

戴泽翰, 郑正, 黄莉舒, 等. 基于深度卷积神经网络的柑橘黄龙病症状识别 [J]. 华南农业大学学报, 2020, 41(4): 111-119.
DAI Zehan, ZHENG Zheng, HUANG Lishu, et al. Recognition of Huanglongbing symptom based on deep convolutional neural network[J]. Journal of South China Agricultural University, 2020, 41(4): 111-119.

基于深度卷积神经网络的柑橘黄龙病症状识别

戴泽翰, 郑 正, 黄莉舒, 赖云燕, 鲍敏丽, 许美容, 邓晓玲
(华南农业大学 农学院, 广东 广州 510642)

摘要:【目的】探究深度学习在柑橘 *Citrus* spp.黄龙病症状识别上的可行性,并评估识别器的识别准确率。【方法】以黄龙病/非黄龙病引起的发病叶片图像及健康叶片图像为训练素材,基于卷积神经网络及迁移学习技术构建二类识别器 (I-2-C 和 M-2-C) 和八类识别器 (I-8-C 和 M-8-C)。【结果】M-8-C 模型的整体识别表现最优,对所有图像的识别准确率为 93.7%,表明构建的神经网络识别器能有效辨别柑橘黄龙病症状;I-8-C 和 M-8-C 对所有类型图像的平均 *F1* 分值分别为 77.9% 和 88.4%,高于 I-2-C(56.3%) 和 M-2-C(52.5%),表明症状细分有利于提高模型的识别能力。同时 M-8-C 比 I-8-C 略高的平均 *F1* 分值表明基于 MobileNetV1 结构的八类识别器识别表现略优于基于 InceptionV3 的八类识别器。基于 M-8-C 改进的识别器 M-8f-C 能够转移到智能手机上,在田间测试中取得较好的识别表现。【结论】基于深度学习和迁移学习开发的识别器对黄龙病单叶症状具有较好的识别效果。

关键词: 柑橘黄龙病; 症状识别; 卷积神经网络; 迁移学习
中图分类号: S436.66 文献标志码: A 文章编号: 1001-411X(2020)04-0111-09

Recognition of Huanglongbing symptom based on deep convolutional neural network

DAI Zehan, ZHENG Zheng, HUANG Lishu, LAI Yunyan, BAO Minli, XU Meirong, DENG Xiaoling
(College of Agriculture, South China Agricultural University, Guangzhou 510642, China)

Abstract: 【Objective】To explore the capability of deploying deep learning to the detection of Huanglongbing (HLB) symptom in *Citrus* spp., and evaluate the classification accuracies of the classifiers. 【Method】Two-class classifiers(I-2-C and M-2-C) and eight-class classifiers(I-8-C and M-8-C) were constructed using images of diseased leaves caused by HLB/non-HLB and healthy leaves based on convolutional neural networks and transfer learning. 【Result】The overall classification performance of M-8-C stood out in all classifiers with accuracy of 93.7%, implying great capability in deep convolutional neural networks for classifying HLB symptoms. The mean *F1* scores of I-8-C and M-8-C were 77.9% and 88.4% respectively, which were higher than those of I-2-C(56.3%) and M-2-C(52.5%). This indicated that subtyping symptoms could help improve the recognition ability of models. The slightly higher mean *F1* score of M-8-C compared with I-8-C indicated that the eight-class model based on MobileNetV1 had better performance than the one based on InceptionV3. An optimized model, namely M-8f-C, was developed based on M-8-C and was successfully mounted on mobile phone. The field tests showed that M-8f-C was of decent performance under field conditions.

【Conclusion】 Classifier based on deep learning and transfer learning has high accuracy for recognizing HLB symptom leaves.

Key words: Citrus Huanglongbing; symptom recognition; convolutional neural network; transfer learning

柑橘 *Citrus* spp. 感染黄龙病 (Huanglongbing, HLB) 后会表现出复杂的症状, 在果实上的症状包括青果、红鼻子果; 在叶部的症状包括斑驳、均匀黄化、叶脉黄化、缺锌状花叶等, 在根部可观察到根部腐烂^[1]。这些复杂的症状常见于同一棵树上, 且在田间常与其他病害症状并发, 极大地增加了黄龙病诊断的难度。因此, 开发一种快速、可靠的检测方法是柑橘黄龙病早期检测的迫切需求。

目前对柑橘黄龙病的诊断、病程分级在一定程度上依赖于肉眼观察。但专业人员需要投入一定时间的专业训练, 且人为判断具有主观性、误差较大、标准不易统一等问题。相反地, 人工智能具有客观、重复性高、判断逻辑数量化等优点, 在上述病理问题的研究上引入人工智能可使得定性问题更准确, 定量问题更精确。目前已有多项基于计算机视觉的柑橘黄龙病症状识别研究。这些研究大多依托传统的机器学习算法 (如支持向量机^[2-5], 神经网络^[3-4] 决策树^[4], k 临近算法^[4-6] 等), 基于可见光^[2, 4-5] 或近红外光谱^[4-6] 技术, 在识别黄龙病发病叶片、遥感识别树冠方面取得较好的效果。但这些研究存在一些问题, 如部分研究只探究了健康和发病柑橘叶片 2 个类别图像的区分, 脱离实际田间症状复杂多样的应用场景, 部分研究的试验材料没有 PCR 确定黄龙病的存在, 研究中用到的发病素材的可靠性难以确定。

由于在个别特定视觉识别任务上具有接近甚至超越人类的表现^[7], 近几年来深度学习被广泛应用于图像识别和分型研究中。在医学领域, 深度卷积神经网络 (Convolutional neural network, CNN) 已被应用于 300 余项研究中, 涉及神经、眼、肺、心脏等脏器或部位的分类、预测、分割、图像配准等研究内容^[8]。其中, 对人类的皮肤癌症状^[9]、骨龄预测^[10]、帕金森颅内成像识别^[11] 识别器的预测水平与经验丰富的临床医生持平。在农业研究上, CNN 被应用于梨、桃子、苹果、葡萄、番茄等多种作物的机器损伤鉴定^[12-13]、病害鉴定^[14-16] 以及表型划分^[13], 并展现出优异的区分效果。这些研究表明, 深度学习在病理图像和序列研究上有着优异的表现, 能作为一种有力的技术手段。目前, 基于深度学习的黄龙病相关研究报道仅有 2 项^[17-18]。Choi 等^[17] 基于 CNN 构建了针对黄龙病、健康、锈蜘蛛为害状、风害的柑橘果实自动鉴别系统, 对 4 种类型果实图像分别能实现 89.7%、100%、94.7% 和 88.9% 的识别率。Mohanty

等^[18] 基于 CNN 构建了针对包括柑橘黄龙病在内的 26 种植物病害图像的识别器, 整体识别准确率达到 99.35%。由于黄龙病症状的复杂性, 上述研究均非针对黄龙病叶片症状的系统研究, 而是对叶片症状人为识别为目前黄龙病早期诊断的主要识别手段。因此, 探究深度学习在黄龙病叶片症状的识别表现具有一定研究价值。

本研究拟通过系统收集大量病叶, 结合症状分型探究深度学习在柑橘黄龙病叶片症状识别上的可行性; 并结合评估和方法论上的优化提升识别器的表现。我们比较了 MobileNetV1(<https://arxiv.org/abs/1704.04861v1>) 和 InceptionV3(<https://arxiv.org/abs/1512.00567>) 2 种神经网络结构的识别器的识别性能, 并探讨了症状分型对识别准确率提升的必要性, 最后将识别器转移到智能手机, 并初步探索识别器在田间的识别表现。

1 材料与方法

1.1 柑橘叶片采集

本研究使用的叶片材料采自广东省惠州市博罗县杨村镇水龙村柑橘黄龙病绿色防控与栽培新模式研发示范基地、华南农业大学柑橘黄龙病研究室的 109 号网室及长岗山网室。所采集的柑橘品种包括沙田柚 *Citrus maxima*、甜橙 *Citrus sinensis*、柠檬 *Citrus limon*、酸桔 *Citrus sunki*。从田间或网室收集具有黄龙病典型症状的叶片、不具有明显黄化症状的健康叶片以及非黄龙病引起叶脉黄化和均匀黄化叶片。用于采集图像的叶片要求为叶片完全展开、叶型完整的单身复叶, 包含叶尖、叶缘、小叶, 以及无明显的虫害、虫害等机械损伤痕迹。

1.2 植物总 DNA 提取及病原 PCR 检测

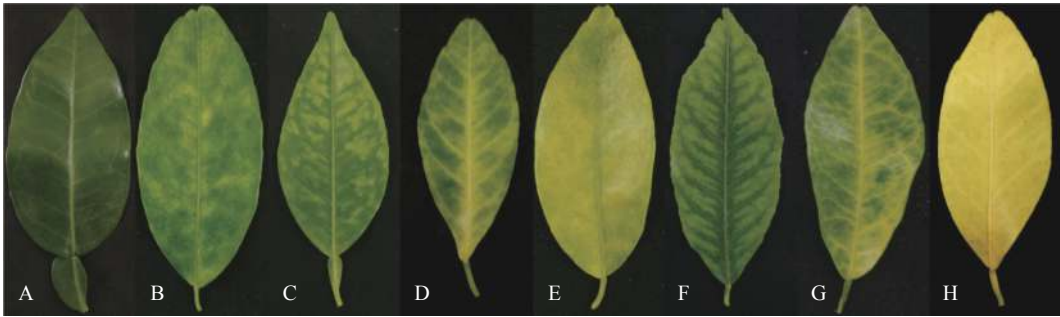
使用 HP-Plant DNA Kit 试剂盒 (Omega Bio-Tek Inc) 提取叶片总 DNA。对所有提取的叶片总 DNA 进行 qPCR 检测, 以确定黄龙病病原菌 *Candidatus Liberibacter asiaticus* (CLAs) 的存在。CLAs 检测使用的引物为 HLBf/HLBasr^[19]。总体积 20 μ L 的反应体系: ddH₂O 7 μ L, Bestar® SybrGreen qPCR mastermix 10 μ L, 正、反向引物 (10 mol/L) 各 1 μ L, 模板 DNA (<1 g/L) 1 μ L。

在 Bio-Rad CFX 96 仪器 (Bio-Rad Laboratories) 上进行实时荧光定量 PCR 反应。反应条件: 预变性阶段 95 $^{\circ}$ C, 3 min; 循环反应阶段 95 $^{\circ}$ C,

10 s; 60 ℃, 30 s, 40 个循环。使用 Biorad CFX 96 仪器自带软件 Bio-Rad CFX Manager(v1.6.541) 生成扩增曲线、溶解曲线及对应的检测结果。

1.3 图像的收集及预处理

在严格控制条件下进行图像的收集。拍摄时采用华为荣耀 9 智能手机的摄像功能对柑橘叶片进行拍摄。具体硬件参数:双摄像头像素分别为 1 200 和 2 000, 光圈 f/2.2, 镜头焦距 27 mm, 相位检测自动对焦, 2 倍无损变焦。摄影参数:选择专业模式, 关闭自动闪光和自动白平衡, 感光度取 200~250, 快门时间 1/20~1/15 s; 将单叶置于黑色摄影布上进行拍摄。将拍摄好的图像进行预处理, 调节裁剪不必要的背景, 使图像只保留完整的叶片及附近区域, 统一调整像素密度为 72 PPI。



A: 健康; B: 黄龙病-斑驳; C: 黄龙病-缺锌状; D: 黄龙病-叶脉黄化; E: 黄龙病-均匀黄化; F: 非黄龙病-缺锌状; G: 非黄龙病-叶脉黄化; H: 非黄龙病-均匀黄化
A: Healthy; B: HLB-mottle; C: HLB-zinc deficiency; D: HLB-vein yellowing; E: HLB-uniform yellowing; F: Non-HLB-zinc deficiency; G: Non-HLB-vein yellowing; H: Non-HLB-uniform yellowing

图 1 本研究收集的 8 种类别柑橘叶片图像示例
Fig. 1 Examples of eight classes of citrus leaves used in this study

1.5 神经网络结构的选取和症状识别器的构建

神经网络的构建采用迁移学习的方式进行训练, 所用的模型结构为 MobileNetV1(<https://arxiv.org/abs/1704.04861v1>) 和 InceptionV3(<https://arxiv.org/abs/1512.00567>)。这 2 个神经网络模型均为 Google Brain 团队 (Google Inc.) 开发的开源神经网络。其中 MobileNetV1 为专门面向智能手机设计的网络结构, 其特点为轻便、响应延迟小, 选取该结构是为了考虑转移模型到智能手机上应用的需要。InceptionV3 相对于 MobileNetV1 有着更多的网络层数, 其理论识别准确率比 MobileNetV1 高。利用开源学习框架 Tensorflow(<https://arxiv.org/abs/1603.04467>) 对模型进行训练, 训练模型用的编程脚本参考 Tensorflow for poet2 流程 (<https://codelabs.developers.google.com/codelabs/tensorflow-for-poets>), 训练迭代次数为 10 000。

1.6 对症状识别器预测性能的评估

利用测试集的图片对训练好的模型进行准确性评估。本研究使用基于准确率、平均 F1 分值、黄龙病识别率以及非黄龙病识别率准确性这几个统

1.4 图像的训练集及测试集结构

根据症状和 PCR 检测结果, 将样品图像分为健康 (共 233 幅)、黄龙病-斑驳 (共 514 幅)、黄龙病-缺锌状 (共 120 幅)、黄龙病-叶脉黄化 (共 115 幅)、黄龙病-均匀黄化 (共 131 幅)、非黄龙病-缺锌状 (共 123 幅)、非黄龙病-叶脉黄化 (共 254 幅)、非黄龙病-均匀黄化 (共 177 幅) 共 8 个分组 (图 1), 作为构建八类识别器的图像素材。另外将 4 个黄龙病分组合并为黄龙病组, 与健康图像作为构建二类识别器的素材。将各分组图像分为训练集和测试集 2 个组。训练集的图像用于训练神经网络, 测试集的图像用于测试训练完成的神经网络的识别效果。训练集与测试集的图像分别为没有数据重叠的不同图像集合。使用同一套测试集图像对不同模型进行比较。

计量对模型作准确性评估。平均 F1 分值为分类问题中评估识别器识别表现的一个统计量, 其计算过程综合了查准率和召回率, 取值范围为 0~1, 值越大说明模型的整体识别表现越好。真阳性率、假阳性率、真阴性率、假阴性率的计算均参考 Fawcett 的方法^[20], 其他统计量和相关的计算公式如下:

$$\text{平均}F1\text{分值} = \frac{2 \times \text{查准率} \times \text{召回率}}{\text{查准率} + \text{召回率}},$$
$$\text{查准率} = \text{真阳性} / (\text{真阳性} + \text{假阳性}),$$
$$\text{召回率} = \text{真阳性} / (\text{真阳性} + \text{假阴性}),$$
$$\text{准确率} = \frac{\text{真阳性} + \text{真阴性}}{\text{真阳性} + \text{真阴性} + \text{假阳性} + \text{假阴性}},$$
$$\text{黄龙病识别率} = \frac{\text{指定类别图像被识别为任意一种黄龙病图像类型的总数}}{\text{该类图像总数}},$$
$$\text{非黄龙病识别率} = \frac{\text{指定类别图像被识别为任意一种非黄龙病图像类型的总数}}{\text{该类图像总数}}。$$

1.7 识别器优化重建及田间识别测试

田间拍摄图像素材: 从广东省惠州市杨村镇柑橘黄龙病和溃疡病综合防控技术与示范果园采集了各类柑橘黄龙病图像共 3 531 幅及健康图像共 3 989 幅。所拍摄图像均为挂在树上的完好叶片。取样植株均经 PCR 检测确定是否感染了病原菌 CLas。

根据上一步识别器评估的结果选择识别性能最优的识别器, 将构建好的识别器通过 Android Studio v3.4.1 编译成 App, 再进一步转移至搭载 Android 操作系统的智能手机。编程脚本参考 TensorFlow for Poets 2 TFMobile(<https://codelabs.developers.google.com/codelabs/tensorflow-for-poets-2>)。所用 TensorFlow 版本为 1.70 识别 App, 在智能手机华为 6 Plus(2014 年产; 处理器: 4x1.8 GHz Cortex-A15; 内存 3.0 GB) 以及华为荣耀 9(2017 年产; 处理器: 4x2.4 GHz Cortex-A73 & 4x1.8 GHz Cortex-A53; 内存 4.0 GB) 均安装成功并能正常运行, 但后续的数据收集均在硬件性能更优、操作系统版本更高的华为荣耀 9 上进行。在智能手机安装好程序后, 对发病植株上的病叶进行检测, 同时取下同一块叶片

在黑色摄影布上拍摄图像, 从而比较复杂背景和简单背景成像的识别率。此外, 将智能手机识别叶片时的图像在计算机上用同样的模型进行验证。为简化数据, 在这一步只记录所有黄龙病/非黄龙病图像的总体识别率, 不呈现所有子分类的识别结果。

2 结果与分析

通过训练集图像构建了共 4 个识别器, 分别为基于 MobileNetV1 网络结构的八类识别器 (MobileNetV1 eight-class classifier, M-8-C) 和二类识别器 (MobileNetV1 two-class classifier, M-2-C) 以及基于 InceptionV3 网络结构的八类识别器 (InceptionV3 eight-class classifier, I-8-C) 和二类识别器 (InceptionV3 two-class classifier, I-2-C)。

2.1 对八类识别器的准确性评估

对 2 种八类识别器的准确性评估结果见表 1 和表 2。总体而言, 只考虑对所有图像类型的识别表现, M-8-C 比 I-8-C 具有更高的识别表现。如 M-8-C 对所有图像类型的准确率和平均 F1 分值分别为 93.7% 和 88.4%(表 1), 而 I-8-C 的对应指标分别为 88.0% 和 77.9%(表 2)。另外, M-8-C 的真阳性率

表 1 使用 MobileNetV1 网络结构的八类识别器的测试结果

Table 1 Test result of eight-class classifier using MobileNetV1 network architecture										
图片类型 Image type	测试集 图片数 No. of images in test set	训练集 图片数 No. of images in training set	真阳性 率/% True positive rate	假阴性 率/% False negative rate	真阴性 率/% True negative rate	假阳性 率/% False positive rate	准确率/% Accuracy	平均F1 分值/% Mean F1 score	黄龙病 识别率/% HLB detection rate	非黄龙病 识别率/% Non-HLB detection rate
黄龙病-斑驳 HLB-mottle	50	514	84.0	16.0	77.4	6.0	91.0	87.9	94.0	6.0
黄龙病-缺锌状 HLB-zinc deficiency	50	120	86.0	14.0	77.1	2.6	94.6	86.1	100.0	0.0
黄龙病-叶脉黄化 HLB-vein yellowing	50	115	74.0	26.0	78.9	3.4	92.1	85.9	90.0	10.0
黄龙病-均匀黄化 HLB-uniform yellowing	50	131	44.0	56.0	83.1	4.3	87.5	89.7	86.0	14.0
健康 Healthy	50	233	92.0	8.0	76.3	2.0	96.1	80.6	6.0	94.0
非黄龙病-缺锌状 Non-HLB-zinc deficiency	50	123	98.0	2.0	75.4	1.4	97.6	95.3	2.0	98.0
非黄龙病-叶脉黄化 Non-HLB-vein yellowing	50	254	84.0	16.0	77.4	2.6	94.3	91.3	6.0	94.0
非黄龙病-均匀黄化 Non-HLB-uniform yellowing	50	177	64.0	36.0	80.3	1.7	92.4	90.9	26.0	74.0
所有黄龙病图像 Total HLB images	200	880	72.0	44.0	79.1	4.1	91.3	87.4	85.5	7.5
所有非黄龙病图像 Total non-HLB images	200	787	84.5	15.5	77.4	1.9	95.1	89.5	10.0	90.0
所有图像 Total images	400	1 667	78.3	21.8	78.3	3.0	93.7	88.4		

表 2 使用 InceptionV3 网络结构的八类识别器的测试结果

Table 2 Test result of eight-class classifier using InceptionV3 network architecture

图片类型	测试集	训练集	真阳性	假阴性	真阴性	假阳性		平均F1	黄龙病	非黄龙病
Image	图片数	图片数	率/%	率/%	率/%	率/%	准确率/%	分值/%	识别率/%	识别率/%
type	No. of	No. of images	True positive	False negative	True negative	False positive	Accuracy	Mean F1	HLB detection	Non-HLB
	images in test set	in training set	rate	rate	rate	rate		score	rate	detection rate
黄龙病-斑驳 HLB-mottle	50	514	56.0	42.0	66.6	6.0	85.7	84.3	82.0	16.0
黄龙病-缺锌状	50	120	84.0	16.0	62.6	10.0	85.4	74.7	94.0	6.0
HLB-zinc deficiency										
黄龙病-叶脉黄化	50	115	68.0	32.0	64.9	4.0	89.2	50.6	78.0	22.0
HLB-vein yellowing										
黄龙病-均匀黄化	50	131	14.0	86.0	72.6	5.7	80.1	89.3	88.0	12.0
HLB-uniform yellowing										
健康 Healthy	50	233	62.0	38.0	65.7	4.9	87.4	74.3	16.0	84.0
非黄龙病-缺锌状	50	123	92.0	8.0	61.4	2.6	94.7	94.2	0.0	100.0
Non-HLB-zinc deficiency										
非黄龙病-叶脉黄化	50	254	84.0	18.0	62.6	4.6	90.7	83.2	10.0	92.0
Non-HLB-vein yellowing										
非黄龙病-均匀黄化	50	177	62.0	38.0	65.7	2.0	90.4	72.7	34.0	66.0
Non-HLB-uniform										
yellowing										
所有黄龙病图像	200	880	55.5	44.0	66.6	6.4	85.1	74.7	85.5	14.0
Total HLB images										
所有非黄龙病图像	200	787	75.0	25.5	63.9	3.5	90.8	81.1	15.0	85.5
Total non-HLB images										
所有图像 Total images	400	1 667	65.3	34.3	65.3	5.0	88.0	77.9		

和真阴性率均为 78.3%，高于 I-8-C(均为 65.3%)，M-8-C 的假阴性率和假阳性率分别为 21.8% 和 3.0%，低于 I-8-C(34.3% 和 5.0%)。

此外，只考虑黄龙病大类和非黄龙病大类的整体识别表现，同样为 M-8-C 优于 I-8-C。M-8-C 对所有黄龙病分类图像的黄龙病识别率与所有非黄龙病分类图像的非黄龙病识别率分别为 92.5% 和 90.0%，这 2 项均优于 I-8-C 对应的识别率 (85.5% 和 85.5%)。此外，除黄龙病-缺锌状图像以外，M-8-C 对各类型图像的识别表现均优于或相仿于 I-8-C。M-8-C 在对各类图像的真阳性率、真阴性率方面均高于 I-8-C，而对各类图像的假阳性率、假阴性率均小于或持平于 I-8-C。综合各种指标，M-8-C 的整体识别表现优于 I-8-C。

进一步对细分症状分析发现，八类识别器对黄龙病/非黄龙病引起的均匀黄化和叶脉黄化区分度较低。如 M-8-C 识别黄龙病和非黄龙病引起的叶脉黄化的真阳性率分别为 74.0% 和 84.0%(表 1)。追溯每幅叶脉黄化类图片的识别结果发现，M-8-C 的假阳性检测结果中，黄龙病-叶脉黄化图像大多被误判为非黄龙病-叶脉黄化图像，反之亦然。在均匀黄化类图像的识别表现也观察到类似的现象，M-

8-C 识别黄龙病和非黄龙病均匀黄化的真阳性率分别为 44.0%和 64.0%(表 1)，对黄龙病-均匀黄化的假阳性结果多为非黄龙病-均匀黄化图像。I-8-C 对应的黄龙病和非黄龙病的均匀黄化真阳性率更低，仅为 14.0% 和 62.0%(表 2)。这说明构建的识别器不能很好地分辨黄龙病和非黄龙病引起的叶脉黄化及均匀黄化，这一点同样跟人眼的识别相似。

2.2 对二类识别器的准确性评估

对 2 种二类识别器的准确性评估结果见表 3 和表 4。为了与八类识别器比较，避免数据集的差异，采用了同样的测试集图像。由于二类识别器解决的为二分类问题，因此只统计识别器对所有图像、黄龙病所有图像、非黄龙病所有图像的识别数据，不统计对细分图像类型的识别数据。此外，由于本研究使用的准确率计算公式应用于二分类问题将得出相同的数值 (表 3、表 4)，因此不纳入本研究的结果分析和讨论中。

2 种二类识别器对所有图像的准确识别率较低，两者的平均 F1 分值仅为 52.5% 和 56.3%。出现此种情况的原因是 2 种识别器对非黄龙病图像的识别表现较低。虽然 M-2-C 和 I-2-C 对所有黄龙病图像的黄龙病识别率均较高 (96.0% 和 84.0%)，

表 3 使用 MobileNetV1 网络结构的二类识别器的测试结果

Table 3 Test result of two-class classifier using MobileNetV1 network architecture

图片类型 Image type	测试集 图片数 No. of images in test set	训练集 图片数 No. of images in training set	真阳性 率/% True positive rate	假阴性 率/% False negative rate	真阴性 率/% True negative rate	假阳性 率/% False positive rate	准确率/% Accuracy	平均F1 分值/% Mean F1 score	黄龙病 识别率/% HLB detection rate	非黄龙病 识别率/% Non-HLB detection rate
所有黄龙病图像 Total HLB images	200	880	96.0	4.0	22.0	78.0	59.0	70.1	96.0	4.0
所有非黄龙病图像 Total non-HLB images	200	787	22.0	78.0	96.0	4.0	59.0	34.9	78.0	22.0
所有图像 Total images	400	1 667	59.0	41.0	59.0	41.0	59.0	52.5		

表 4 使用 InceptionV3 网络结构的二类识别器的测试结果

Table 4 Test result of two-class classifier using InceptionV3 network architecture

图片类型 Image type	测试集 图片数 No. of images in test set	训练集 图片数 No. of images in training set	真阳性 率/% True positive rate	假阴性 率/% False negative rate	真阴性 率/% True negative rate	假阳性 率/% False positive rate	准确率/% Accuracy	平均F1 分值/% Mean F1 score	黄龙病 识别率/% HLB detection rate	非黄龙病 识别率/% Non-HLB detection rate
所有黄龙病图像 Total HLB images	200	880	84.0	16.0	34.0	66.0	59.0	67.2	84.0	16.0
所有非黄龙病图像 Total non-HLB images	200	787	34.0	66.0	84.0	16.0	59.0	45.3	66.0	34.0
所有图像 Total images	400	1 667	59.0	41.0	59.0	41.0	59.0	56.3		

但 M-2-C 和 I-2-C 对非黄龙病图像误判为黄龙病图像的识别率高达 78.0% 和 66.0%，表明存在着大量图像被错误识别为黄龙病的情况。以上结果表明，2 种二类识别器均不能有效识别非黄龙病症状图像。

与八类识别器相比，二类识别器无论是对所有图像整体识别表现，还是对黄龙病图像、非黄龙病的正确识别的能力均较低。这表明八类识别器的识别表现优于二类识别器，基于症状细分构建的识别器能有效提高识别表现。

2.3 识别器的重建及田间识别测试

根据上一步分类器的评估结果，剔除黄龙病/非黄龙病引起的均匀黄化图像重新训练了基于 MobileNetV1 结构的识别器。由于评估的实际环境为复杂的田间条件，光线、背景、叶型残缺的因素使得成像条件比原来的图像拍摄条件复杂，因此在本次测试加入大量田间拍摄的图像。新加入的图像为黄龙病-斑驳图像共 3 531 幅和健康类图像共 3 980 幅，最终的训练集结构为包括黄龙病-斑驳、黄龙病-斑驳 (田间拍摄)、黄龙病-缺锌状、黄龙病-叶脉黄化、健康、健康 (田间拍摄)、非黄龙病-缺锌状、非黄龙病-叶脉黄化的 8 类图像。将重构的模型命名为

M-8f-C，与之前构建的模型作区分。

为了估计图像数量对提升准确率的影响，探究了训练集图像数量和测试准确率的关系 (图 2)。结果表明，随着训练集图像数量的增加，平均 F1 分值逐渐升高，在达到约 400 幅之后 F1 分值维持在 95% 以上。这表明 400 幅训练集图像就足够使模型准确率达到最大化。因此在本步骤重构模型使用的图像数量是足够的。

M-8f-C 识别田间叶片的结果汇总为表 5。在智能手机上，M-8f-C 对在树上叶片的识别效果比在摄影布上的识别效果好。对树上拍摄的黄龙病和非黄龙病叶片图像的平均 F1 分值分别为 87.2% 和

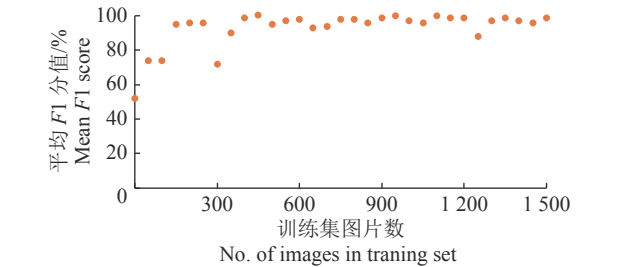


图 2 训练集图片数与平均 F1 分值动态关系图

Fig. 2 Dynamic relationship of the number of mages in training set and mean F1 score

82.4%, 而在摄影布上拍摄的平均 $F1$ 分值则分别为 68.0% 和 77.6%, 而准确率同样为在树上拍摄的比摄影布上的高。将智能手机拍摄的图像转移到计算机上进行验证的结果表明, 在计算机测试的图像识别准确率有所提升。如在计算机上对黑布背景的健康叶和病叶的平均 $F1$ 分值分别为 91.6% 和 93.1%, 在智能手机上对应的平均 $F1$ 分值分别为 77.6% 和 68.0%, 表明同一识别器在计算机上的识别表现明显提高。考虑到构建 M-8f-C 的训练集图像的种类较多, 而图像识别与硬件计算力密切相

关, 推测这是因为计算机有着更高的计算力所致。此外, 在计算机对摄影布背景的黄龙病、非黄龙病叶片图像的准确率 (92.4%、92.4%) 和平均 $F1$ 分值 (93.1%、91.6%) 高于在树上拍摄图像背景的准确率 (88.3%、88.3%) 和平均 $F1$ 分值 (86.4%、89.8%), 这一点与在智能手机上识别结果相反。

总体而言, 识别器在智能手机对树上拍摄的叶片图像有着最优的识别表现, 对黄龙病和非黄龙病图像的准确率以及平均 $F1$ 分值均达到 80% 以上。识别器在计算机上对摄影布为背景拍摄图像的识

表 5 识别器的田间测试结果
Table 5 Field test result of classifier

测试平台	图像背景	图片类别	测试集	训练集	真阳性	假阴性	真阴性	假阳性	准确率	平均F1
Test platform	Image Background	Image type	图片数 No. of images in test set	图片数 No. of images in training set	率/% True positive rate	率/% False negative rate	率/% True negative rate	率/% False positive rate	准确率/% Accuracy	分值/% Mean F1 score
手机 Mobile phone	树上 On the tree	黄龙病	150	4 280	100	0	68.7	29.3	85.2	87.2
		HLB								
		非黄龙病	147	4 534	68.7	29.3	100	0	85.2	82.4
	摄影布 Photographic backdrop	Non-HLB								
		黄龙病	150	4 280	56.0	44.0	91.3	8.7	73.7	68.0
		HLB								
计算机 Computer	树上 On the tree	非黄龙病	151	4 534	91.3	8.7	56.0	44.0	73.7	77.6
		Non-HLB								
		黄龙病	149	4 280	76.5	23.5	99.4	0.6	88.3	86.4
	摄影布 Photographic backdrop	HLB								
		非黄龙病	159	4 534	99.4	0.6	76.5	23.5	88.3	89.8
		Non-HLB								
		黄龙病	157	4 280	98.7	1.3	85.7	13.2	92.4	93.1
		HLB								
		非黄龙病	147	4 534	85.7	14.3	98.7	1.3	92.4	91.6
		Non-HLB								

别表现较好, 对黄龙病和非黄龙病图像实现的准确率以及平均 $F1$ 分值均达 90% 以上, 但同时为树上为背景拍摄图像的准确率以及平均 $F1$ 分值也能达到 85% 以上。

3 讨论与结论

目前对柑橘黄龙病的相关研究多基于传统的机器学习算法, 在深度学习上的研究处于起步阶段。基于深度学习对黄龙病症状进行系统深入的探讨有着重要的研究价值和必要性。本研究初步探索了深度学习在构建柑橘黄龙病的症状识别上的应用, 以 Tensorflow 开源学习框架作为操作平台, 基于大量高质量的黄龙病和健康图片成功构建了具有较优识别表现的 4 种识别器。其中 M-8-C 对各

类图像识别的准确率、平均 $F1$ 分值分别达到了 93.7%、88.4%, 充分说明利用深度学习开发黄龙病识别器是可行的, 且能够达到较高的准确识别率。

前人对柑橘黄龙病相关的深度学习研究报道中, Choi 等^[17]的研究为针对果实分型的研究, 旨在探究黄龙病病果、健康果以及风害、虫害的鉴别, 未涉及黄龙病叶片症状的探讨, Mohanty等^[18]的研究侧重于多种病害的叶片症状的鉴别, 涉及 13 种作物共 26 种病害类别, 并非专门针对柑橘黄龙病的研究。与之相比, 本研究对症状进行了系统分型, 增设健康对照, 并结合 PCR 检测降低叶片被人为误判的可能性。因此本研究为利用深度学习识别叶片图像较为系统的研究。以上研究使用 CNN 构建识别器, 最优模型的总体识别准确率均达到了 90% 以

上,本研究构建的 CNN 模型的识别准确率为 93.7%,达到了相仿的水平。此外,本研究的总体识别准确率与上述 2 项基于传统机器学习的近似研究的准确率 (92.0%、91.9%) 持平^[2, 21]。

基于 MobileNetV1 结构的识别器的识别表现性能优于 InceptionV3 结构的模型。MobileNetV1 是以在智能手机上应用为目的开发的模型,特点是轻量化、网络层数少,以减少计算资源为首要目的。InceptionV3 相对于 MobileNetV1 有着更多的神经元层数,理论上其学习能力和识别性能应当更高 (<https://arxiv.org/abs/1704.04861v1>, <https://arxiv.org/abs/1512.00567>), 但本研究的实际结果为 MobileNetV1 的识别表现优于 InceptionV3。有研究表明 CNN 模型往往针对复杂问题设计,隐含层较多、参数量大,在模型拟合能力远远高于问题复杂程度时,容易产生过拟合的问题^[22-23]。因此模型容量和复杂度与黄龙病叶片识别任务不相适可能是 InceptionV3 识别表现更低的原因。因此,在本研究的控制条件下, MobileNetV1 更适于黄龙病症状的识别。在未来的研究中可以对网络结构的调整、参数的优化,借鉴或使用其他神经网络来进行进一步探究。

八类识别器的识别表现优于二类识别器,这表明症状细分对提高黄龙病识别率有着重要作用,同时也从侧面支持了人为分型的正确性和必要性。目前对黄龙病症状的鉴别及病程分级尚无一套公认的标准,对症状的鉴别主要依靠有经验的技术人员对其进行症状描述和分型。八类识别器对细分症状的成功鉴别说明从计算机视觉的角度症状分型是可行的,这为以后以相关技术为基础构建统一标准提供了理论依据。

识别器对黄龙病和非黄龙病造成的均匀黄化和叶脉黄化症状的识别效果较差,这一点和人眼识别能力相仿。在病原尚未确认前,黄龙病曾被怀疑是水害因素引起^[24-25],原因是水害会引发受害柑橘出现极其相似的均匀黄化症状,直到有研究通过实验排除了水害引发黄龙病的可能^[1]。此外,环割、天牛蛀蚀、根系受害同样会导致整株柑橘出现均匀黄化、叶脉黄化的现象^[26],这些因素引起的黄化即使是专业人员亦无法单凭肉眼区分。考虑到人眼识别的可见光谱范围较窄,而本研究涉及的图像同样为可见光成像,因此推测人眼和模型识别应具有相似的局限性。在使用识别器对均匀黄化和叶脉黄化进行病因辨别时应结合其他技术或方法来判断,如 PCR 检测、观察同一病株上的典型症状等,在未来可依托可见光光谱成像作进一步研究。

识别器能够快速布置至智能手机上,且在大田环境下的识别应用取得优秀的效果,表明本研究采用的研究方法具有一定的应用价值。在计算机上使用 M-8f-C 模型对同一批图像的识别表现有所提升,推测这是由于 M-8f-C 需要更多的计算资源,而手机计算力的不足所致。因此,在未来的研究中需要从智能手机 App 的优化并进一步提高硬件计算力等方面进行探究。

本研究的结果表明迁移学习是一种构建病害识别器的有力工具。迁移学习是一种结合已训练模型和新素材快速构建和布置新识别器的深度学习技术。CNN 行使判断功能的神经元层位于最后一层,迁移学习通过利用新的图像素材重塑这一层,从而使 CNN 适用于新的识别任务^[13]。旧模型对新素材的适用性在于物体的颜色和纹理具有通用性。以 MobileNetV1 为例,它是以 ImageNet 数据库 (<https://arxiv.org/abs/1409.0575>) 构建的模型,该数据库包括多种花卉在内的 1 000 多种图像。本研究的数据证明了此模型在植物病害识别上的适用性。相对于传统的机器学习,深度学习结合迁移学习有着不可比拟的优势,如不需要耗费大量的人力和时间资源从头训练模型,仅需掌握必要的编程知识即可使用迁移学习训练模型,这降低了开展相关研究的技术门槛,使得广大的农业学科研究人员都能采用。

综上,本研究的结果表明深度学习和迁移学习是一种能够用于黄龙病症状识别的有力研究手段,结合症状分型、训练集结构调整等优化能实现较高的识别效果。结合深度学习和迁移学习能作为柑橘病害图像研究的新工具,为促进后续黄龙病早期诊断、自动化识别、症状分型等工作提供新的思路和理论依据。

参考文献:

- [1] 林孔湘. 柑桔黄梢(黄龙)病研究 I: 病情调查[J]. 植物病理学报, 1956, 2(1): 1-11.
- [2] DENG X, LAN Y, HONG T, et al. Citrus greening detection using visible spectrum imaging and C-SVC[J]. Comput Electron Agr, 2016, 130(2016): 177-183.
- [3] GARCIA-RUIZ F, SANKARAN S, MAJA J M, et al. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees[J]. Comput Electron Agr, 2013, 91: 106-115.
- [4] MISHRA A, KARIMI D, EHSANI R, et al. Evaluation of an active optical sensor for detection of Huanglongbing (HLB) disease[J]. Biosyst Eng, 2011, 110(3): 302-309.
- [5] MISHRA A, KARIMI D, EHSANI R, et al. Identification of citrus greening (Hlb) using a vis-nir spectroscopy technique[J]. T ASABE, 2012, 55(2): 711-720.
- [6] SANKARAN S, MISHRA A, MAJA J M, et al. Visible-

near infrared spectroscopy for detection of Huanglong-bing in citrus orchards[J]. *Comput Electron Agr*, 2011, 77(2): 127-134.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Commun ACM*, 2017, 60(6): 84-90.

[8] ZHANG Y C, KAGEN A C. Machine learning interface for medical image analysis[J]. *J Digit Imaging*, 2017, 30(5): 615-621.

[9] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115-118.

[10] LEE H, KIM K G. Applying deep learning in medical images: The case of bone age estimation[J]. *Healthc Inform Res*, 2018, 24(1): 86-92.

[11] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. *Med Image Anal*, 2017, 42: 60-88.

[12] GHOSAL S, BLYSTONE D, SINGH A K, et al. An explainable deep machine vision framework for plant stress phenotyping[J]. *Proc Natl Acad Sci USA*, 2018, 115(18): 4613-4618.

[13] TAPAS A. Transfer learning for image classification and plant phenotyping[J]. *Int J Res Appl Sci Eng Technol*, 2016, 5(11): 2664-2668.

[14] SLADOJEVIC S, ARSENOVIC M, CULIBRK A A D, et al. Deep neural networks based recognition of plant diseases by leaf image classification[J]. *Comput Intel Neurosc*, 2016: 3289801.

[15] ZHANG K, WU Q, LIU A, et al. Can deep learning identify tomato leaf disease?[J]. *Advances in Multimedia*, 2018, 2018: 1-10.

[16] RAMCHARAN A, BARANOWSKI K, MCCLOSKEY P, et al. Deep learning for image-based cassava disease detection[J]. *Front Plant Sci*, 2017, 8: 1-7.

[17] CHOI D, LEE W, SCHUELLER J K, et al. A precise fruit inspection system for Huanglongbing and other common citrus defects using GPU and deep learning technologies[C]// ISPA. Proceedings of 13th International Conference on Precision Agriculture, St Louis, Missouri: ISPA, 2016, 1-6.

[18] MOHANTY S P, HUGHES D P, SALATHE M. Using deep learning for image-based plant disease detection[J]. *Front Plant Sci*, 2016, 7: 1-7.

[19] LI W, HARTUNG J S, LEVY L. Quantitative real-time PCR for detection and identification of Candidatus Liberibacter species associated with citrus huanglong-bing[J]. *J Microbiol Meth*, 2006, 66(1): 104-115.

[20] FAWCETT T. An introduction to ROC analysis[J]. *Pattern Recogn Lett*, 2006, 27(8): 861-874.

[21] DENG X, LAN Y, XING X, et al. Citrus huanglongbing detection based on image feature extraction and two-stage back propagation neural network modeling[J]. *Int J Agr Biol Eng*, 2016, 9(6): 20-26.

[22] 李俭川, 秦国军, 温熙森, 等. 神经网络学习算法的过拟合问题及解决方法[J]. *振动、测试与诊断*, 2002, 22(4): 260-264.

[23] 陶砾, 杨朔, 杨威. 深度学习的模型搭建及过拟合问题的研究[J]. *计算机时代*, 2018(2): 14-21.

[24] REINKING O A. Diseases of economic plants in southern China[J]. *Philipp Agric*, 1919, 8: 109-135.

[25] TU C. Notes on disease of economic plants in South China[J]. *Lingnan Science Journal*, 1932, 11: 489-504.

[26] 吴定尧. 柑橘黄龙病及综合防治[M]. 广州: 广东科技出版社. 2010: 12-20.

【责任编辑 庄 延】