

彭思绘,汪宇玲. 基于 P-MobileViT 网络的小麦病害分类研究[J]. 江苏农业科学,2024,52(20):260-267.

doi:10.15889/j.issn.1002-1302.2024.20.031

# 基于 P-MobileViT 网络的小麦病害分类研究

彭思绘,汪宇玲

(东华理工大学信息工程学院,江西南昌 330013)

**摘要:**针对小麦病害图像分类方法的识别准确率不理想、模型参数量大等问题,提出一种基于 P-MobileViT 的小麦病害分类模型。首先对小麦图像进行健康和病害二分类,融合 Grabcut 算法、大律法对小麦病害图像的病斑区域进行分割;然后将病斑图像输入 P-MobileViT 分类模型,在其 block 的局部表征模块中引入深度卷积提取病斑图像的局部特征,在全局表征模块使用 PoolFormer 结构提取全局特征,以减少模型计算量和参数量;将输入特征图和全局特征叠加后与局部特征进行融合,从而强化模型对特征的分类能力。与经过迁移学习的轻量级深度学习模型 MobileViT、ShuffleNet v2、MobileNet v3、GhostNet、EfficientNet v2 在公开小麦病害数据集上进行试验对比,结果表明,P-MobileViT 模型的准确率达到 97.2%,比 MobileViT 模型高出了 2.0 百分点,同时参数量、推理时间分别减少了 23.1%、31.6%;与其中准确率较高的模型 MobileNet v3、GhostNet 相比,P-MobileViT 模型的准确率也分别提高 3.1、3.3 百分点,参数量分别减少 58.3%、61.5%。在小麦病害分类任务中,P-MobileViT 模型实现了识别精度的提升,有效减少了识别时间开销,且降低了模型复杂度。

**关键词:**小麦病害;病害分类;MobileViT;图像分割;PoolFormer

**中图分类号:**S126;TP391.41 **文献标志码:**A **文章编号:**1002-1302(2024)20-0260-07

小麦是全球产量最大的主要粮食作物之一,决定全球大部分国家的粮食安全。小麦病害不仅是限制其产量的主要因素之一<sup>[1]</sup>,还是阻碍优质、高效农业持续发展的重要原因;其中叶部病害等问题对小麦产量有显著不利影响<sup>[2-3]</sup>。在传统农业中,对小麦病害发生情况的评估主要依赖于日常的田间巡查和经验判断,容易导致问题发现不及时而延误最佳防治时机,而诊断范围不准也会导致过度喷洒化学药剂等问题<sup>[4]</sup>。目前基于深度学习和计算机视觉的方法被广泛应用于农作物病害检测<sup>[5]</sup>,以高效、智能的方式实现病害检测自动化,在农作物经济效益提升方面具有重大价值意义<sup>[6]</sup>。

近年来,基于深度学习的作物病害识别技术得到快速发展,在作物病害检测方面取得了许多成果<sup>[7-8]</sup>。在复杂背景下的分割领域中,顾博等针对

玉米的 3 种病害类型,结合 SLIC 算法、GrabCut 算法进行分割,得到较好的分割效果<sup>[9]</sup>。陈燕等提出基于 ResNet50、Unet 网络的分割模型,对小麦茎秆截面的分割精度有较大提升<sup>[10]</sup>。病斑区域被有效分割后,还需对其病害类型进行识别。冯晓等基于 MobileNet v2 网络构建小麦叶部病害识别模型,利用迁移学习方法得到较高的识别精度<sup>[11]</sup>。Nanni 等基于不同结构的 CNN 模型在 3 个病害数据集上进行训练,发现不同 Adam 的 CNN 组合模型表现最好,并且节约了训练时间和资源<sup>[12]</sup>。Khan 等提出基于深度迁移学习模型的病害识别系统,结果表明,这种模型在大数据集上的性能优于现有的分类算法<sup>[13]</sup>。Li 等在相同的数据集上比较学习速率、迁移学习、数据增强方法在不同组合下的分类精度,发现使用迁移学习模型的分类精度优于新学习的模型<sup>[14]</sup>。相较于传统农业生产中判断植物病害的方法,深度神经网络在病虫害识别中具有准确、及时、客观的优势<sup>[15-16]</sup>。对于区分度明显的分类任务来说,卷积神经网络的识别效果显著,但是对于农作物病害分类识别而言,其识别精度还不够理想,这与农作物病害种类多且区分度较低等有关<sup>[17-18]</sup>。卷积神经网络大多参数量较大,对硬件设备的需求较高,不利于模型在实际应用中的部署。

收稿日期:2023-11-24

基金项目:国家自然科学基金(编号:62066003);国家留学基金(编号:CSC202208360143);江西省研究生创新专项资金(编号:YC2022-s626)。

作者简介:彭思绘(1997—),女,安徽池州人,硕士研究生,主要研究方向为计算机视觉。E-mail:pengsh1002@163.com。

通信作者:汪宇玲,博士,教授,硕士生导师,主要从事模式识别与图像处理、计算机视觉研究。E-mail:wangyuling\_119@vip.163.com。

综上,本研究提出一种基于 P-Mobile ViT 的小麦病害分类模型。本研究模型针对小麦病害图像,将大律法融合到 GrabCut 算法中实现病斑区域的分割,从而提升病害识别精度;在 P-MobileViT block 中分别引入深度卷积、PoolFormer 结构,用于降低模型复杂度,增加融合局部特征,提升模型对小麦病害特征的感知能力。该模型在提升准确率的同时节省了算力资源,在小麦病害的智能诊断方面具有一定的参考价值。

## 1 材料与方法

### 1.1 数据获取

为了保证数据正确可靠,本研究小麦病害数据来源于 2 个部分:kaggle 网站(<https://www.kaggle.com>)中的 Wheat Leafs Diseases 数据集以及 CGIAR Computer Vision for Crop Disease 数据集。去除重复

数据和无效数据后,共获取数据图像 2 807 幅,分别为黑穗病、叶枯病、白粉病、条锈病、秆锈病、叶锈病、健康 7 种类型,将上述各类数据均按照 80% 训练集和 20% 测试集进行划分。详细数据信息见表 1,部分可视化样本如图 1 所示。

表 1 小麦病害图像数据

病害类型	训练集 (幅)	测试集 (幅)
黑穗病	278	70
叶枯病	279	70
白粉病	129	32
条锈病	166	42
秆锈病	229	57
叶锈病	572	143
健康	592	148

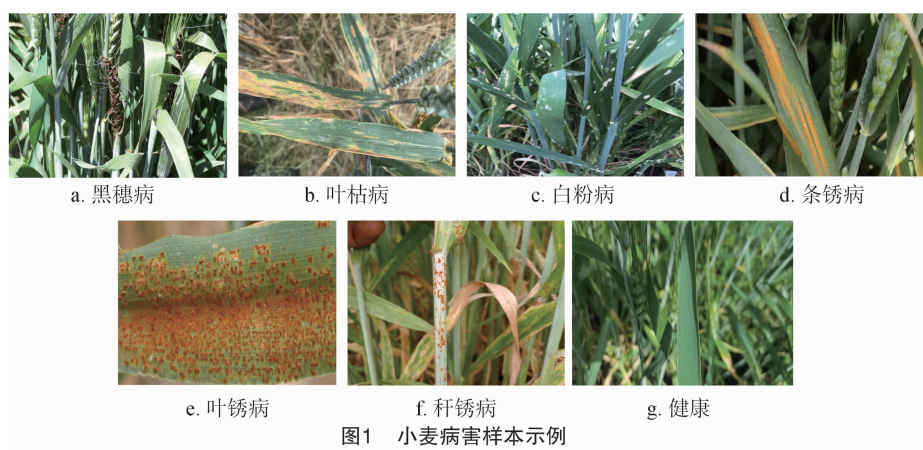


图1 小麦病害样本示例

### 1.2 MobileViT 模型

MobileViT 是一种适用于移动设备的轻量级通用视觉 Transformer 模型<sup>[19-20]</sup>,其主要构成包括 MobileNet v2 网络和 MobileViT block,如图 2 所示。其中 MobileViT block 结合了卷积神经网络(CNN)和 Vision Transformer(ViT)的优点<sup>[21]</sup>,通过使用  $N$  个 Transformer 模块提取数据的全局信息,同时保持 CNN 的空间归纳偏置,以适用于移动视觉任务的轻量级设备部署。相比于传统的 CNN 模型,MobileViT 模型具有更少的参数量和计算量,因此更加适合在移动设备上部署。在 ImageNet-1k 数据集上,MobileViT 实现了 78.4% 的 Top-1 精度,拥有约 600 万个参数;与类似的参数量模型相比,其精度比 MobileNet v3(基于 CNN)、DeiT(基于 ViT)分别提升 3.2、6.2 百分点<sup>[22-23]</sup>。

## 2 P-MobileViT 模型

MobileViT 是 ViT 的一个轻量级版本,其计算高效且参数量较少,因此成为移动设备上应用视觉任务的较优选择。然而,在小麦病害识别应用中,MobileViT 的性能仍然需要进一步提升。基于此,本研究提出 P-MobileViT 模型,针对 MobileViT block 在局部表征模块、全局表征模块、融合模块 3 个方面进行改进,得到 P-MobileViT block 结构(图 3)。

### 2.1 局部特征提取

MobileViT block 的局部表征模块包含 1 个  $3 \times 3$  卷积层和 1 个  $1 \times 1$  卷积层,用于提取局部特征。但  $3 \times 3$  卷积层通常需要大量的计算资源(图 4-a),不适用于轻量级模型。因此,P-MobileViT block 引入深度卷积(depthwise convolution)来提取局部特

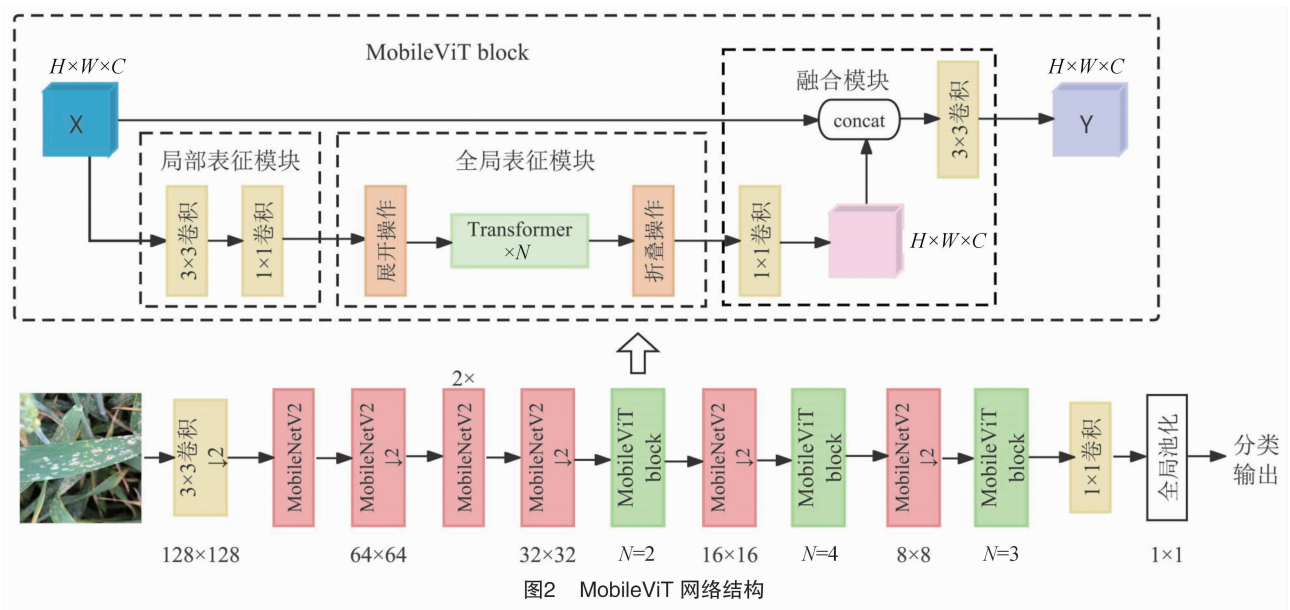


图2 MobileViT 网络结构

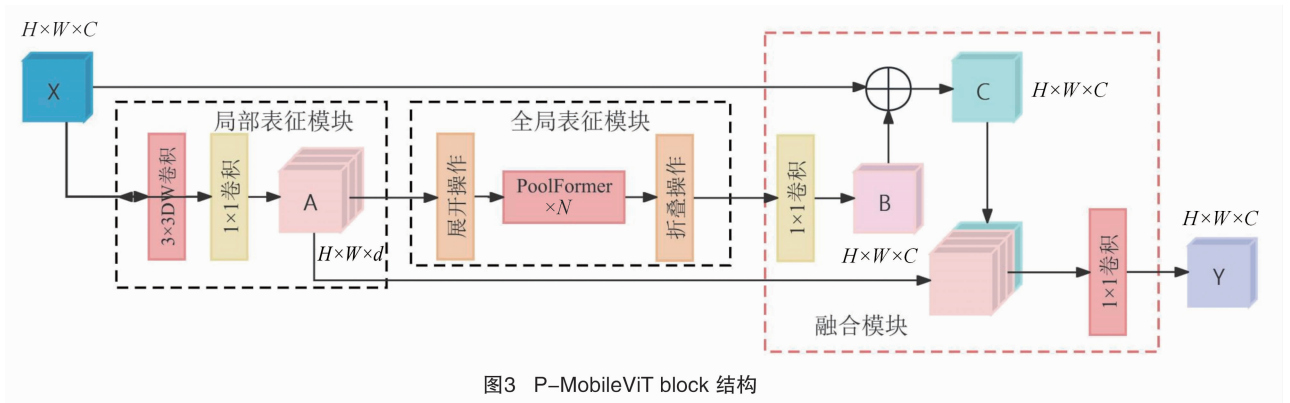


图3 P-MobileViT block 结构

征<sup>[24]</sup>,深度卷积将输入特征图根据通道分离开,每个通道仅与1个卷积核相对应,分别对该通道的特征图进行卷积操作(图4-b),使得该卷积在保证精度的同时,有效减少计算量;接着通过 $1\times 1$ 卷积层

将张量映射到高维空间d,以获取局部特征A,如图3局部表征模块所示。在本研究中,深度卷积设置为 $3\times 3$ 卷积核。

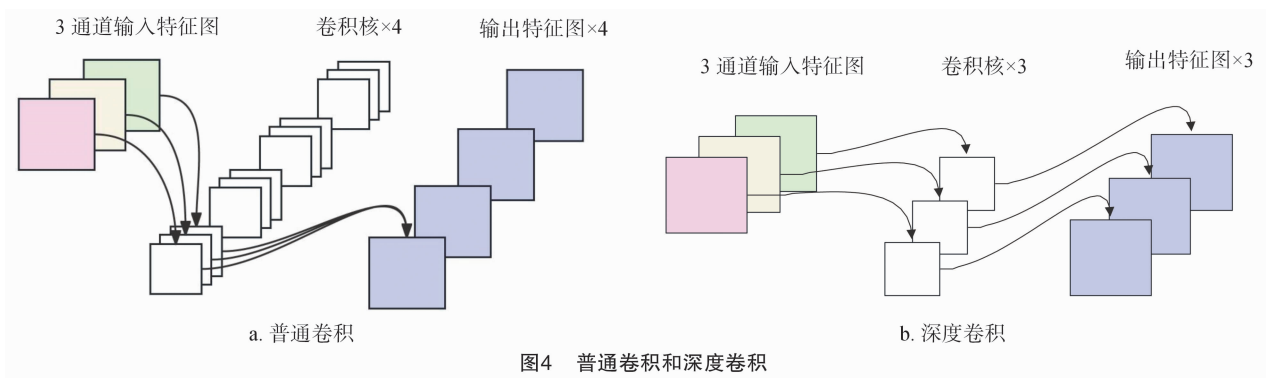


图4 普通卷积和深度卷积

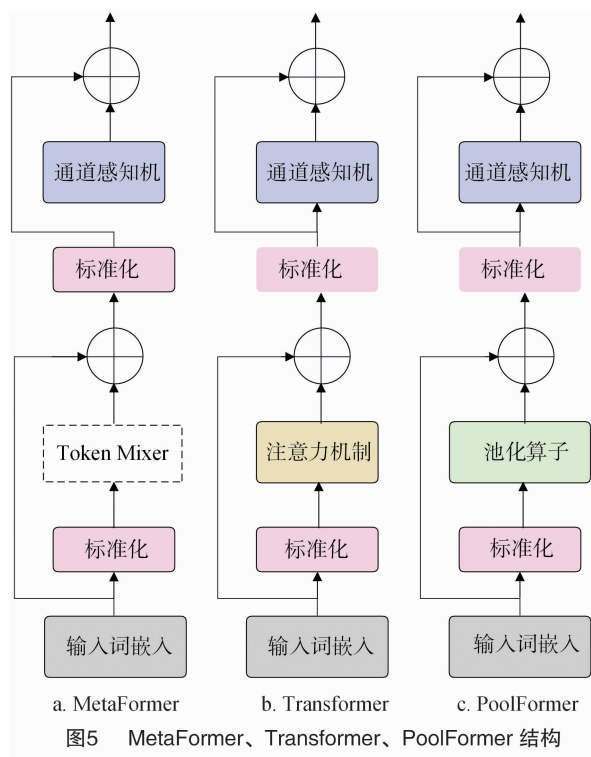
## 2.2 全局特征提取

获取全局特征可以捕捉到整个小麦病害图像的信息,帮助模型更好地理解病害类别,以提高识别精度。MobileViT block 的全局表征模块包含多个

transformers 模块,用于提取全局特征。transformers 模块中使用多头自注意力机制 (multi-head attention, MHA) 来计算每个位置的注意力分布,以掌握全局信息。然而, MHA 的计算复杂度较高,需

要大量的计算资源,且推理时间较长,不利于轻量级模型的部署。

为了解决该问题,P-MobileViT block 采用 PoolFormer 来提取全局特征,进一步减少计算资源的占用。PoolFormer 是一种用池化算子来作为 Token Mixer 的 MetaFormer 结构,而 MetaFormer 可以抽象成 Transformer 的通用结构<sup>[25]</sup>(图 5-a)。这一结构已被证实是实现 Transformer(图 5-b)高效性能的主要原因,且 Token Mixer 模块的改变,对该结构整体性能的影响不大。因此,通过使用无参数的池化算子替代原始的自注意力机制作为 Token Mixer(图 5-c),可以明显节约模型的算力资源,更有利于实现对移动设备的部署。



将局部特征 A 展开为多个不重叠的扁平 patch,然后输入 PoolFormer 模块编码全局信息,编码完成后,通过折叠操作还原大小,最后使用  $1 \times 1$  卷积层将张量映射到低维空间 C,得到全局特征 B(图 3)。

### 2.3 特征融合

MobileViT block 中仅将输入特征与全局特征融合,但对于一些区分度较低的小麦病虫害(如白粉病、蚜虫等),更需要局部特征来丰富语义信息。因此,P-MobileViT block 在融合模块上增加局部特征,以提高模型对图像全局和局部信息的感知能力,在处理小麦病害图像时可以更好地捕捉病害区

域的特征。具体操作如下:将输入特征图 X 和全局特征 B 进行叠加后,与局部特征 A 进行拼接,再使用  $1 \times 1$  卷积层进行特征融合,从而获取融合特征 Y,融合流程见图 3。相较于普通卷积,使用  $1 \times 1$  卷积层避免了模型缩放时参数和浮点运算数(FLOPs)的大幅增加。

## 3 试验与分析

### 3.1 试验环境及参数设置

硬件环境:试验配置中央处理器(CPU)为 Intel® Xeon® Platinum 8255C CPU @ 2.50 GHz,内存 43 GB;GPU 为 NVIDIA GeForce RTX 3090,显存 24 GB。软件环境:操作系统为 Ubuntu 22.04,深度学习框架采用 Pytorch 1.13.0,编程语言使用 Python 3.9,CUDA 版本为 11.7,在 Pycharm 2022.3 平台搭建模型。

为保证模型的训练能力和泛化能力,从每种小麦病害种类中随机抽出平均 80% 作为训练集,图片输入尺寸统一为 224 像素  $\times$  224 像素。优化器采用 Adam(Adaptive moment estimation)自动优化学习率,共选择 3 种初始学习率,分别为 0.001、0.000 1、0.000 01。通过试验对比选定其他超参数的值,将识别准确率最高的 1 组超参数应用到分类模型中,其中批次大小(Batchsize)设置为 32,迭代轮次(Epoch)设置为 100,训练集在每次迭代训练前会被随机打乱。

### 3.2 评价指标

选择当前主流的评价指标对分类模型进行评估,主要为准确率(Accuracy)、 $F_1$  分数、精确率(Precision)、召回率(Recall)、参数量(params)。计算如公式(1)至公式(5)所示。

$$Accuracy = \frac{TP + FN}{TP + FN + FP + FN} \times 100\%; \quad (1)$$

$$F_1 \text{ 分数} = \frac{2 \times Precision \times Recall}{Precision + Recall}; \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\%; \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%。 \quad (4)$$

式中:TP 表示模型预测正确的正样本个数,FP 表示模型预测错误的正样本个数,TN 表示模型预测正确的负样本个数,FN 表示模型预测错误的负样本个数。

对于每个卷积层而言,参数量计算如公式(5)



所示:

$$params = C_o \times (k_w \times k_h \times C_i + 1)。(5)$$

式中:  $C_o$  表示输出通道数,  $C_i$  表示输入通道数,  $k_w$  表示卷积核宽,  $k_h$  表示卷积核高。

### 3.3 试验方案

在预处理阶段,首先使用支持向量机(support vector machines, SVM)将小麦图像预先分为健康图像、病害图像 2 类;目的是为了将病害图像提取出来,并仅对病害图像分割处理后输入分类模型进行分类。相较于直接将所有数据输入分类模型,该步骤节省了算力资源。

**3.3.1 图像分割** 针对复杂背景下难以精确划分小麦病害区域的问题,本研究改进传统分割算法,通过使用大津法辅助 GrabCut 算法,以达到较好的分割效果,实现运用少量计算资源且有效提高分割精度。

**3.3.1.1 大津法(Otsu)** 大津法是一种基于图像灰度信息的自适应阈值分割算法<sup>[26]</sup>,其核心是通过寻找最佳的阈值将图像二值化分割,从而最大化图像中目标和非目标像素之间的类间方差。这一算法以其计算简便、对亮度和对比度变化不敏感等特

点而被广泛应用,尤其在前景和背景分离中表现较好。尽管该算法能够较好地分割出目标叶片边缘,但同时也会粘连较多的背景像素(图 6-b),病斑位置定位不够精准,导致识别精度不理想。

**3.3.1.2 GrabCut 算法** GrabCut 算法是一种半自动图像分割方法,2004 年由英国微软研究院的 Carsten Rother 提出<sup>[27]</sup>,该算法使用高斯混合模型(GMM)建模和图割优化迭代,实现前景和背景的分割。与 Otsu 算法相比,该算法能更好地分离目标与非目标的粘连部分且分割速度快。然而,要获取更精确的分割结果,需要通过手动交互标记 ROI 区域才能实现,且存在目标边缘过分割的问题(图 6-c)。

**3.3.1.3 结合算法** 基于上述问题,本研究提出一种改进方法,将经过大津法分割后的图像,使用腐蚀形态学来分离与目标相连的区域,并保留最大连通域,然后将因腐蚀操作缺失的像素通过膨胀操作还原;接着将经过处理后的二值图作为 ROI 区域进行框选,并作为 GrabCut 算法的标记框进行分割操作,从而省去手动标记步骤;最后将 Otsu 算法和改进的 Grabcut 算法的分割结果进行求交集处理,得到最终分割结果,如图 6-d、图 6-e 所示。



图6 白粉病叶片分割结果

**3.3.2 迁移学习训练** 在小麦病害数据集上训练全新的模型时,在训练过程中出现了准确率较低且无法稳定收敛的问题,因此本研究采用迁移学习的方法来进行训练。常用的迁移学习方法一般为 2 类,一是冻结预训练模型除分类器外的所有层,仅在目标数据集上对新分类器进行训练;二是不冻结预训练模型的任何网络层,直接在目标数据集上进行训练。由于 P-MobileViT 属于轻量级模型,模型参数量相对较小,故采用第 2 种迁移学习方法进行训练。首先在 ImageNet 数据集上对 P-MobileViT 模型进行预训练,得到预训练权重;然后在小麦病害数据集上,加载预训练权重并使用训练所有网络层的迁移学习方法进行训练。

### 3.4 试验结果分析

**3.4.1 不同初始学习率对比** 本研究将使用 3 种不同初始学习率的模型进行训练,识别准确率如图 7-A、图 7-B 所示,可以看出,初始学习率为 0.000 1 的模型的识别准确率较高,收敛快且较稳定。而初始学习率为 0.001 的模型虽然在训练集上表现优异,但在验证集上振荡较大,模型不够稳定;初始学习率为 0.000 01 的模型在训练集和验证集上均表现稳定,但是收敛速度过慢。这是由于模型已经通过迁移学习获得良好的权重参数,初始学习率选取过大会导致参数更新幅度过大,容易在最优解附近徘徊从而产生振荡;而学习率过小会导致参数更新幅度小,收敛时间长且易陷入局部最优。

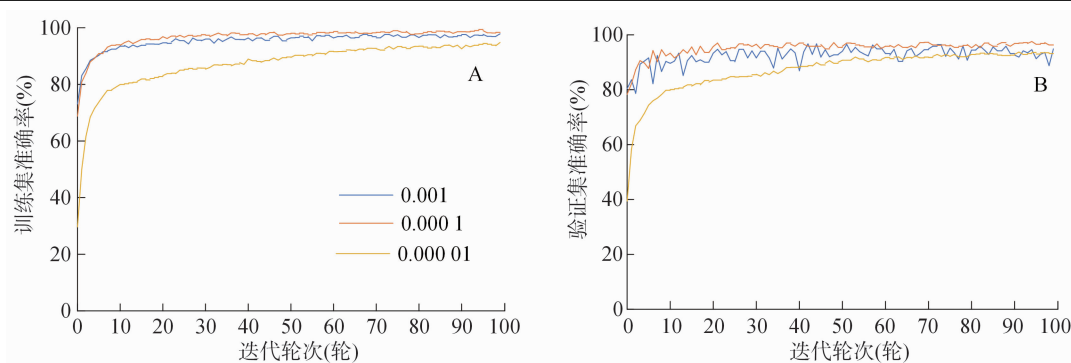


图7 不同初始学习率在训练集和验证集上的准确率曲线

3.4.2 不同改进方法对比 针对 4 种不同改进方法的 MobileViT 模型(均为 xxs 型号),分别为仅增加图像分割的 MobileViT 模型(MobileViT\_seg)、仅引入深度卷积的 MobileViT 模型(MobileViT\_DW)、仅增加局部特征的 MobileViT 模型(MobileViT\_LF)、仅使用 PoolFormer 的 MobileViT 模型(MobileViT\_PF)。将经过预训练后的上述 4 种模型、MobileViT 模型、P-MobileViT 模型分别在小麦病害数据集上进行训练,试验对比结果如表 2 所示。可以看出,本研究提出的 P-MobileViT 的准确率要明显高于其他模型,比基准模型高出 2.0 百分点,且浮点运算数和参数量分别减少了 33.5%、23.1%,该模型实现识别准确率提升的同时,明显降低了模型复杂度。

表2 不同改进方法对比

模型	精确率 (%)	召回率 (%)	识别准确率 (%)	浮点运算数 (M)	参数量 (M)
MobileViT	95.1	95.4	95.2	364	1.3
MobileViT_seg	96.4	95.7	96.0	382	1.3
MobileViT_DW	95.8	94.1	95.0	328	1.2
MobileViT_LF	97.0	95.2	96.1	395	1.4
MobileViT_PF	94.4	96.0	95.1	213	0.9
P-MobileViT	97.0	97.5	97.2	242	1.0

3.4.3 不同轻量级网络模型对比 将 4 种轻量级网络模型 ShuffleNet v2(1.0 × 型号)、MobileNet v3(Small 型号)、GhostNet、EfficientNet v2(B0 型号)进行预训练后,分别在小麦病害数据集上进行训练,对比结果如表 3 所示。可以看出,相较于准确率较高的 MobileNet v3、GhostNet 模型,P-MobileViT 的准确率分别提升 3.1、3.3 百分点,参数量分别减少 58.3%、61.5%;并且在相似的浮点运算数下,P-MobileViT 模型的识别时间开销更少,与 MobileViT 模型相比降低 31.6%。

表3 轻量级网络模型对比

模型	识别准确率 (%)	浮点运算数 (M)	参数量 (M)	推理时间 (min)
ShuffleNet v2	93.6	239	1.4	6.23
MobileNet v3	94.1	321	2.4	8.36
GhostNet	93.9	267	2.6	10.65
EfficientNet v2	93.5	390	5.3	7.81
MobileViT	95.2	364	1.3	7.56
P-MobileViT	97.2	242	1.0	5.17

3.4.4 测试样本识别结果 对测试样本的识别结果进一步统计,结果如图 8 所示,各类别样本识别准确率分别为黑穗病 99%、叶枯病 93%、白粉病 98%、条锈病 98%、秆锈病 94%、叶锈病 97%、健康 98%,其中叶枯病、秆锈病的识别准确率明显低其他 6 类,考察样本的特征表现,发现因黑穗病呈明显黑褐色,而健康样本几乎无病斑区域,故容易区分;条锈病和叶锈病的病斑颜色、形态明显不同,也较易区分;而秆锈病样本中往往伴随叶片呈枯黄色,和叶枯病病斑较为相似,使其区分难度增加,从而导致误判。

## 4 讨论与结论

本研究以小麦病害为研究对象,基于 P-MobileViT 和迁移学习构建小麦病害分类模型,对 7 种(含健康)小麦病害进行分类研究。通过图像分割,引入深度卷积、PoolFormer 结构和增加融合局部特征等,进一步优化模型的识别精度和推理时间。与 MobileViT 模型相比,P-MobileViT 的准确率提升了 2.0 百分点,达到了 97.2%,浮点运算数和参数量分别减少了 33.5%、23.1%,推理时间减少 31.6%。与轻量级网络 MobileNet v3、GhostNet 模型相比,P-MobileViT 的准确率分别提升了 3.1、3.3

百分点,参数量分别降低 58.3%、61.5%,表明本研究模型取得较好的效果,验证了改进方案的可行性。在实际应用中,本模型更有利于移动设备的部署,为农业病害分类等实际应用提供有力的技术支持。在本研究的图像分割过程中,对近端的主病害区域进行了分割,但对远端的病害区域未予以考

虑,这可能会对病害的识别准确率产生一定的影响。在下一步的研究中,将进一步优化和改进分割策略,以提升模型的性能和鲁棒性。同时,也将探索小麦病害分类任务与其他计算机视觉任务(如病害检测、预测等)相结合等方法,以提高任务的多样性和应用范围。

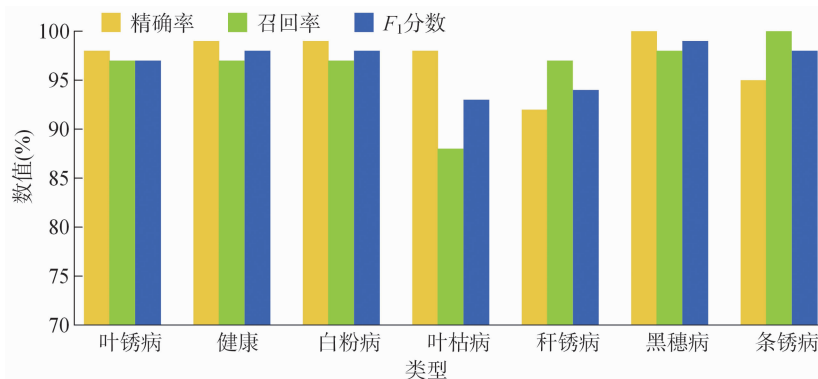


图8 小麦病害分类评估

#### 参考文献:

- [1] Liu W, Wang B, Zhao Z, et al. Historical review and countermeasures of wheat stripe rust epidemics in China[J]. China Plant Prot, 2022, 42:21-7.
- [2] Picon A, Seitz M, Alvarez - Gila A, et al. Crop conditional Convolutional Neural Networks for massive multi - crop plant disease classification over cell phone acquired images taken on real field conditions[J]. Computers and Electronics in Agriculture, 2019, 167: 105093.
- [3] 焦悦, 王思曼, 赵喜兰, 等. 50 个国外小麦品种(系)抗叶锈性鉴定[J]. 河南农业科学, 2019, 48(12): 79-88.
- [4] 刁智华, 袁万宾, 罗雅雯, 等. 基于图像处理的小麦白粉病斑生长模型构建[J]. 中国农机化学报, 2019, 40(6): 158-161.
- [5] Boulent J, Foucher S, Théau J, et al. Convolutional neural networks for the automatic identification of plant diseases[J]. Frontiers in Plant Science, 2019, 10:941.
- [6] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报, 2017, 33(19): 209-215.
- [7] Xiong Y H, Liang L F, Wang L, et al. Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset[J]. Computers and Electronics in Agriculture, 2020, 177: 105712.
- [8] Singh P, Verma A, Alex J S R. Disease and pest infection detection in coconut tree through deep learning techniques[J]. Computers and Electronics in Agriculture, 2021, 182: 105986.
- [9] 顾博, 邓蕾蕾, 李巍, 等. 基于 GrabCut 算法的玉米病害图像识别方法研究[J]. 中国农机化学报, 2019, 40(11): 143-149.
- [10] 陈燕, 朱成宇, 胡小春, 等. 基于改进 Unet 的小麦茎秆截面参数检测[J]. 农业机械学报, 2021, 52(7): 169-176.
- [11] 冯晓, 李丹丹, 王文君, 等. 基于轻量级卷积神经网络和迁移学习的小麦叶部病害图像识别[J]. 河南农业科学, 2021, 50(4): 174-80.
- [12] Nanni L, Manfè A, Maguolo G, et al. High performing ensemble of convolutional neural networks for insect pest image detection[J]. Ecological Informatics, 2022, 67: 101515.
- [13] Khan M K, Ullah M O. Deep transfer learning inspired automatic insect pest recognition[C]//Proceedings of the 3rd international conference on computational sciences and technologies. Jamshoro, Pakistan: Mehran University of Engineering and Technology, 2022.
- [14] Li C, Zhen T, Li Z H. Image classification of pests with residual neural network based on transfer learning[J]. Applied Sciences, 2022, 12(9): 4356.
- [15] 杨非凡, 徐伟诚, 陈盛德, 等. 融合 Focal Loss 与典型卷积神经网络结构的水稻病害图像分类[J]. 江苏农业科学, 2023, 51(14): 198-204.
- [16] 惠巧娟, 孙婕. 基于多尺度特征度量元学习的玉米叶片病害识别模型研究[J]. 江苏农业科学, 2023, 51(9): 199-206.
- [17] Khalifa N E M, Loey M, Taha M H N. Insect pests recognition based on deep transfer learning models[J]. J Theor Appl Inf Technol, 2020, 98(1): 60-68.
- [18] Pattnaik G, Shrivastava V K, Parvathi K. Transfer learning - based framework for classification of pest in tomato plants[J]. Applied Artificial Intelligence, 2020, 34(13): 981-993.
- [19] Mehta S, Rastegari M. MobileViT: light - weight, general - purpose, and mobile - friendly vision transformer[EB/OL]. (2021 - 10 - 05) [2023 - 11 - 01]. <https://arxiv.org/abs/2110.02178>.
- [20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017 - 06 - 12) [2023 - 11 - 01]. <https://arxiv.org/abs/1706.03762>.
- [21] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16 x 16 words: transformers for image recognition at scale[EB/OL]. (2020 - 10 - 22) [2023 - 11 - 01]. <https://arxiv.org/abs/2010.11929>.

李 帅,薄敬东,龚瑞昆,等. 基于多尺度特征增强的轻量化黄瓜病害识别模型[J]. 江苏农业科学,2024,52(20):267-276.  
doi:10.15889/j.issn.1002-1302.2024.20.032

# 基于多尺度特征增强的轻量化黄瓜病害识别模型

李 帅,薄敬东,龚瑞昆,崔传金  
(华北理工大学电气工程学院,河北唐山 063210)

**摘要:**在复杂的背景环境下对农作物病害进行准确识别与分类,为农作物病害的诊断及防治提供可靠依据,具有重要经济意义。提出了一种新的网络模型——MeNet(multiscale enhance on me),用于对大田中黄瓜的 8 种形态(其中包含 6 种病害和鲜黄瓜、鲜叶)进行精准识别。该模型的设计包括适用于网络前端的特征增强模块,对原始图像进行像素级多尺度特征增强,从而提升模型的特征表达效率;运用特征挑选的思想进行后续的特征提取和增强,再加入基于空域抑制的 SimAM 注意力,进一步突出了显著特征,提高特征效用;运用逐点卷积对特征图进行通道间信息交互,再以全局平均池化总结特征图。结果表明,相较于其他模型,本研究的 MeNet 性能更为优越,在复杂背景病害数据集上,平均准确率达到 92.38%,最高准确率达到了 92.92%,而模型的参数量仅为 0.33 M,浮点运算量仅为 0.30 G,证明 MeNet 模型在图像识别领域具有实际应用的潜力和继续研究的价值。

**关键词:**黄瓜病害;图像识别;卷积神经网络;轻量化;多尺度特征增强;空域抑制

**中图分类号:**S126;S436.421.1;TP391.41 **文献标志码:**A **文章编号:**1002-1302(2024)20-0267-10

生物防治是农业可持续发展模式的重要内容,生物防治学科为农业可持续发展提供了重要的科学支撑。近年来,我国已进入加快推进农业可持续发展的历史阶段,这为生物防治学科的发展提供了新的机遇,同时也对生物防治学科的发展提出了更高的要求<sup>[1]</sup>。

在深度学习兴起之前,人们主要使用传统的机

器学习方法来解决图像分类问题。这些方法通常涉及人工设计的特征提取器和分类器,如支持向量机(SVM)、随机森林等。这些方法在小规模问题上效果较好,但在复杂图像数据上的分类任务中表现不佳。但是,关于深度学习的探索一直都在进行,一直到 2012 年 Alex 等的 AlexNet 在 ImageNet 图像分类比赛中获胜,标志着卷积神经网络(CNN)正式崭露头角。随着卷积提取特征的特性被认可,深度学习开始向着更深的网络进发,此后的 VGG 和 ResNet 都将网络推向了更深的方向,而 GoogleNet 则提出了多尺度卷积的概念,又让网络变得更宽,通过大量卷积的叠加,追求网络的识别性能,因为人们普遍认为卷积神经网络中特征的复杂性是随网络深度加深而提高的<sup>[2]</sup>。在卷积不能满足需求后,又发展出了叠加注意力机制的网络模型,进一步提升识别准确度。

收稿日期:2023-11-09

基金项目:河北省自然科学基金(编号:F2015209308-PT);唐山市科技计划项目(编号:20150212C);河北省高等学校科学技术研究项目(编号:ZD2016070);河北省省级研究生示范课程建设项目(编号:KCJSX2021061)。

作者简介:李 帅(1996—),男,安徽阜阳人,硕士,主要从事检测技术及智能装置研究。E-mail:ncstlishuai@163.com。

通信作者:龚瑞昆,博士,教授,硕士生导师,主要从事检测技术及智能装置研究。E-mail:ncstgongruikun@163.com。

[22]Howard A, Sandler M, Chen B, et al. Searching for MobileNet v3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 1314-1324.

[23]Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention [EB/OL]. (2020-12-23) [2023-11-01]. <https://arxiv.org/abs/2012.12877>.

[24]Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/

OL]. (2017-04-17) [2023-11-01]. <https://arxiv.org/abs/1704.04861>.

[26]Otsu N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[27]Rother C, Kolmogorov V, Blake A. GrabCut: interactive foreground extraction using iterated graph cuts [M]//Seminal Graphics Papers: Pushing the Boundaries, Volume 2. New York, NY, USA: ACM, 2023: 593-598.