

多尺度融合注意力机制的番茄叶病识别网络^①



王 斌¹, 余本国^{1,2}

¹(中北大学 软件学院, 太原 030051)

²(海南医学院 生物医学信息与工程学院, 海口 571199)

通信作者: 余本国, E-mail: yubg@hainmc.edu.cn

摘 要: 针对普通神经卷积网络对番茄叶病的识别精准度, 先提出一种新型的多尺度融合注意力机制的网络 (MIPSA Net), 在该网络中采用轻量级网络作为主要框架, 减少了网络的参数, 为了增加网络的深度和宽度, 加入了 Inception 结构, 用于提取数据的多尺度特征信息, 同时, 在这个过程中使用更加精细的双重注意力机制, 极化自注意力 (polarized self-attention, PSA), 作为一个即插即用的模块, 将其嵌入整个模型中, 提高了重要特征点的表达能力, 同时 PSA 模块的轻量化也符合本模型的使用. 在卷积后加入全连接层, 进行分类. 使用提出的网络在 Kaggle 公开数据集 tomato leaves dataset 上进行实验, 对其进行 30 批次的训练, 取得了 91.05% 的准确率, 与其他方法进行对照, 取得良好的效果. 试验结果表明该网络对番茄叶病的分类有很好的效果, 为分类网络的网络结构和参数配置方面提供一些参考价值.

关键词: 多尺度; Inception 结构; 番茄叶疾病; 识别网络; 极化自注意力

引用格式: 王斌, 余本国. 多尺度融合注意力机制的番茄叶病识别网络. 计算机系统应用, 2023, 32(7): 202-210. <http://www.c-s-a.org.cn/1003-3254/9163.html>

Identification Network of Tomato Leaf Disease Based on Multi-scale Fusion Attention Mechanism

WANG Bin¹, YU Ben-Guo^{1,2}

¹(School of Software, North University of China, Taiyuan 030051, China)

²(College of Biomedical Information and Engineering, Hainan Medical University, Haikou 571199, China)

Abstract: To improve the identification accuracy of ordinary neural convolutional networks for tomato leaf disease, a new network based on the multi-scale fusion attention mechanism (MIPSA Net) is proposed. The lightweight network is used as the main framework to reduce the network parameters in this network. To increase the depth and width of the network, the Inception structure is added to extract multi-scale feature information of data. Meanwhile, a more elaborate dual attention mechanism, polarized self-attention (PSA), is used in this process as a plug-and-play module to be embedded in the whole model, which improves the expressive power of important feature points. The lightweight PSA modules are also suitable for this model. A full connection layer is added after the convolution for classification. The proposed MIPSA Net is applied to conduct experiments on Kaggle public dataset, tomato leaves dataset, with 30 batches of training, achieving an accuracy rate of 91.05%. The results show that this network is strikingly effective in the classification of tomato leaf diseases compared with other networks, which provides some reference value for the network structure and parameter configuration of the classification network.

Key words: multi-scale; Inception structure; tomato leaf disease; identification network; polarized self-attention (PSA)

① 基金项目: 海南省自然科学基金 (822RC713)

收稿时间: 2022-12-09; 修改时间: 2023-01-17; 采用时间: 2023-02-13; csa 在线出版时间: 2023-05-19

CNKI 网络首发时间: 2023-05-22

随着计算机视觉的发展,逐渐出现了很多卷积神经网络,在1998年LeCun等人提出了LeNet5网络,采用了基于梯度的反向传播的算法,利用了二维卷积,对图像进行处理,CNN网络逐渐走入大众视野^[1].在2012年AlexNet网络的问世,首次在CNN中应用ReLU,Dropout和LRN等Trick,把CNN网络的基本原理应用到了很深很宽的卷积网络中^[2].在AlexNet网络出现不久后,出现了NIN网络,相对于AlexNet网络加入了 1×1 的卷积核,增强了模型在感受野内对局部区域的辨别能力,在NIN网络中,利用卷积取代了全连接层,从而使空间信息能够自然地传递到之后的层中,增加了网络的非线性能力^[3].2014年,由牛津大学工程系提出VGG网络,从网络的深度入手,VGG网络采用连续的 3×3 的卷积核来代替大的卷积核,不仅使得网络的层数更加深厚,也大大减少了网络的参数^[4].同时,在2014年,Google团队吸收了NIN网络中的串联思路,并在此基础上进行改进,提出了GoogLeNet网络^[5],Inception模块就是GoogLeNet网络核心,不仅使用了卷积核为 1×1 的卷积来进行升降维,同时加入并联结构,可以使网络在多个尺度上进行卷积.2015年ResNet^[6]网络的出现,提出了残差块,解决了模型因为深度原因的退化问题.之后的网络,大部分都是对ResNet网络的变种.2018年Google团队提出了MobileNetV2网络^[7],对残差块进行改变,提出了inverted residuals(倒残差结构),采用了新的激活函数ReLU6,并且采用了深度可分离卷积,减少了网络参数.

注意力机制源于对人类视觉的研究,2014年,Bahdanau等人提出了注意力(attention)机制,并广泛地应用于深度学习的各个领域^[8].多头注意力(multihead attention),允许注意力机制组合使用查询、键和值的不同组合^[9].自注意力(self-attention),在每一个查询中都会关注所有的键值对,并生成一个注意力输出^[9,10],同时出现了位置编码(positional encoding),为输入的数据中添加绝对或相对的位置信息.2021年提出了多尺度注意力机制,金字塔分割注意力模块,在图像分类中取得显著的性能提升^[11].

为了更好地提高番茄叶病的准确率,更好地提取有效特征,本文提出一种多尺度融合注意力机制的识别网络,本文工作大致如下.

(1) 减少网络的参数,采用深度可分离卷积,提高模型的准确率,防止数据的过拟合.

(2) 增加多尺度Inception模块,增加了网络的深度和宽度,更加优秀的提取有效特征.

(3) 为了引导网络具有他更好的识别准确度,提出了极化自注意力(PSA),提高有效特征的权重,有效提高了网络的识别分类能力.

1 多尺度融合注意力机制的识别网络

1.1 MobileNetV2网络

MobileNetV2网络是在2018年由Google团队提出的,网络中加入了倒残差模块,是残差模块的变种,两模块的对比如图1所示.

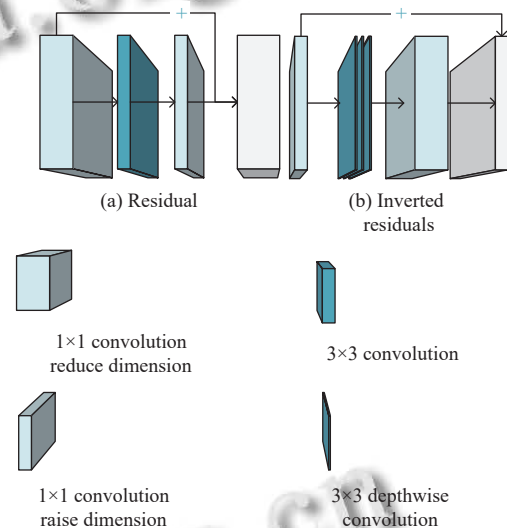


图1 倒残差与残差模块对比

图1(a)为残差模块,图1(b)为倒残差模块两者的区别就在于卷积核 1×1 的卷积的作用,在残差块中先用 1×1 的卷积进行降维,在模块最后在用 1×1 的卷积进行升维;而在倒残差块中正好相反,先用 1×1 的卷积进行升维,在模块最后在用 1×1 的卷积进行降维.其中,在倒残差模块中,使用 3×3 的逐通道卷积代替普通的 3×3 卷积,残差块中的激活函数ReLU,可能会造成数据的丢失或破坏,所以替换为了ReLU6,ReLU6函数能够更好地使用在移动端中,更好地降低精度的损失.

在MobileNet网络中,利用了深度可分离卷积,深度可分离卷积由深度卷积和逐点卷积组成,深度卷积为每一个通道提供过滤器,点卷积用于创建深度层输出的线性组合.深度可分离卷积与传统卷积的对比如图2所示.

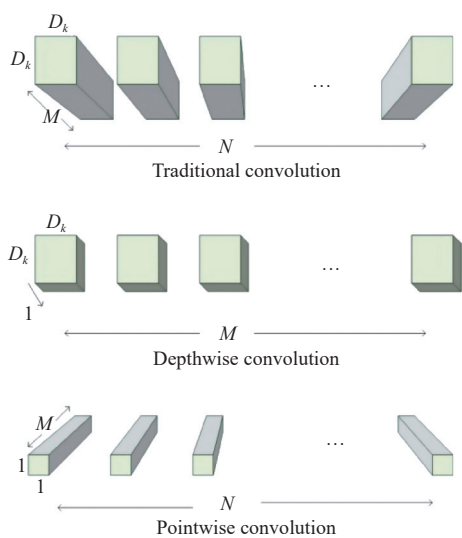


图2 卷积对比

在深度可分离卷积中的参数为深度卷积的参数+逐点卷积参数. 深度卷积的参数为: $D_k \times D_k \times M \times 1$, 逐点卷积的参数为: $1 \times 1 \times N \times M$, 所以深度可分离卷积中的参数总数为 $(D_k \times D_k + N) \times M$. 而在传统卷积中的参数总数为 $D_k \times D_k \times M \times N$. 其中, D_k 为卷积核的大小, M 为卷积的深度, N 为卷积的数量, 由此看出深度可分离卷积的参数和计算量都小于传统卷积.

1.2 Inception 模块

Inception 模块是 GoogLeNet 网络的重要模块, 吸收了 NIN 网络的串联思路, 并在此基础上进行改进, 在 Inception 模块中加入卷积之间的并联, 重点解决了什么样的卷积更好的训练网络, 并提出了使用不同大小的卷积组合可以更有利于网络的训练^[12], Inception 模块的结构如图 3 所示.

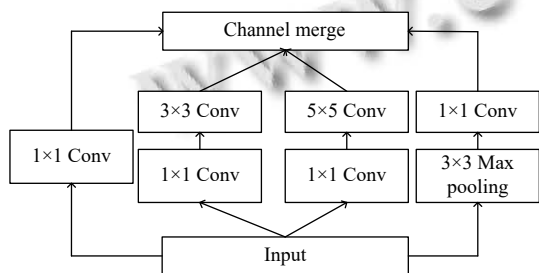


图3 Inception 模块架构

Inception 模块由 4 条并行路径组成, 前 3 条路径由 1×1 , 3×3 和 5×5 卷积构成, 从不同的空间大小中提取特征信息, 中间的两条路径在输入上执行了 1×1 的卷积层来减少通道数, 从而降低模型的复杂度, 在对后

的路径中使用 3×3 的最大汇聚层. 4 条路径中, 对图像的高度和宽度不会进行改变, 在最后的通道合并层中, 只需要将 4 条路径的输出相加合并即可.

本文使用 Inception 模块, 其目的是在保证模型质量的前提下, 解决卷积层的堆积问题, 避免冗余计算, 减少参数数量, 提取高维特征. 其中不同尺度的卷积组合, 使得特征进入不同大小的卷积组合进行提取, 对特征进行多尺度提取, 增强了网络的泛化能力和表征能力, 同时增加了网络的非线性, 大大地提高了网络的学习效率.

1.3 PSA 极化自注意力机制

PSA 注意力机制, 用于解决像素级的回归任务, 相比于其他注意力机制, 极化自注意力机制在通道和空间维度保持比较高的分辨率, 这一步能够减少降维度造成的信息损失, 其次采用细粒度回归输出分布的非线性函数.

PSA 注意力机制的主要操作就是将输入从通道上分成 S 组. 每一组进行不同卷积核大小的卷积, 以获取不同尺度的感受野, 提取不同尺度的信息. 再通过 SE 模块, 提取每组的通道的加权值, 最后对 S 组的加权值进行 Softmax 归一化并加权, 如图 4 所示.

1.4 番茄叶病识别网络

根据对深度卷积网络的学习, 以及通过文献 [13,14] 的启发, 本文提出了一种新型的多尺度注意力机制融合 的识别网络, 网络框架如图 5 所示.

在输入图片后, 对图片进行卷积的一系列操作, 提取图像特征信息, 最终通过全连接层进行分类. 采用轻量级 MobileNetV2 网络作为基础架构, 因为深度可分离卷积优于普通卷积, 使得网络的参数以及计算量都得到了不同程度的下降, 使得网络的性能更加优秀, 同时轻量级网络可以更好地嵌入在移动端. 采用 Inception 模块, 增加网络的深度和宽度, 输入图片经过不同尺度的特征提取以及池化操作, 获得输入图片的不同尺度信息, 不仅使得参数个数减少, 同时又增强了提取到的特征. 由于不同类别中的图像, 在局部特征中有更明显的差别, 因此为了使得网络更加关注图像中的局部不同区域, 提取到的特征有更高的分辨率, 所以引入注意力机制, 提高网络对局部特征的识别, 提高网络准确率.

整体模型参数如表 1 所示, 在识别网络的第 4 层之后, 整体卷积开始进行升维, 为了保证数据冗余以及卷积的堆叠所造成的问题, 在第 4 层之后加入 Inception 模块, 提取高维特征, 提高网络的泛化性和表征能力.

并且将最后的 3×3 卷积用于升维, 提高维度后接入 Inception 模块, 增加了网络的深度和宽度, 同时采用多尺度特征进行处理, 提高特征提取的有效性, 并且将其

中的 ReLU 激活函数换为 ReLU6. 在最后的 Inception 模块之后加入一个全局自适应池化层, 以及用于分类的全连接层.

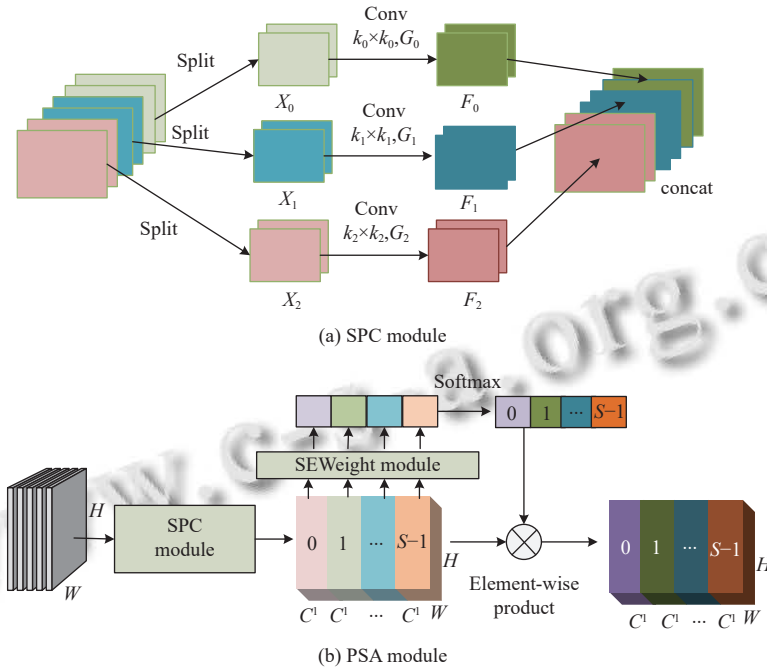


图4 PSA 注意力机制模块

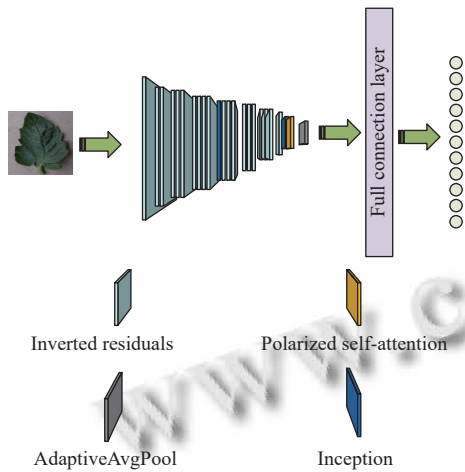


图5 网络架构图

PSA 注意力机制, 运用了金字塔分割原理, 实现多尺度特征融合, 该注意力机制可以融合不同尺度的上下文信息, 并产生更好的像素级注意力. 其次 PSA 机制具备了即插即用, 轻量级的特点, 可以更好地嵌入在网络之中. 由于 PSA 注意力机制第 1 次的出现是在 ResNet 网络中^[12], 在 ResNet 网络中具有良好的效果. 在进行试验中, 将 PSA 注意力机制直接加入在深度可

分离卷积网络中, 进行模型训练后, 该网络的识别效果没有达到预期, 所以对其进行改动, 使它更好地融入网络, 结构如图 6 所示. 首先利用 1×1 进行降维, 接入 PSA 模块, 左右后通过 1×1 升维. 也就是嵌入在残差模块中, 使得 PSA 模块的性能得以发挥. 在自适应池化层后加入改动后的 PSA 极化自注意力机制, 实现多尺度注意力机制的融合, 提高网络的识别率, 卷积层后依次加入批量归一化和 ReLU6 激活函数, 提取到的特征进入全连接层进行分类.

表1 模型参数

输入类型	步幅	重复次数	输出尺寸
3×3 Convolution	1	1	$128 \times 128 \times 32$
Inverted residuals	1	1	$128 \times 128 \times 16$
Inverted residuals	2	2	$64 \times 64 \times 24$
Inverted residuals	2	3	$32 \times 32 \times 32$
Inverted residuals	2	4	$16 \times 16 \times 64$
Inception	—	1	$16 \times 16 \times 256$
Inverted residuals	1	3	$16 \times 16 \times 96$
Inverted residuals	2	3	$8 \times 8 \times 160$
Inverted residuals	1	1	$8 \times 8 \times 320$
3×3 Convolution	—	1	$8 \times 8 \times 832$
Inception	1	1	$8 \times 8 \times 1024$
PSA	—	1	$8 \times 8 \times 1024$
AdaptiveAvgPool	—	1	$1 \times 1 \times 1024$

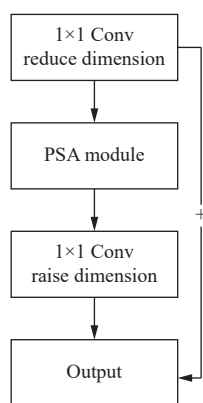


图6 PSA嵌入方式

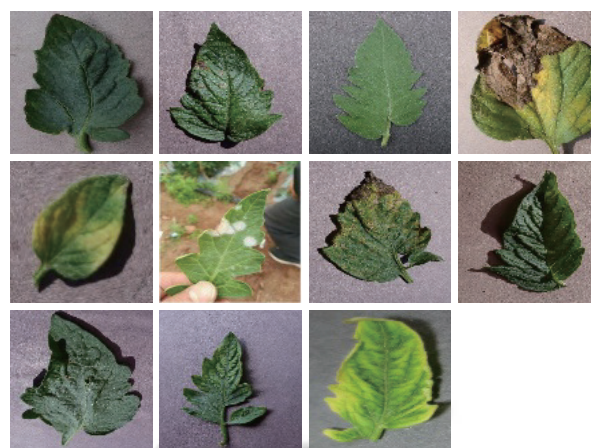


图7 数据集不同病类

2 实验结果及分析

本文对注意力机制的选择进行实验对比,同时也将每一个模块的加入进行实验对照。

2.1 实验环境

实验环境为 Windows 11 64 位系统, 16 GB 内存和 8 GB 显存, 采用了 Jupyter Notebook 和 PyCharm 2021 64 bit, 搭载 AMD Ryzen 7 5800H with Radeon Graphics CPU @3.20 GHz 处理器以及 RTX3060 GPU 显卡. 操作环境为 PyTorch+Python 3.7.

2.2 数据集展示

在该实验中, 使用了 Kaggle 图像分类中的关于番茄叶病的数据集, 共包含了 32 535 张番茄疾病图像, 其中训练集为 25 851 张图片, 测试集为 6 684 张, 并且所有图片尺寸已被固定为 256×256, 共有 11 种标签, 分别为细菌斑病、早疫病、健康型、晚疫病、叶霉菌、白粉病、败血症叶斑病、双斑蜘蛛螨、目标斑病、番茄花叶病毒、番茄黄叶卷曲病毒. 图7为该数据中各种病类的样例, 根据 11 种类别依次进行展示, 在不同环境, 不同疾病下, 番茄叶的形态也不尽相同。

2.3 数据操作

为了减少训练过程中的数据过拟合, 增强模型的鲁棒性, 是模型有更好的识别能力以及训练效果, 在该实验中的数据进行数据增强, 首先是对训练集进行随机翻转, 使得数据扩充为原来的 2 倍, 同时对数据的颜色进行调整改变图片的亮度, 颜色, 饱和度同样扩充数据集. 在训练集中, 因为要是模型识别现实中的实例, 因此不对训练集进行任何操作, 才能更好地体现模型的实用性。

2.4 实验结果分析

将番茄叶病的数据进行数据预处理后, 用该文中提出的网络模型进行训练, 损失函数为 CrossEntropy-Loss 交叉熵损失函数, 交叉熵损失函数为式 (1) 所示:

$$Loss = - \sum_{c=1}^M y_c \log(p_c) \quad (1)$$

其中, M 为分类类别的数量, 在本文中为 10, y_c 为指示变量, 如果该类别和样本的类别相同时为 1, 反之则为 0, p_c 为观测样本属于该类别的预测概率。

模型的准确率认定为如果输出相同则为 1, 反之则为零, 叠加之后与数据总数相除, 准确率为式 (2) 所示:

$$Accuracy = \frac{n_{correct}}{n_{total}} \quad (2)$$

其中, $n_{correct}$ 代表分类正确的记录总数, n_{total} 代表全部测试数据的总数. 准确率很好地表现出模型在验证集中的表现状态。

由于设备的限制, 显存不足, 数据 batch size 设置为 16, 学习率为 0.05, epoch 批次为 30, 优化器选择为 SGD, 对模型进行训练, 训练结果及准确率如图 8 所示。

图 8 中横坐标为训练的数据量, 纵坐标为模型训练的数值, 图 8(a) 为 Loss 函数的变化曲线, 在第 1-10 批次中, Loss 函数下降趋势较为明显, 图 8(b) 为识别率的趋势, 变化较为平稳. 对测试集进行测试验证, 根据混淆矩阵验证模型性能, 如图 9 所示。

根据图 9 混淆矩阵可知, 在所有类别中, 该模型对于 powdery mildew 类别的识别率的错误率较高, 可能 powdery mildew 类别中的图片与健康类别中的图片存在一定的相似性, 在 powdery mildew 类别中的图

片,发生病状的图片没有特别的明显,在局部特征中的区分存在误差,导致模型没有很好地识别,但在其他病例的图片中,模型的识别性能还是很好地体现了出来。

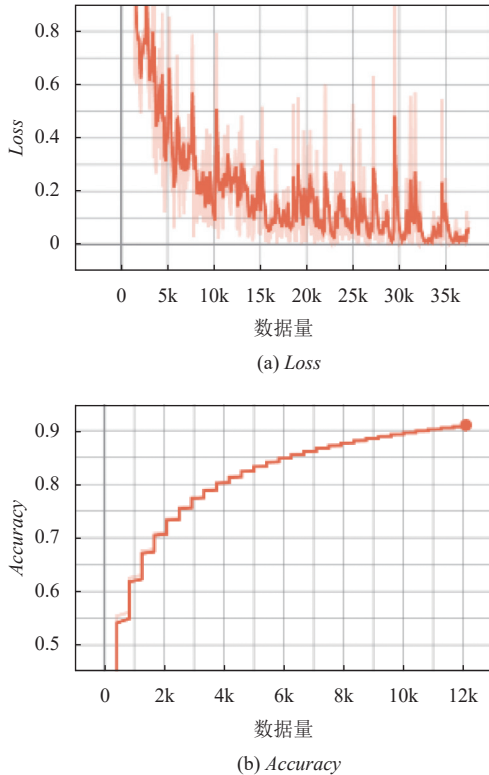


图8 实验结果

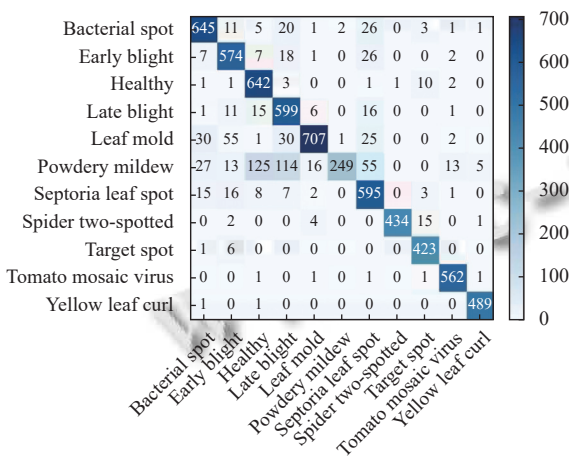


图9 混淆矩阵

由于对该模型准确率的不确定性,以及对它的性能的怀疑,增加模型与原模型 MobileNetV2 网络以及与图像识别中较为准确的 ResNet50 网络进行比较,同时在 ResNet50 网络中加入 PSA 注意力机制,定义为 EPSANet^[15]. 实验环境以及参数与以上模型一致,训练

结果如图 10 所示. 对比结果如表 2 所示.

图 10 横坐标为训练的数据量,纵坐标为模型训练的数值,图 10(a),图 10(b)为 MobileNetV2 网络训练的损失函数和准确率;图 10(c),图 10(d)为 ResNet50 网络模型的损失函数以及准确率,图 10(e),图 10(f)则为 ResNet50 网络嵌入了 PSA 注意力机制模型 (EPSANet) 的损失函数以及准确率,图 10(g),图 10(h)为本文提出所提出的网络 MIPSANet 训练的损失函数和准确率.

由图 10 可知,该模型的参数和计算量大大减小,缩短了训练和测试的时间,体现了该模型的轻量性.大大地提高了 Loss 函数的下降趋势以及识别率,该方法中在很多方面运用了多尺度融合,不仅使用了 Inception 模块的多尺度特征训练,同时也加入了 PSA 多尺度融合注意力机制,大大地提高了该模型的识别率,在目前实验中,该方法获得良好的识别效果.

该实验中,根据表 2 可知,该模型相对于 MobileNetV2 原网络、经典网络 ResNet50 网络和在 ResNet50 网络中嵌入 PSA 注意力机制的 EPSANet50 网络来说,本文提出的网络模型在准确率上都优于这些网络.

2.5 注意力机制的选择

为了确保本文网络中的注意力机制使用得当且具有良好的效果,在同样的位置,将本文网络的 PSA 自注意力机制,分别更换为 SE, CBAM 注意力机制进行对比训练,实验结果如图 11 所示,对比结果如表 3 所示.

在图 11 中横坐标为训练的数据量,纵坐标为模型训练的数值,提供了嵌入 CBAM 和 SE 注意力机制的准确率趋势图,而 MIPSANet 根据图 8(b),进行对照,识别率提高了 1 个百分点,同时都对原网络有所提高.

选取一张数据中的图片,根据 Grad-CAM^[16]方法生成 PSA 注意力机制嵌入与没有嵌入的热力图,观察网络对于图片局部特征的注意点更侧重于的区域,如图 12 所示,图 12(a)为没有嵌入 PSA 的热力图,图 12(b)为嵌入 PSA 的热力图.为了更好地实现该模型的性能,应该将关注点放在叶子的病害区域,加入 PSA 注意力机制后,使得模型更多的关注于叶子的有效特征区域,提高了模型对于有效特征的识别.

2.6 模块有效性验证

为了进一步验证本文提出方法的有效性,本文进行消融实验:(1)首先去除所有添加模块,只对原框架 MobileNetV2 网络进行训练,定义为 Base.(2)在 Base 的基础上,加入 Inception 模块,定义为 Base_Ince.

(3) 在 Base 的基础上引入 PSA 注意力机制, 不加入 Inception 模块, 定义为 Base_PSA. (4) 在 Base 的基础

上加入 Inception 模块以及引入 PSA 注意力机制, 为本文提出网络. 分别对这些进行训练对比, 如图 13, 表 4.

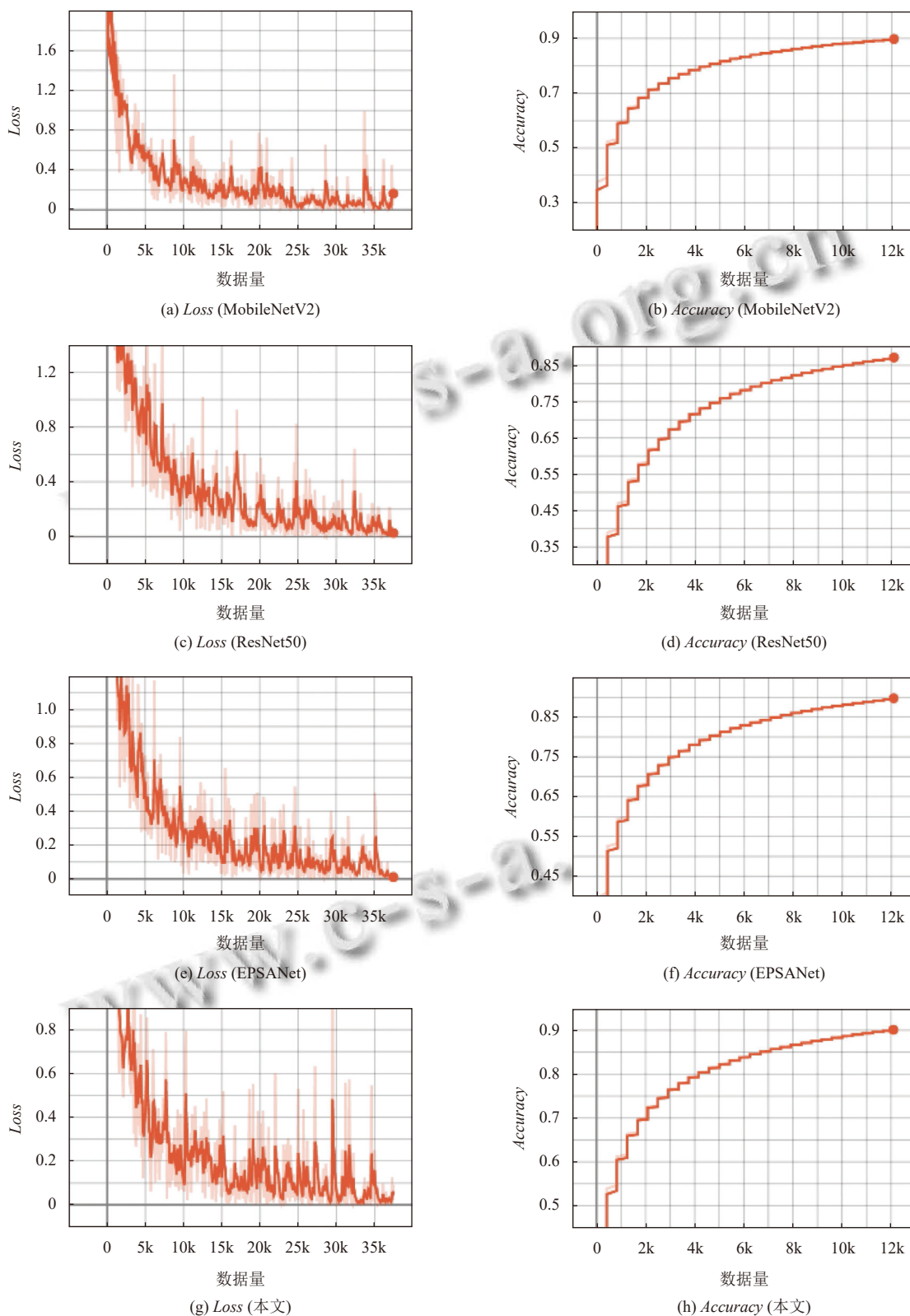


图 10 训练结果图

表 2 各方法对比结果

模型	损失函数	识别率 (%)
ResNet50 ^[6]	0.009 2	87.21
EPSANet ^[14]	0.006 7	89.75
MobileNetV2	0.008 2	89.82
MIPSA Net	0.002 3	91.05

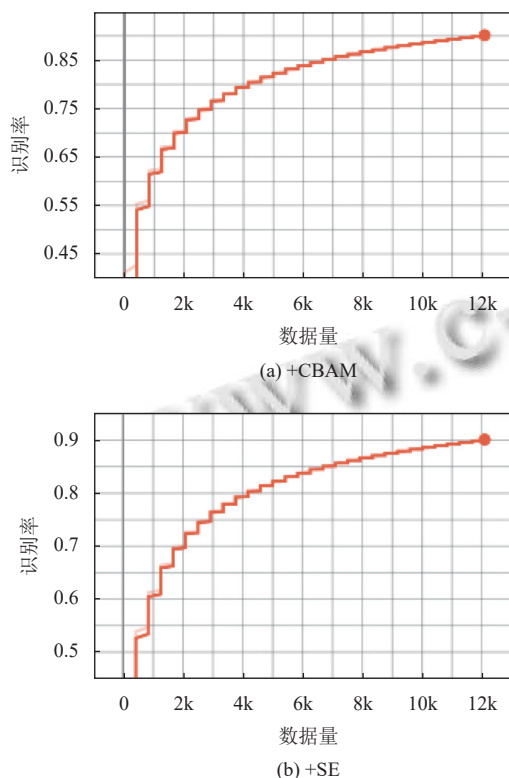
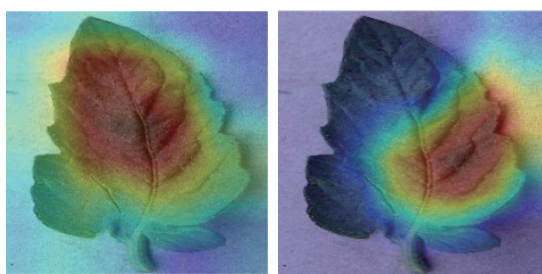


图 11 注意力机制训练结果

表 3 各注意力机制对比结果

注意力机制	识别率 (%)
+SE	90.15
+CBAM	90.13
MIPSA Net	91.05



(a) 没有嵌入 PSA (b) 嵌入 PSA

图 12 热力图分析

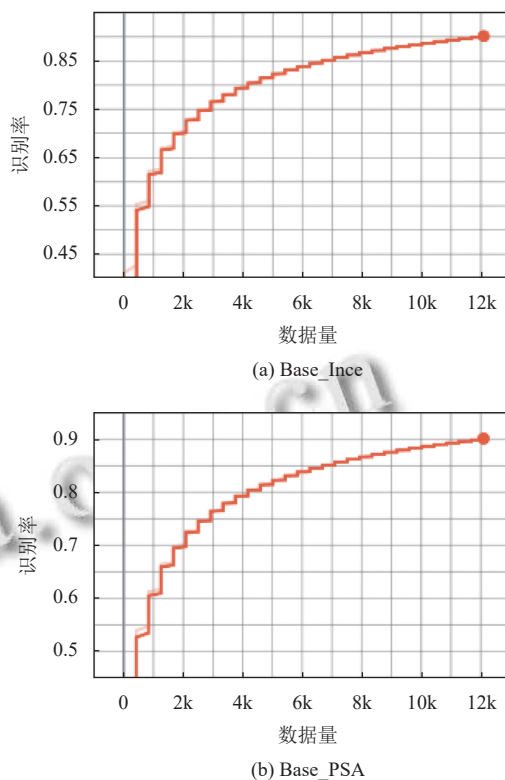


图 13 消融实验准确率结果

表 4 消融实验结果

模块	识别率 (%)
Base	89.82
Base_Ince	90.03
Base_PSA	90.19
MIPSA Net	91.05

图 13 中横坐标为训练的数据量, 纵坐标为模型训练的数值, 其中图 13(a) 为 Base_Ince 模块, 图 13(b) 为 Base_PSA 模块, Base 模块为图 9(b) 中所示. 根据以上数据显示, Inception 模块和 PSA 注意力机制都对模型进行了一定程度上的提高, 说明了该模块的有效性. 通过 Inception 模块, 通过不同尺度的卷积, 对数据进行多尺度特征提取, 嵌入 PSA 注意力机制, 也实现了多通道注意力机制融合, 增加了网络对有效特征区域的识别, 提高了网络的准确率.

3 结束语

本文提出的 MIPSA Net 网络, 在 MobileNetV2 网络框架的基础上, 加入了 Inception 模块以及嵌入了极化自注意力机制 (PSA). 在一定程度上增加了网络的深度和宽度, 结合了多尺度的特征和注意力机制, 增加了

网络的识别率,但是由于数据的单一以及实验环境的性能较低,使得网络模型存在一定的误差.在未来的实验中,有想法加入 FPN 模块,实现多尺度特征提取,更好地针对局部特征的不同,增加网络的识别性能.

参考文献

- 1 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
- 2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc., 2012. 1097–1105.
- 3 Lin M, Chen Q, Yan SC. Network in network. *arXiv:1312.4400*, 2013.
- 4 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.
- 5 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2014. 1–9.
- 6 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2015. 770–778.
- 7 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4510–4520.
- 8 Bahdanau D, Cho KH, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv:1409.0473*, 2014.
- 9 Vaswani A, Shazeer N, ParmarN, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017. 6000–6010.
- 10 Lin ZH, Feng MW, Dos Santos CN, *et al.* A structured self-attentive sentence embedding. *arXiv:1703.03130*, 2017.
- 11 Liu HJ, Liu FQ, Fan XY, *et al.* Polarized self-attention: Towards high-quality pixel-wise regression. *arXiv:2107.00782*, 2021.
- 12 Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco: AAAI, 2017. 4278–4284.
- 13 罗思诗, 李茂军, 陈满. 多尺度融合注意力机制的人脸表情识别网络. *计算机工程与应用*, 2023, 59(1): 199–206. [doi: [10.3778/j.issn.1002-8331.2203-0170](https://doi.org/10.3778/j.issn.1002-8331.2203-0170)]
- 14 魏甫豫, 张振宇, 梁桂珍. 基于卷积神经网络下昆虫种类图像识别应用研究. *河南师范大学学报(自然科学版)*, 2022, 50(6): 96–105.
- 15 Zhang H, Zu KK, Lu J, *et al.* EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. *Proceedings of the 16th Asian Conference on Computer Vision*. Macao: Springer, 2022. 541–557.
- 16 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 618–626.

(校对责编: 孙君艳)