

# 基于注意力机制和多尺度轻量型网络的农作物病害识别

王泽钧<sup>1</sup>, 马凤英<sup>1</sup>, 张 瑜<sup>2</sup>, 张芳芳<sup>1</sup>, 纪 鹏<sup>1</sup>, 曹茂永<sup>1\*</sup>

(1. 齐鲁工业大学(山东省科学院)电子电气与控制学部, 济南 250353;

2. 齐鲁工业大学(山东省科学院)数学与人工智能学部, 济南 250353)

**摘 要:** 准确识别农作物病害并及时防护是保障农作物产量的重要措施。针对传统农作物病害识别模型体积大、准确率不高的问题, 该研究提出一种基于注意力机制和多尺度特征融合的轻量型神经网络模型(Lightweight Multi-scale Attention Convolutional Neural Networks, LMA-CNNs)。首先, 为减少参数量, 使模型轻量化, 网络主体结构采用深度可分离卷积; 其次, 在深度可分离卷积基础上设计出残差注意力模块和多尺度特征融合模块; 同时引入 Leaky ReLU 激活函数增强负值特征的提取。残差注意力模块通过嵌入通道和空间注意力机制, 增强有用特征信息的权重并减弱噪声等干扰信息的权重, 残差连接能够有效防止网络退化。多尺度特征融合模块利用其不同尺度的卷积核提取多种尺度的病害特征, 提高特征的丰富度。试验结果表明, LMA-CNNs 模型在 59 类公开农作物病害图像测试集上的准确率为 88.08%, 参数量仅为  $0.14 \times 10^7$ , 优于 ResNet34、ResNeXt、ShuffleNetV2 等经典神经网络模型。通过比较不同研究者在同一数据集下所设计的网络模型, 进一步验证 LMA-CNNs 模型不仅拥有更高的识别精度, 还具有更少的参数。该研究提出的 LMA-CNNs 模型较好地平衡模型复杂程度和识别准确率, 为移动端农作物病害检测提供参考。

**关键词:** 农作物; 模型; 图像识别; 注意力机制; 多尺度卷积

doi: 10.11975/j.issn.1002-6819.2022.z.020

中图分类号: TP391.4

文献标志码: A

文章编号: 1002-6819(2022)-Supp.-0176-08

王泽钧, 马凤英, 张瑜, 等. 基于注意力机制和多尺度轻量型网络的农作物病害识别[J]. 农业工程学报, 2022, 38(增刊): 176-183. doi: 10.11975/j.issn.1002-6819.2022.z.020 http://www.tcsae.org

Wang Zejun, Ma Fengying, Zhang Yu, et al. Crop disease recognition using attention mechanism and multi-scale lightweight network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(Supp.): 176-183. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.z.020 http://www.tcsae.org

## 0 引 言

农作物病害给农业造成巨大损失。因此, 早期实现农作物病害快速精准识别, 对于后期防治具有重大意义。长期以来, 农作物病害的鉴定通常依赖于农业专家判断<sup>[1]</sup>, 这种方法基于人的主观感知, 很容易造成错判, 并且无法保证时效性, 可能会错过农作物病害最佳治理时间从而造成经济损失。随着计算机技术和图像处理技术的发展, 越来越多的研究者使用智能识别技术鉴定农作物病害。传统的图像识别方法首先对图像进行预处理, 提取病害图像中一些人工标记的特征, 如病斑、病害叶片边缘等特征信息, 并将提取到的信息传入分类器中进行分类。Rumpf 等<sup>[2]</sup>使用高光谱技术和支持向量机(Support Vector Machine, SVM)方法识别甜菜疾病, 试验结果表明, 甜菜健康类和疾病类的准确率均达到 97%; 王健仑<sup>[3]</sup>使用 Canny 算子和 OTSU 算子提取复杂背景中的病害叶片信息, 并使用 K-means 聚类算法提高识别准

确率; Qin 等<sup>[4]</sup>使用苜蓿叶片病害数据集来验证各种聚类算法(如 K\_median 聚类)和监督分类方法(如线性判别分析法)的性能, 来确定识别苜蓿病害的最佳方法。然而, 这种传统图像识别方法需要手工定制冗长的特征, 需要花费大量的时间和人力, 并且只能应用于少数农作物的病害检测, 泛化能力较差。

随着计算机技术进一步发展, 人工智能开始崛起, 深度学习被广范应用在各行各业<sup>[5-7]</sup>。研究表明利用深度学习泛化能力强的优势, 可以对多类农作物病害进行识别和检测<sup>[8-10]</sup>, 识别准确率与传统方法相比有较大提升。黄林生等<sup>[11]</sup>通过对 ResNet<sup>[12]</sup>改进, 增加模型的鲁棒性, 在田间收集的 8 种农作物病害数据集上的平均识别准确率达到 95.62%。张善文等<sup>[13]</sup>在 VGG-16<sup>[14]</sup>模型的基础上构建一种可形变 VGG-16 模型, 实现对不规则田间害虫图像的特征提取, 在实际数据集中的检测准确率达到 91.14%。Gao 等<sup>[15]</sup>提出一种 DECA\_ResNet 网络模型, 对 3 种不同的农作物病害数据集进行分类, 提高病害图像识别准确率。

尽管上述研究表明卷积神经网络应用在农作物病害检测的可行性, 但是仍然面临一些问题。首先多数研究使用的数据集农作物病害种类较为单一, 只针对一种或几种农作物病害进行识别检测, 缺少在多类别大规模农作物病害数据集上的应用。其次采集的同一类别农作物图像仅有健康和患病两类, 并未关注农作物的病害程度,

收稿日期: 2022-03-28 修订日期: 2022-05-16

基金项目: 国家自然科学基金项目(61903207); 山东省重点研发计划项目(重大科技创新工程)(2019JZZY010731, 2020CXGC010901); 齐鲁工业大学国际合作研究项目(QLUTGJHZ2018020)

作者简介: 王泽钧, 研究方向为图像识别、深度学习。

Email: 823031251@qq.com

\*通信作者: 曹茂永, 教授, 博士生导师, 研究方向为数字图像处理, 声光电检测。Email: cmy@qlu.edu.cn

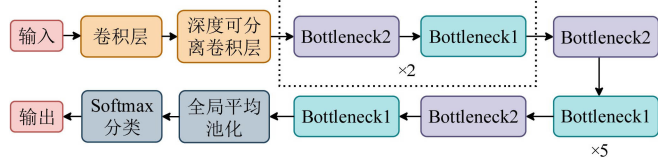
难以为智能施药提供可靠参考。并且大型神经网络会消耗大量的计算资源和存储资源，通常只能在高算力的GPU上运行，不利于网络模型的部署。

针对农作物病害图像种类多且颜色纹理相近、同种农作物不同病害差异较小，以及现有网络模型体积大等问题，本文提出一种基于注意力机制和多尺度轻量型网络模型（Lightweight Multi-scale Attention Convolutional Neural Networks, LMA-CNNs）。LMA-CNNs的创新点如下：1）添加注意力机制对通道特征进行权重分配，增强有用特征信息并减弱无用特征信息的干扰；2）设计多尺度融合机制提取病害特征，解决病害特征信息不丰富的问题，以期更好地平衡模型复杂程度和识别准确率，为移动端农作物病害检测提供参考。

## 1 算法设计

以搭建高精度、低消耗的农作物病害网络为目标，本文设计的 LMA-CNNs 模型如图 1 所示。图中，Bottleneck1 为轻量残差注意力模块，Bottleneck2 为多尺度特征融合模块。两个模块均采用深度可分离卷积代替标准卷积，详细的网络结构参数如表 1 所示。

LMA-CNNs 是在轻量化模型 MobileNet<sup>[16]</sup>的基础上进行改进，改进点包括：1）提出基于注意力机制的轻量残差模块 Bottleneck1，注意力机制增强网络对有用特征的提取，残差连接解决因模型深度增加所导致的网络退化问题；2）提出针对农作物病害性状的多尺度特征融合模块 Bottleneck2，提高网络对不同尺度病斑的识别能力；3）使用 Leaky ReLU 激活函数代替 ReLU 激活函数，增加对图像负值特征的提取，提高模型识别准确率。



注：Bottleneck1 为轻量残差注意力模块；Bottleneck2 为多尺度特征融合模块。  
Note: Bottleneck1 is lightweight residual attention module; Bottleneck2 is multi-scale feature fusion module.

图 1 LMA-CNNs 模型结构

Fig.1 Model structure of LMA-CNNs (Lightweight Multi-scale Attention Convolutional Neural Networks)

表 1 LMA-CNNs 网络参数表

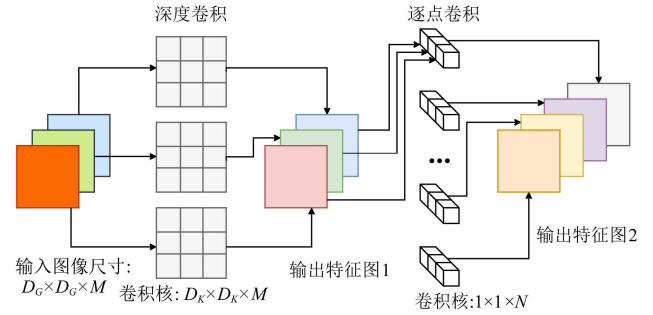
Table 1 LMA-CNNs network parameters table

层序号 Layer No.	结构 Structure	输入尺寸 Input size	重复次数 Repeat times	步长 Stride
Conv1	卷积	224 <sup>2</sup> ×3	1	2
Conv2	深度可分离卷积	112 <sup>2</sup> ×16	1	1
Conv3	Bottleneck2+Bottleneck1	112 <sup>2</sup> ×32	2	2+1
Conv4	Bottleneck2	28 <sup>2</sup> ×128	1	2
Conv5	Bottleneck1	14 <sup>2</sup> ×256	5	1
Conv6	Bottleneck2	14 <sup>2</sup> ×256	1	2
Conv7	Bottleneck1	7 <sup>2</sup> ×512	1	1
Average pooling	全局平均池化	7 <sup>2</sup> ×512	1	-
Softmax	全连接层	1 <sup>2</sup> ×59	-	-

### 1.1 深度可分离卷积

深度可分离卷积<sup>[16]</sup>可在移动端和嵌入式设备中使

用。与标准卷积相比，深度可分离卷积采用分解的思想，将标准卷积分解为深度卷积和逐点卷积，深度卷积用于滤波，逐点卷积用于组合。分解过程如图 2 所示。首先深度卷积利用卷积核对输入图像的每一个通道进行卷积，其次利用 1×1 大小的逐点卷积对输出特征图 1 进行卷积运算，进而得到输出特征图 2。这种运算方式相比于传统的标准卷积可以有效减少模型的参数和计算量。



注： $D_K \times D_K$  表示卷积核的大小， $M$  和  $N$  分别表示输入输出通道数。

Note:  $D_K \times D_K$  represent size of convolution kernel;  $M$  and  $N$  represent input and output channels, respectively.

图 2 深度可分离卷积分解过程

Fig.2 Depthwise separable convolution decomposition process

在标准卷积中输入通道数等于每个卷积核的通道数，然而在深度卷积中每个卷积核通道数都为 1，为保证输出通道数为  $N$ ，逐点卷积采用  $N$  个卷积核大小为 1×1 的标准卷积。假设图像的尺寸为  $D_G \times D_G$ ，深度卷积的计算量为  $D_K \cdot D_K \cdot M \cdot D_G \cdot D_G$ ，逐点卷积的计算量为  $M \cdot N \cdot D_G \cdot D_G$ ，传统标准卷积的计算量为  $D_K \cdot D_K \cdot M \cdot N \cdot D_G \cdot D_G$ ，深度可分离卷积与传统标准卷积的计算量比值为

$$\frac{\text{深度可分离卷积}}{\text{传统标准卷积}} = \frac{1}{N} + \frac{1}{D_K \cdot D_K} \quad (1)$$

在提取图像特征时，通常卷积核的尺寸设置为 3×3。由于通道数  $N$  比较大，故深度可分离卷积的参数量和计算量均约为传统标准卷积的 1/9。

### 1.2 基于注意力机制的轻量化残差模块

农作物病害图像分类属于细粒度图像分类。由于不同类别的农作物病害图像纹理与颜色非常接近，提取到的图像全局特征存在一些特征冗余。由文献[17]可知，在 59 类公开农作物病害数据集上，同种类别农作物病害的病害程度差异较小，容易造成网络模型的错判，影响模型的识别准确率。因此在模型中引入注意力机制<sup>[18-19]</sup>，其核心思想是通过网络的前向传递和反向传播来学习各个通道的特征权重，通过不同的特征权重区分特征图通道的重要程度，得到网络模型需要重点关注的特征通道，并抑制作用较小的特征通道，使模型得到更好的训练效果。本研究所设计的轻量残差注意力模块如图 3 所示。

Bottleneck1 中使用深度可分离卷积可以减少模型训练过程的参数量，嵌入通道空间注意力机制在特征提取中更加侧重病害叶片细微的颜色及纹理变化，引入残差结构缓解因网络深度的增加而出现的梯度爆炸现象或梯度弥散现象。假设输入特征图为  $I$ ，图像特征尺寸为  $C \times H \times W$ 。输入特征图  $I$  首先经过 1 个深度可分离卷积，

得到特征图  $F$ , 如式 (2) 所示。

$$F = PW^{1 \times 1}[DW^{3 \times 3}(I)] \quad (2)$$

式中  $PW^{1 \times 1}$  表示卷积核大小为  $1 \times 1$  的逐点卷积;  $DW^{3 \times 3}$  表示卷积核大小为  $3 \times 3$  的深度卷积。

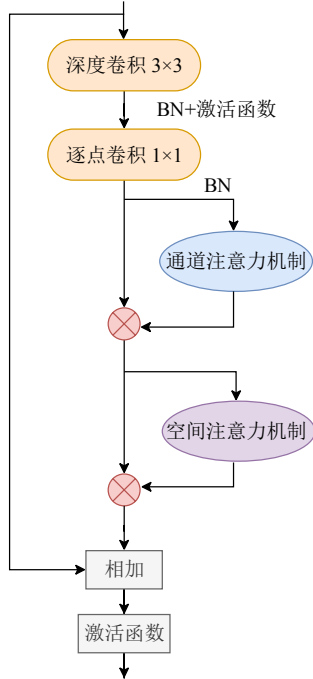


图3 轻量残差注意力模块 (Bottleneck1)

Fig.3 Lightweight residual attention module (Bottleneck1)

特征图  $F$  的不同通道特征之间缺少关联, 会降低模型的特征提取能力。通道注意力机制 (图 4a 所示) 通过最大池化操作和平均池化操作获取不同通道在相同空间位置上的特征信息, 可以更好地融合通道特征, 如式 (3) 所示。

$$H_c(F) = s[\text{Conv}(\text{MaxPool}(F)) + \text{Conv}(\text{AvgPool}(F))] \\ = s[W_2(W_1(F_{\max})) + W_2(W_1(F_{\text{avg}}))] \quad (3)$$

式中  $H_c(F)$  表示特征图  $F$  经过通道注意力模块后的特征映射结果;  $s$  为 sigmoid 激活函数;  $\text{Conv}$  表示卷积操作;  $W_1$  和  $W_2$  表示 2 个卷积层的权重;  $F_{\max}$  和  $F_{\text{avg}}$  分别表示基于通道进行最大池化操作和平均池化操作后的结果。

空间注意力机制 (图 4b 所示) 是对通道注意力机制的补充, 空间注意力主要聚焦于“哪里”是输入特征图最具信息量的部分, 池化操作是沿着通道维度进行的, 即每次池化时比较不同通道之间的数值, 如式 (4) 所示。

$$H_s(F) = s[\text{Conv}(\text{cat}(F_{\max}, F_{\text{avg}}))] \quad (4)$$

式中  $H_s(F)$  表示特征图  $F$  经过空间注意力模块后的特征映射结果;  $\text{cat}$  表示特征融合; 特征图  $F$  通过通道注意力机制和空间注意力机制的过程如下所示。

$$F_1 = H_c(F) \otimes F \quad (5)$$

$$F_2 = H_s(F_1) \otimes F_1 \quad (6)$$

式中  $F_1$  表示特征图  $F$  通过通道注意力机制的结果;  $F_2$  表示特征图  $F_1$  通过空间注意力机制的结果;  $\otimes$  表示特征图矩阵之间的乘法。

在运算过程中, 输入特征图通过通道注意力映射得到  $C \times 1 \times 1$  的维度矩阵, 该维度矩阵的值代表着不同通道

之间的重要性。 $C \times 1 \times 1$  的维度矩阵与原特征图  $F$  相乘得到  $C \times H \times W$  的维度矩阵。同理, 输入特征图通过空间注意力映射得到  $1 \times H \times W$  的维度矩阵, 该维度矩阵的值代表着特征图在空间域上的重要程度。 $1 \times H \times W$  的维度矩阵原特征图  $F$  相乘得到  $C \times H \times W$  的维度矩阵。最后输出特征图  $O$  是由特征图  $F_2$  与原特征图  $F$  进行相加得到的最终输出, 如式 (7) 所示。

$$O = \sigma(I + F_2) \quad (7)$$

式中  $\sigma$  表示 Leaky ReLU 激活函数; 输出特征图  $O$  的特征维度为  $C \times H \times W$ , 和输入特征图  $I$  的维度相同。

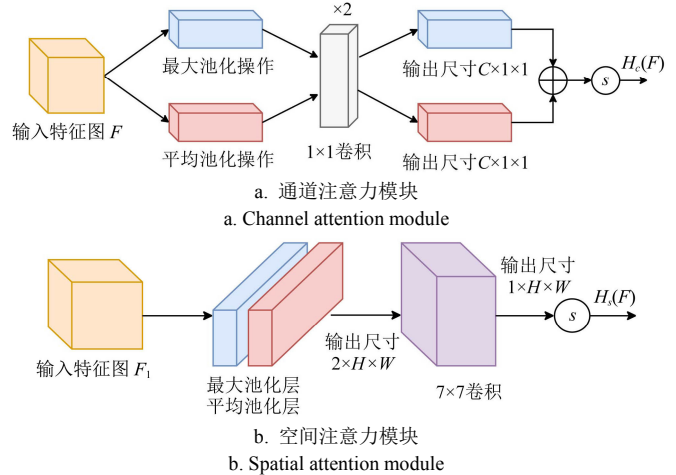


Fig.4 Attention mechanism structure diagram

### 1.3 多尺度特征融合模块

农作物病害性状具有差异性和多样性, 例如不同农作物的叶片轮廓、叶片纹理不同, 同种农作物不同类别病害叶片的性状差异较小, 极易引起混淆, 导致网络模型发生错判。固定尺寸的卷积核在感知农作物不同病害特征时效果不佳, 针对此问题引入多尺度特征融合模块, 结构如图 5 所示。

Bottleneck2 使用多个尺度的深度可分离卷积代替普通神经网络中大尺寸卷积核和池化层, 减少计算量。本研究使用 3 条并行分支进行通道升维和特征融合, 通过设置  $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$  尺寸大小的卷积核组合的形式进行图像特征提取, 改善网络模型对不同病害特征的提取。考虑到拼接操作会增加输出特征图的通道数, 导致模型参数量大幅增加, 本研究使用相加操作代替拼接操作, 最终使得输出特征图通道数为输入特征图通道数的 2 倍。

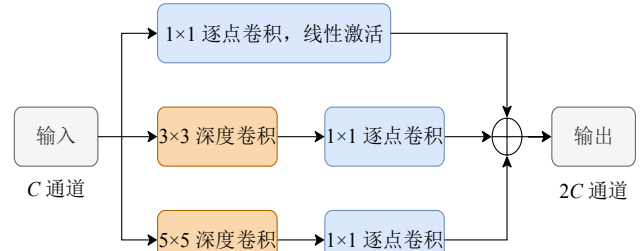


图5 多尺度特征融合模块 (Bottleneck2)

Fig.5 Multi-scale feature fusion module (Bottleneck2)

### 1.4 引入 Leaky ReLU 函数作为激活函数

在神经网络中, 激活函数的非线性功能是神经网络



进行学习的关键。ReLU 函数<sup>[20-21]</sup>可以有效解决神经网络传播过程中梯度消失的问题，成为激活函数的主流选择。然而 ReLU 函数在前向传播过程中，其负半轴的输出恒为 0，对应神经元保持非激活状态，在反向传播中无法更新梯度，导致部分神经元坏死。针对以上 ReLU 函数的缺陷，Leaky ReLU 函数<sup>[22-23]</sup>在负半轴有一个较小斜率，使得负半轴的输出不恒为 0，数学表达式如式（8）所示。

$$y_j = \begin{cases} x_j & x_j \geq 0 \\ \frac{x_j}{a_j} & x_j < 0 \end{cases} \quad (8)$$

式中  $x_j$  第  $j$  层的输出； $y_j$  第  $j$  层经过该激活函数做非线性变换后的输出； $a_j$  代表第  $j$  层的固定参数。

Leaky ReLU 函数不仅解决 ReLU 函数存在的负半轴神经元不学习的问题，还对病害图像的负值特征进行提取，得到更加丰富的特征信息。因此本研究采用 Leaky ReLU 函数作为激活函数。

## 2 试验与结果

### 2.1 试验数据集及试验环境

本研究所用的数据集为全球 AI 挑战赛提供的农作物病害图像。该数据集按照“物种-病害-程度”来划分，总共包含 61 个分类，27 种病害，10 个物种，按照分类排序依次为苹果、樱桃、玉米、葡萄、柑橘、桃子、辣椒、马铃薯、草莓、番茄。由于第 44 类和 45 类在训练集中仅有 1 张病害图像，测试集中分别有 1 张和 0 张病害图像，数据严重缺失，因此在试验中去除这两类数据图像，最终有 59 类共计 36 255 张病害图像。将数据集图像按照训练集：测试集按 7：1 划分，得到 31 716 张训练集图片，4 539 张测试集图片。

试验使用的深度学习框架为 PyTorch。计算机环境为 Linux 系统，CPU 型号为 Intel (R) Xeon Gold 5118 CPU@2.30GHz；GPU 型号为 GeForce RTX 2080 Ti 11GB。试验设置每次取出样本数量为 64，共取 496 次，进行 100 次迭代，初始学习率为 0.001，采用动态更新学习率的方法：每迭代 10 次，模型学习率缩小为原来的 1/5，使用 Adam 优化器。

### 2.2 分类模型性能对比

为验证 LMA-CNNs 模型的有效性，将该模型与常用的图像分类网络 ResNet34<sup>[12]</sup>及其改进版 ResNeXt<sup>[24]</sup>，轻量级网络 ShuffleNetV2<sup>[25]</sup>和 MobileNetV3<sup>[26]</sup>，及最新提出的 ViT (Vision Transformer)<sup>[27]</sup>进行对比试验，上述网络的代码均来自官方 GitHub 库。在相同的训练条件下，每训练一个 epoch，测试集会对模型进行一次分类精度验证，输出并记录该模型在这一轮中的分类准确率。LMA-CNNs 模型在引入注意力机制和多尺度融合后，模型参数增长较少，模型识别准确率有较大地提升。不同网络的分类准确率、参数数量和模型规模如表 2 所示。为方便对比分析，绘制出不同网络模型在农作物病害测试集上的准确率变化曲线，如图 6 所示。

由表 2 可知，LMA-CNNs 的识别准确率高于其他模

型，达到 88.08%。ResNeXt 模型识别准确率仅次于本研究模型，但因其使用大量卷积层进行分组堆叠，参数量过大，且该模型的模型规模达到 88.49 MB，不利于在移动端的部署。轻量级卷积神经网络 MobileNetV3 的参数数量和模型规模不仅略多于本研究的 LMA-CNNs 模型，因其使用倒残差结构搭建网络，相比于本研究残差块的改进特征表达有限，使得准确率相比于 LMA-CNNs 模型低 1.81 个百分点。由图 6 可以直观看出不同网络模型的准确率都呈现上升趋势，网络模型在不断循环迭代中提取农作物病害特征，提高分类能力。网络模型在 40 个 epoch 后达到平稳，LMA-CNNs 模型准确率上升趋势较快，在 10 个 epoch 后准确率高于其他模型。最新的 ViT 模型由于使用迁移学习，因此在前 5 个 epoch 中，准确率明显高于其他模型，但由于 ViT 模型使用类似于循环神经网络 (Recurrent Neural Network, RNN) 的编码器结构，在 10 个 epoch 后识别准确率不如其他模型。虽然 ViT 模型在 ImageNet 图像分类任务中取得不错的成绩，然而在农作物病害分类（细粒度分类）上明显不如其他卷积神经网络模型。此外，ViT 模型的参数量和模型规模都是最大的，不利于在移动端的应用。

表 2 不同网络模型性能对比

Table 2 Performance comparison of different network models

模型 Model	准确率 Accuracy/%	参数量 Parameters	模型规模 Model weights/MB
ResNet34	86.96	$2.13 \times 10^7$	81.40
ResNeXt	87.16	$2.31 \times 10^7$	88.49
ShuffleNetV2	86.89	$0.14 \times 10^7$	5.18
MobileNetV3	86.27	$0.16 \times 10^7$	6.15
ViT	85.17	$8.64 \times 10^7$	327.52
LMA-CNNs	88.08	$0.14 \times 10^7$	5.37

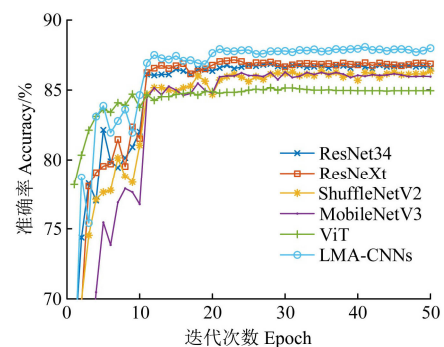


图 6 不同网络模型准确率结果对比

Fig.6 Comparison of accuracy results of different network models

综上所述，本研究构建的 LMA-CNNs 模型分类效果更好，模型参数量更少。主要原因有：1) 模型主体结构采用深度可分离卷积，在保证准确率没有大幅降低的情况下大量减少网络参数；2) 设计的残差注意力模块增强重要特征信息的权重，减少与分类特征信息无关的权重，减弱噪声等干扰因素对网络模型性能的影响。设计的多尺度融合模块使用不同尺寸的卷积核提取病害图像特征，改善网络模型提取不同大小病斑特征的能力，从而获得更加全面的特征信息，提升模型对不同程度的病害图像的分类准确率；3) 引入 Leaky ReLU 激活函数获取图像负值特征信息，提高模型的分类能力。

### 2.3 消融试验

为证明 LMA-CNNs 所设计的模块对模型性能的提升都有贡献,进行消融试验。在消融试验中,以模型在测试集的准确率、参数量、模型规模作为指标。为证明深度可分离卷积作为模型的主体结构可以减少参数量,设计对比试验,将模型中的深度可分离卷积全部替换为标准卷积;为证明轻量残差注意力模块(Bottleneck1)和多尺度特征融合模块(Bottleneck2)能够提升识别准确率,将两个模块分别替换为等通道数的深度可分离卷积;为

证明 Leaky ReLU 函数对模型的贡献,将 Leaky ReLU 函数替换成 ReLU 函数。试验结果如表 3 所示。

表 3 的试验结果表明:1)使用深度可分离卷积相较于标准卷积可以大幅减少参数量;2)在增加可接受范围内参数量的前提下,使用 Bottleneck1 模块、Bottleneck2 模块和 Leaky ReLU 函数的模型性能优于单独使用其中一种的模型性能;3) LMA-CNNs 模型的 3 种结构能够有效地融合,农作物病害识别准确率达到 88.08%,参数量仅为  $0.14 \times 10^7$ 。

表 3 消融试验结果  
Table 3 Ablation experiment results

卷积的选择 Choice of convolution	Bottleneck1	Bottleneck2	Leaky ReLU 函数 Leaky ReLU function	准确率 Accuracy/%	参数量 Parameters	模型规模 Model weights/MB
标准卷积 Standard convolutions	-	-	-	85.39	$0.71 \times 10^7$	27.20
	√	-	-	87.24	$0.73 \times 10^7$	27.80
	-	√	-	85.94	$0.96 \times 10^7$	36.89
	-	-	√	85.50	$0.71 \times 10^7$	27.20
深度可分离卷积 Depthwise separable convolutions	-	-	-	86.38	$0.08 \times 10^7$	3.34
	√	-	-	87.09	$0.10 \times 10^7$	3.94
	-	√	-	86.83	$0.12 \times 10^7$	4.77
	-	-	√	86.69	$0.08 \times 10^7$	3.34
	√	√	√	88.08	$0.14 \times 10^7$	5.37

### 2.4 不同研究者分类方法对比

为进一步验证 LMA-CNNs 模型的有效性,本研究在使用相同数据集的前提下与不同研究者提出的方法(IncRes-V2、改进 ResNet18、DECA 和 I-CBAM)进行比较。IncRes-V2<sup>[17]</sup>方法结合残差结构和分块卷积的优点,以提升图像分类准确率;改进 ResNet18<sup>[28]</sup>方法通过引入焦点损失函数使模型更加专注于难以分类的样本;DECA 模型<sup>[15]</sup>使用双分支一维卷积操作来过滤有效的特征信息,并引入自适应卷积核参数和自适应参数  $\alpha$  参与模型的反向训练,使模型能够独立选择有效特征并建立通道之间的依赖关系;I-CBAM<sup>[29]</sup>通过空间注意力和通道注意力的并行连接,解决串行连接两种注意力产生干扰的问题。比较结果如表 4 所示。

表 4 不同研究方法比较  
Table 4 Comparison of different research methods

方法 Method	年份 Year	图像数量 Image number	初始学习率 Initial learning rate	优化器 Optimizer	准确率 Accuracy/%	模型规模 Model weights/MB
IncRes-V2*	2019	36 258	0.01	RMSprop	86.10	237.2
改进 ResNet18 <sup>[28]</sup>	2020	36 255	0.001	Adam	86.83	-
DECA <sup>[15]</sup>	2021	35 861	0.05	SGD	86.35	185.4
I-CBAM <sup>[29]</sup>	2021	36 258	0.000 1	Adam	86.39	28.3
LMA-CNNs	2022	36 255	0.001	Adam	88.08	5.37

注: \*代表使用迁移学习。

Note: \* represents the use of transfer learning.

农作物病害分类任务的难点是如何提高模型识别多类病害的准确率并降低网络参数,可以看出 LMA-CNNs 模型在没有使用迁移学习的情况下具有最高准确率、最小的模型规模,很好地平衡模型复杂程度和模型识别精度。可能的原因是 LMA-CNNs 模型使用深度可分离卷积减少参数量,引入注意力机制和多尺度融合结构来弥补

深度可分离卷积分类能力不足的缺陷并解决网络模型退化的问题,使用 Leaky ReLU 函数相比于 ReLU 函数提取到图像负值特征,达到高精度与低消耗兼容的效果。

### 2.5 不同程度农作物病害识别结果

为分析 LMA-CNNs 模型对不同程度农作物病害的识别效果,选取未参与神经网络训练过程的番茄类别病害图片进行评估,生成 3 种模型的混淆矩阵如图 7 所示。其中混淆矩阵对角线代表被正确分类的数量,行代表真实标签,列代表预测标签,其中对角线上的值越大说明分类效果越好。在本试验所使用的数据集中,番茄类有 12 种病害类别,且部分类别“一般”图像和类别“严重”图像肉眼差别不大,容易发生错误判断,故采用番茄病害数据集进行结果展示。从图 7 可以看出 LMA-CNNs 模型在番茄病害程度判别上取得不错的结果,性能表现优于 MobileNetV3 和 ViT。

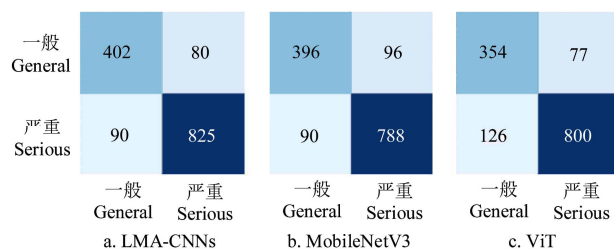


图 7 番茄病害程度识别结果的混淆矩阵

Fig.7 Confusion matrix of tomato disease degree recognition results

通过混淆矩阵计算出 59 类各类图像的精确率 (Precision)、召回率 (Recall) 和特异度 (Specificity) 作为 LMA-CNNs 模型性能的评价指标。指标的计算公式如下所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (9)$$



$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (10)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (11)$$

式中 TP 表示正样本被正确预测为正样本的数量；FP 表示负样本被错误预测为正样本的数量；FN 表示正样本被错误预测为负样本的数量；TN 表示负样本被正确预测为负样本的数量。精确率、灵敏度和特异度的值越接近 1 则说明模型的分类性能好。结果如表 5 所示，可以看出表中各类别的精确率、灵敏度和特异度均在 70%以上，部分试验结果达到 100%，表明 LMA-CNNs 模型具有较好的识别精度。

表 5 LMA-CNNs 模型部分数据分类结果

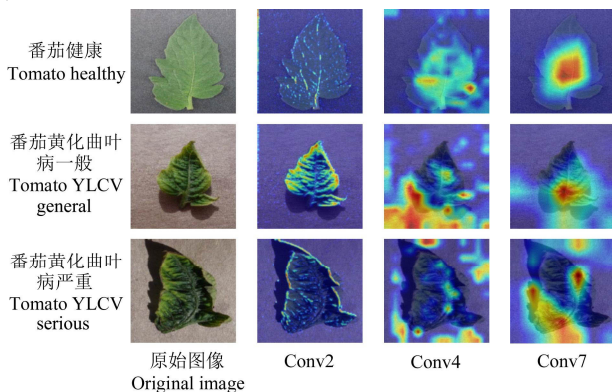
Table 5 LMA-CNNs model partial data classification results %

类别 Category	等级 Grade	精确率 Precision	召回率 Recall	特异度 Specificity
苹果雪松锈病 Cedar apple rust	一般	90.0	90.0	100
	严重	83.3	83.3	100
樱桃白粉病 Cherry powdery mildew	一般	80	100	99.9
	严重	100	94.4	100
葡萄轮斑病 Grape black measles fungus	一般	85.5	87.8	99.8
	严重	87.5	83.1	99.8
番茄晚疫病 Tomato late blight	一般	90.0	71.1	99.9
	严重	91.6	96.2	99.7
番茄斑枯病 Tomato septoria leaf spot fungus	一般	88.3	88.3	99.8
	严重	93.8	92.2	99.8

## 2.6 特征可视化结果

卷积神经网络的特征可视化有助于理解分类模型的训练过程。本研究采用 Grad-CAM<sup>[30]</sup>的方法，通过对模型中间层所提取到的特征进行可视化操作，了解 LMA-CNNs 模型在不同层之间学习的特征并进行对比与分析。

图 8 为模型训练完成后，中间层的输出特征图可视化结果。可以看出 LMA-CNNs 模型的学习具有层次性。通过 Conv2 的可视化结果可知，模型在神经网络浅层学习到图像边缘、颜色、叶脉纹理等特征信息。由 Conv4 和 Conv7 的可视化结果可知，随着网络层数的增加，模型更关注病害区域，提取到的特征信息多数来自叶片表面。

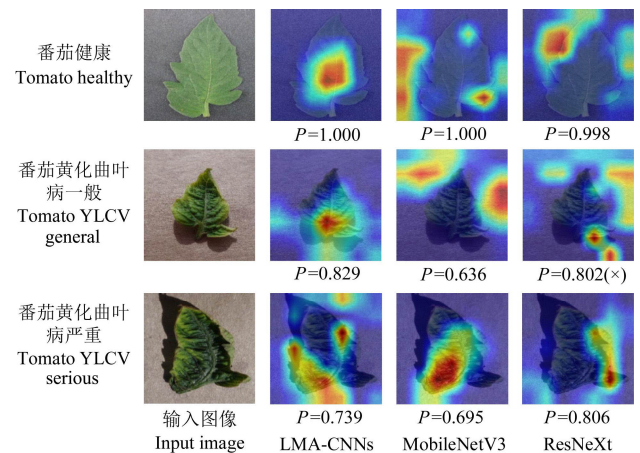


注(Note): YLCV : Yellow Leaf Curl Virus.

图 8 网络中间层特征可视化结果

Fig.8 Visualization results of network middle layer features

同时，提取 ResNeXt 和 MobileNetV3 网络最后卷积层的输出特征进行可视化处理并与 LMA-CNNs 模型进行比较，结果如图 9 所示。可以看出，ResNeXt 模型和 MobileNetV3 模型在提取特征时容易受到背景的影响，聚焦于叶片边缘和叶片阴影处，并没有特别关注叶片病害区域。因此，ResNeXt 模型发生错判，MobileNetV3 模型识别正确，然而预测概率不高，这也是 2.2 节 ResNeXt 和 MobileNetV3 相较于 LMA-CNNs 模型识别准确率低的合理解释。LMA-CNNs 模型由于添加注意力机制和多尺度融合机制，相比于其他模型可以更准确地定位到农作物病害区域，更精准地区分输入图像有无病害及病害程度，对于正确分类的判断效果也优于其他模型。



注：P 为预测概率；×表示发生误判。

Note: P represents prediction probability; × represents misjudge.

图 9 各网络特征提取可视化对比结果

Fig.9 Visual comparison results of feature extraction of each network

## 3 结 论

1) 针对传统卷积神经网络模型对多类别病害图像分类效果差、收敛慢、处理能力不足的缺陷，本文设计出 LMA-CNNs 模型。模型的主体结构采用深度可分离卷积，减少卷积核的个数，大幅缩减构建网络的参数量，减少储存空间。模型添加多尺度融合机制和注意力机制，使模型更关注对病害分类有用的信息并降低无用信息的影响。

2) 本研究所设计的 LMA-CNNs 模型最终对 59 类农作物病害的识别准确率达到 88.08%，并且模型占用内存仅为 5.37 MB，分类效果优于其他模型。

3) 通过对比经典神经网络模型和不同研究者的方法，LMA-CNNs 模型识别准确率更高，模型结构更加精简。利用混淆矩阵和 Grad-CAM 可视化证明 LMA-CNNs 模型对农作物病害图像特征提取的有效性。

综上，本研究提出的 LMA-CNNs 模型能够出色地识别 59 类农作物病害，为农业病害检测模型的部署提供新的方法和思路。

## [参 考 文 献]

- [1] Barbedo J G A. A review on the main challenges in automatic plant disease identification based on visible range

- images[J]. *Biosystems Engineering*, 2016, 144: 52-60.
- [2] Rumpf T, Mahlein A K, Steiner U, et al. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance[J]. *Computers and Electronics in Agriculture*, 2010, 74(1): 91-99.
  - [3] 王建仑. 田间叶片图像分割与单幅三维重建的机器视觉算法研究[D]. 北京: 中国农业大学, 2013.  
Wang Jianlun. Study on Field Leaf Image Segmentation and 3D Reconstruction From a Single Image Machine Vision Algorithms[D]. Beijing: China Agricultural University, 2013. (in Chinese with English abstract)
  - [4] Qin F, Liu D, Sun B, et al. Identification of alfalfa leaf diseases using image recognition technology[J]. *PLoS One*, 2016, 11(12): e0168274.
  - [5] 杜文圣, 王春颖, 朱衍俊, 等. 采用改进 Mask R-CNN 算法定位鲜食葡萄疏花夹持点[J]. *农业工程学报*, 2022, 38(1): 169-177.  
Du Wensheng, Wang Chunying, Zhu Yanjun, et al. Fruit stem clamping points location for table grape thinning using improved mask R-CNN[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(1): 169-177. (in Chinese with English abstract)
  - [6] 甘雨, 郭庆文, 王春桃, 等. 基于改进 EfficientNet 模型的作物害虫识别[J]. *农业工程学报*, 2022, 38(1): 203-211.  
Gan Yu, Guo Qingwen, Wang Chuntao, et al. Recognizing crop pests using an improved EfficientNet model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(1): 203-211. (in Chinese with English abstract)
  - [7] 袁山, 汤浩, 郭亚. 基于改进 Mask R-CNN 模型的植物叶片分割方法[J]. *农业工程学报*, 2022, 38(1): 212-220.  
Yuan Shan, Tang Hao, Guo Ya. Segmentation method for plant leaves using an improved Mask R-CNN model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(1): 212-220. (in Chinese with English abstract)
  - [8] Nagaraju M, Chawla P. Systematic review of deep learning techniques in plant disease detection[J]. *International Journal of System Assurance Engineering and Management*, 2020, 11(3): 547-560.
  - [9] 翟肇裕, 曹益飞, 徐焕良, 等. 农作物病虫害识别关键技术研究综述[J]. *农业机械学报*, 2021, 52(7): 1-18.  
Zhai Zhaoyu, Cao Yifei, Xu Huanliang, et al. Review of key techniques for crop disease and pest detection[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(7): 1-18. (in Chinese with English abstract)
  - [10] 贾少鹏, 高红菊, 杭潇. 基于深度学习的农作物病虫害图像识别技术研究进展[J]. *农业机械学报*, 2019, 50(增刊): 313-317.  
Jia Shaopeng, Gao Hongju, Hang Xiao. Research progress on image recognition technology of crop pests and diseases based on deep learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(Supp): 313-317. (in Chinese with English abstract)
  - [11] 黄林生, 罗耀武, 杨小冬, 等. 基于注意力机制和多尺度残差网络的农作物病害识别[J]. *农业机械学报*, 2021, 52(10): 264-271.  
Huang Linsheng, Luo Yaowu, Yang Xiaodong, et al. Crop disease recognition based on attention mechanism and multi-scale residual network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(10): 264-271. (in Chinese with English abstract)
  - [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE Computer Society, 2016: 770-778.
  - [13] 张善文, 许新华, 齐国红, 等. 基于可形变 VGG-16 模型的田间作物害虫检测方法[J]. *农业工程学报*, 2021, 37(18): 188-194.  
Zhang Shanwen, Xu Xinhua, Qi Guohong, et al. Detecting the pest disease of field crops using deformable VGG-16 model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(18): 188-194. (in Chinese with English abstract)
  - [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *Computer Science*, 2014, 18(2): 1409-1556.
  - [15] Gao R, Wang R, Feng L, et al. Dual-branch, efficient, channel attention-based crop disease identification[J]. *Computers and Electronics in Agriculture*, 2021, 190: 106410.
  - [16] Howard A, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE, 2017: 1704-1712.
  - [17] 魏超, 范自柱, 张泓, 等. 基于深度学习的农作物病害检测[J]. *江苏大学学报(自然科学版)*, 2019, 40(2): 190-196.  
Wei Chao, Fan Zizhu, Zhang Hong, et al. Detection of crop diseases based on deep learning[J]. *Journal of Jiangsu University( Natural Science Edition)*, 2019, 40(2): 190-196. (in Chinese with English abstract)
  - [18] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
  - [19] Li Y, Liu Y, Cui W G, et al. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020, 28(4): 782-794.
  - [20] Shen Z, Yang H, Zhang S. Optimal approximation rate of ReLU networks in terms of width and depth[J]. *Journal de Mathématiques Pures et Appliquées*, 2022, 157: 101-135.
  - [21] Parisi L, Neagu D, Ma R, et al. Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19[J]. *Expert Systems with Applications*, 2022, 187: 115892.
  - [22] Liang X L, Xu J. Biased ReLU neural networks[J]. *Neurocomputing*, 2021, 423: 71-79.
  - [23] 高淑萍, 赵清源, 齐小刚, 等. 改进 MobileNet 的图像分类方法研究[J]. *智能系统学报*, 2021, 16(1): 11-20.  
Gao Shuping, Zhao Qingyuan, Qi Xiaogang, et al. Research on the improved image classification method of MobileNet[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1): 11-20. (in Chinese with English abstract)
  - [24] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE, 2017: 1492-1500.
  - [25] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient CNN architecture design[C]//

- Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 116-131.
- [26] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019: 1314-1324.
- [27] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//International Conference on Learning Representations. 2020.
- [28] 史冰莹, 李佳琦, 张磊, 等. 基于 CNN 的农作物病虫害图像识别模型[J]. 计算机系统应用, 2020, 29(6): 89-96. Shi Bingying, Li Jiaqi, Zhang Lei, et al. Recognition model of crop pests and diseases images based on CNN[J]. Computer Systems & Applications, 2020, 29(6): 89-96. (in Chinese with English abstract)
- [29] 王美华, 吴振鑫, 周祖光. 基于注意力改进 CBAM 的农作物病虫害细粒度识别研究[J]. 农业机械学报, 2021, 52(4): 239-247. Wang Meihua, Wu Zhenxin, Zhou Zuguang. Fine-grained identification research of crop pests and diseases based on improved CBAM via attention[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 239-247. (in Chinese with English abstract)
- [30] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

## Crop disease recognition using attention mechanism and multi-scale lightweight network

Wang Zejun<sup>1</sup>, Ma Fengying<sup>1</sup>, Zhang Yu<sup>2</sup>, Zhang Fangfang<sup>1</sup>, Ji Peng<sup>1</sup>, Cao Maoyong<sup>1\*</sup>

(1. Department of Electronics, Electrical and Control, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China;

2. Department of Mathematics and Artificial Intelligence, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China)

**Abstract:** In recent years, diseases and pests have caused a huge loss in agricultural production. Accurate identification of crop diseases and timely protection are important measures to ensure crop yield. Traditional methods of diagnosing agricultural diseases typically depend on the expertise and judgment of specialists. This approach is dependent on human subjective perception, which is prone to error and cannot ensure timeliness. The optimal time to cure agricultural diseases may be missed by traditional methods, resulting in financial losses. The neural networks and the development of deep learning have brought new technologies to the appraisal of agricultural diseases. However, certain large-scale neural networks cannot be implemented on mobile terminals to accomplish crop disease detection in realistic settings due to the low identification accuracy and a huge number of parameters. To address the problems of large size and low accuracy of traditional crop disease recognition models, we proposed a Lightweight Multi-scale Attention Convolutional Neural Networks (LMA-CNNs) to solve the above problems. First, in order to reduce the number of parameters and make the model lightweight, depthwise separable convolution was adopted as the main structure of the network; secondly, the residual attention module and multi-scale feature fusion module were designed on the basis of depthwise separable convolution; at the same time, the Leaky ReLU activation function was introduced to enhance the extraction of negative-valued features. The residual attention module enhanced the weight of useful feature information and weakened the weight of interference information such as noise by embedding channels and spatial attention mechanisms, and improved the recognition of important features by the network model. Residual connections could effectively prevent network degradation. The multi-scale feature fusion module used its convolution kernels of different scales to extract disease features of multiple scales, which improved the richness of features. The experimental results showed that the accuracy of the LMA-CNNs model on the test set of 59 types of disease images was 88.08%, and the number of parameters was only  $0.14 \times 10^7$ . Through comparative experiments, the LMA-CNNs model outperformed ResNet34, ResNeXt, ShuffleNetV2, MobileNetV3, and the more popular Vision Transformer recently. This study further verified the effectiveness of the LMA-CNNs model by comparing the network models designed by different researchers under the same dataset. Comparative experiments showed that the LMA-CNNs model reduced the number of model parameters on the premise of improving accuracy. Because of the problem of poor interpretability of the neural network model, this study used Grad-CAM to visualize the features extracted by the middle layer of the model and explained the model through the visualization results to obtain different feature information on different convolutional layers. As the number of layers increased, the LMA-CNNs model paid more attention to the diseased area. In summary, the LMA-CNNs model could extract more disease feature information, better balance the model complexity and model recognition accuracy, and provide a reference for mobile crop disease recognition. In the future, we will continue to optimize the algorithm, deploy the model to the mobile terminal to detect crop diseases in real-field scenarios, and improve detection accuracy and efficiency.

**Keywords:** crops; models; image identification; attention mechanism; multi-scale convolution