

苏航, 陈旭昊, 寿德荣, 等. 基于注意力机制轻量化模型的植物病害识别方法[J]. 江苏农业学报, 2024, 40(8): 1389-1399.
doi: 10.3969/j.issn.1000-4440.2024.08.004

基于注意力机制轻量化模型的植物病害识别方法

苏航^{1,2,3}, 陈旭昊^{1,2}, 寿德荣¹, 张朝阳¹, 许彪³, 孙丙宇²

(1.重庆三峡学院机械工程学院, 重庆 404100; 2.中国科学院合肥物质科学研究院智能机械研究所, 安徽 合肥 230000; 3.中国工程物理研究院, 四川 成都 610000)

摘要: 针对现有植物病害识别模型存在响应速度慢、参数量多、计算机内存资源消耗大等问题, 本研究提出了一种轻量化神经网络模型, 该模型由特征提取层、特征增强层和分类器组成。为了减小模型大小并提高网络响应速度, 在特征提取层中使用深度可分离卷积进行特征提取。为了防止网络传播过程中的梯度消失并增强病害像素特征融合, 在特征提取层中引入了大卷积核倒置残差结构(IRBCKS)模块。此外, 在特征增强层集成了轻量级卷积块注意力模块(CBAM)注意力机制, 以捕捉植物病害相关图像中像素之间的关系, 增强关键信息的提取。最后, 采用剪枝技术剔除模型中冗余特征信息, 从而再次减少模型参数量, 形成最终的轻量级网络模型 Cut-MobileNet。为验证该模型的先进性, 将其与轻量化模型(MobileNet V2、SqueezeNet、GoogLeNet)和非轻量化模型(Vision Transformer、AlexNet)进行性能对比, 研究结果表明, Cut-MobileNet 在浮点运算量、准确率、单张图片推理时间、参数量、F1值和模型大小等性能指标上都取得了较优的效果。

关键词: 模型剪枝; 卷积块注意力模块(CBAM)注意力机制; 大卷积核倒置残差结构(IRBCKS)模块; 植物病害; 轻量化网络

中图分类号: S608 文献标识码: A 文章编号: 1000-4440(2024)08-1389-11

Plant disease recognition method based on lightweight model with attention mechanism

SU Hang^{1,2,3}, CHEN Xuhao^{1,2}, SHOU Derong¹, ZHANG Chaoyang¹, XU Biao³, SUN Bingyu²

(1.School of Mechanical Engineering, Chongqing Three Gorges University, Chongqing 404100, China; 2.Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230000, China; 3.China Academy of Engineering Physics, Chengdu 610000, China)

Abstract: In light of the issues associated with slow response speed, numerous parameters, and high computational memory requirements in existing plant disease recognition models, we proposed a lightweight neural network model. The model consisted of feature extraction layer, feature enhancement layer, and classifier. To reduce model size and increase network response speed, we utilized deep separable convolution in the feature extraction layer. To prevent gradient disappearance during network propagation and enhance the fusion of disease pixel features, we introduced the inverted residual block convolution kernel structure (IRBCKS) module into the feature extraction layer. Furthermore, we integrated a lightweight convolutional

block attention module (CBAM) attention mechanism into the feature enhancement layer to capture the relationships between pixels in plant disease-related images and enhance key information extraction. Finally, we employed a pruning technique to eliminate redundant feature information from the base model, thereby reducing the number of model parameters once again, yielding this lightweight network model, Cut-MobileNet. In order to verify the progressiveness of

收稿日期: 2023-11-29

基金项目: 国家自然科学基金项目(61773360); 2019年重庆市人工智能+智慧农业学科群开放基金项目(ZNNYKFA201901)

作者简介: 苏航(1996-), 男, 陕西渭南人, 硕士研究生, 研究方向为智能制造、农业信息化研究、图像处理等。(E-mail) 935659856@qq.com

通讯作者: 孙丙宇, (E-mail) bysun@iim.ac.cn

this model, it was compared with lightweight models (MobileNet V2, SqueezeNet, GoogLeNet) and non-lightweight models (Vision Transformer, AlexNet). The results show that better results have been achieved by Cut-MobileNet in floating-point operation, accuracy, single image inference time, parameter count, F1 value, and model size.

Key words: model pruning; convolutional block attention module (CBAM) attention mechanism; inverted residual block convolution kernel structure (IRBCKS) module; plant diseases; lightweight networks

作物在生产过程中易受到病虫害的影响^[1],特别是一些病毒、真菌、细菌、霉菌的感染会导致农产品产量和质量降低。在作物生长过程中,病害的及早发现至关重要,及时对病害进行精准判别再实施救治,可以大大减少作物生产损失。目前植物病害诊断方法主要是通过人们的经验对病害种类做出识别,然而,在生产实践过程中,凭借经验对病害种类做出分类会耗费更多的人力和物力资源,由于病害种类繁多,症状相似,即使是有经验的农民或专家也可能无法正确识别病害^[2],这也带来了效率和可靠性的问题,所以人工判别不适用于对大规模种植区域内的植物病害识别^[3]。

随着科技的不断发展,人工智能(AI)视觉技术已经逐渐应用到农业生产中。早期人们利用支持向量机法^[4-5]、随机森林法^[6]、K-均值法^[7]等机器学习算法进行植物病害的分类^[8],然而,这些方法的构建过程相当繁琐,且泛化能力差。近年来,深度学习似乎引领着计算机视觉领域的研究^[9]。

自 2012 年 AlexNet 网络^[10]在 ImageNet^[11]上成功验证了深度学习方法的可行性后,深度学习领域便迅速发展起来。Jia 等^[12]使用 Keras/Tensorboard 框架搭建了 VGG16 深度学习网络模型,对番茄病虫害进行了分类,研究结果表明,该模型对番茄病虫害分类的准确率为 89.00%。Yadav 等^[13]改进了 ResNet18 模型,在原始模型内添加了 SE (Squeeze-and-Excitation) 模块^[14]和修改了层结构,试验结果显示,改进后的模型平均准确率为 99.63%,比原始模型的准确率提高了 0.53%,取得了相当高的准确率。Yang 等^[15]和 Szegedy 等^[16]设计了 1 个 22 层的深度

学习网络 GoogLeNet, GoogLeNet 网络提出了 Inception 的网络架构,提高了计算资源利用率,相应地增加了网络的深度和宽度。然而,这些模型在复杂环境中仍然存在特征提取不充分、植物病害识别准确率较低、推理速度较慢等问题。

综上,本研究拟提出一种轻量化的植物病害识别模型 Cut-MobileNet,该模型可以在短时间内对植物病害精准识别,可部署于资源有限的移动端,农户可利用该模型高效诊断病害,提高种植的效率 and 效益。

1 材料与方法

1.1 数据集的建立

本研究数据集源自 Kaggle 竞赛公开数据集,根据试验要求收集了 4 种作物的数据,包括番茄、芒果、小麦和苹果,共 32 种病害,由于 Kaggle 公开数据集中存在部分病害分类错误的问题,研究开始前进行了筛选,经筛选后的数据集图像总计 7 274 张,其中 80% 的数据被划分为训练集,20% 的数据被划分为验证集。模型在训练过程中,数据集图片调整为 224×224 像素的大小。

1.2 数据预处理

图像预处理的方法一般有图像白化、图像缩放、图像裁剪、图像翻转、图像旋转、图像仿射、图像加噪以及调整图像亮度、色度、饱和度等方式,在本模型训练过程中采用的是随机图像明暗度的调整、图像大小缩放、裁剪与反转、添加高斯噪声、添加遮挡的图像处理方式。图 1 是以番茄叶片样本为例展示的图像预处理的方法。



图 1 图像预处理方式

Fig.1 Image preprocessing methods

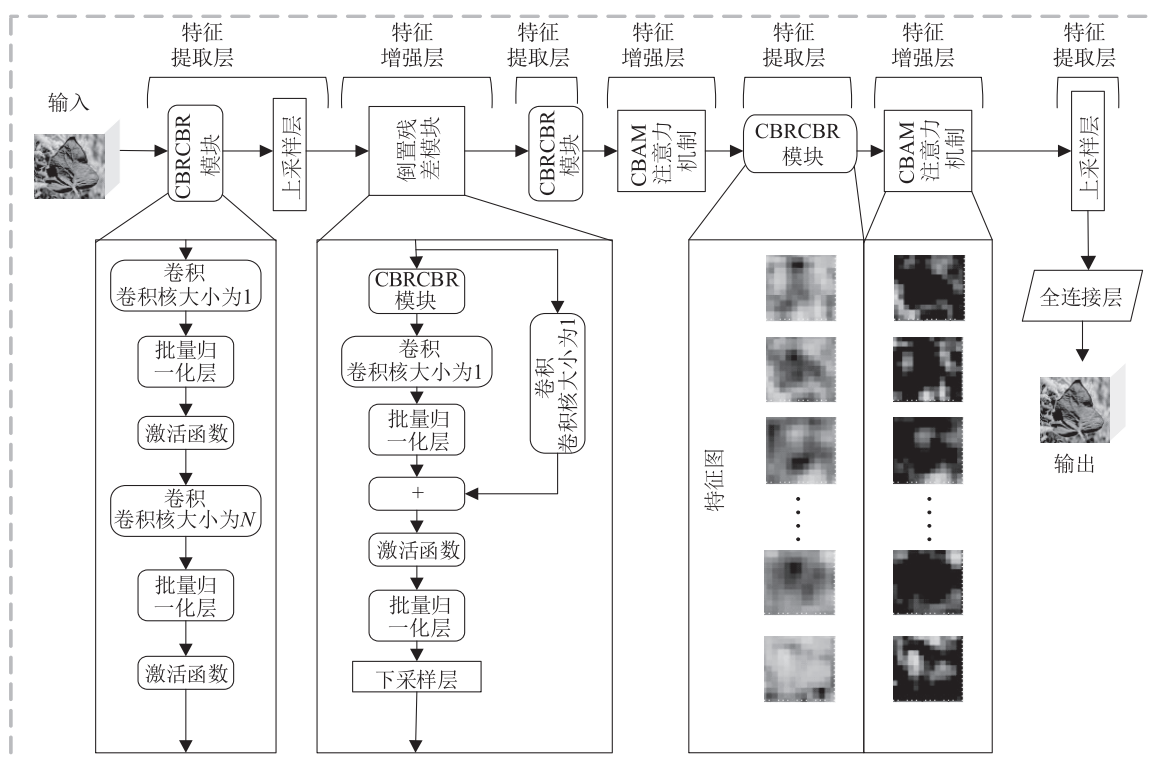
1.3 试验环境

本研究使用 Windows 11 操作系统,计算平台环境:中央处理器(CPU)型号为 12th Gen Intel(R) Core(TM) i7-12700F 2.10 GHz,图形处理器(GPU)型号为 NVIDIA GeForce RTX 3060 Ti,随机存取存储器(RAM)容量为 16 GB,cuda GPU 加速模型 CUDA 版本为 11.3,深度学习框架为 pytorch1.10,编程语言为 python3.8。本模型训练参数:Learning rate

设置为 0.000 1,优化器为 Adam、Batch,数量为 32,epoch 为 600,稀疏因子为 0.000 1。

1.4 Cut-MobileNet 模型构建

本研究提出的模型结构如图 2 所示,先根据轻量化构建方法得到 1 个准确率较高、参数量较少的神经网络模型 MobieNet-Chase,然后再对该模型进行剪枝处理,得到最终的轻量化模型 Cut-MobileNet,以上 2 种模型结构相同,唯一不同之处在于通道数。



CBRCBR 表示双重卷积、批量归一化和激活操作;CBAM 表示卷积块注意力模块。

图 2 Cut-MobileNet 网络结构图

Fig.2 Cut-MobileNet network structure diagram

模型的主干结构由 3 个主要部分组成:特征提取层、特征增强层和分类器。

模型层结构及层操作如表 1 所示。首先,在第一个特征提取层(CBRCBR)中,主要任务是提取叶片区域的颜色信息。这一步操作可以获取叶片的基本颜色特征,为后续处理提供基础数据。其次,在首个特征增强层中,引入大卷积核倒置残差结构(IRBCKS)模块^[17]。该模块有助于加强网络对病害区域的关注度,同时防止在传播过程中梯度消失的问题。通过这一层的处理,模型能够更好地识别和区分健康与病变的叶片区域。再次,在特征增强层的基础上,再次引入一个特征提取层,以提取浅层过滤后的高维信息。这一步操作

有助于丰富模型的视觉特征信息,提升模型的识别能力。接着,引入一个轻量化的卷积块注意力模块(CBAM)注意力机制作为一个特征增强层^[18-19]。该机制能够增强高级特征的提取能力,进一步提升模型对叶片病害区域的识别精度。为了使网络能够提取更多的高维信息,在网络深层再次引入了 1 个特征提取层、1 个特征增强层和 1 个特征提取层。这些额外的层次有助于网络更深入地分析和提取叶片图像中的复杂特征。最后,将全连接层作为分类器,将前面处理得到的特征输入并转化为最终的分类结果。全连接层根据前面步骤得到的特征信息,对叶片图像进行分类判断,从而得到最终的输出结果。通过上述流程,该模型能够

有效地对叶片图像进行特征提取、病害识别和分类,为 农业智能化提供有力的支持。

表 1 MobileNet-Cbase 网络模型层结构及层操作

Table 1 MobileNet-Cbase network model architecture and layer operations

层结构	层操作
CBRCBR 模块层	卷积核大小为 1,卷积操作,通道数 48;卷积核大小为 11,卷积操作,通道数 48
池化层	池化操作
IRBCKS 模块层	卷积核大小为 1,卷积操作,通道数 192;卷积核大小为 11,卷积操作,通道数 192;卷积核大小为 1,卷积操作,通道数 128
	卷积核大小为 1,卷积操作,通道数 128
池化层	池化操作
CBRCBR 模块层	卷积核大小为 1,卷积操作,通道数 192;卷积核大小为 3,卷积操作,通道数 192
注意力机制层	CBAM 注意力机制
CBRCBR 模块层	卷积核大小为 1,卷积操作,通道数 128;卷积核大小为 3,卷积操作,通道数 128
注意力机制层	CBAM 注意力机制
池化层	池化操作
全连接层	全连接操作,向量长度 915;全连接操作,向量长度 32

IRBCKS:大卷积核倒置残差结构;CBRCBR 表示双重卷积、批量归一化和激活操作;CBAM 表示卷积块注意力模块。

1.4.1 嵌入大卷积核倒置残差结构和深度可分离卷积 卷积核大小对模型性能尤其是分割、检测等下游任务至关重要^[20]。Cut-MobileNet 模型的残差网络改进了传统 3×3 的卷积核,采用了大卷积核倒置残差结构 (IRBCKS),如图 3 所示,先用 1×1 的卷积核对输入特征图升维卷积,其次用 11×11 的卷积核提取特征,再用 1×1 的卷积核降维,最后,将得到的特征图再与输入特征图融合。11×11 的大卷积核拥有比 3×3 卷积核更大的感受野,与叠加 5 个 3×3 的小卷积核相比,

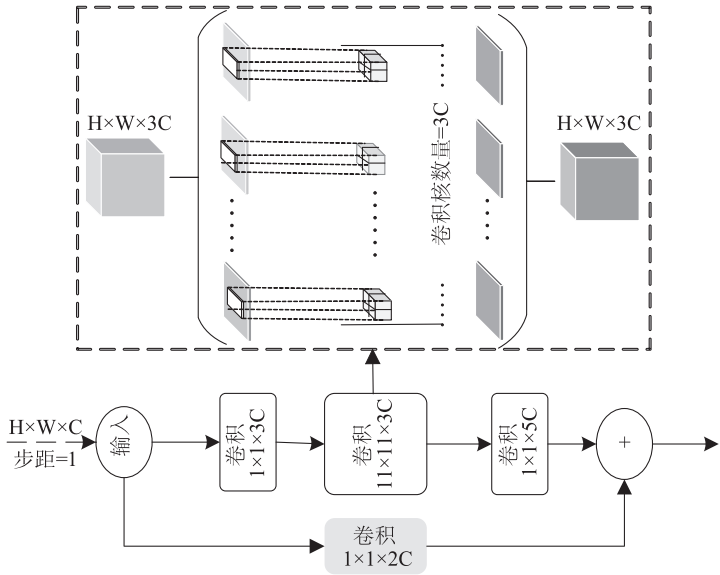
同等感受野的条件下,结果准确率更高。 IRBCKS 使用了深度可分离卷积方法,同等输出条件下,深度可分离卷积方法大大减少了参数量。使用标准卷积核时参数量(N_{std})计算公式:

$$N_{std}=D_k\times D_k\times M\times N\tag{1}$$

使用深度可分离卷积时参数量(N_{dwise})计算公式:

$$N_{dwise}=D_k\times D_k\times M\times 1+M\times N\tag{2}$$

公式(1)和公式(2)中, D_k 为卷积核大小, M 代表输入通道数, N 为输出通道数。



H 表示图像高度;W 表示图像宽度;C 表示通道数。

图 3 使用 11×11 卷积核的倒置残差网络图

Fig.3 Diagram of an inverted residual network with a 11×11 convolutional kernel

1.4.2 CBAM 注意力机制 模型在神经网络的倒数第 1 个和倒数第 3 个卷积层后分别添加了 CBAM 注意力机制模块,在浅层网络传播过程中获取的都是一些低维特征,故只在网络深层添加注意力机制,加强模型对高维特征的关注。该机制能够自适应地学习图像的通道和空间关系,更好地捕捉图像的关键特征,从而提高模型的性能。

CBAM 注意力机制有 2 个主要步骤,即先对特征图进行通道注意力加权,再考虑特征图空间信息,如图 4 和图 5 所示。具体流程如下:

首先,对输入特征图 F 进行全局最大池化和全局平均池化后生成 2 个 $1 \times 1 \times C$ (C 为通道数) 的特征

图。然后,将这 2 个特征图输入到神经网络 (MLP) 中,对 MLP 神经网络输出的特征图进行融合和激活操作,得到通道注意力特征。最后,将通道注意力特征和输入特征相乘得到空间注意力机制模块所需要的输入特征 F' 。该输入特征经过最大池化后再进行平均池化操作得到的 2 个特征图,其维度大小为 $H \times W \times 1$ (H 为图像高度; W 为图像宽度),再经过 Concat 拼接方法将 2 个同维度、同尺寸的特征图拼接,之后通过 7×7 的卷积核卷积得到通道数为 1 的特征图,进而再经过 1 个激活函数得到输出特征图,最后将输出特征与输入特征相乘变回 $C \times H \times W$ 大小的维度。

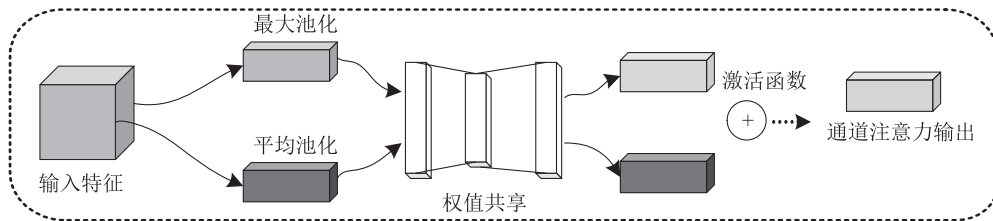


图 4 通道注意力机制模块

Fig.4 Channel attention mechanism module

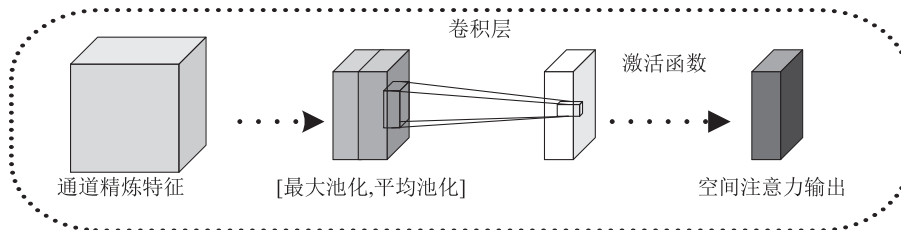


图 5 空间注意力机制模块

Fig.5 Spatial attention mechanism module

通道注意力公式:

$$Mc(F) = \sigma \{ MLP[AvgPool(F)] + MLP[MaxPool(F)] \} \quad (3)$$

式中, F 表示原始特征图, Mc 表示通道注意力机制输出特征, MLP 表示权值共享神经网络, $AvgPool$ 表示平均池化, $MaxPool$ 表示最大池化, σ 表示倍率因子。

空间注意力公式:

$$Ms(F) = \sigma \{ f^{7 \times 7} [AvgPool(F); MaxPool(F)] \} \quad (4)$$

式中, F 表示原始特征图, Ms 表示空间注意力机制输出特征, $f^{7 \times 7}$ 表示 7×7 的卷积操作, $AvgPool$ 表示平均池化, $MaxPool$ 表示最大池化, σ 表示倍率因子。

1.4.3 模型剪枝与批量归一化 Cut-MobileNet 模型采用了批量归一化 (BN) 层剪枝策略,如图 6 所

示,即根据 BN 层正则化后的 γ 系数的相对大小进行局部剪枝,将 γ 系数相对较小的通道去除,达到模型压缩的目的^[21]。当 γ 系数大于 1 时,输出的参数缩放到较大的范围,则表示轻度归一化,当 γ 系数小于 1 时,输出的参数缩放到较小的范围,则表示过度归一化。一般情况下,网络在训练完之后, γ 的数值很少分布在 0 附近,难以划分冗余通道。为了有效去除网络冗余通道,在 BN 层网络中引入稀疏因子 λ 进行训练,使得 BN 层变得稀疏。公式 (5) 是网络模型的目标函数,在目标函数中添加正则项可以使模型更加稀疏,提高系数 γ 的稀疏性。经过稀疏后的模型 γ 数值分布会有向 0 靠近的趋势,选择合适的稀疏速率,就可对通道剪枝。

公式(6)是卷积神经网络的常规损失函数,目标函数和常规损失函数的区别是增加了第二项,第二项为正则项,用来惩罚模型中的权重,抑制模型的复杂性。

$$L = \sum_{(x,y)} l[f(x,W),y] + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (5)$$

$$L = \sum_{(x,y)} l[f(x,W),y] \quad (6)$$

公式(5)和公式(6)中, (x,y) 是训练的输入坐标, W 是网络中的参数, λ 是稀疏因子, γ 是权重系

数, Γ 是一个集合,包含所有正则项的参数。

上述目标函数和常规损失函数的区别是增加了第二项,第二项为正则项,用来惩罚模型中的权重,抑制模型的复杂性,在目标函数中添加正则项可以使模型更加稀疏,提高系数 γ 的稀疏性。经过稀疏后的模型 γ 数值分布会有向 0 靠近的趋势,选择合适的稀疏速率,就可对通道剪枝。

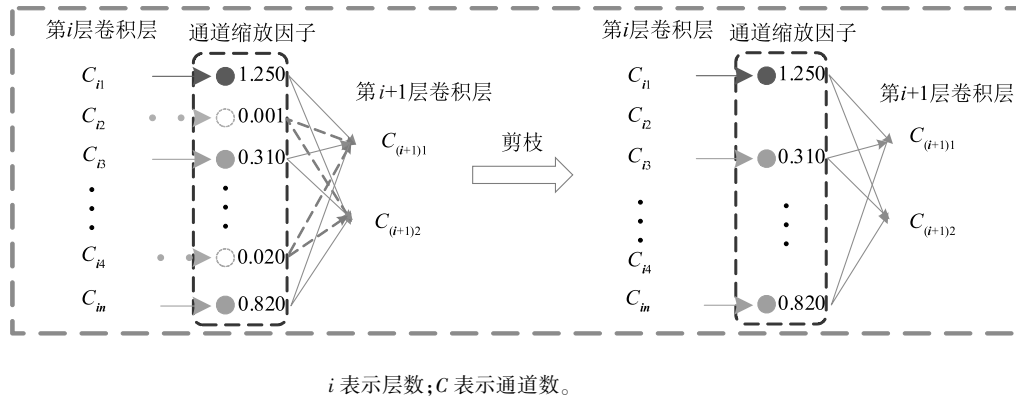


图6 模型剪枝示意图

Fig.6 Model pruning schematic diagram

具体剪枝流程如图7所示,首先对稀疏化训练后的模型中所有BN层的权重进行统计排序,指定保留BN层数量,以此设定权重阈值,剪去低于阈值的冗余通道。剪枝后的模型,对其进行训练和微调,

使其达到较高的精度。本研究使用循环迭代法对模型剪枝,即训练、剪枝、微调循环,直到获得最优的模型为止,该方法避免了大规模剪枝对模型的破坏,同时使得模型性能逐渐趋于最优。

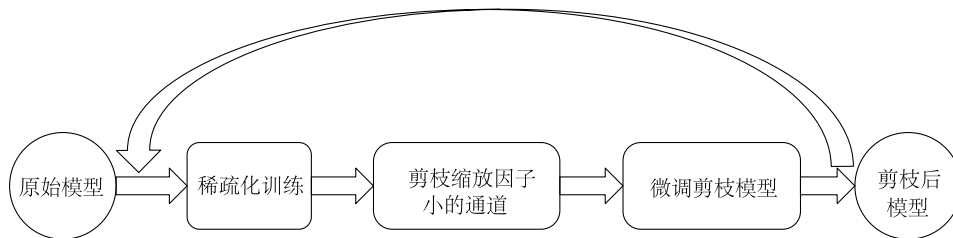


图7 模型剪枝流程

Fig.7 Model pruning flowchart

2 结果与分析

2.1 评价指标

本研究中模型的评价指标有精确率 (Precision)、召回率 (Recall)、F1 值、准确率 (Accuracy)、模型参数量、模型大小 (内存占用)、浮点运算量 (FLOPs)、每秒传输帧数 (FPS)。

准确率: 预测正确的样本占总样本的百分比:

$$Accuracy = \frac{TP+TN}{TP+TN+FP} \quad (7)$$

式中, TP 代表真阳性: 被模型预测为正类的正样本; TN 代表真阴性: 被模型预测为负类的负样本; FP 代表假阳性: 被模型预测为正类的负样本。

召回率: 实际为正的样本中被预测为正样本的概率:

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

式中, FN 代表假阴性: 被模型预测为负类的正样本。

精确率: 所有预测为正的样本中实际为正样本

的概率:

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

F1 值即 $F1_Score$:精确率和召回率的调和平均值:

$$F1_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

浮点运算量 ($FLOPs$) 衡量算法的复杂度:

$$FLOPs = 2HW(C_{in}K^2 + 1)C_{out} \quad (11)$$

$$FLOPs = (2I - 1)O \quad (12)$$

每秒传输帧数 (FPS):

$$FPS = \frac{1}{inference} \quad (13)$$

模型参数量计算方法参考公式(1)和公式(2);公式(11)为卷积层 $FLOPs$ 的计算公式,其中 H 为图像高度, W 为图像宽度, C_{in} 为输入通道数, C_{out} 为输出通道数, K^2 为卷积核大小;公式(12)为全连接层 $FLOPs$ 的计算公式,其中 I 为输入维度, O 为输出维度;公式(13)中, $inference$ 表示单张图片推理时间。

2.2 初步形成轻量化模型与剪枝结果

首先,以轻量化结构思想为基础,初步构建了轻量化模型 MobieNet-Cbase,训练后模型性能如表 2 所示,识别准确率为 92.8%,具有较高的识别精度,模型参数量为 5.3 M,通道数为 1 376,较为轻量化,因此可以将此模型作为预剪枝的基础网络。

表 2 模型剪枝前与迭代剪枝后各模型性能比较结果

Table 2 Comparison results of model performance before and after iterative pruning

模型	剪枝前通道数	剪枝后通道数	剪枝前准确率(%)	剪枝后准确率(%)	参数量(M)
MobieNet-Cbase	1 376	1 376	92.8	92.8	5.3
MobieNet-Cbase1th	1 376	1 018	92.8	94.1	4.3
MobieNet-Cbase2th	1 018	788	94.1	94.0	4.0
MobieNet-Cbase3th	788	636	94.0	94.0	2.8
Cut-MobileNet	636	600	94.0	93.8	2.2
MobieNet-Cbase5th	600	546	93.8	92.1	1.9

MobieNet-Cbase 表示剪枝前模型;MobieNet-Cbase1th 表示第 1 次剪枝后的模型;MobieNet-Cbase2th 表示第 2 次剪枝后的模型;MobieNet-Cbase3th 表示第 3 次剪枝后的模型;Cut-MobileNet 表示第 4 次剪枝后的模型;MobieNet-Cbase5th 表示第 5 次剪枝后的模型。

其次,根据 BN 层的权重分布来设定阈值,对稀疏训练后的 MobieNet-Cbase 进行通道剪枝,模型首次剪枝后,得到通道数为 1 018 的 MobieNet-Cbase1th 模型,模型进行微调后的准确率为 94.1%,比 MobieNet-Cbase 模型的准确率有微小的提升,参数量减少了 19%,模型大小也降低。模型接下来的 2 次剪枝方法和首次相同,第 2 次和第 3 次剪枝微调训练后的模型分别为 MobieNet-Cbase2th、MobieNet-Cbase3th,其准确率均为 94.0%,模型参数量分别为 4.0 M 和 2.8 M。在准确率几乎不降低的情况下,模型的参数量明显减少。模型在进行第 4 次剪枝微调后的准确率降低 0.2 个百分点,参数量也随之减少了 0.6 M,最后一次剪枝模型为 MobieNet-Cbase5th,相比于 MobieNet-Cbase1th,其参数量明显减少,但识别准确率减少 2.0 个百分点。

综上所述,将模型第 4 次剪枝得到的 Cut-MobileNet 作为最终模型,其参数量仅为 MobieNet-Cbase 的 41.5%,准确率相比于 MobieNet-Cbase 提高了 1.0 个百分点。

2.2.1 对比试验 为了评估 Cut-MobileNet 的性能

表现,采用准确率、模型大小、浮点运算量、精确率、召回率、参数量等指标进行评估。由表 3 可知,Cut-MobileNet 与非轻量化的 AlexNet、Vision Transformer 相比,准确率分别提高了 7.3 个百分点和 0.6 个百分点,其 F1 值也分别提高了 7.3 个百分点和 0.5 个百分点,相较于轻量化的 GoogLeNet 和 SqueezeNet, Cut-MobileNet 的识别准确率和 F1 得分上表现更优,且具有更少的参数量和更小的模型大小。尽管 Cut-MobileNet 相比于 MobileNet V2 的参数量稍高一些,但其准确率相较于 MobileNet V2 有了明显提升。本研究所提的方法在工程应用中具有更高的价值。

综上所述,Cut-MobileNet 在识别性能、模型大小和推理速度等方面表现良好,满足快速准确识别病害的要求,在模型的识别准确度和计算复杂度上达到了相应的平衡,满足嵌入到资源有限的移动设备的要求。

2.2.2 混淆矩阵对比 为了直观地验证本模型的有效性,可以通过混淆矩阵可视化查看模型的分类输出,如图 8 和图 9 所示,横轴数字表示病害类别及其真实标签,纵轴数字表示病害类别及其预测标签,

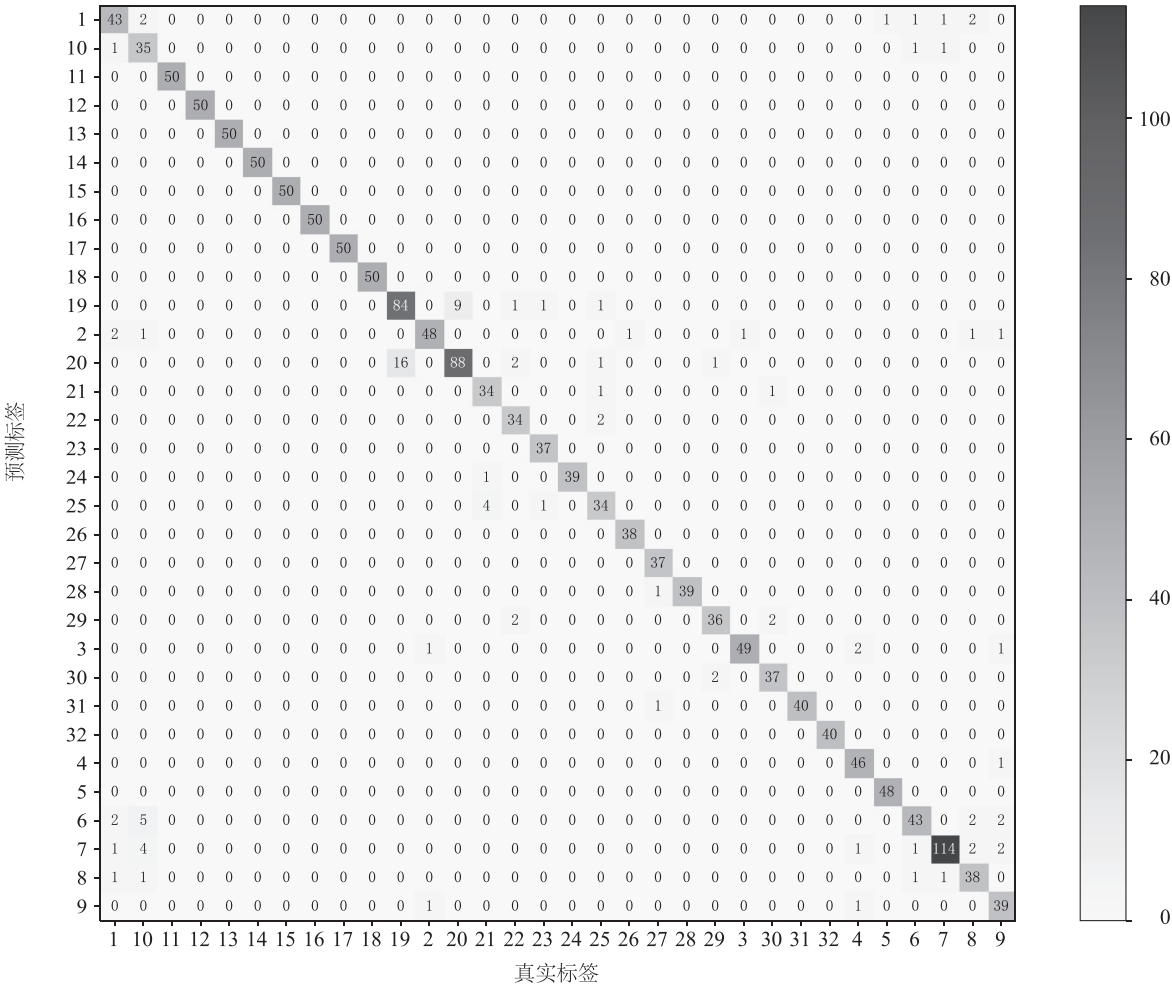
当模型预测为正的样本越多,则表明模型的识别准确率越高,颜色的深浅则代表了模型对某类病害分类效果的优劣。病害类别及其数字标签如表 4 所示,从列举的 Cut-MobileNet 和 MobileNet V2 的混淆矩阵可知,病害种类 7 颜色最深,代表 7 的分类效果

最好,分类准确率最高。MobileNet V2 在病害种类 1 中效果最差,而使用 Cut-MobileNet 后这种现象得以改善。本研究构建的 Cut-MobileNet 模型通过改变网络结构结合模型剪枝技术,在植物病害分类识别上具有显著优势。

表 3 模型性能对比

Table 3 Performance comparison of models

模型	模型大小 (M)	参数量 (M)	浮点运算量 (GFLOPs)	准确率 (%)	精确率 (%)	召回率 (%)	单张图片推理 时间 (ms)
MobileNet V2	8.8	2.3	0.30	87.8	88.8	88.8	4.40
SqueezeNet	15.2	4.0	2.70	91.9	93.1	92.1	3.30
GoogLeNet	39.7	6.0	1.60	89.3	90.2	89.7	5.30
Vision Transformer	1 161.1	304.2	59.70	93.2	94.3	93.6	15.50
Cut-MobileNet	8.7	2.4	0.03	93.8	95.0	94.0	1.65
AlexNet	55.9	14.7	0.30	86.5	87.5	86.8	1.00



坐标轴上刻度代表病害类别,具体见表 4。颜色的深浅代表分类性能的优劣,对角线值代表真阳性值。

图 8 Cut-MobileNet 在 32 类病害中的混淆矩阵
Fig.8 Confusion matrix of Cut-MobileNet in 32 diseases



Fig.9 Confusion matrix of MobileNet V2 in 32 diseases

Table 4 Disease categories and their labels[illegible]

3 讨论

在植物病害识别的领域中,深度学习方法为我们提供了新的解决方案^[22-25]。然而,现有深度学习模型存在参数庞大、推理速度慢或识别准确率不高等问题,导致其难以在存储和计算资源有限的设备上应用。目前,虽然 MobileNet V2、GoogLeNet、SqueezeNet 等轻量化模型在参数量和推理速度上有所优化,但其识别准确率相对较低;而 AlexNet、Vision Transformer 的参数量大,同样不适用于植物病害识别领域。轻量化模型架构与模型剪枝方法相结合,为我们提供了一种全新的解决思路。这种方法能够更好地平衡模型参数量、推理速度和识别准确率之间的关系。当前,众多研究聚焦于轻量化模型在移动端部署的优化,其中主流且实用的方法包括:轻量化模型架构协同蒸馏、知识蒸馏协同模型剪枝、参数共享和量化等^[26-27]。这些方法为植物病害的轻量化模型精准识别提供了有效的思路。

本研究通过试验对比了 3 个轻量化模型(MobileNet V2、GoogLeNet、SqueezeNet)和 2 个非轻量化(Vision Transformer、AlexNet)模型,结果表明,本研究提出的模型在大多数性能指标上均优于对比试验中的模型。这进一步验证了本研究所采用的模型构建方法的优越性和先进性。

4 结论

针对现有植物病害分类模型存在的病害分类准确率与模型大小不平衡的问题,本研究提出了一种基于轻量化结构的神经网络模型 Cut-MobileNet。

Cut-MobileNet 较好地平衡了病害分类准确率和模型大小之间的关系,相比于 MobileNet V2、GoogLeNet、SqueezeNet 等轻量化模型,该模型在植物病害识别任务中具有更高的识别准确率和更小的模型参数量,相比于 Vision Transformer、AlexNet 等非轻量化模型,该模型在识别准确率提升的同时大大减少了模型参数量。

本研究基于模型剪枝方法使模型瘦身,去除模型冗余通道,使得模型参数量和计算量明显减少,加快了植物病害分类模型的响应速度,使其能够应用到资源有限的移动端设备。

参考文献:

- [1] TIAN H K, WANG T H, LIU Y D, et al. Computer vision technology in agricultural automation —a review[J]. *Information Processing in Agriculture*, 2020, 7(1): 1-19.
- [2] 刘拥民,刘翰林,石婷婷,等. 一种优化的 Swin Transformer 番茄叶片病害识别方法[J]. *中国农业大学学报*, 2023, 28(4): 80-90.
- [3] 马 丽,周巧黎,赵丽亚,等. 基于深度学习的番茄叶片病害分类识别研究[J]. *中国农机化学报*, 2023, 44(7): 187-193.
- [4] 牛学德,高丙朋,南新元,等. 基于改进 DenseNet 卷积神经网络的番茄叶片病害检测[J]. *江苏农业学报*, 2022, 38(1): 129-134.
- [5] YAZDIAN H, SALMANI-DEHAGHI N, ALIJANIAN M. A spatially promoted SVM model for GRACE downscaling: using ground and satellite-based datasets[J]. *Journal of Hydrology*, 2023, 626: 130214.
- [6] KHALIFA F, ABDELKADER H, ELSAID A. An analysis of ensemble pruning methods under the explanation of Random Forest[J]. *Information Systems*, 2024, 120: 102310.
- [7] GUAN X, TERADA Y. Sparse kernel K-means for high-dimensional data[J]. *Pattern Recognition*, 2023, 144: 109873.
- [8] 崔兆亿,耿秀丽. 基于随机森林和量子粒子群优化的 SVM 算法[J]. *计算机集成制造系统*, 2023, 29(9): 2929-2936.
- [9] 汤文亮,黄梓锋. 基于知识蒸馏的轻量级番茄叶部病害识别模型[J]. *江苏农业学报*, 2021, 37(3): 570-578.
- [10] ABADE A, FERREIRA P, DE BARROS VIDAL F. Plant diseases recognition on images using convolutional neural networks: a systematic review[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106125.
- [11] SMIRNOV E, TIMOSHENKO D, ANDRIANOV S. Comparison of regularization methods for imageNet classification with deep convolutional neural networks[J]. *AASRI Procedia*, 2014, 6: 89-94.
- [12] JIA S J, JIA P Y, HU S P, et al. Automatic detection of tomato diseases and pests based on leaf images: 2017 Chinese automation congress (CAC)[C]. Jinan: IEEE, 2017.
- [13] YADAV D, JALAL A, GARLAPATI D, et al. Deep learning-based ResNeXt model in phycological studies for future[J]. *Algal Research*, 2020, 50: 102018.
- [14] KHAN M, UDDIN M, PARVEZ M, et al. A squeeze and excitation ResNeXt-based deep learning model for Bangla handwritten compound character recognition[J]. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(6): 3356-3364.
- [15] YANG L, YU X Y, ZHANG S P, et al. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases[J]. *Computers and Electronics in Agriculture*, 2023, 204: 107543.
- [16] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//IEEE. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA, 2015: 1-9.
- [17] RAZA M, NOSHEEN A, YASMIN H, et al. Application of aquatic plants alone as well as in combination for phytoremediation of household and industrial wastewater[J]. *Journal of King Saud Uni-*

- versity-Science,2023,35(7):102805.
- [18] LU Q, YE W X, YIN L F. ResDenIncepNet-CBAM with principal component analysis for wind turbine blade cracking fault prediction with only short time scale SCADA data[J]. Measurement,2023,212:112696.
- [19] CHEN L J, YAO H D, FU J Y, et al. The classification and localization of crack using lightweight convolutional neural network with CBAM[J]. Engineering Structures,2023,275:115291.
- [20] LAU K, PO L, REHMAN Y. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications,2023,236:121352.
- [21] 杨佳昊,左昊轩,黄祺成,等. 基于 YOLO v5s 的作物叶片病害检测模型轻量化方法[J]. 农业机械学报,2023,54(增刊1):222-229.
- [22] 陈智超,汪国强,李 飞,等. 基于 Bi-LSTM 与多尺度神经网络模型的番茄病害识别[J]. 江苏农业科学,2023,51(15):194-203.
- [23] 曹 林,周 凯,申 鑫,等. 智慧林业发展现状与展望[J]. 南京林业大学学报(自然科学版),2022,46(6):83-95.
- [24] 张会敏,谢泽奇. 基于知识图谱与深度学习的黄瓜叶部病害识别方法[J]. 江苏农业科学,2023,51(15):173-178.
- [25] 鲍 彤,罗 瑞,郭 婷,等. 基于 BERT 字向量和 TextCNN 的农业问句分类模型分析[J]. 南方农业学报,2022,53(7):2068-2076.
- [26] 刘媛媛,王定坤,邬 雷,等. 基于知识蒸馏和模型剪枝的轻量化模型植物病害识别[J]. 浙江农业学报,2023,35(9):2250-2264.
- [27] 邵仁荣,刘宇昂,张 伟,等. 深度学习中知识蒸馏研究综述[J]. 计算机学报,2022,45(8):1638-1673.

(责任编辑:陈海霞)