分割。Tian 等^[24]通过改进的 YOLO-V3 模型检测不同生长阶段的苹果, F1 值达到 81.7%。

本文提出一种基于 SSD 深度学习模型的采后柑橘实时分类检测方法, 利用该模型可达到同时识别多类多个柑橘的目的, 为实现生产线上实时分拣有缺陷柑橘提供

收稿日期: 2019-10-06 修订日期: 2019-12-07

基金项目: 现代农业(柑橘)产业技术体系建设专项资金项目(CARS-26); 国家重点研发计划(2017YFD0202001); 柑橘全程机械化科研基地建设项目(农计发[2017]19号); 湖北省农业科技创新行动项目

作者简介: 李善军, 副教授, 博士, 主要从事水果生产机械化技术与智能装备研究。Email: shanjunlee@163.com

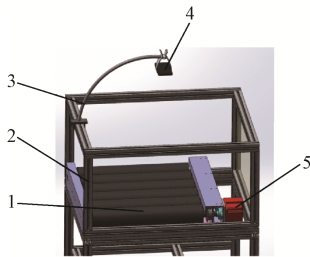
*通信作者: 朱明, 研究员, 博士生导师, 主要从事农产品加工与智能农业装备研究。Email: 13801392760@163.com

技术支持。

1 材料与方法

1.1 数据采集及分类

本文试验所采用的柑橘品种为纽荷尔, 采集于宜昌市秭归县柑橘果园, 相机使用小米 8SE 1200 万像素的后置摄像头, 于自然光照条件下, 模拟产线背景, 在经改装的小型打蜡机上拍摄, 在打蜡机上方固定相机, 采取俯视视角拍摄, 图 1 为图像采集装置示意图。每次拍摄前在打蜡机的滚筒间放置数量不等的柑橘, 之后滚轮以 0.72 r/s 的转速自动旋转, 以此带动柑橘的旋转; 相机每隔 1 秒拍摄 1 张照片, 因此在柑橘旋转过程中可拍摄到柑橘的不同位面, 以此获取更加全面的数据信息, 增加数据量, 拍摄到的图像分辨率为 2 448×2 448 像素, 共拍摄图像样本 2 500 张, 拍摄图像示例如图 2 所示。



1.滚筒 2.框架 3.手机固定夹具 4.高清手机 5.电机
1. Roller 2. Framework 3. Mobile phone fixture 4. HD mobile phone 5. Motor

图 1 图像采集装置

Fig.1 Image acquisition device



图 2 图像样本示例

Fig.2 Example of image

本文将柑橘分为 3 类, 分别为正常柑橘、表皮病变柑橘、机械损伤柑橘。正常柑橘表皮基本无病变斑纹, 可进行正常贩卖; 表皮病变柑橘多有病变斑纹, 其外观受到损伤, 但大部分该类柑橘内质并未受到损坏, 通常可贩卖给榨汁厂与罐头厂进行加工, 存在一定价值; 机械损伤柑橘为表皮已破裂柑橘, 该类柑橘极易腐烂, 通常会被丢弃, 不存在价值。图 3 为 3 类柑橘示意图。



a. 正常
a. Normal
b. 表皮病变
b. Skin lesions
c. 机械损伤
c. Mechanical damage

图 3 3 类柑橘示意图

Fig.3 Examples of 3 citrus species

1.2 生成数据集及其预处理

本文试验在 2 500 张图像样本中随机挑选 2 000 张作为训练集, 剩余 500 张作为测试集, 在 2 500 张图像样本中各类柑橘数量的信息如表 1 所示。

表 1 柑橘数据集种类及其数量

Table 1 Category and quantity of citrus data sets

数据集 Dataset	正常 Normal	表皮病变 Skin lesions	机械损伤 Mechanical damage
训练集 Training set	15 553	7 353	3 455
测试集 Test set	3 954	1 744	872
总计 Total	19 507	9 097	4 327

本文使用 LabelImg 对图像样本进行标定, 采用 COCO 数据集格式。在标定时, 部分柑橘可能既存在表皮病变, 又存在机械损伤, 对于该类柑橘将其标定为机械损伤柑橘; 为增加训练后模型的容错率, 对于柑橘表皮病变或裂纹特征不明显的, 如只在柑橘边缘处有少量裂纹或病变特征的柑橘, 此类柑橘将被标定为正常柑橘。

为了提高训练效果, 使模型的泛化性得以提升, 对数据集使用数据增强方法。鉴于柑橘更换不同方向角度观察都不会改变其特征的特点, 本文使用水平翻转与垂直翻转 2 种数据增强方法。

1.3 试验平台

本文试验基于 Windows10 操作系统, GPU 为 GeForce GTX 1060 (6 GB 显存), 处理器为 Inter(R) Core(TM) i5-8500 CPU@3.00GHz, 运行内存为 16 G。模型的搭建与训练验证通过 Python 语言实现, 基于 PyTorch 深度学习框架, 并行计算框架使用 CUDA 10.1 版本。

2 柑橘分类检测模型与评价指标

2.1 SSD-ResNet18 模型

SSD^[25]是深度学习目标检测中经典且有效的单阶段检测模型, 其首先通过特征提取网络 (backbone) 得到特征图 (feature map), 再从特征图中预测出众多的边界框, 最后通过非极大值抑制 (non-maximum suppression) 最终确定图像中物体的类别与位置。SSD 训练时输入的图像分辨率一般为 300×300 像素或 512×512 像素, 本文中后续与原 SSD 模型相关的试验都选用 512×512 像素的图像分辨率。

2.1.1 ResNet18 特征提取网络

原 SSD 模型以 VGG16^[26]作为特征提取网络, 但 VGG16 网络参数量庞大, 计算速度缓慢, 无法达到生产线上实时分类检测柑橘的要求。因此本文将 SSD 的特征提取网络更改为 ResNet18 深度残差网络, 该网络仅有 18 层, 且具有更快的计算速度, 其浮点计算量仅为 VGG16 网络的 1/10^[27], 这样可以更好满足实时分类检测柑橘的要求, 且训练时可以使模型更快地收敛, 从而减少训练时间。

2.1.2 特征图选取

通过特征提取网络得到的特征图通常会选择每次下采样后, 下次下采样前的一层所得到的结果。本文以 C_n 代表在网络第 n 次下采样后所得到的特征图, 如第 1 次下采样后所得到的特征图就称为 C_1 , 在现阶段的特征提取网络中一般都会选择 C_3 、 C_4 、 C_5 这 3 个特征图。在

目标检测模型中，对于不同的特征图会先安排固定类型的先检框（Anchor），通常为每个特征图 3 种。本文使用 K -means 聚类算法设置 $K=9$ ，通过 2 500 张数据集中的手动标框得到 9 种不同的先检框，其相对百分比范围为 0.084 5~0.151 1，若以 768×768 像素的分辨率为例，则其先检框的尺寸范围为 64.90~116.04 像素，但 C3 的有

有效感受野（Effective receptive field）一般为 32，而 C4 为 64，C5 为 128^[28]，所以 C3 特征图并不一定适用于本课题的柑橘的分类检测。因此，本文选取 C4、C5 这 2 个特征图，并使用 K -means 聚类算法重新得到 6 种不同的先检框，分别应用于 C4、C5。SSD-ResNet18 的模型结构如图 4 所示。

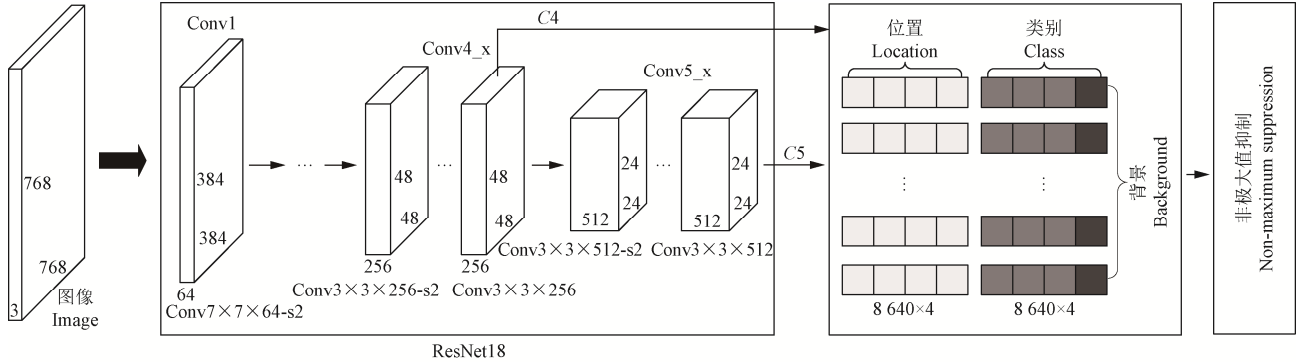


图 4 SSD-ResNet18 模型结构
Fig.4 Architecture of SSD-ResNet18

2.2 损失函数

SSD-ResNet18 模型训练时所使用的损失函数由置信度损失 (L_{conf}) 和位置损失组成 (L_{loc})，其公式为

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

式中 x 表示预测框与真实标签框是否匹配，若匹配则为 1，不匹配则为 0； c 代表 softmax 函数对每一个类别的置信度， l 代表预测框， g 代表真实标签框， N 代表匹配默认框的数量， α 代表置信度损失与位置损失的加权系数。关于该损失函数的详细定义可参考文献[18]。

2.3 评价指标

本文采用 mAP 作为模型的检测精度的评价指标，AP 作为每一类别的检测精度的评价指标。mAP 和 AP 与准确率 (precision, P)、召回率 (recall, R) 有关，准确率和召回率的计算公式如下

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

式中 TP 代表被正确划分到正样本的数量，FP 代表被错误划分到正样本的数量，FN 代表被错误划分到负样本的数量。

通过计算所得准确率与召回率可以绘制出准确率—召回率曲线，该曲线以召回率为横坐标，以准确率为纵坐标，代表某一类别的准确率与召回率情况，而 AP 则是对该曲线进行积分所得，其积分公式为

$$AP = \int_0^1 PR dR \quad (4)$$

mAP 则为所有类别 AP 的平均值，其公式为

$$mAP = \frac{1}{M} \sum_{k=1}^M AP(k) \quad (5)$$

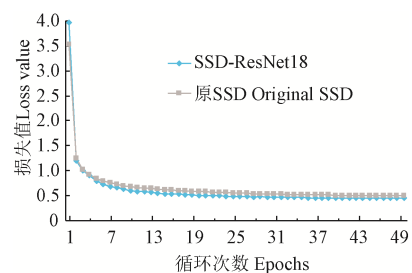
式中 M 代表类别总数， $AP(k)$ 代表第 k 类的 AP 值。

速度评估指标采用模型检测一幅图所耗费的平均时

间，即平均检测时间，单位为 ms。

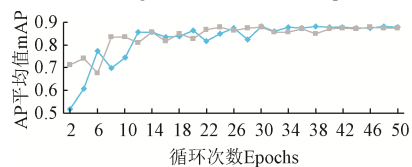
3 模型训练

为加快模型训练收敛速度和提高模型训练效果，本文所搭建的所有模型都加载了基于 ImageNet 的预训练参数。在模型的训练过程中，优化器使用 SGD (stochastic gradient descent)^[29] 算法，batch_size 设置为 12，动量 (momentum) 设置为 0.9，初始学习率 (learning rate) 设置为 0.001，学习率调度器为余弦学习率衰减 (cosine decay)^[29]，权重衰减系数 (weight decay) 设置为 0.000 1，训练 50 个大循环 (epoch)。由于本文训练使用余弦学习率衰减，学习率在最后会下降到 0，因此在训练期间损失值会不断下降，在训练后期，模型的精度变化会趋于平稳，所以在训练结束后使用最后保存的模型，即第 50 次大循环所得的模型，在测试集上进行进一步的验证分析。图 5 为在训练过程中原 SSD 与 SSD-ResNet18 的训练损失与测试集 mAP。



a. 损失与循环次数曲线图

a. Curve diagram of loss value and epochs



b. 测试集 mAP 与循环次数曲线图

b. Curve diagram of mAP of test set and epochs

图 5 原 SSD 与 SSD-ResNet18 训练损失和测试集 mAP
Fig.5 Training loss and mAP of test set of original SSD and SSD-ResNet18

4 结果与分析

4.1 不同特征图对模型分类检测效果的影响

从上文可知特征图 C3 不一定适用于 SSD-ResNet18 模型检测本文的柑橘数据集, 本文对用 C3、C4、C5 特征图的试验结果进行比较研究, 结果如表 2 所示。不使用 C3 的模型的 mAP 比使用 C3 的模型高出 4.2 个百分点, 平均检测时间比其少 1ms, 这可能是因为 C3 设置了符合橘子尺寸的先检框后, 有一部分柑橘会被分配到 C3 进行检测, 而 C3 的有效感受野仅为 32, 无法有效地检测到整个柑橘的特征, 因而会造成误判, 所以去除掉 C3 后, mAP 有明显提升, 而且减少了模型的计算量, 检测速度也得到了提升。

表 2 使用不同特征图的试验结果

特征图 Feature map	mAP/%	平均检测时间 Average detection time/ms
C3、C4、C5	83.69	21.72
C4、C5	87.89	20.72

4.2 不同分辨率对模型分类检测效果的影响

选择较为合适的分辨率进行训练, 有利于模型分类检测效果的提升, 本文分别选取 512×512 像素, 640×640 像素, 768×768 像素, 896×896 像素, 1 024×1 024 像素 5 种分辨率进行对比试验, 结果如表 3 所示。从表 3 可以看出, 分辨率从 512×512 像素提升至 768×768 像素时, 虽然模型的平均检测时间增加了 7.33 ms, 但 mAP 增长了 2.28 个百分点, 有明显提升。而分辨率从 768×768 像素提升至 1 024×1 024 像素的过程中 mAP 已在 88% 上下波动, 变化很小, 且平均检测时间增加 10.04 ms, 说明此时分辨率的增加已经没有意义。综合来看, 768×768 像素的分辨率比较适合该模型。

4.3 各类柑橘的分类检测效果

将测试集在以 768×768 像素为分辨率, 使用 C4、C5 特征图的 SSD-ResNet18 模型与原 SSD 模型上分别进行测试, 得到两种模型对各类柑橘的分类检测效果, 结

果如表 4 所示。原 SSD 与 SSD-ResNet18 的 mAP 相差不多, 但 SSD-ResNet18 的检测时间却是原 SSD 的 1/5, 而且对于机械损伤类别的柑橘其 AP 值比原 SSD 的高出 1.56 个百分点, 说明 SSD-ResNet18 对于该类柑橘的分类检测效果相比于原 SSD 有了一定的提升。另外, 两种模型对于正常柑橘的识别效果最好, 对于机械损伤的柑橘识别效果最差, 这应该与各类柑橘在数据集中的数据量有关。若增加表皮病变与机械损伤这 2 类柑橘的数据量或者实现对柑橘个体的目标追踪, 通过柑橘各个位面的表皮信息进行综合判定, 可能可以进一步地提升模型的识别效果, 特别是对于病变与损伤的柑橘。

表 3 不同分辨率的分类检测效果对比

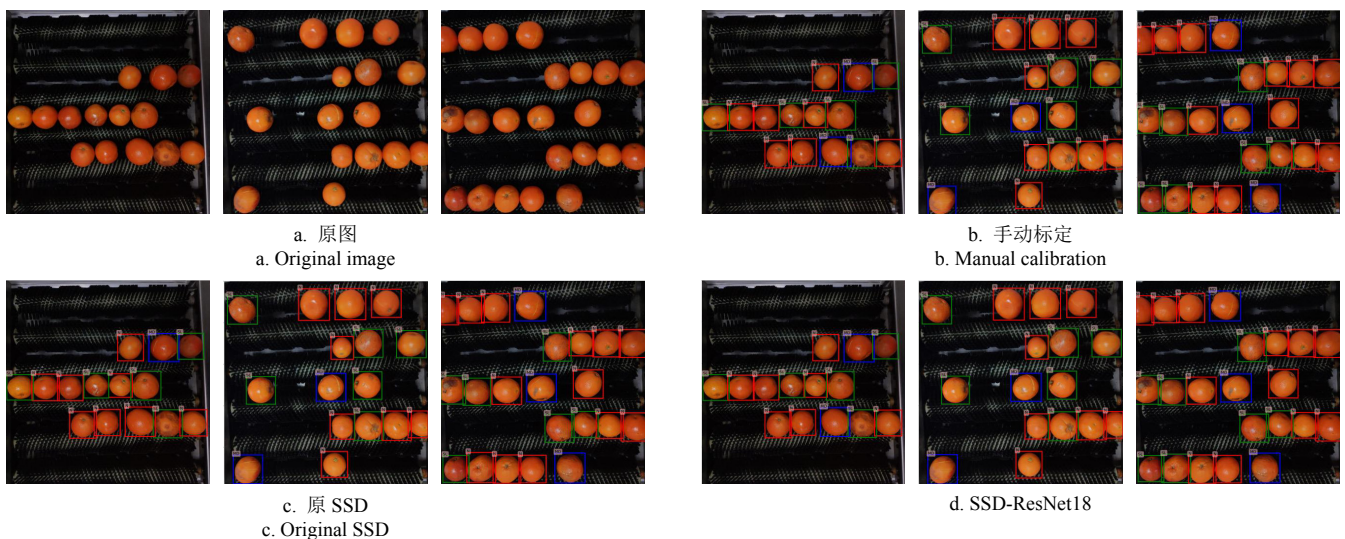
分辨率 Resolution/pixels	mAP/%	平均检测时间 Average detection time/ms
512×512	85.61	13.39
640×640	86.82	17.40
768×768	87.89	20.72
896×896	87.81	26.40
1 024×1 024	88.04	30.76

表 4 原 SSD 与 SSD-ResNet18 的分类检测结果

模型 Models	平均精度 AP/%			mAP/%	平均检测时间 Average detection time/ms
	N	SL	MD		
原 SSD Original SSD	95.38	85.66	81.61	87.55	108.83
SSD-ResNet18	94.72	85.79	83.17	87.89	20.72

注: N: 正常; SL: 表皮病变; MD: 机械损伤
Note: N: Normal; SL: Skin lesions; MD: Mechanical damage

图 6 为 2 种模型分类检测的具体效果。从图 6 可以看出, 2 种模型对于柑橘位置的识别都非常准确, 没有出现漏检柑橘的情况。对于分类情况, SSD-ResNet18 与原 SSD 效果也相差不多, 分类正确率较高, 而且随着 1 张图中柑橘数目的增多, 2 种模型的误检率并没有上升, 说明 2 种模型均有较好的分类效果。



注: 红色方框代表正常; 绿色方框代表表皮病变; 蓝色方框代表机械损伤。
Note: The red box represents normal; The green box represents skin lesions; The blue box represents mechanical damage.

图 6 原 SSD 与 SSD-ResNet18 的柑橘分类检测效果

Fig.6 Detection effect of original SSD and SSD-ResNet18 of citrus

4.4 不同特征提取网络的分类检测效果对比

MobileNetV3^[30]、ESPNetV2^[31]、VoVNet39^[32]是当前阶段优秀的特征提取网络。其中 MobileNetV3 与 ESPNetV2 为轻量级网络, 参数量少, 适合应用于移动端; VoVNet39 为重量级网络, 层数较多, 参数量大, 对复杂困难的分类检测问题有较好的效果。本文在保证模型其他部分不改动的前提下, 分别更换了这 3 种特征提取网络与 ResNet18 进行分类检测效果的对比试验, 结果如表 5 所示。通过 4 种特征提取网络所得的 mAP 差别很小, 说明在分类检测的精准度上, 4 种特征提取网络的效果相近, 但在检测时间上, ResNet18 优于其他的特征提取网络, 比 MobileNetV3 快 10.52 ms, 比 ESPNetV2 快 16.78 ms, 比 VoVNet39 快 36.76 ms, 说明 ResNet18 在提升实时检测的速度上有明显优势。

表 5 4 种特征提取网络的效果对比

Fig.5 Comparison of four feature extraction networks

特征提取网络 Feature extraction networks	mAP/%	平均检测时间 Average detection time/ms
ResNet18	87.89	20.72
MobileNetV3	87.21	31.24
ESPNetV2	87.70	37.50
VoVNet39	87.35	57.48

5 结 论

本文提出了一种基于 SSD-ResNet18 深度学习模型的柑橘生产线实时分类检测方法。可区分正常柑橘、表皮病变柑橘、机械损伤柑橘。本文选取浮点计算量少的 ResNet18 作为特征提取网络, 加快了模型的检测速度, 使用 C4、C5 特征图进行预测, 同时调整数据集分辨率为 768×768 像素, 增加了模型分类检测精度。最终本文模型的 mAP 达到 87.89%, 平均检测时间为 20.72 ms。与原 SSD 模型相比, 分类检测精度相近, 但平均检测时间减少了 88.11ms, 检测速度明显提高。该模型有较高的分类检测精度, 检测速度提高明显, 从而为生产线上自动化分拣表面缺陷柑橘提供参考。

[参 考 文 献]

[1] 应义斌, 饶秀勤, 赵匀, 等. 机器视觉技术在农产品品质自动识别中的应用 (I) [J]. 农业工程学报, 2000, 16(1): 103—108.
Ying Yibin, Rao Xiuqin, Zhao Yun, et al. Application of machine vision technique to quality automatic identification of agricultural products(I) [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2000, 16(1): 103—108. (in Chinese with English abstract)

[2] 李江波, 饶秀勤, 应义斌. 基于照度-反射模型的脐橙表面缺陷检测[J]. 农业工程学报, 2011, 27(7): 338—342.
Li Jiangbo, Rao Xiuqin, Ying Yibin, et al. Detection of navel surface defects based on illumination-reflectance model [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(7): 338—

342. (in Chinese with English abstract)

[3] 赵杰文, 刘少鹏, 邹小波, 等. 基于支持向量机的缺陷红枣机器视觉识别[J]. 农业机械学报, 2008, 39(3): 113—115.
Zhao Jiewen, Liu Shaopeng, Zou Xiaobo, et al. Recognition of defect chinese dates by machine vision and support vector machine [J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(3): 113—115. (in Chinese with English abstract)

[4] 章海亮, 高俊峰, 何勇. 基于高光谱成像技术的柑橘缺陷无损检测[J]. 农业机械学报, 2013, 44(9): 177—181.
Zhang Hailiang, Gao Junfeng, He Yong. Nondestructive detection of citrus defection using hyper-spectra imaging technology [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(9): 177—181. (in Chinese with English abstract)

[5] Dong C, Yang Y, Zhang J, et al. Detection of thrips defect on green-peel citrus using hyperspectral imaging technology combining PCA and B-spline lighting correction method [J]. Journal of Integrative Agriculture, 2014, 13(10): 2229—2235.

[6] Zou X B, Zhao J W, Li Y X, et al. In-line detection of apple defects using three color cameras system [J]. Computers and Electronics in Agriculture, 2010, 70(1): 129—134.

[7] Sharif M, Khan M A, Iqbal Z, et al. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection [J]. Computers and Electronics in Agriculture, 2018, 150: 220—234.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[9] Kestur R, Meduri A, Narasipura O. MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard [J]. Engineering Applications of Artificial Intelligence, 2019, 77: 59—69.

[10] Li Y, Cao Z, Lu H, et al. In-field cotton detection via region-based semantic image segmentation [J]. Computers and Electronics in Agriculture, 2016, 127: 475—486.

[11] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C] // Advances in Neural Information Processing Systems. 2015: 91-99.

[12] 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. 2016: 779—788.

[13] 孙哲, 张春龙, 葛鲁镇, 等. 基于 Faster R-CNN 的田间西兰花幼苗图像检测方法 [J]. 农业机械学报, 2019, 50(7):

- 216—221.
- Sun Zhe, Zhang Chunlong, Ge Luzhen, et al. Image detection method for broccoli seedlings in field based on Faster R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 216—221. (in Chinese with English abstract)
- [14] 刘慧, 张礼帅, 沈跃, 等. 基于改进 SSD 的果园行人实时检测方法[J]. 农业机械学报, 2019, 50(4): 29—35.
- Liu Hui, Zhang Lishuai, Shen Yue, et al. Real-time pedestrian detection in orchard based on improved SSD[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 29—35. (in Chinese with English abstract)
- [15] 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): 181—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): 181—186. (in Chinese with English abstract)
- [16] Tian Yunong, Yang Guodong, Wang Zhe, et al. Detection of apple lesions in orchards based on deep learning methods of CycleGAN and YOLOV3-Dense[J]. Journal of Sensors, 2019: 1—13.
- [17] 彭红星, 黄博, 邵园园, 等. 自然环境下多类水果采摘目标识别的通用改进 SSD 模型[J]. 农业工程学报, 2018, 34(16): 155—162.
- Peng Hongxing, Huang Bo, Shao Yuanyuan, et al. General improved SSD model for picking object recognition of multiple fruits in natural environment[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(16): 155—162. (in Chinese with English abstract)
- [18] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.
- [19] Qiao Y, Trueman M, Sukkariyah S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming[J]. Computers and Electronics in Agriculture, 2019, 165: 1-9.
- [20] 高云, 郭继亮, 黎焯, 等. 基于深度学习的群猪图像实例分割方法[J]. 农业机械学报, 2019, 50(4): 186—194.
- Gao Yun, Guo Jiliang, Li Xuan, et al. Instance-level segmentation method for group pig images based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 186—194. (in Chinese with English abstract)
- [21] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164—173.
- Zhao Dean, Wu Rendi, Liu Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164—173. (in Chinese with English abstract)
- [22] 王丹丹, 何东健. 基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别[J]. 农业工程学报, 2019, 35(3): 156—163.
- Wang Dandan, He Dongjian. Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 156—163. (in Chinese with English abstract)
- [23] Dias P A, Tabb A, Medeiros H. Apple flower detection using deep convolutional networks[J]. Computers in Industry, 2018, 99: 17—28.
- [24] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.
- [25] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. 2016: 21-37.
- [26] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 2014, 1409.1556v6.
- [27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770—778.
- [28] Zhang S, Zhu X, Lei Z, et al. S3fd: Single shot scale-invariant face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 192—201.
- [29] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts[J]. arXiv: 2016, 1608.03983.
- [30] Andrew H, Mark S, Grace C, et al. Searching for mobilenetv3[J]. arXiv: 2019, 1905.02244.
- [31] Mehta S, Rastegari M, Shapiro L, et al. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 9190—9200.
- [32] Lee Y, Hwang J, Lee S, et al. An energy and GPU-Computation efficient backbone network for real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019.

Real-time classification and detection of citrus based on improved single short multibox detector

Li Shanjun^{1,2,3,4,5}, Hu Dingyi^{1,2}, Gao Shumin^{1,2}, Lin Jiahao^{1,2}, An Xiaosong^{1,2}, Zhu Ming^{1,2*}

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China; 2. Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China; 3. China Agriculture (Citrus) Research System, Wuhan 430070, China; 4. National R&D Center for Citrus Preservation, Wuhan 430070, China; 5. Citrus Mechanization Research Base, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China)

Abstract: Manually classifying citrus based on its surface defects is tedious and time-consuming and a new real-time method is proposed in this paper based on the improved SSD deep learning model. In the testing bench of the waxing machine, 2 500 images of a variety of citrus species were taken, of which 2 000 were randomly selected as training set and 500 as testing set. Among them, the method classified 19 507 as normal, 9 097 skin defects and 4 327 mechanically damaged. Considering that traditional methods using near-infrared spectra, support vector machines, HSV and RGB color space model are inefficient to detect surface defects of citrus and can only identify one, we proposed an improved method to calculate the image using the one-stage detection model - SSD-ResNet18. The method gets the feature maps through backbone first, and then predicts the number of boundary boxes from the feature maps before determining the location and category of citrus using confidence and non-maximum suppression. This can detect a batch of citrus. In the proposed method, we used the mAP (mean average precision) as the precision index and the mean detection time as the speed index. Optimization in the proposed method was solved using the SGD (stochastic gradient descent) algorithm. The learning scheduler was based on cosine decay, enabling the learning rate to drop to 0 at the end of the training period. This ensures the lost value during the training period to continuously decline. As the model was stable at the end of the training period, it can be saved at the end of the training for further use. While the VGG16 was used as the original SSD backbone, it needs a multitude of parameters and is hence computationally inefficient. We replaced it with the ResNet18, which is approximately 100 times more efficient than the VGG16. An improved feature map was obtained from the analysis of the effective sensory field of different feature maps and the size of citrus in the map, the anchor in which was obtained using the K-means clustering algorithm from the manual label box. The suitable image resolution for the proposed model was obtained by comparing images taken at five resolutions: 512×512 pixels, 640×640 pixels, 768×768 pixels, 896×896 pixels and 1024×1024 pixels. The results showed that the accuracy of the mAP of SSD-ResNet18 was 87.89%, improving 0.34 percentage points higher than the original SSD. The average detecting time of the SSD-ResNet18 was 20.72 ms, reduced by 436.90% compared to the original SSD's 108.83 ms. The accuracy of the AP of SSD-ResNet18 was 94.72%, 85.79% and 83.17%, respectively, for detecting normal, skin lesion and mechanical damage. We compared MobileNetV3, ESPNetV2, VoVNet39 and ResNet18 as backbones and did not find significant difference between their accuracy, but ResNet18 was 10.52 ms, 16.78 ms and 36.76 ms less than MobileNetV3, ESPNetV2 and VoVNet39 in detection time, respectively. The method proposed in the paper meets the requirement on detecting speed in real-time citrus production line and can effectively classify and detect a multitude of citrus simultaneously.

Keywords: object recognition; models; nondestructive detection; citrus; surface defects; deep learning; SSD; ResNet18