

杨德龙,李 婧. 基于注意力与小平方核的 ConvNeXt 农业杂草识别方法[J]. 江苏农业科学,2024,52(14):207-214.  
doi:10.15889/j.issn.1002-1302.2024.14.029

# 基于注意力与小平方核的 ConvNeXt 农业杂草识别方法

杨德龙,李 婧

(上海电力大学计算机科学与技术学院,上海 200120)

**摘要:**针对复杂自然环境下杂草识别准确率不高、泛化和拟合能力较差等问题,提出基于注意力与小平方核的 ConvNeXt 杂草图像识别方法。首先,在 ConvNeXt 模块中加入 GRN 正则化策略,有效减少识别过程的过拟合风险,提高模型的泛化能力;其次,提出在所有 Block 中把  $7 \times 7$  的深度可分离卷积分解为有 4 个平行分支的小平方核,提升对杂草图像的特征提取能力;最后,在 ConvNeXt 结合上述方法下,引入 SENet 通道注意力模块,进一步提高模型在通道方向的特征融合,强化杂草特征,构建出杂草识别模型。为验证模型的识别性能,使用公开的 9 类杂草图像样本进行对比试验,结果表明,与主流模型相比,模型在准确率、精确率、召回率、 $F_1$  分数上均表现优异,分别达到 96.172%、95.556%、96.478%、96.014%;消融试验结果表明,与基准模型 ConvNeXt 相比,GRN、小平方核分别提高 8.639%、5.691%,SENet 在前二者基础上提高了 5.174 个百分点;可视化分析证明,引入的通道注意力能更好关注到杂草特征。该模型可提高杂草识别准确率和对真实环境的泛化能力,为精准防控杂草提供有效的解决方法。

**关键词:**杂草识别;ConvNeXt;GRN 正则化策略;小平方核;通道注意力

**中图分类号:**S126;TP391.41 **文献标志码:**A **文章编号:**1002-1302(2024)14-0207-07

世界人口快速增长,预计 2050 年将达到 90 亿人,对粮食的需求量将大幅增加;发展精准农业对解决粮食增长至关重要<sup>[1]</sup>,有效防控杂草是其中关键措施之一。杂草通过争夺资源、释放化学物质而干扰作物生长,杂草的自动检测和识别在杂草防控中具有重要作用<sup>[2]</sup>,其精准识别能有效防控杂草。

卷积神经网络在图像识别领域已取得不错的成果,并被研究者引入杂草识别领域。赵辉等提出以农作物田间杂草为研究对象,以 DenseNet-121 网络为基础,在每个卷积层后引入通道注意力机制,强化杂草特征并抑制背景特征,然后加入 dropout 正则化,并且和 VGG16、ResNet-50、未改进的 DenseNet-121 模型进行对比试验<sup>[3]</sup>。疏雅丽等提出基于 Resnet-50 的网络模型,改变残差块下采样的位置,引入注意力机制,结合迁移学习,使杂草识别准确率达到 96.31%<sup>[4]</sup>。曲福恒等为了提高作物杂草图像识别的准确率,在 Resnet-50 网络模型中提出在 Retinex 算法上加入颜色恢复函数,将残差

网络中的激活函数换位 Leaky ReLU,并引入 CBAM 注意力机制模块,识别准确率达到 95.3%,高于一些主流算法<sup>[5]</sup>。Olsen 等提出的 DeepWeeds 数据集,为杂草物种贡献了第 1 个大型、公共、多类图像数据集,并用 Inception-v3、ResNet50 介绍该数据集分类性能的基线<sup>[6]</sup>。综上所述,卷积神经网络可以极好地解决图像识别问题,在识别准确率上有较高的提升,但仍存在以下问题:(1)杂草分类的主要挑战是杂草和作物可能具有非常相似的颜色或纹理,现有的深度学习分类模型难以准确提取特征,导致结果泛化能力不强;(2)因杂草图像多样且图像像素占比高,在特征提取过程中,卷积神经网络会提取大量无效的背景信息,导致过拟合。

针对上述问题,本研究提出改进的 ConvNeXt 农业杂草识别模型 GSI\_ConvNeXt<sup>[7]</sup>,通过加入 GRN 正则化策略<sup>[8]</sup>,提高模型的泛化能力,降低过拟合风险。将 ConvNeXt Block 中的深度可分离卷积分解为 4 个平行分支的小平方核,提高特征提取能力。同时加入 SENet 通道注意力机制<sup>[9]</sup>,进一步加强图像特征之间的融合,从而实现对不同杂草的高效准确识别。

## 1 农业杂草识别模型

### 1.1 模型构建

针对杂草图像识别准确率较低以及泛化能力

收稿日期:2023-08-17

基金项目:国家自然科学基金(编号:U1936123)。

作者简介:杨德龙(1999—),男,河南信阳人,硕士研究生,主要从事深度学习、图像识别研究。E-mail:delongyang@mail.shiep.edu.cn。

通信作者:李 婧,博士,副教授,主要从事深度学习、区块链技术研究。E-mail:lijing@shiep.edu.cn。

弱等问题,在原有 ConvNeXt 网络的基础上,增加 GRN 正则化策略和通道注意力机制 SENet,并把  $7 \times 7$  深度可分离卷积分解成小平行核,从而构建农业杂草识别网络模型。具体来说,在 ConvNeXt 模块中添加 GRN 正则化策略和 SENet 注意力模块(图 1),同时把 ConvNeXt 网络的 Layer Scale 的功能直接封装到 GRN 里面,这样在减少代码冗余的同时,也不会降低网络的精度。然后将每个模块中第一部分

DWConv 卷积分解成 4 个平行分支,分别为 DWConv、DWConv、DWConv、Identity。最后经过添加 dropout、激活函数、全局平均池化等操作,可以精准实现杂草特征的提取。所构建的 GSI\_ConvNeXt 网络模型以不同种类的杂草图像为研究对象,进行一系列试验验证,杂草分类准确度和泛化能力对比基准模型和现有主流的网络相比都有一定的提升,证明本模型对杂草识别的有效性。

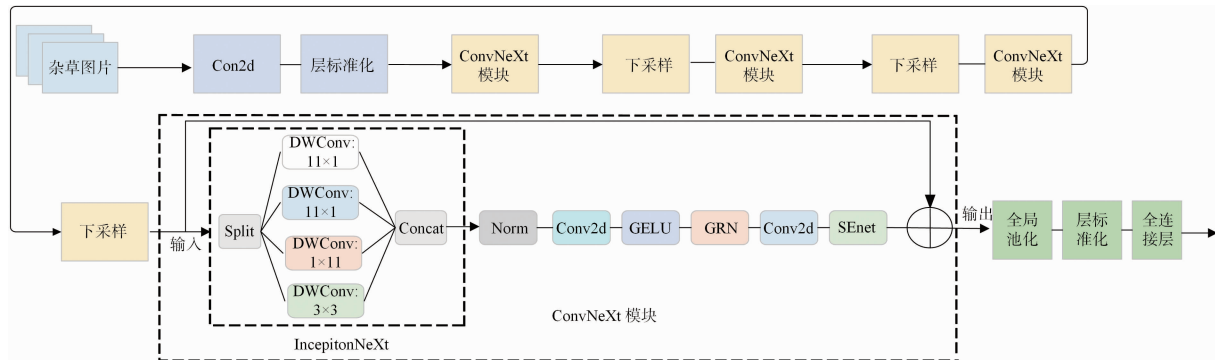


图1 GSI\_ConvNeXt 结构

## 1.2 ConvNeXt 模型

GSI\_ConvNeXt 是在基准模型 ConvNeXt 基础上进行改进的。ConvNeXt 网络是由 Facebook 团队于 2022 年提出的一种纯卷积神经网络架构模型,ConvNeXt 的整体结构是基于 ResNet 结构进行设计的。如图 2 所示,ConvNeXt 模型输入图像的高、宽、

深为  $H \times W \times C$ 。首先通过 1 个普通卷积和层标准化,之后通过 1 次核心结构 ConvNeXt 模块,再经过多次下采样和 ConvNeXt 模块(这一步是整个网络的核心),最后对输出的特征图经过全局池化、层标准化以及全连接层输出最终的结果。表 1 所展示的是 ConvNeXt - Tiny 各部分详细参数。

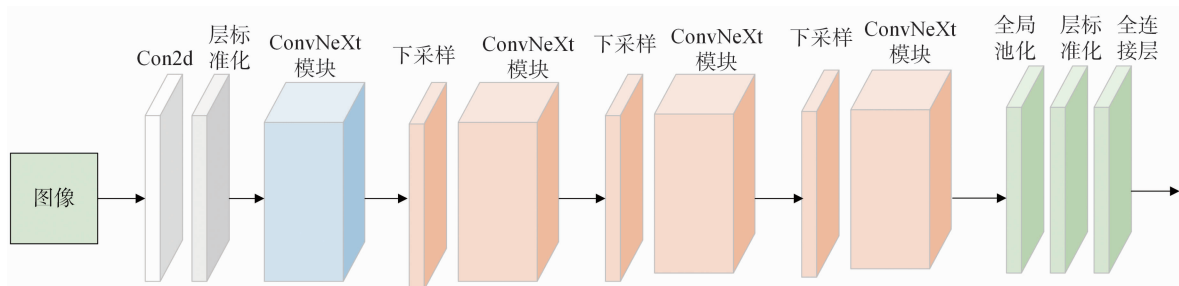


图2 ConvNeXt 模型结构

## 1.3 全局响应归一化

GSI\_ConvNeXt 加入响应归一化层 GRN,其目的是提高模型通道的对比度和选择性,防止过拟合,提高模型的泛化能力。全局响应归一化(global response normalization, GRN)是一种神经网络的有效正则化方法,旨在缓解深度卷积神经网络中的过拟合问题。它会对同一特征图中的所有位置和通道同时进行归一化,而不是针对单个位置或者通道进行归一化处理。这种全局归一化方法可以保证在

输入数据的范围发生较大变化时,神经网络的表现仍然稳定,应用在卷积神经网络中时,可以大大减少过拟合现象的发生。

输入时给定 1 个输入特征  $X \in \mathbf{R}^{H \times W \times C}$ ,在 GRN 中会经过 3 个步骤:

全局特征聚合,使用一个全局函数  $G(\cdot)$  将一个空间特征图  $X_i$  聚合成一个向量  $\mathbf{g}x$ :

$$G(X) := X \in \mathbf{R}^{H \times W \times C} \rightarrow \mathbf{g}x \in \mathbf{R}^C. \quad (1)$$

公式(1)可以看作是一个简单的池化层。其中,

表 1 ConvNeXt - Tiny 具体结构

层名	输入	ConvNeXt - T	输出
Conv1	$224 \times 224 \times 3$	$4 \times 4, 96, \text{stride}4$	$56 \times 56 \times 96$
Conv2_x	$56 \times 56 \times 96$	Layer Norm $\begin{bmatrix} d7 \times 796 \\ 1 \times 1 \ 384 \\ 1 \times 196 \end{bmatrix} \times 3$	$56 \times 56 \times 96$
Conv3_x	$56 \times 56 \times 96$	下采样 $\begin{bmatrix} d7 \times 7 \ 192 \\ 1 \times 1 \ 768 \\ 1 \times 1 \ 192 \end{bmatrix} \times 3$	$28 \times 28 \times 192$
Conv4_x	$28 \times 28 \times 192$	下采样 $\begin{bmatrix} d7 \times 7 \ 384 \\ 1 \times 11 \ 563 \\ 1 \times 1 \ 384 \end{bmatrix} \times 9$	$14 \times 14 \times 384$
Conv_5	$14 \times 14 \times 384$	下采样 $\begin{bmatrix} d7 \times 7 \ 384 \\ 1 \times 13 \ 072 \\ 1 \times 1 \ 768 \end{bmatrix} \times 3$	$7 \times 7 \times 768$
	$7 \times 7 \times 768$	全局池化 层标准化 全连接层	1 000

$G(X) = \mathbf{g}x \mid \|X_1\|, \|X_2\|, \dots, \|X_c\| \in \mathbf{R}^C$ , 是聚合第  $i$  个通道统计信息的标量。

特征归一化, 将一个响应归一化函数  $N(\cdot)$  应用于聚合值。具体来说, 使用一个标准的除法归一化操作:

$$N(\|X_i\|): \|X_i\| \in \mathbf{R} \rightarrow \frac{\|X_i\|}{\sum_{j=1,2,3,\dots,c} \|X_j\|} \in \mathbf{R}. \quad (2)$$

式中:  $\|X_i\|$  表示第  $i$  通道的 L2 范数。

特征校准, 使用计算得到的特征归一化分数来校准原始输入响应:

$$X_i = X_i \times N[G(X)_i] \in \mathbf{R}^{H \times W}. \quad (3)$$

#### 1.4 InceptionNeXt 模块

InceptionNeXt 是把基准模型 ConvNeXt 所有的深度可分离卷积分解为有 4 个平行分支的小平方核<sup>[10]</sup>, 分解后的卷积核可以更好地提取杂草图像特征。在一些著名的神经网络如 AlexNet<sup>[11]</sup>、Inceptionv1<sup>[12]</sup> 中, 已经使用  $11 \times 11$ 、 $7 \times 7$  的大卷积核, 虽然可以明显提高一些图像的识别准确率, 但是在一些数据集上的效果反而不好。为了改变这个缺陷, VGG 提出大量的  $3 \times 3$  卷积堆叠<sup>[13]</sup>, Inceptionv3 将  $k \times k$  卷积分解为  $1 \times k$ 、 $k \times 1$  的序列

叠加<sup>[14]</sup>。本研究使用的基准模型 ConvNeXt 默认采用  $7 \times 7$  的卷积核。对于大卷积核而言, 如 MixConv 将卷积核分成从  $3 \times 3$  到  $k \times k$  的几个组, 又受到 ShuffleNet v2 的启发<sup>[15]</sup>, 发现对部分通道进行处理时, 单个深度可分离卷积层已经足够, 因此本研究引入 InceptionNeXt 模块代替  $7 \times 7$  的卷积核, 将大的  $k_h \times k_w$  卷积核分解为  $1 \times k_w$ 、 $k_h \times 1$  的卷积核。具体来说, InceptionNeXt 模块会有以下 3 个步骤:

首先, 将输入的  $X$  沿着通道维度分为 4 组, 即

$$\begin{aligned} X_{hw}, X_w, X_h, X_{id} &= \text{Split}(X) \\ &= X_{:, :g}, X_{:, :2g}, X_{:, :3g}, X_{:, :3g:} \circ \end{aligned} \quad (4)$$

式中:  $g$  是卷积分支的通道数。可以设置比率  $r_g$  来确定分支通道数, 其中  $g = r_g C$ 。

接下来, 将分裂输入馈送到不同的并行分支中:

$$\begin{aligned} X_{hw}' &= \text{DWConv}_{k_s \times k_s}^{g \rightarrow g}(X_{hw}); \\ X_w' &= \text{DWConv}_{1 \times k_b}^{g \rightarrow g}(X_w); \\ X_h' &= \text{DWConv}_{k_b \times 1}^{g \rightarrow g}(X_h); \\ X_{id}' &= X_{id}. \end{aligned} \quad (5)$$

式中:  $k_s$  表示小方形卷积核大小, 默认设置为 3;  $k_b$  代表带卷积核大小, 默认设置为 11。

最后, 从每个分支把得到的各个输出 Concat 起来:

$$X' = \text{Concat}(X_{hw}', X_w', X_h', X_{id}'). \quad (6)$$

上述过程如图 3 所示。

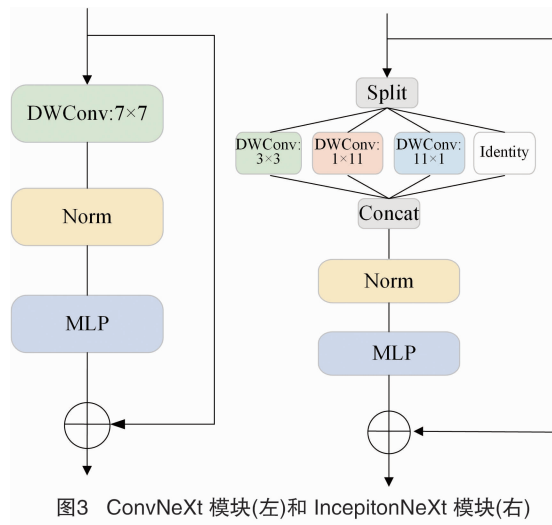


图3 ConvNeXt 模块(左)和 InceptionNeXt 模块(右)

#### 1.5 通道注意力

SENet 是经典使用通道注意力机制的网络, 是 2017 届 ImageNet 分类比赛的冠军, 它提出一种通道方向的注意力机制。通过在网络模型中插入 SE 模块, 可以允许网络执行特征再校准机制, 从而学习

使用全局信息有选择地强调有信息量的特征,抑制不太有用的特征。简单来说,就是评估各个特征通道的重要程度,再对不同的特征通道进行增强或者

抑制。SE 模块主要是压缩 (squeeze)、激励 (excitation) 这 2 个核心部分,其核心 (SENet) 结构如图 4 所示。

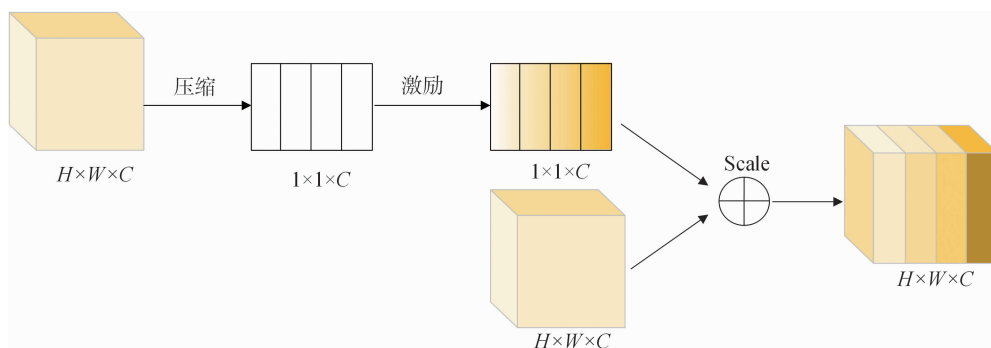


图4 SENet 结构

SENet 总体可以分为压缩操作、激励操作、Scale 操作 3 个部分。

(1) 压缩操作。作用于 1 组通道数据,将它们进行压缩,生成每个通道的全局特征表示。这样,全局特征就可以表示为整个网络提供通道响应的全局信息。

$$z = F_{sq}^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W v^k(i, j) \quad k = 1, 2, 3, \dots, C. \quad (7)$$

式中: $z$  是指输入特征执行压缩操作后的结果; $v^k(i, j)$  为经过一系列卷积之后的特征层; $C$  为  $v$  的通道数。

(2) 激励操作。对每个通道进行计算,以得到该通道的权重,用这个权重进行调制,使得网络可以自适应地对不同通道信息进行加权,以提高网络对于不同通道特征的敏感度。

$$s = F_{ex}(z, W) = \sigma[g(z, W)] = \sigma[W_2 \delta(W_1 z)]. \quad (8)$$

式中: $\sigma$  表示 Sigmoid 函数; $\delta$  表示 ReLU 激活函数; $W_1 \in \mathbf{R}^{\frac{C}{\gamma} \times C}$ 、 $W_2 \in \mathbf{R}^{C \times \frac{C}{\gamma}}$  是 2 个全连接层的参数, $\gamma$  用于减少全连接层的维度。

(3) Scale 操作。Scale 操作是在激励操作的基础上使用 1 个标量运算,对每个通道的特征进行缩放,以达到对不同通道特征值范围的动态平衡。

$$\bar{X} = F_{scale} = s_k \times v_k. \quad (9)$$

式中: $F_{scale}$  表示标量  $s_k$  和特征层  $v_k$  的逐通道乘法。

## 1.6 数据来源

本研究所采用的数据是公开数据集 DeepWeeds,该数据集是针对杂草自动识别和分类任务而建立的,包含 17 509 幅具有较高复杂度农业

区域图像,主要是原产于澳大利亚的 9 种不同杂草物种与邻近植物群,包括中国苹果、蛇麻草、香蒲、刺槐、暹罗草、苦参、橡胶藤、百日草、阴性草共 9 类农业杂草图像,这些图像类别大多都有相似的颜色、形状、大小,其中极个别的杂草非常相似,使得准确识别难度加大。该数据集中每个图像大小均为  $256 \times 256$ ,这可能会增加模型的计算量。图 5 展示了该数据集的部分杂草样本。

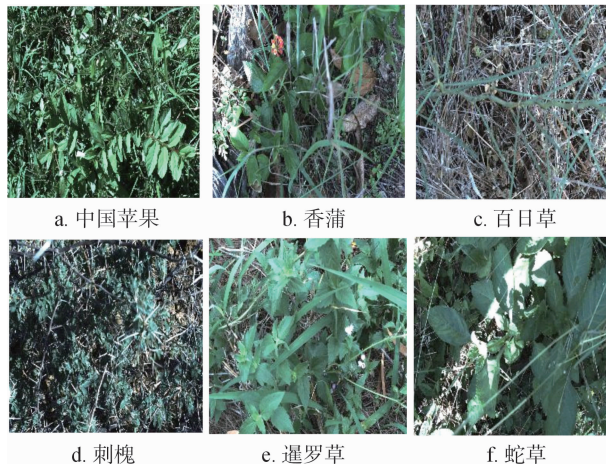


图5 部分杂草样本示例

## 1.7 数据预处理

在深度学习图像识别中,当数据集的数量较大并且各类别分配均匀的情况下,卷积神经网络才能发挥出较好的特征提取能力。在 DeepWeeds 杂草数据集中,阴性草类的图像有 9 106 幅,其余 8 类杂草每类样本数量大约为 1 000 幅;对比其他 8 类样本图像数量,需要减少阴性草的比例,使其达到正常图像识别范畴的类别比例。本研究采取随机去除 70% 的阴性草样本的做法,最终总样本数量达到 11 103 幅图像。然后将样本随机划分为 3 个数据

集,三者划分的比例大约为训练集:验证集:测试集=3.74:1:1,训练集为7 232 幅,验证集1 938 幅,测试集1 933 幅。DeepWeeds 数据集图像识别难度可归结为颜色相同、光照反光、角度偏移、遮挡物遮挡,并且类别间差异较小。因此在试验开始之前,要先对数据集预处理,例如本研究将图像大小设置为 $224 \times 224$ ,进行中心裁剪、随机水平翻转、标准化等预处理。

## 2 试验方法

### 2.1 试验环境

本试验于2023年5月在上海电力大学计算机科学与技术学院使用Autodl上的服务器运行,所有试验均是在同一个服务器上开发。试验使用Ubuntu 20.04,CPU型号为12 vCPU Intel® Xeon® Platinum 8255C CPU @ 2.50 GHz,GPU型号为RTX 2080 Ti(显存为11 GB),深度学习框架为PyTorch 1.11.0,Cuda版本为11.3。网络试验参数设置如下:优化器采用AdamW<sup>[16]</sup>,学习率调整使用余弦退火策略,在模型训练初期使用Warm up 预热学习率的方式<sup>[17]</sup>,加速模型的收敛速度。模型的迭代次数为100,批次大小为16,初始学习率为0.000 5,采用权重衰减的方法抑制过拟合<sup>[18]</sup>,衰减系数设为0.02。损失函数采用交叉熵损失函数CrossEntropyLoss<sup>[19]</sup>。

### 2.2 评估指标

本研究选取准确率(accuracy,  $A$ )、召回率(recall,  $R$ )、精确率(precision,  $P$ )、 $F_1$ 分数( $F_1$  scores)作为判断模型好坏的评价方法。

准确率指分类正确样本数与总样本数的比例,其计算公式为:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

精确率指预测为正的样本中是真正的正样本的比例。其计算公式为:

$$P = \frac{TP}{TP + FP} \quad (11)$$

召回率指样本中的正例被预测正确的比例。其计算公式为:

$$R = \frac{TP}{TP + FN} \quad (12)$$

$F_1$ 分数指精确率和召回率的调和平均数。其计算公式为:

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (13)$$

式中: $TP$ (真阳性)表示将正例预测为正例的数量; $FP$ (假阳性)表示将负例预测为正例的数量; $FN$ (假阴性)表示将负例预测为负例的数量; $TN$ (真阴性)表示将正例预测为负例的数量。一般而言,准确率、精确率、召回率、 $F_1$ 分数的值越高,模型的分类性能越好。

## 3 试验结果分析

### 3.1 模型性能分析

本研究利用验证集准确率、测试集准确率来评估GSI\_ConvNeXt网络分类模型的训练效果。其中,验证集是用于调整模型的超参数和进行模型选择,验证集的准确率高说明超参数的选择比较优;而测试集用于评估真实世界数据的准确率和泛化能力,对模型进行客观的、独立的评估,最能说明模型的性能。

为了验证本研究模型的有效性,在一定条件(数据集与处理方法相同)下迭代100次,在验证集上分别与3种典型图像识别网络进行准确率和Loss曲线的对比,其结果如图6所示。图6-A为不同模型准确率对比,而图6-B则为验证集的损失值在不同模型中的对比。由图6-A可知,本研究模型准确率最高,达到96.172%;ResNet50达到95.200%;而ShuffleNet v2、MoblieNet v2<sup>[20]</sup>接近90%。由图6-B可知,本研究模型验证集的Loss曲线比其他3个模型更加平滑,说明模型拟合得更好。因此,相比主流的经典模型,本研究构建的模型在验证集上训练得最好。

对于多分类而言,验证集的准确率不能完全说明该模型在真实数据中拟合得好,还需要看测试集的准确率,也就是看该模型对真实数据的泛化能力。本研究模型采用准确率、精确率、召回率、 $F_1$ 分数来说明模型的性能。由表2可知,本研究模型在准确率、精确率、召回率、 $F_1$ 分数上的数值均为最高,识别准确率比Resnet50高约1.3百分点;本研究模型在测试集的准确率仅仅比验证集的准确率低0.528百分点,说明本研究模型对数据的拟合能力好。

### 3.2 混淆矩阵

混淆矩阵在多分类任务中是最常见且最有效的评价指标,可以直观地看到类别的误识别情况。采用验证集训练好的权重对测试集的1 933 幅图像进行分类结果测试,9类杂草的混淆矩阵结果如图7



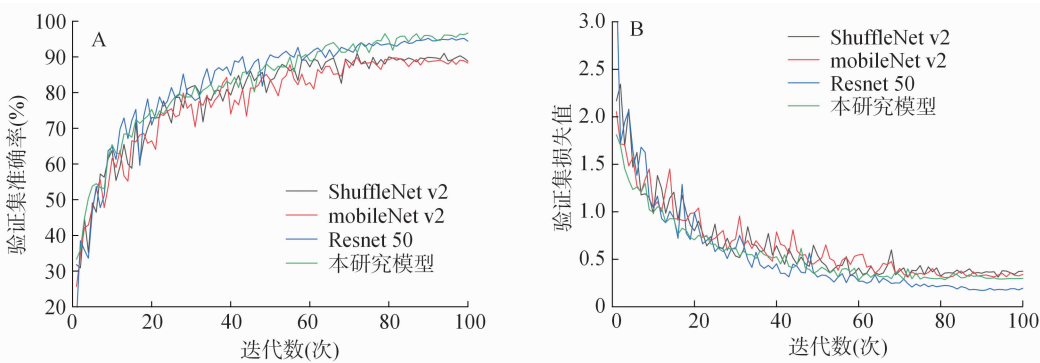


图6 验证集准确率对比(A)和验证集 Loss 值对比(B)

表 2 测试集准确率结果对比

模型	准确率 (%)	精确率 (%)	召回率 (%)	F <sub>1</sub> 分数 (%)
ShuffleNet v2	88.722	87.322	88.311	87.814
MoblieNet v2	88.464	87.144	88.233	87.686
ResNet50	94.878	94.111	94.833	94.471
本研究模型	96.172	95.556	96.478	96.014

所示(颜色的深浅代表每个类别识别正确图像数量)。由图 7 可以看出,大部分杂草类别预测错误的图像数量都较少,其中,中国苹果、蛇草这 2 个类别误识别数量较多。中国苹果有 9 幅图像误识别为阴性草,有 6 幅图像误识别为蛇草;蛇草有 10 幅图像误识别为中国苹果。这是因为这 2 个类别图像特征较为相似,造成这 2 类互识别错误相较于其他较多。

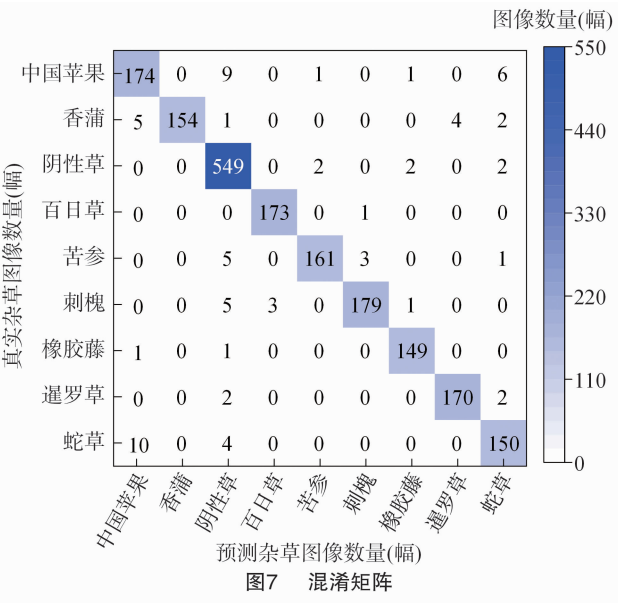


图7 混淆矩阵

3.3 消融试验

为了验证本研究所提出的优化方法对基准模型性能的提高能力,在相同试验条件下,使用 ConvNeXt 模型对测试集采用不同优化方法进行测

试,其结果如表 3 所示。在基准模型上,添加 GRN 正则化策略方法,可以在准确率、精确率、召回率、F<sub>1</sub> 分数这 4 项指标上都有大约 8 个百分点左右的性能提升;而把 7×7 的深度可分卷积更改为 4 个平行分支的小平方核 IncepitonNeXt,其结果在 4 个指标下也有 5 个百分点的性能提升;在只加入 SENet 注意力机制的准确率也大概提升了 0.5 百分点。本研究所提出的 GSI\_ConvNeXt 模型,取得了约 13 百分点的性能提升。从试验结果上看,基准模型 ConvNeXt 准确率较低,其原因首先可能是 Deepweeds 数据集的特征与 ConvNeXt 模型所能捕捉的特征不匹配,模型可能无法准确地对其特征进行识别;把深度可分卷积更改为 IncepitonNeXt,可更好地捕捉 DeepWeeds 数据集的特征。其次,单独使用 ConvNeXt 模型很容易导致结果严重过拟合,需要加入 GRN 正则化策略,可以明显提升模型精度。单独加注意力机制效果提升甚微,但是在加入 GRN 正则化策略和 IncepitonNeXt 的基础上再加入 SENet,经过不断优化超参数,可以在前二者基础上再提升 5 百分点的精度,使本研究模型在测试集上达到 96.172% 的识别准确率,结果明显高于基准模型。因此,本研究所提出的改进 ConvNeXt 模型的方法对农业杂草的识别是有效的。

表 3 不同优化方法对于 ConvNeXt 模型的影响

模型	准确率 (%)	精确率 (%)	召回率 (%)	F <sub>1</sub> 分数 (%)
ConvNeXt	80.445	78.733	80.011	79.367
GRN_ConvNeXt	89.084	88.256	88.378	88.317
SE_ConvNeXt	80.600	79.365	80.944	80.142
Inxt_ConvNeXt	86.136	85.144	85.778	85.460
GRN_SE_ConvNeXt	90.895	90.256	90.311	90.283
GRN_Inxt_ConvNeXt	90.998	89.744	90.456	90.099
本研究模型	96.172	95.556	96.478	96.014

### 3.4 杂草图像特征图可视化

卷积神经网络的特征可视化可以有效判别分类模型的效果。本研究利用 Grad - CAM 可视化杂草图像的特征图<sup>[21]</sup>,主要是将本研究所提出的网络模型的最后一层特征层利用 Grad - CAM 进行可视化输出,观察杂草图像中的有效区域,可以直观看出模型能否正确提取杂草特征。如图 8 所示,从上到下使用 3 种杂草图像分别是中国苹果、苦参、暹罗草。热力图展示的是基准模型和本研究模型之间特征图的对比(图中颜色越深,代表模型对于所在区域关注度越高,对识别结果的影响越大),可以明显看出本研究提出的杂草识别模型正确关注到了图像中的杂草特征,说明本研究添加的注意力机制可以有效增加模型对图像特征提取能力。进一步看出,热力图主要以杂草的叶子及其颜色为感兴趣区域,也从侧面说明本研究模型对于杂草特征的提取效果较好。

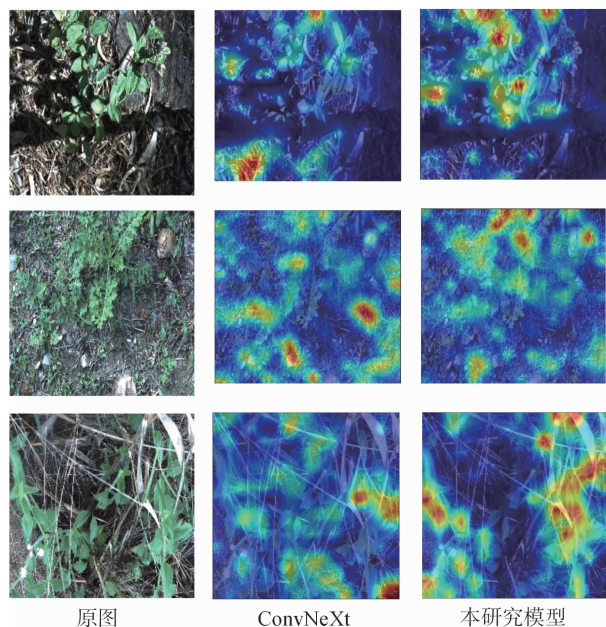


图8 特征可视化对比结果

## 4 结论

为了精准防控杂草,本研究提出 GSI\_ConvNeXt 模型,在基准 ConvNext 基础上,利用 9 类杂草图像,加入 GRN 正则化策略和通道注意力模块,把深度可分卷积分解成 4 个平行分支的小平方核,构建杂草识别模型,实现对农业杂草分类的高效准确识别,并得出以下结论:(1)在 ConvNeXt 模块中加入 GRN 正则化方法,降低模型过拟合;针对不同杂草图像自身特点,加入通道注意力机制、更改深度可

分卷积,提高矿物图像特征融合的能力。(2)与一些主流网络模型相比,GSI\_ConvNeXt 对 9 类杂草图像识别中训练过程收敛更快,准确率、精度、召回率、 $F_1$  分数 4 项数值均最佳。消融试验进一步证明了模型的有效性。(3)利用 Grad - CAM 方法,通过与基准模型的图像特征图的可视化结果进行对比,进一步证明本研究模型在杂草识别领域的有效性。

本研究成果对于农业杂草识别领域具有一定的借鉴意义。如何进一步提高杂草识别的准确率和泛化能力,以及如何将模型应用到实际场景中,仍然是重要的研究方向。

### 参考文献:

- [1] Tauber M, Gollan B, Schmittner C, et al. Passive precision farming reshapes the agricultural sector[J]. Computer, 2023, 56(1): 120 - 124.
- [2] 李东升, 胡文泽, 兰玉彬, 等. 深度学习在杂草识别领域的研究现状与展望[J]. 中国农机化学报, 2022, 43(9): 137 - 144.
- [3] 赵辉, 曹宇航, 岳有军, 等. 基于改进 DenseNet 的田间杂草识别[J]. 农业工程学报, 2021, 37(18): 136 - 142.
- [4] 疏雅丽, 张国伟, 王博, 等. 基于深层连接注意力机制的田间杂草识别方法[J]. 计算机工程与应用, 2022, 58(6): 271 - 277.
- [5] 曲福恒, 李婉婷, 杨勇, 等. 基于图像增强和注意力机制的作物杂草识别[J]. 计算机工程与设计, 2023, 44(3): 815 - 821.
- [6] Olsen A, Kononov D A, Philippa B, et al. DeepWeeds: a multiclass weed species image dataset for deep learning[J]. Scientific Reports, 2019, 9: 2058.
- [7] Liu Z, Mao H Z, Wu C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966 - 11976.
- [8] Woo S, Debnath S, Hu R H, et al. ConvNeXt v2: co - designing and scaling ConvNets with masked autoencoders[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 16133 - 16142.
- [9] Hu J, Shen L, Sun G. Squeeze - and - excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132 - 7141.
- [10] Yu W H, Zhou P, Yan S C, et al. InceptionNeXt: when inception meets ConvNeXt[EB/OL]. [2022 - 11 - 12]. <http://arxiv.org/abs/2303.16900>.
- [11] Alom M Z, Taha T M, Yakopcic C, et al. The history began from AlexNet: a comprehensive survey on deep learning approaches[EB/OL]. [2022 - 11 - 12]. <http://arxiv.org/abs/1803.01164>.
- [12] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818 - 2826.
- [13] Yu W, Yang K, Bai Y, et al. Visualizing and comparing AlexNet and VGG using deconvolutional layers[C]//Proceedings of the 33

姜月明,王 健,董光辉,等. 基于改进卷积神经网络的苹果叶片病害识别[J]. 江苏农业科学,2024,52(14):214-221.  
doi:10.15889/j.issn.1002-1302.2024.14.030

# 基于改进卷积神经网络的苹果叶片病害识别

姜月明,王 健,董光辉,胡彭元

(东北林业大学计算机与控制工程学院,黑龙江哈尔滨 150040)

**摘要:**为了提高真实条件下苹果叶片病害识别准确度和识别速度,提出了一种基于改进的卷积神经网络苹果叶部病害识别方法,该方法是在卷积神经网络 VGG16 的基础上进行改进完成的。首先针对 5 类常见苹果叶片病害图片样本集合,采用数字图像处理算法(如旋转照片角度、增强降低图像亮度和锐度、添加高斯噪声等)进行数据集增强完成原有数据集的扩充,扩充后获得 26 377 张苹果叶片病害图像,以增加样本多样性,提高模型的泛化能力。通过对叶片病斑特征的差异进行研究,比较了多种高效的卷积神经网络模型架构,最终选出 VGG16 网络模型作为基础模型,并对其进行改进,通过添加 SK 模块以及将全连接层改为全局平均池化,提升了模型的识别准确率以及网络稳定性,同时也加快了模型的收敛速度,提升了苹果叶片病害识别速度。试验表明,改进后的 VGG16 模型识别准确率高达 96.17%,相对于 VGG16 模型提升了 3.55 百分点。试验结果表明,本研究为苹果叶片病害识别提供了一种可行的高性能解决方案,可有效提升苹果叶片病害的识别准确度和速度,也为深度学习和人工智能技术在农业信息化领域的应用探索了新的途径。

**关键词:**病害识别;卷积神经网络;迁移学习;图像识别;VGG16 模型

**中图分类号:**TP391.41 **文献标志码:**A **文章编号:**1002-1302(2024)14-0214-08

苹果种植技术在我国拥有悠久的历史。近年来,我国苹果产业稳定发展并且成为世界上苹果产量最大的国家,种植面积和产量均占世界 50% 以上,作为最畅销的水果之一,苹果每年的消费量大约在 1 900 万 t<sup>[1]</sup>。大面积种植苹果时如果病害不能及时发现并得到治理,会给苹果生产带来重大影响。特别是在早春初夏季节,苹果花叶病等病害肆

虐流行,造成重大损失。“预防为主,综合防治”是防治病害的重要原则。因此早发现病害和快速精准识别,就能够及时为防治病害提供有效手段<sup>[2-3]</sup>。

苹果叶片病害症状复杂,传统的疾病防治技术是通过人工实时观察病情动态发展,然后根据病害提出防治策略。由于工作人员识别经验不同,若采取肉眼识别病害的办法准确率不稳定,农药等剂量使用也会不合理,严重影响苹果的生长及质量。此外,基于生物传感器及光谱技术识别方法存在成本过高,对使用人员要求高等弊端,无法进行大规模推广<sup>[4]</sup>。因此,深度学习技术与现代农业相结合已经成为当代智慧农业的主流。卷积神经网络(convolutional neural networks, CNN)是一种具有深

收稿日期:2023-07-24

基金项目:中央高校基本科研业务费专项资金(编号:2572022BH03)。

作者简介:姜月明(1989—),女,黑龙江哈尔滨人,博士,讲师,硕士生导师,研究方向为人工智能、智能检测和诊断。E-mail:jym\_nefu@nefu.edu.cn。

rd International Conference on Machine Learning,2016.

[14] Xia X L, Xu C, Nan B. Inception - v3 for flower classification [C]//2017 2nd International Conference on Image, Vision and Computing. Chengdu:IEEE,2017:783-787.

[15] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet v2: practical guidelines for efficient CNN architecture design [C]//Computer Vision - ECCV 2018 of Munich:ACM,2018:122-138.

[16] Loshchilov I, Hutter F. Decoupled weight decay regularization[EB/OL]. [2022-11-12]. <http://arxiv.org/abs/1711.05101>.

[17] You Y, Gitman I, Ginsburg B. Large batch training of convolutional networks[EB/OL]. [2022-11-12]. <http://arxiv.org/abs/>

1708.03888.

[18] 李文举,苏 攀,崔 柳. 基于随机扰动的过拟合抑制算法[J]. 计算机仿真,2022,39(5):134-138.

[19] Ho Y, Wookey S. The real - world - weight cross - entropy loss function; modeling the costs of mislabeling[J]. IEEE Access,2020,8:4806-4813.

[20] 王改华,翟乾宇,曹清程,等. 基于 MoblieNet v2 的图像语义分割网络[J]. 陕西科技大学学报,2022,40(1):174-181.

[21] 朱炳宇,刘 朕,张景祥. 融合 Grad - CAM 和卷积神经网络的 COVID-19 检测算法[J]. 计算机科学与探索,2022,16(9):2108-2120.