

孙道宗, 刘锦源, 丁郑, 等. 基于改进 EfficientNetv2 模型的多品种南药叶片分类方法[J]. 华中农业大学学报, 2023, 42(1): 258-267.
DOI: 10.13300/j.cnki.hnlkxb.2023.01.031

基于改进 EfficientNetv2 模型的多品种 南药叶片分类方法

孙道宗^{1,2}, 刘锦源¹, 丁郑¹, 刘欢¹, 彭家骏¹, 谢家兴^{1,2}, 王卫星^{1,2}

1. 华南农业大学电子工程学院(人工智能学院), 广州 510642;
2. 广东省农情信息监测工程技术研究中心, 广州 510642

摘要 为提高南药叶片的分类和分拣效率, 本研究对 EfficientNetv2 网络模型进行改进, 引入迁移学习机制训练模型, 选取自适应矩估计优化算法, 通过多次测试进行超参数优化, 确定学习率; 采用 MultiMarginLoss 损失函数改善复杂背景信息对识别效果的影响。应用改进后的 EfficientNetv2 模型与其他轻量级模型对实地采集的复杂背景下的 8 种南药叶片进行分类效果对比试验, 试验结果显示, 改进模型对复杂背景下的南药叶片图像样本识别准确率为 99.12%, 相较于初始模型 EfficientNetv2-S, 准确率提高 1.17%, 并且参数数量和模型大小均下降约 85%, 平均训练时间下降 47.62%。与 DenseNet121、ShuffleNet 和 RegNet 等模型相比, 改进模型在模型存储空间大小、准确率和训练时间 3 个指标上有明显优势。研究结果表明, 在多品种南药叶片分类任务中, 改进模型取得优良表现, 模型的轻量化程度和性能得到进一步的提升。

关键词 图像识别; 南药叶片; 种类识别; 卷积神经网络; 改进 EfficientNetv2 网络; 超参数优化

中图分类号 S571.1 **文献标识码** A **文章编号** 1000-2421(2023)01-0258-10

南药是指长江以南、南岭以北地区包括湖南、江西、福建、台湾等省区的全部或大部分地区所产的地道药材, 是我国传统中药材的重要组成部分。目前广东省种植的药材有 60 多种, 其中具有鲜明地域特色的南药种类有 40 余种, 种植面积共约 21 万 hm^2 , 年产量约 115 万 t, 农业产值约 160 亿元^[1]。在实际种植生产过程中, 由于叶片的纹理复杂、背景的复杂性及药材叶片与树叶颜色和纹理的高度相似性, 给南药材的分类和分拣工作带来一定困难。

对于复杂自然场景下的植物叶片目标检测和识别, 近年来已取得一定进展。朱黎辉等^[2]将基于形状和纹理特征的中药图像检索方式融合模式识别技术及人机交互技术运用到对中药材图像的检索上, 提取 BLBP 纹理特征和 IHOG 形状特征并将 2 种特征进行组合, 该方法的平均准确率达到 93.56%。杨万里等^[3]提取了颜色、形态和纹理等 85 个表型参数进

行归一化处理后, 将表型数据与作物的生长质量作线性回归分析, 建立最优预测模型。凌秀华等^[4]对麦冬药材的几何形状、颜色、纹理、横切面各组织的面积分数等特征进行提取, 实现有效鉴别不同产地麦冬药材。陶欧等^[5]通过分析 18 种中药材不同样本图像的 26 个纹理特征数, 采用朴素贝叶斯及 BP 神经网络 2 种建模方法和十折交叉验证, 建立 18 种中药材的判别模型, 其准确率达到 90%。钱丹丹等^[6]利用计算机视觉系统提取中药饮片图像的特征并输入到朴素贝叶斯分类器, 其综合分类准确率达到 94%。药材通过加工制成后的中药饮片也可以利用深度学习进行分类和识别训练, 并建立模型^[7]。模型提取到足够的特征后, 利用卷积神经网络对中药材图像进行识别。

此前手工提取特征的方法容易受图像噪声、背景复杂性等影响, 造成提取特征出现偏差。随着深

收稿日期: 2022-04-20

基金项目: 广东省现代农业关键技术模式集成与示范推广项目(粤财农【2021】37号-200011); 国家自然科学基金项目(31671591; 31971797); 广州市科技计划项目(202002030245); 广东省科技专项(2020020103); 广东省现代农业产业技术体系创新团队建设专项(2021KJ108); 广东省教育厅特色创新类项目(2019KTSCX013); 2020年广东省科技创新战略专项(pdjh2020a0084); 广东省大学生创新创业项目(S202010564150; 202110564042)

孙道宗, E-mail: sundaozong@scau.edu.cn

通信作者: 王卫星, E-mail: weixing@scau.edu.cn

度学习理论的发展,基于学习的方法能够自动化提取数据特征,因此鲁棒性更好^[8]。Xu等^[9]提出了1种新的用于中药材识别的注意金字塔网络 APN(attentional pyramid networks)。APN可以自适应地对不同特征尺度的中草药图像进行建模,将其运用至构建好的中药数据集上开展试验并获得2%的性能提升。除了与图像识别结合,深度学习也能够和电子鼻技术进行结合,在原有的深度卷积神经网络(DCNN)上进行优化,在卷积层和池化层上使用特殊的一维内核,使其适合于电子鼻的数据,实现对多种类型的中药材分类^[10]。孙鑫等^[11]利用卷积神经网络研究中药饮片图像识别,运用 Softmax 函数来优化 CNN 识别模型,在 50 种中药饮片图像中可实现 70% 的平均识别精度。Anvarkhah 等^[12]利用机器视觉技术,针对药用植物种子的形态和颜色 2 种特征参数的组合,进行识别和检测研究,找出准确率最优的特征组合。基于深度学习的草本植物自动分类识别系统是根据叶片形状来识别植物类型,其方法可以应用到中药材图像信息库,进而建立更高效的、快速的药材信息检索方法^[13]。Husin 等^[14]设计一种便捷高效的草本植物分类系统来提高分类效率,研究的重点在于中药叶片的形状和纹理特征的识别方法,识别准确率达到 98.9%。Sladojevic 等^[15]利用深度卷积网络进行植物病害识别,该模型的试验精度在 91%~98%,单独分类测试的平均精度为 96.3%。

现有研究对药材识别的方法多为提取特征后再输入到卷积网络得到分类结果,但是仍然存在识别准确率不高、生成模型存储空间较大的问题,对硬件设备要求较高而难以应用到嵌入式或移动式设备上。Tan 等^[16]在 2019 年发布了基于 CNN 的主干特征提取网络 EfficientNetv1,其使用基于强化学习^[17-18]的神经架构搜索,将图像输入分辨率 r 、网络的深度 $depth$ 以及卷积核个数 $width$ 这 3 个参数进行合理化配置,从而设计一个性能较好的主干网络。相较于 AlexNet、VGGNet、GoogleNet 等经典网络, EfficientNetv1 模型在 ImageNet 上有更高的精度,并且卷积网络体量更小,速度更快。但 EfficientNetv1 网络也存在一些不足,比如训练图像要求的尺寸较大时,训练速度会非常慢,训练迭代周期较多时对硬件资源损耗较大;对每个 stage 的深度和宽度都是统一且同等放大,然而不同的 stage 在网络的训练速度以及参数数量的贡献并不相同,因此,这种同等缩放存在不合理之处。EfficientNetv2^[19]针对 EfficientNetv1

的不足进行改进,提出了改进后的渐进学习方法,该方法根据训练图像的尺寸动态调节正则方法。在模型缩放中采用非均匀的策略来解决同等缩放带来的不合理问题,通过试验获得结果与此前的网络相比,训练速度和参数数量上都有一定的优势。

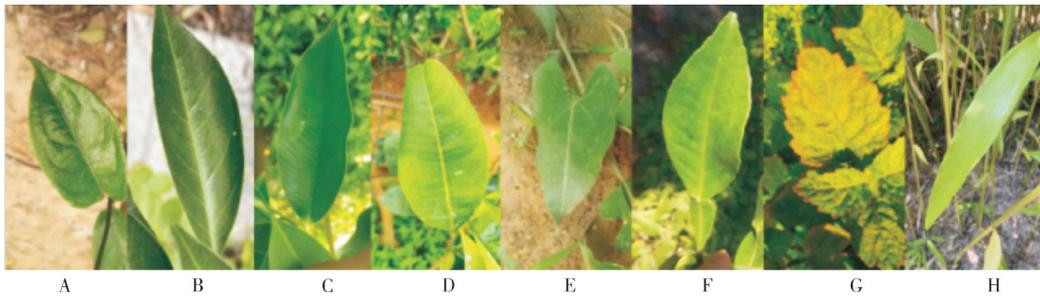
为进一步提高分类网络的泛用性,优化网络结构,解决在复杂场景下南药材快速、高精度识别问题,本研究选择基于结构复杂但精密度较高的 EfficientNetv2 网络模型进行改进,在保持高识别准确率前提下进一步提升网络的性能,旨在进一步提高南药材的分类准确率和分拣效率。

1 材料与方法

1.1 南药图像数据集的制作和预处理

1) 试验数据采集。采样的药材基地包括位于广州市天河区柯木塍的广东省农业技术推广总站以及位于肇庆市高要区的南药省级现代农业产业园,涵盖广藿香(*Pogostemon cablin*)、沉香(*Aquilaria sinensis* (Lour.) Gilg)、巴戟天(*Morinda officinalis* How)、何首乌(*Fallopia multiflora*)、化橘红(*Citri Grandis Exocarpium*)、广佛手(*Citri sarcodactylis* Fructus)、陈皮(*Pericarpium Citri Reticulatae*)和阳春砂(*Amomum villosum* Lour)共 8 种南药,如图 1 所示。使用 SONY IMX386 1 200 万像素带 $f/2.0$ 光圈的摄像头,距离每种叶片 5~10 cm、拍摄时取景角度在 $-49^\circ \sim 66^\circ$,拍摄周期内采集 10 次以上的数据,涵盖晴天、阴天和雨天 3 种常见气象情况,并且全部为室外情况,其光照强度范围为 $50 \sim 1 \times 10^8$ lx,共采集南药叶片图像 2 466 张。图像采集期间包含晴天、多云、阴天及小雨 4 种天气情况,采集时段集中于 09:00—12:00,拍摄方位涵盖顺光、侧光、逆光、侧顺光和侧逆光。

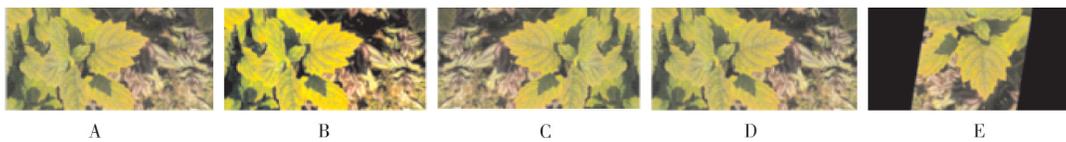
2) 数据扩增。为了提高训练过程的模型精确率和泛用性,本研究对训练样本进行数据扩增处理,扩增方法包含对比度增强、高斯噪声、仿射变换、镜面翻转 4 种手段,扩增效果如图 2 所示。通过对训练样本的扩增处理,将广藿香、沉香、巴戟天、何首乌、化橘红、广佛手、陈皮、阳春砂样本数分别从 309、304、308、304、304、297、320、320 张扩增至 945、1 020、1 040、1 020、1 020、980、1 100、1 100 张,整体数据集由原来的 2 466 张增加至 8 225 张图片,数量相较于扩增前约提升 3 倍,防止模型出现过拟合现象,提高模型的泛化能力。



A:巴戟天 *Morinda officinalis*; B:沉香 Eaglewood; C:广佛手 Fingered citron; D:陈皮 Tangerine peel; E:何首乌 *Polygonum multiflorum*; F:广藿香 Cablin patchouli; G:化橘红 Tomentose pummelo peel; H:阳春砂 Villous amomum fruit.

图1 南药叶片图像样本

Fig.1 Leaf of southern medicine image samples



A:原图 Original image; B:对比度增强 Contrast enhancement; C:镜面翻转 Mirror flip; D:高斯模糊 Gaussian blur; E:仿射变换 Affine transformation.

图2 数据扩增处理后的效果图

Fig.2 The effect picture after data amplification processing

3)数据预处理。由于扩增后图像样本数量较大,为了进一步提升网络模型的训练速度、收敛性以及准确率,对训练图像数据集进行预处理。本研究采用来自 Pytorch 官方的算法工具库所提供 transforms.Compose 组件对南药叶片图像进行张量化处理、归一化处理和正态分布处理,其中 transform.ToTensor()将输入图像数据原本(W(宽度),H(高度),C(通道数))的形状转换为(C,H,W)形状的 Tensor 格式,在将所有数据除以 255,归一化到[0,1.0];正态分布处理则调用 transforms.Normalize 方法。将 RGB 3个通道上的均值都设为 0.5,3个通道的标准差也设置为 0.5,参数 mean 和 std 均以 0.5 的形式传递,使图像的灰度归一化在[0,1],处理后的数据符合标准正态分布,模型更容易收敛。

1.2 改进 EfficientNetv2 网络模型的构建

1) EfficientNetv2 相较 EfficientNetv1 的改进点。EfficientNetv2 网络中在原本 EfficientNetv1 网络的浅层中引入了 Fused-MBConv 模块,模块结构变化如图 3 所示。由图 3 可知,Fused-MBConv 是在原来的 MBConv 结构基础上,用 1×1 的卷积层替换了原本结构中的 1×1 的卷积层和深度可分离 3×3 卷积层,该网络模块替换 v1 网络浅层的 MBConv 层后进行训练,能够显著提升训练速度。EfficientNetv2 模型将原本 EfficientNetv1 中卷积核大小为 5×5 的部分全

部用 3×3 代替、降低输入/输出通道增加倍率值的大小、移除 EfficientNetv1 中第 8 个 stage,进一步降低内存访问产生的开销并提高训练速度。

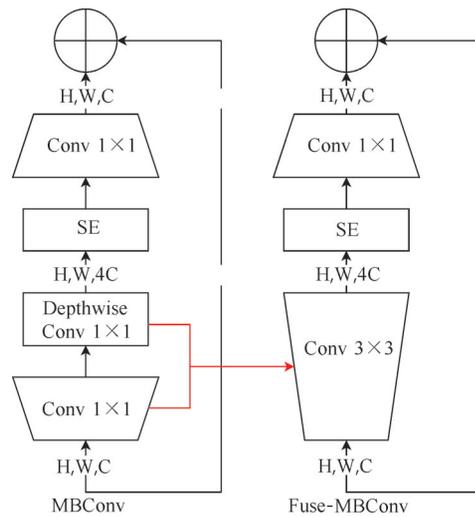


图3 MBConv 和 Fused-MBConv 的网络结构

Fig.3 Structure of MBConv and Fused-MBConv

2) 改进 EfficientNetv2 网络模型的构建。EfficientNetv2 网络提供了 S、M 以及 L 共 3 种初始模型,本研究选取最轻量的初始模型 EfficientNetv2-S 作为改进基础,该模型有 44 层网络结构,共分为 7 个阶段。初始模型 EfficientNetv2-S 是原作提供的几个初始模型中最轻量的 1 个,但是在训练时也存在运算量较大、单次训练时间较长、模型空间占用大的问题,因此对原有的网络结构层实行改进操作。改进后的

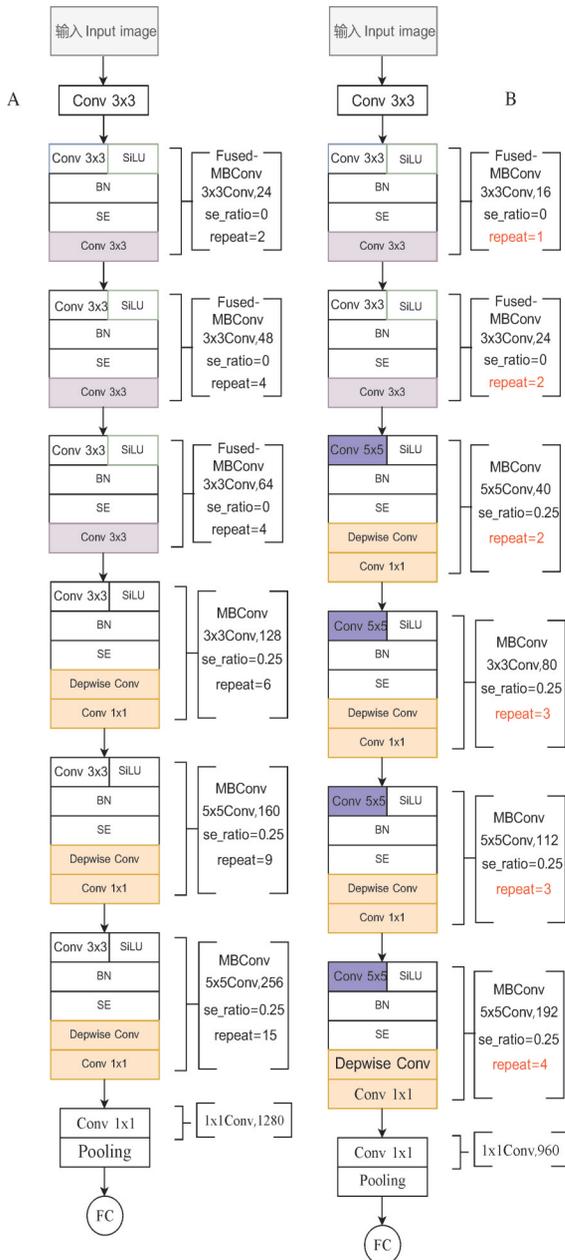
EfficientNetv2-S 模型结构在模型的 7 个阶段里, 第一阶段和最后一阶段与原网络结构保持一致, 通道数发生变化, 其余的主要针对中间的第 2~6 阶段部分进行重新设计。网络结构改进前后如图 4 所示。图 4A 中, Conv3×3 即 3×3 卷积和 SiLU 激活函数; BN 即批归一化层, SE=0.25 表示使用了 SE 模块, 0.25

表示 SE 模块中首个全连接层的节点个数是输入该 MBConv 模块特征矩阵通道数 1/4, repeat 即重复层数。将原本在前 3 个调用 Fused-MBConv 的阶段调整为前 2 个阶段使用 Fused-MBConv, 余下 3~6 这 4 个阶段使用 MBConv 模块, 该改动的原因是 MBConv 比 Fused-MBConv 有更小的扩展比, 而更小的扩展比往往具有较少的内存访问开销。因此, 调整了 Fused-MBConv 的使用架构, 提高模型的感受野大小, 降低模型复杂度和深度, 从而进一步提高模型的轻量化。

第二步是将第 5~17 层的 3×3 卷积核替换为 5×5 卷积核, 即图 4B 中蓝色标记的部分, 这一步是结合了 EfficientNetv1 的思想到改进模型中。在初始的 EfficientNetv2-S 模型中大量使用 3×3 卷积核, 导致感受野减小, 从而迫使整个网络需要增加更多卷积层进行弥补来保证模型精度。将部分 3×3 卷积层替换为 5×5 卷积核既可以提高精度, 也能减少卷积层数量。

在前两步中, 由于调整 MBConv 模块数量和部分卷积层替换更大的 5×5 卷积核, 使得整个网络的感受野范围得到提升, 因此不需要依赖更多的卷积层来维持模型精度。根据初始模型的比例对网络层数进行缩减。按初始模型的比例将第 4~6 阶段的 6、9、15 层缩减至 3、3、4 层, 即图 4B 中红色字体标记的 repeat 部分, 进一步降低模型的复杂度和冗余。经过上述调整, 整个网络结构从原来的 44 层缩减至 19 层。其中, 将 1~6 阶段共 40 层的 3 个 Fused-MBConv+3 个 MBConv 的网络结构, 优化为 17 层的 2 个 Fused-MBConv+4 个 MBConv 的网络结构, 初始通道数由初始的 24 调整为 16, 最后在 1×1 卷积的全连接层作用下将特征图进行降维处理, 降维后特征的分辨率没有发生变化, 但是通道数变为 960, 改进后的网络结构如表 1 所示。

3) 引入 Adam 优化算法。在 EfficientNetv2-S 中, 采用的是固定步长随机梯度下降法 (stochastic gradient descent, SGD)^[20]。虽然该优化算法对梯度要求相对较低, 但在随机选择梯度的过程中会增加噪声, 导致权值更新比率会出现偏差。因此, 引入自适应矩估计 (adaptive moment estimation), 即 Adam^[21] 优化算法替代原先的 SGD 优化算法。可以用训练数据的迭代来更新所用神经网络中的各个权重, 在计算梯度的一阶矩估计和二阶矩估计时, Ad-



A: 改进前的 EfficientNetv2-S 网络结构图 The structure of EfficientNetv2-S network before improvement; B: 改进后的 EfficientNetv2-S 网络结构图 The structure of EfficientNetv2-S network after improvement.

图 4 改进前后的 EfficientNetv2-S 网络模型结构对比图
Fig.4 Comparison of the structure of EfficientNetv2-S network model before and after improvement

表1 改进的EfficientNetv2-S的结构
Table 1 Improvement of EfficientNetv2-S architecture

阶段 Stage	模块 Operator	步长 Stride	通道数 Channels	层数 Layers
0	Conv3×3	2	16	1
1	Fused-MBCConv k3×3	1	16	1
2	Fused-MBCConv k3×3	2	24	2
3	MBCConv, k5×5, SE0.25	2	40	2
4	MBCConv, k3×3, SE0.25	2	80	3
5	MBCConv, k5×5, SE0.25	1	112	3
6	MBCConv, k5×5, SE0.25	2	192	4
7	Conv1×1 & Pooling&FC	—	960	1

am会为各个参数设置独立的学习率。其算法更新规则如式(1)~(3)所示。

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g^2 \quad (2)$$

$$\text{variable} = \text{variable} - r_t m_t / (v_t + \epsilon) \quad (3)$$

式(1)~(3)中, m_t 是计算历史梯度的一阶指数平滑值, v_t 是计算历史梯度平方的一阶指数平滑值, g 指的是上一轮刚更新的梯度值, β_1, β_2 是矩估计的指数衰减率, variable是指计算变量的更新值, r_t 为学习率, ϵ 为除零误差项, 目的是避免除数变为0。文中的Adam优化算法里, 初始学习率是 10^{-2} , 指数衰减率 β_1 和 β_2 分别是0.9和0.999, 除零误差项 ϵ 是 10^{-8} 。

4)损失函数的选取。MultiMarginLoss函数适用于相似且复杂背景下的多分类任务, 在具体训练过程中对网络的收敛性有一定的提升。其表达式如式(4)所示。

$$\text{Loss} = \frac{1}{N} \max(0, \text{margin} - x_n[y_n] + x_n[i])^p \quad (4)$$

在式(4)中, x 为神经网络的输出, y 是真实的类别标签, N 是类别标签数, y_n 的取值范围为 $[0, C-1]$, p 值默认取1, margin值代表边距, 可人为设定阈值, 一般取默认值1。

1.3 改进 EfficientNetv2-S 分类模型的训练与评估

1)模型的训练环境。基于Python3.8编程语言的Pytorch深度学习框架, Ubuntu操作系统, 8核16线程的Intel E5-2620的CPU, GPU为内存12GB的GTX TITAN X显卡4张, 调用CUDA平台网络训练过程实行加速, 集成开发环境为Pycharm。

2)模型评价指标。为了衡量模型的性能, 本研究分别与初始提供的3类模型进行参数量(parameters)和浮点运算次数(FLOPs)指标的对比, 其中参数量用于衡量模型复杂度, FLOPs用于衡量模型运

算性能。除此之外, 模型还采用单次验证法进行模型性能的评估。具体操作如下: 将数据集按8:2的比例划分为训练集和测试集, 然后使用训练集对模型进行指定迭代周期的训练, 接着在测试集上进行客观评价指标的计算。为了对训练出的模型进行客观评价, 采用测试准确率(test accuracy)、损失值(loss)、模型存储空间大小(model storage space)和平均训练时间(training time)等评价指标。

3)网络训练过程。初始模型EfficientNetv2-S在大型数据集ImageNet上进行预训练后产生的模型权重文件, 选取pre_EfficientNetv2-S.pth进行迁移模型训练, 通过迁移学习能够提升神经网络模型的收敛性, 进一步提高模型分类的精确度。对于改进EfficientNetv2-S模型在不同种类和不同超参数模型上的对比, 采用单次验证法来训练和评估模型的性能。模型的训练流程如图5所示。由图5可知, 整个网络的训练过程具体描述为: 在扩增后的样本量基础上划分训练集和测试集, 并在Pytorch深度学习框架中, 加载预训练模型, 获取与设定的批大小处理量相等数量的图像。利用Adam优化算法和学习率衰减机制对每个迭代次数(epoch)的学习率(learning rate)进行相应地调整, 使其随着迭代次数逐渐减小。每批次样本在梯度下降时更新1次参数, 迭代次数设置为100, 批大小为32。在训练集上对改进EfficientNetv2-S进行训练, 并且MultiMarginLoss对分类概

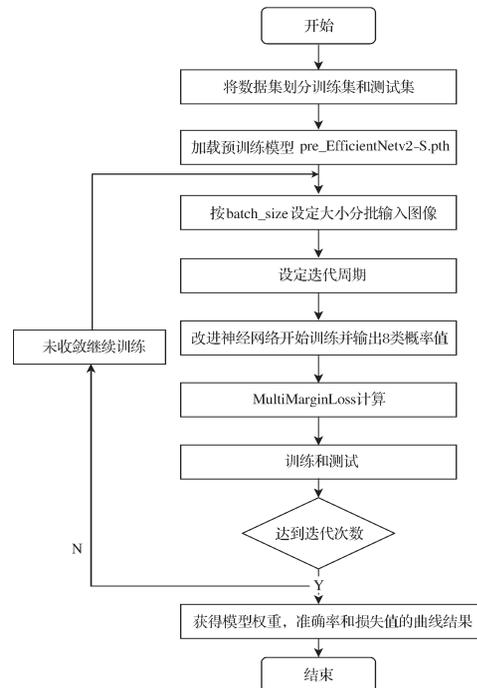


图5 改进 EfficientNetv2-S 模型训练流程图

Fig.5 Improved EfficientNetv2-S training flow chart

率值进行优化,在测试集上获得每个迭代次数的模型测试结果。

4)超参数优化。学习率的大小会对模型的收敛速度有较大的影响,在 Adam 优化算法中,学习率设定过低会使模型收敛速度放慢,甚至降低模型精确率。如果设定太高,梯度有一定概率会在最小值的范围内上下波动,以至于可能出现无法收敛的情况。因此,需要对多个量级的学习率进行试验和比较。试验中初始批大小设置为 32^[22],尽可能通过较大的批大小处理量来改善收敛不稳定的问题。选取比较的学习率值分别为 0.01、0.001、0.000 1 和 0.000 01,

观察这 4 种学习率在模型训练中产生的损失值和准确率的效果(图 6)。由图 6 可见,当学习率为 0.000 01 时,模型收敛速度相比较其他学习率是最慢的,第 100 个周期时的准确率为 88.6%,均低于另外 3 种学习率的准确效果。而学习率为 0.01 和 0.000 1 时则收敛速度较慢,到第 20 个周期,2 个模型分别是 0.422 和 0.429 才开始趋近收敛。相较之下,学习率为 0.001 时,损失值下降地较快,收敛效果最好,损失值较低,说明学习率为 0.001 时具有更佳的效果,因此选择学习率为 0.001。

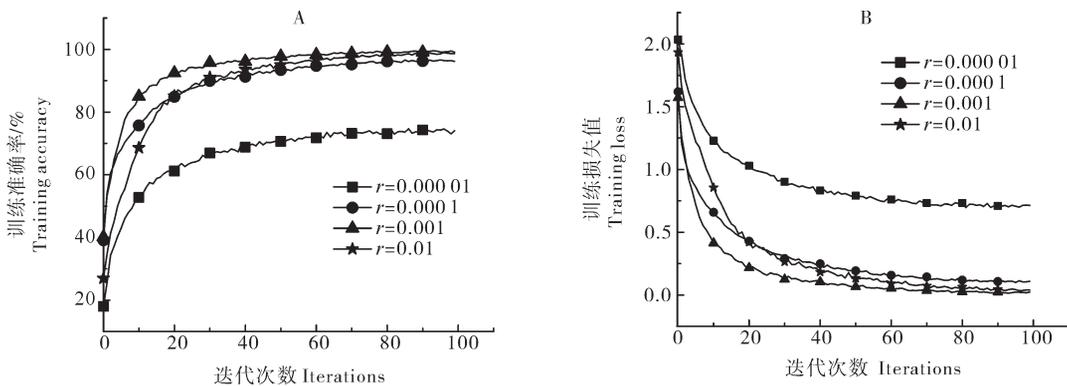


图 6 学习率调节对准确率(A)和损失值(B)的影响

Fig.6 Influence of learning rate tuning on accuracy(A)and loss value(B)

2 结果与分析

2.1 不同优化器处理的优化效果

构建复杂背景图像的 8 种南药叶片图像数据集,通过超参数优化确定了 Adam 优化器的学习率为 0.001 后,设定迭代周期为 100 次,批大小为 32,通过训练横向对比 Rprop、RMSprop、Adagrad 以及 SGD 等 4 种优化器的优化效果,观察不同优化器下的训练

准确率和损失值结果(图 7)。由图 7 可见,Adam 优化器在收敛性方面要好于另外 4 个优化器,Adam 优化器的准确率为 99.12%,Adagrad、RMSprop 和 SGD 分别为 96.84%、97.40% 和 97.97%,三者较为接近,Rprop 的准确率最低,为 41.23%。损失值方面,Adam 优化器为 0.002 17,Adagrad、RMSprop、SGD 和 Rprop 分别为 0.105 5、0.075 36、0.002 59 和 1.502,结果表明 Adam 优化器优化效果均优于其他 4 个优化器。

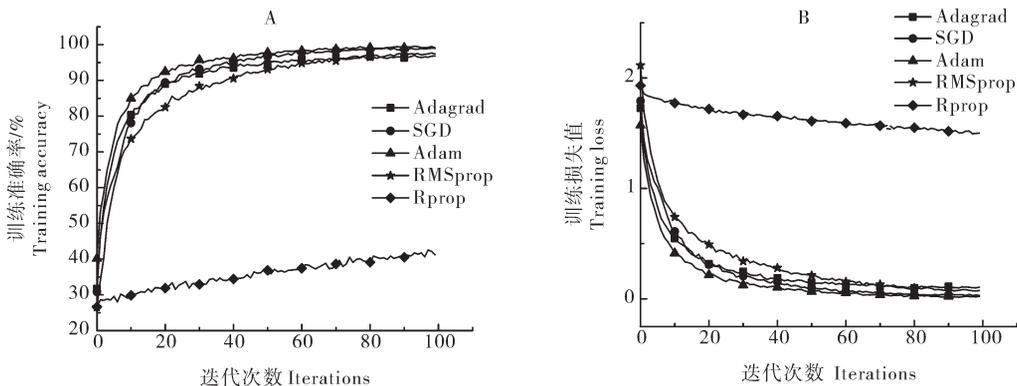


图 7 不同优化器下模型的准确率(A)和损失值(B)

Fig.7 Comparison of the results of accuracy(A) and loss values(B) of the model under different optimizers

2.2 网络模型改进后的效果

利用 torchstat 工具计算每个模型参数量和浮点运算次数以及训练时所占用内存大小,将改进 EfficientNetv2-S 模型与 EfficientNetv2-S、EfficientNetv2-M、EfficientNetv2-L 这 3 个模型进行比较,其结果如表 2 所示。由表 2 可知,改进后的 EfficientNetv2-S 模型在参数量、运算次数以及占用内存大小等 3 个指标都明显优于初始的 3 个模型,相较初始最轻量的 EfficientNetv2-S 模型在参数量和运算次数方面都下降了约 85%,表明改进模型结构在复杂度上有所下降,模型性能明显提升。除了上述参数量和计算次数,在保持其他超参数、批大小、学习率、迭代周期等变量一致的基础上,还比较了原始模型 EfficientNetv2-S 和改进后的 EfficientNetv2-S 网络模型在测试准确率、损失值、模型空间存储大小和训练时间等 4 个指标的差异,进一步观察改进后的网络模型性能变化情况。本研究将南药叶片图像数据集按照 8:2 比例划分为训练集和测试集,迭代次数为 100,学习率和批大小分别取 0.001 和 32,则改进前后模型的种类识别结果如图 8 所示。由图 8 可知,改进后的

表 2 不同模型的性能参数对比

Table 2 Comparison of performance parameters of different models

模型 Model	参数量/Mb Parameters	浮点运算次数/G FLOPs	占用内存/Mb Memory size
改进 EfficientNetv2-S Improved Efficient-Netv2-S	2.91	0.79	81.52
EfficientNetv2-S	19.25	5.35	270.26
EfficientNetv2-M	50.41	10.2	438.49
EfficientNetv2-L	111.81	23.2	773.26

EfficientNetv2-S 模型在测试训练集的准确率为 99.12%,相较于初始网络模型 EfficientNetv2-S 的 97.97%,提升了约 1.17%。损失值方面,改进模型和初始模型分别为 0.002 17 和 0.002 59,降低 16.21%。从结果上看,改进后的 EfficientNetv2-S 模型在准确率和损失值两方面比 EfficientNetv2-S 模型有小幅领先。从图 8 也可看出,改进后的 EfficientNetv2-S 模型的准确率和损失值在收敛性上优于初始的 EfficientNetv2-S 模型。

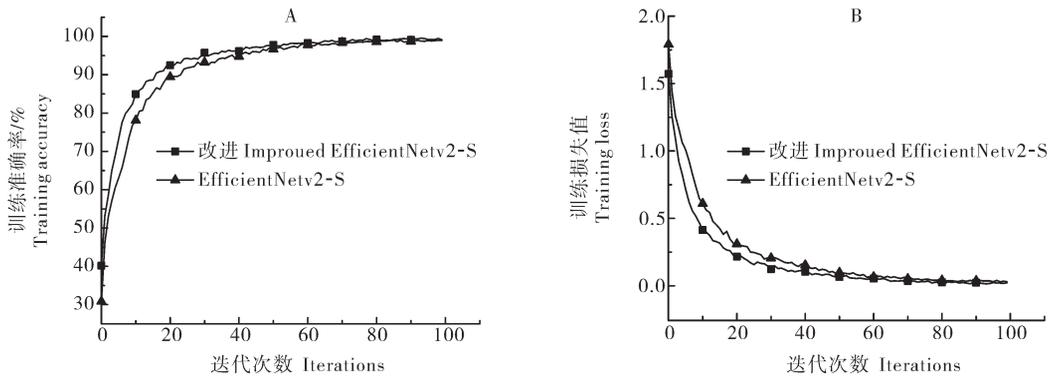


图 8 改进前后模型的准确率(A)和损失值(B)

Fig.8 Accuracy rate(A) and loss value(B) of the model were compared before and after improvent

2.3 不同神经网络模型的对比结果

利用复杂背景图像的 8 种南药叶片图像数据集对改进的 EfficientNetv2-S 模型进行训练,并与其他轻量级的网络模型进行对比。运用迁移学习训练模型,以 8:2 比例划分为训练集 6 580 张和测试集 1 645 张,不同模型的预处理方式、优化算法等基本设置保持相同,迭代次数 100,学习率和批大小分别取 0.001 和 32,不同模型的种类识别结果对比见表 3。

由表 3 可知:(1)改进的 EfficientNetv2-S 模型在单次的平均训练时间上与 DenseNet121、ShuffleNet、RegNet 3 个轻量级分类网络相比差距在 1.29%~

5.19%,并且均低于 80 s;与 EfficientNetv1 模型相比,训练时间减少 79 s,缩短 50.64%。在训练过程中还发现改进后的 EfficientNetv2-S 模型收敛性和准确性明显高于其他模型,特别是相较 RegNet 的准确性有明显的优势(图 9)。(2)经过改进的 EfficientNetv2-S 模型在模型存储空间上得到大幅压缩,相比 DenseNet121、RegNet 和 ShuffleNet 模型存储空间大小分别降低 55.30%、21.26% 和 19.31%,模型存储空间的轻量程度有一定的明显优势。(3)改进的 EfficientNetv2-S 模型在模型存储空间和平均训练时间这 2 个指标保持一致或领先的条件下,相较于 EfficientNetv1 模型,准确率提高 0.69%,损失值降低

表 3 不同模型的种类识别结果比较

Table 3 Comparison of the results with different kinds of different mode

模型 Model	准确率/% Accuracy	损失值 Loss	模型存储空间/kb Model storage space	训练时间/s Training time
DenseNet121	97.98	0.022 30	27 294	76
RegNet	76.88	0.414 96	15 496	73
ShuffleNet	98.52	0.006 43	15 120	75
EfficientNetv1	98.44	0.004 49	42 383	156
EfficientNetv2-S	97.97	0.002 59	79 742	147
改进 Improved EfficientNetv2-S	99.12	0.002 17	12 201	77

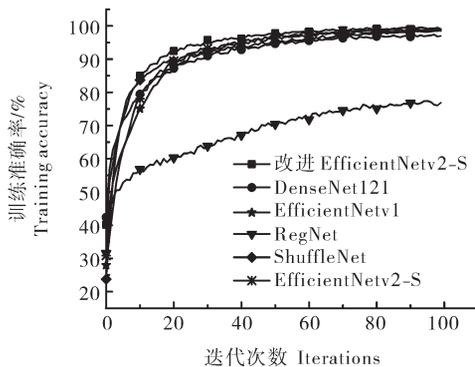


图 9 不同模型的训练准确率结果对比

Fig.9 Comparison of the training accuracy of different models for species recognition

51.67%。改进后的 EfficientNetv2-S 模型在模型存储空间上相较于初始模型从 79 742 kb 下降到 12 201 kb,降幅为 84.70%,存储空间优化效果明显,单次训练所耗费的平均时间从 147 s 下降到 77 s,模型训练时间下降 47.62%,表明改进的 EfficientNetv2-S 模型在保持高准确率基础上,能够做到对空间存储大小和训练时间的优化,实现复杂分类网络模型的轻量化。说明改进的 EfficientNetv2-S 模型在保持高准确度的情况下,还能进一步降低模型损失值,并且在此条件下降低训练时间和模型存储空间大小,其模型性能要明显优于 EfficientNetv1 模型。

3 讨论

本研究在解决目标多分类问题时,原始模型采用交叉熵损失函数进行模型优化,使模型输出逼近理想情况下的输出结果。虽然交叉熵损失的收敛效果良好,但是仍然存在局限性:对于包含复杂背景信息的图像,即叶片像素数量明显小于背景像素数量时,会导致交叉熵损失函数中背景信息成分会占据主导,使网络模型明显偏向背景,导致识别效果有所下降^[23]。Goutham (http://cs230.stanford.edu/projects_fall_2021/reports/103171653.pdf)分别比较了 MultiMar-

gin 损失函数和交叉熵损失函数对模型的影响,其结果显示在准确性方面 MultiMarginLoss 比交叉熵损失函数高 0.62%~1.00%,并且,具有复杂背景的信息熵较大,容易导致交叉熵不确定性增加,影响模型准确性。而 MultiMarginLoss 是基于边距的多分类合页损失函数,不容易受到影响,因此采用 MultiMarginLoss 作为改进 EfficientNetv2-S 模型的损失函数。在深度学习的卷积神经网络领域中,要获取最优的权值结果,就需要优化算法在网络模型中损失函数,从而得到最小的损失函数值,而求解最优化问题最常用的方法就是梯度下降算法。深度学习中常用的优化算法包括 SGD、RMSprop、Adam、Adguard 等^[24]。本研究提出引入 Adam 优化器并进行超参数优化,获取精度较高的学习率对模型进行训练。

本研究提出一种基于改进 EfficientNetv2-S 模型的多品种南药叶片种类识别方法,通过调整 MBConv 模块结构,替换部分卷积核,加载预训练权重进行迁移学习,对模型的网络结构复杂度和冗余度进行一定精简,实现轻量化目标。将改进后模型对多品种南药叶片数据集进行多组对照试验,结果显示:与初始模型 EfficientNetv2-S 相比,改进后的 EfficientNetv2-S 在参数量和计算次数都明显下降,在对南药叶片种类进行识别的过程中,准确率、损失值、训练时间、存储空间大小等指标明显优于初始模型,说明该改进模型的识别准确度较高,并且模型的复杂度和参数量下降,模型性能得到进一步提升。与其他轻量级的分类网络模型进行横向对比,结果表明,在准确率方面,改进 EfficientNetv2-S 要比另外 3 个轻量级分类网络模型都有所提升,其中对比 DenseNet121 提升效果较为显著。改进 EfficientNetv2-S 分类模型能够保持高精度识别效果的同时减少模型参数量,达到轻量化与高精度之间的平衡。

本研究采用卷积神经网络并对网络模型进行轻量化重构,实现了在保持高精确率的同时进一步降低模型复杂度和存储空间,对于嵌入到移动式设备或单

片机等对存储空间和实时传输条件要求较高的环境而言有较好的适用性。该模型仍有进一步改进的空间,比如测试多种不同学习率衰减机制来确定更适合的更新梯度算法组合。对数据集方面如果能考虑到南药叶片在更多气候条件下的成像效果,再进行训练和测试能够进一步提高模型的泛用性和鲁棒性。

参考文献 References

- [1] 詹若挺,刘军民,陈立凯,等.广东省南药生产发展现状调查[J].广州中医药大学学报,2020,37(9):1836-1843.ZHAN R T, LIU J M, CHEN L K, et al. Investigation on the production status of Lingnan herbal medicine in Guangdong Province[J]. Journal of Guangzhou University of Traditional Chinese Medicine, 2020, 37(9):1836-1843(in Chinese with English abstract).
- [2] 朱黎辉,李晓宁,张莹,等.基于形状特征及纹理特征的中药材检索方法[J].计算机工程与设计,2014,35(11):3903-3907. ZHU L H, LI X N, ZHANG Y, et al. Image retrieval method for Chinese herbal medicine based on shape features and texture features[J]. Computer engineering and design, 2014, 35(11):3903-3907(in Chinese with English abstract).
- [3] 杨万里,段凌凤,杨万能.基于深度学习的水稻表型特征提取和穗质量预测研究[J].华中农业大学学报,2021,40(1):227-235. YANG W L, DUAN L F, YANG W N. Deep learning-based extraction of rice phenotypic characteristics and prediction of rice panicle weight[J]. Journal of Huazhong Agricultural University, 2021, 40(1):227-235(in Chinese with English abstract).
- [4] 凌秀华,卢文彪,王耐,等.基于图像处理技术的麦冬药材特征提取与识别[J].辽宁中医杂志,2017,44(7):1460-1462. LING X H, LU W B, WANG N, et al. Feature extraction and recognition of ophiopogonis *Radix* based on image processing[J]. Liaoning journal of traditional Chinese medicine, 2017, 44(7):1460-1462(in Chinese with English abstract).
- [5] 陶欧,林兆洲,张宪宝,等.基于饮片切面图像纹理特征参数的中药辨识模型研究[J].世界科学技术-中医药现代化,2014,16(12):2558-2562. TAO O, LIN Z Z, ZHANG X B, et al. Research on identification model of Chinese herbal medicine by texture feature parameter of transverse section image[J]. World science and technology/Modernization of traditional Chinese medicine and materia medica, 2014, 16(12):2558-2562(in Chinese with English abstract).
- [6] 钱丹丹,周金海.基于计算机视觉的中药饮片检测与分级研究[J].时珍国医国药,2019,30(1):203-205. QIAN D D, ZHOU J H. Study on detection and grading of Chinese herbal pieces based on computer vision[J]. Lishizhen medicine and materia medica research, 2019, 30(1):203-205(in Chinese with English abstract).
- [7] 王耐,卢文彪,凌秀华,等.牛膝和川牛膝药材的特征提取与图像识别[J].中国药房,2017,28(12):1670-1673. WANG N, LU W B, LING X H, et al. Feature extraction and image recognition of *Achyranthes bidentata* and *Cyathula officinalis*[J]. China pharmacy, 2017, 28(12):1670-1673(in Chinese with English abstract).
- [8] SUN X, QIAN H. Chinese herbal medicine image recognition and retrieval by convolutional neural network [J/OL]. PLoS One, 2016, 11(6):e0156327 [2022-04-20]. <https://doi.org/10.1371/journal.pone.0024357>.
- [9] XU Y X, WEN G H, HU Y, et al. Multiple attentional pyramid networks for Chinese herbal recognition [J/OL]. Pattern recognition, 2021, 110:107558 [2022-04-20]. <https://doi.org/10.1016/j.patcog.107558>.
- [10] WANG Y, DIAO J W, WANG Z, et al. An optimized deep convolutional neural network for *Dendrobium* classification based on electronic nose [J/OL]. Sensors and actuators A: physical, 2020, 307: 111874 [2022-04-20]. <https://doi.org/10.1016/j.sna.2020.111874>.
- [11] 孙鑫,钱会南.基于深度卷积网络的中药饮片图像识别[J].世界科学技术-中医药现代化,2017,19(2):218-222. SUN X, QIAN H N. Identification of the images of Chinese herb slices with deep convolutional network [J]. World science and technology/Modernization of traditional Chinese medicine and materia medica, 2017, 19(2):218-222(in Chinese with English abstract).
- [12] ANVARKHAH S, HAJEH-HOSSEINI M K, DAVARIEDALAT-PANAH A, et al. Medicinal plant seed identification using machine vision [J]. Seed science and technology, 2013, 41(1):107-120.
- [13] BIEN D. Deep convolutional neural network based plant species recognition through features of leaf [J]. Multimedia tools and applications, 2021, 80(4):6443-6456.
- [14] HUSIN Z, SHAKAFF A Y M, AZIZ A H A, et al. Embedded portable device for herb leaves recognition using image processing techniques and neural network algorithm [J]. Computers and electronics in agriculture, 2012, 89:18-29.
- [15] SLADOJEVIC S, ARSENOVIC M, ANDERLA A, et al. Deep neural networks based recognition of plant diseases by leaf image classification [J/OL]. Computational intelligence and neuroscience, 2016, 2016: 3289801 [2022-04-20]. <https://doi.org/10.1155/2016/3289801>.
- [16] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [J/OL]. arXiv, 2019, 1905.11946 [2022-04-20]. <https://arxiv.org/abs/1905.11946>.
- [17] ZOPH B, LE Q V. Neural architecture search with reinforcement learning [J/OL]. arXiv, 2016, 1611.01578 [2022-04-20]. <https://arxiv.org/abs/1611.01578>.
- [18] BAKER B, GUPTA O, NAIK N, et al. Designing neural network architectures using reinforcement learning [J/OL]. arXiv, 2016, 1611.02167 [2022-04-20]. <https://arxiv.org/abs/1611.02167>.
- [19] TAN M X, LE Q V. EfficientNetV2: smaller models and faster training [J/OL]. arXiv, 2021, 2104.00298 [2022-04-20]. <https://arxiv.org/abs/2104.00298>.
- [20] MOULINES E, BACH F. Non-asymptotic analysis of stochastic approximation algorithms for machine learning [J]. Advances in neural information processing systems, 2011, 24: 451-459.

- [21] KINGMA D P, BA J. Adam: a method for stochastic optimization [J/OL]. arXiv, 2014, 1412.6980 [2022-04-20]. <https://arxiv.org/abs/1412.6980>.
- [22] SMITH S L, KINDERMANS P J, YING C, et al. Don't decay the learning rate, increase the batch size [J/OL]. arXiv, 2018, 1711.00489 [2022-04-20]. <http://arxiv.org/abs/1711.00489>.
- [23] ZHANG Z L, SABUNCU M R. Generalized cross entropy loss for training deep neural networks with noisy labels [J/OL]. arXiv, 2018, 1805.07836 [2022-04-20]. <https://arxiv.org/abs/1805.07836>.
- [24] 王天翔. 深度学习中梯度下降算法的应用研究 [D]. 金华: 浙江师范大学, 2019. WANG T X. Applied research on gradient descent algorithm in deep learning [D]. Jinhua: Zhejiang Normal University, 2019 (in Chinese with English abstract).

Classification of leaves of multi-variety southern traditional Chinese medicine based on improved EfficientNetv2 model

SUN Daozong^{1,2}, LIU Jinyuan¹, DING Zheng¹, LIU Huan¹,
PENG Jiajun¹, XIE Jiaying^{1,2}, WANG Weixing^{1,2}

1. College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China;

2. Guangdong Engineering Research Center for Monitoring Agricultural Information, Guangzhou 510642, China

Abstract For the application of deep learning in the field of Chinese herbal medicine, especially in the field of southern Chinese medicine, the complex background will reduce the accuracy of recognition. If a complex network structure is used, high computing power is required to support training and detection, but the actual embedded or mobile devices are difficult to meet, which affects the effect of on-site detection. This article proposed to improve the EfficientNetv2 network model to classify and identify 8 kinds of southern Chinese medicine leaves in the complex background collected in the field. The network structure was redesigned. The scope of the Fused-MBConv and MBConv architectures was adjusted. Some 3×3 convolutional kernels were replaced with 5×5 convolutional kernels to increase the perceptual field size, reduce the number of convolutional layers of the network, and to further reduce the network complexity. The transfer learning was introduced to train the model. The adaptive moment estimation optimization algorithm was used to optimize the hyperparameters with multiple tests to determine the learning rate. MultiMarginLoss was selected as the loss function to solve the problem of complex background information affecting the accuracy of recognition. The diversity of the data set was increased and the problem of model over-fitting was avoided by adopting data augmentation methods including affine transformation and Gaussian blur and other methods to the experimental data set to improve the stability of the model training process. The results showed that the model improved achieved 99.12% accuracy in recognizing the image of southern Chinese medicine leaves with complex backgrounds, 1.17% more accurate than the baseline model EfficientNetv2-S. The size of parameters was reduced by 85% approximately. The average training time was reduced by 47.62%. The improved model had significant advantages in model storage space, accuracy, and training time, comparing with lightweight models including DenseNet121, ShuffleNet, and RegNet. It is indicated that the model proposed performs well in classifying leaves of multi-variety southern Chinese medicine. The lightweight degree and performance of the model are further improved.

Keywords image recognition; leaves of southern traditional Chinese medicine; variety identification; convolutional neural network; improved EfficientNetv2 network; hyperparameter optimization

(责任编辑: 陆文昌)