

基于多尺度特征引导的细粒度野生菌图像识别

张志刚, 余鹏飞*, 李海燕, 李红松

云南大学信息学院, 云南 昆明 650500

摘要 鉴于国内野生菌中毒事件频发,拟采用深度学习技术来解决这一社会问题。但由于野生菌图像的类间差异较小,图像背景复杂,细粒度识别精度较低。针对这一问题,提出了一种改进的 ResNeXt50 网络。首先设计了一种多尺度特征引导(MSFG)模块,通过短连接引导网络充分学习和利用低级特征与高级特征;接着采用改进的注意力机制模块来减少网络对复杂背景的学习;最后对模型中的不同层级特征进行融合,利用得到的联合特征进行识别。实验结果表明,所提改进模型在测试集上的准确率可以达到 96.47%,较未改进的 ResNeXt50 网络,在准确率上提升了 2.64 个百分点。对比结果显示,相较 VGG19、DenseNet121、Inception_v3、ResNet50、ShuffleNet_v2 这 5 个网络模型,所提改进模型在准确率上分别提升了 8.10 个百分点、5.13 个百分点、3.24 个百分点、3.30 个百分点、4.25 个百分点。

关键词 图像识别; 细粒度; 多尺度; 特征引导; 注意力机制; 联合特征

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1210016

Fine-Grained Image Recognition of Wild Mushroom Based on Multiscale Feature Guide

Zhang Zhigang, Yu Pengfei*, Li Haiyan, Li Hongsong

School of Information Science & Engineering, Yunnan University, Kunming 650500, Yunnan, China

Abstract Deep learning technology is proposed to solve the social problem of the frequent occurrences of wild mushroom poisoning in China. However, due to the small difference between classes and complex image backgrounds, fine-grained recognition accuracy is low. To solve this problem, this paper proposes an improved ResNeXt50 network. First, a multiscale feature guide (MSFG) module is designed, which guides the network to learn and use low and high-level features fully through short connections. Then, the improved attention mechanism module is used to reduce the network's learning for complex backgrounds. Finally, the different hierarchical features in the model are fused, and the obtained joint features are used for recognition. Experimental results show that the accuracy of the proposed network on the test set can reach 96.47%, which is 2.64 percentage points higher than the unimproved ResNeXt50 network. Comparison results show that the accuracy of the improved network model is 8.10 percentage points, 5.13 percentage points, 3.24 percentage points, 3.30 percentage points, and 4.25 percentage points better than VGG19, DenseNet121, Inception_v3, ResNet50, and ShuffleNet_v2, respectively.

Key words image recognition; fine-grained; multi-scale; feature guide; attention mechanism; joint feature

1 引言

野生菌因味道鲜美、营养丰富而深受人们喜

爱,但是由于许多有毒野生菌和可食用菌在外观上极为相似,对于普通人群而言,不具备鉴别野生菌是否有毒的相关知识,容易误食有毒野生菌而中

收稿日期: 2021-07-02; 修回日期: 2021-07-28; 录用日期: 2021-08-17

基金项目: 国家自然科学基金(62066046)

通信作者: *pfyu@ynu.edu.cn

毒。据文献[1]统计显示,仅2011年到2017年,云南省共报告野生菌中毒事件1789起,中毒7966人,死亡219人。因此利用深度学习技术对野生菌进行识别,帮助人们鉴别野生菌的种类,将对改善野生菌中毒问题起到重要的作用。

传统的野生菌识别方法大致分为动物实验检验法、化学检测法、真菌分类学鉴定法和形态特征识别方法等^[2]。文献[3-6]分别使用动物实验检验法、化学检测法、真菌分类学鉴定法和形态特征识别方法对毒性进行检测,但这些方法实验器材难以获取,且耗时较长,实用性较差,不易推广。文献[7-9]则分别基于支持向量机、贝叶斯网络以及stack算法进行野生菌的识别,但这些方法依赖人们手工标注特征,操作繁琐,识别率不高,不利于推广和使用。随着深度学习技术的发展,基于卷积神经的图像识别方法成为图像识别的重要途径,被一些学者应用到了野生菌识别研究中。例如文献[10]基于Xception与ResNet50网络进行野生菌的识别,实验数据集是国外的kaggle数据集,此数据集仅有9类野生菌,类别较少;文献[11-12]则基于互联网采集的野生菌图片数据集进行野生菌的识别。但是由于这两种数据集都是在互联网上进行采集的,样本数量少,图片清晰度不足,部分图片还存在水印,容易丢失物体的一些细节信息,不能较好地保留图像信息,不利于增强模型的泛化能力。同时,不同于粗粒度图像各类别之间的差异,例如人和树的差异,野生菌作为一个大类,其子类繁多,类间差异相对较小,因此野生菌的识别属于细粒度识别问题,跟普通的识别问题不同。细粒度识别^[13-14]在一个种类的基础上进行更精细的分类,因为类间差异较小,所以识别难度较大。例如,袁培森等^[15]从细粒度图像分类角度出发,基于Inception-ResNet-v2模型对野生菌进行识别,但没有考虑到图片复杂背景的影响,识别精度相对较低。

上述基于深度学习的野生菌识别方法存在的问题主要包括几个方面。一、现有的野生菌识别方法受复杂背景的干扰较大。二、野生菌图像的低级特征往往包括很多细节信息,例如纹理等,这些低级细节特征对识别外形较为相似的野生菌细粒度具有重要的作用,高级特征则包括较多的语义信息,能够帮助定位显著性物体所在位置,有利于减少背景干扰。现有方法并没有充分利用网络提取过程中的低级特征和高级特征,也没有对低级特征

和高级特征进行融合,没有利用多尺度特征来提升细粒度识别精度。三、文献[10]使用的数据集来源于国外,跟国内野生菌种类存在一定差异,在国内适用性不强。文献[11-12]使用的数据集是通过互联网采集的,图片质量参差不齐,图片数量少,分别为1675张和3500张,不利于推广使用。

本文提出了一种改进的ResNeXt50模型。该模型首先利用提出的多尺度特征引导(MSFG)结构进行特征引导,用多尺度的低级特征引导网络学习细节特征,再使用多尺度的高级语义特征来引导网络关注图像中显著性区域所在位置。其次,在网络结构中添加改进的Inverse Document Frequency-SE(IDF-SE)注意力机制,再构建特征金字塔结构进行特征融合,利用联合特征进行识别。实验识别率和热力图结果表明,改进的ResNeXt50能够有效地提升细粒度识别精度,减少网络对复杂背景的学习,识别率明显优于现有的传统模型。

2 改进的ResNeXt50模型

2.1 ResNeXt50

由于野生菌的外形特征丰富多样,并且部分野生菌在外形上存在极大的相似性,为了提升细粒度图像的识别精度,这就要求网络模型能够学习到更丰富的特征。ResNeXt50综合了ResNet^[16]和Inception^[17]网络的优点,在不增加网络参数量的前提下,提升了网络的分类精度。ResNeXt50采用了拆分、变换和融合的思想,实现分组卷积,将ResNet中较大的卷积核替换成一系列卷积核尺寸更小、数量更多的平行拓扑结构,对输入同时进行多组卷积来提升识别精度。ResNeXt的残差块结构如图1所示。

ResNeXt50的输入 \mathbf{X} 通过拆分,分为32个部分,可以表示为 $\mathbf{X} = [x_1, x_2, \dots, x_{32}]$ 。传统的ResNet残差结构没有并行结构,只有一组包含三个卷积层的通路,正是并行结构的增加让网络能够学习到更加丰富的特征。接着对拆分后的数据进行变换和融合,变换和融合过程为

$$\mathbf{Y} = \mathbf{X} + \sum_{c=1}^C T_c(\mathbf{X}), \quad (1)$$

式中: T 是指一系列变换操作的集合,如图1任意一条通路所示,变换包括了三个卷积层,卷积核大小分别为 1×1 、 3×3 、 1×1 ; C 是分组卷积中的分组数量; \mathbf{Y} 表示分组卷积融合得到的输出。分组卷积得

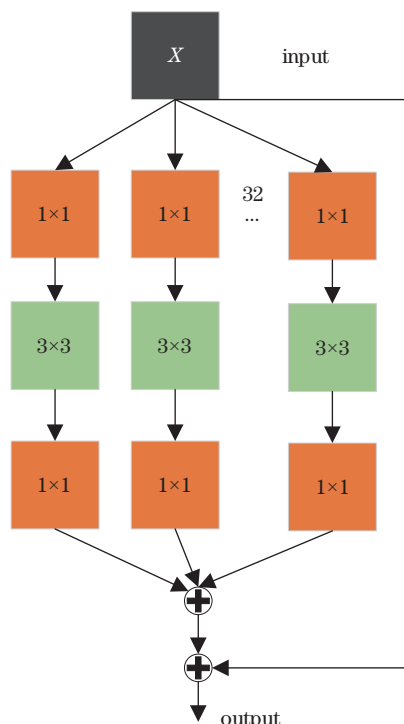


图 1 ResNeXt 残差块结构

Fig. 1 Structure of the ResNeXt residual block

到的结果融合之后再加上残差块中的短连接输入 X , 就得到了残差块的最终输出。ResNeXt50 在网络结构的搭建上继续沿用 ResNet 的重复层结构, 这使得 ResNeXt 网络只需要堆叠残差块, 搭建方便。在性能方面, ResNeXt50 和 ResNet50 的参数量分别是 25.0×10^6 和 25.5×10^6 , 可以发现 ResNeXt50 模型的参数量比 ResNet50 少, 但 ResNeXt50 在 ImageNet 数据集上的精度却比 ResNet50 高, 并且 ResNeXt50 的参数量比 ResNeXt101 少, 识别速度更快, 更适合应用到生产环境, 故使用 ResNeXt50 作为基础网络对野生菌进行识别。

2.2 MSFG

MSFG 为所提的多尺度特征引导模块。特征引导在图像显著性区域检测方面有重要的应用^[18-19], 能够有效地提升网络性能。MSFG 通过构建多尺度特征, 利用短连接来引导网络对重要特征的学习和利用, MSFG 结构如图 2 所示。

MSFG 结构首先使用指数加权激活的 SoftPool 池化^[20]对特征图进行处理, SoftPool 具有更精细的下采样, 能够提升识别精度, 本结构采用大小为 $1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8$ 的池化核来获取不同尺度感受野的特征图, 增加网络对全局信息的获取。不同于传统的最大池化和平均池化, SoftPool 利用 softmax 函数对池化区域 R 内的值进行激活, 得到池化区域里的每一个值 p_m 对应的激活权重 w_m , 这能够很好地保留特征图中的重要信息, 有利于提升网络的识别率。softmax 函数为

$$w_m = \frac{\exp(p_m)}{\sum_{j \in R} \exp(p_j)} \quad (2)$$

再将激活权重 w_m 和池化区域里对应的值 p_m 相乘, 对得到的结果进行求和, 将其作为这一个池化区域对应的池化值。每一个池化区域对应的池化结果 s 表示为

$$s = \sum_{m \in R} w_m \times p_m \quad (3)$$

经过 SoftPool 池化处理之后, 会得到具有不同感受野的多尺度特征图, 首先利用最邻近插值对多尺度特征图进行上采样, 让四个特征图的尺寸保持一致; 接着对四个特征图进行拼接, 再利用 1×1 卷积进行通道间的缩减和信息的交融, 减少混叠效应; 然后进行批量归一化, 增加网络收敛速度, 抑制梯度爆炸, 利用 Relu 激活函数进行激活; 得到最终的

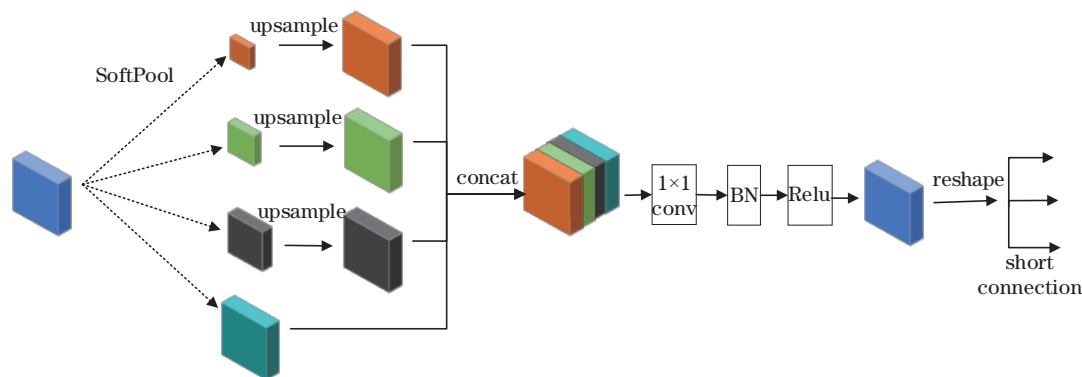


图 2 多尺度特征引导结构

Fig. 2 Multi-scale feature guiding structure

多尺度特征图之后,再对这一特征图进行尺寸的变化,利用短连接对变化的结果和其他特征图进行融合,引导网络学习关键特征。

2.3 IDF-SE 注意力机制

注意力机制主要是模拟人眼的视觉注意力机

制,在训练过程中关注图像中的显著性区域,以便更好地学习图像中的重要特征,其被广泛应用于诸多领域^[21-22]。针对野生菌复杂背景会影响识别精度,提出了一种 IDF-SE 注意力机制,来减少对复杂背景的学习。IDF-SE 结构如图 3 所示。

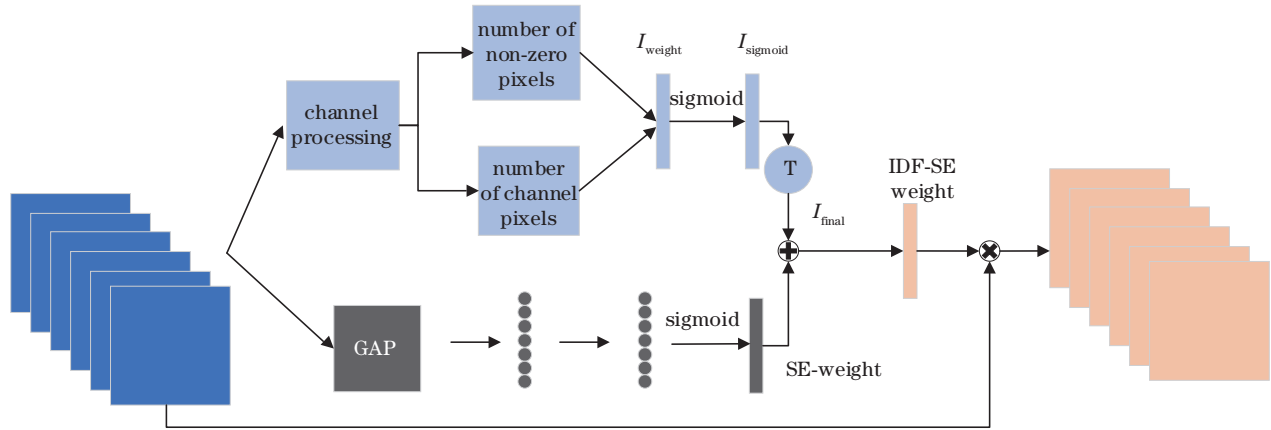


图 3 IDF-SE 结构

Fig. 3 Structure of the IDF-SE

IDF-SE 注意力机制由两部分组成,一部分是改进的 SE 网络,为了防止 SE 网络^[23]中多层感知机降维产生的信息丢失,本文不再进行降维,而是采用两个神经元数量相同的全连接层进行非线性关系建模;另一部分是 IDF 辅助权重模块,用来对 SE 注意力机制得到的权重进行补充。考虑到一些通道特征图中非零像素的数量少,即白色区域少,黑色区域较多,而往往这类通道特征图可能具有关键的特征,但随着卷积神经网络的不断加深,卷积和池化过程会导致这类通道特征信息丢失,为了在一定程度上减少这类通道特征图的信息丢失,使其能够尽可能多地被网络所关注,采用逆文档频率 (IDF) 算法思想,增强白色区域的通道特征权重。IDF 常被用作信息检索的加权,常用于自然语言处理^[24],IDF 的主要思想是包含词条的文档越少,逆向文本指数越高。于是利用 IDF 的思想,将其应用于图像处理,白色区域越少的通道图,权重指数越高,再对这一权重值增加一个温度系数,控制该加权系数的影响程度。具体算法思路如下:首先计算每一个通道特征图中白色区域占比 D_i ,其次将所有通道的白色区域占比相加,再将总和除以每一个通道特征图的白色区域占比,最后将结果取对数得到加权系数 I_{weight} ,再利用 sigmoid 函数进行激活得到 $I_{sigmoid}$ 。为了控制 $I_{sigmoid}$ 对网络的影响程度,利用温度系数 T 来控制, T 的取值范围为 0 到 1,将 $I_{sigmoid}$ 与 T 相乘得

到 I_{final} ,处理过程的公式为

$$D_i = N_i / (W \times H), i \in 0, 1, \dots, k, \quad (4)$$

$$I_{weight} = \log \frac{\sum_{i=0}^k D_i}{D_i}, \quad (5)$$

$$I_{sigmoid} = \sigma(I_{weight}), \quad (6)$$

$$I_{final} = T \times I_{sigmoid}, T \in [0, 1], \quad (7)$$

式中: N_i 表示每一个通道特征图中非零像素的个数; W 和 H 表示每一个通道特征图的长和宽; k 表示通道数的大小; σ 表示 sigmoid 函数。最后将 IDF 模块的 I_{final} 系数和改进之后的 SE 网络得到的权重相加,得到最终的 IDF-SE 注意力机制权重。再利用该权重对特征图进行加权,增强网络对有用信息的提取能力,减少背景信息的干扰。

2.4 模型结构

基于 ResNeXt50 主干网络进行改进,改进之后的整体网络结构如图 4 所示。

首先对输入的野生菌图片进行预处理操作,将图片的尺寸调整为 224×224 ,进行标准化处理,并转化为张量,依次通过 ResNeXt50 网络中卷积核大小为 7×7 的卷积层、4 个 Layer 模块来提取特征,4 个 Layer 模块包含的残差块数量分别是 3, 4, 6, 3。野生菌图片经过 ResNeXt50 网络的第一层进行特征提取,会得到具有较多底层信息的特征图,例如纹理、颜色等,考虑到一些底层特征对野生菌的识

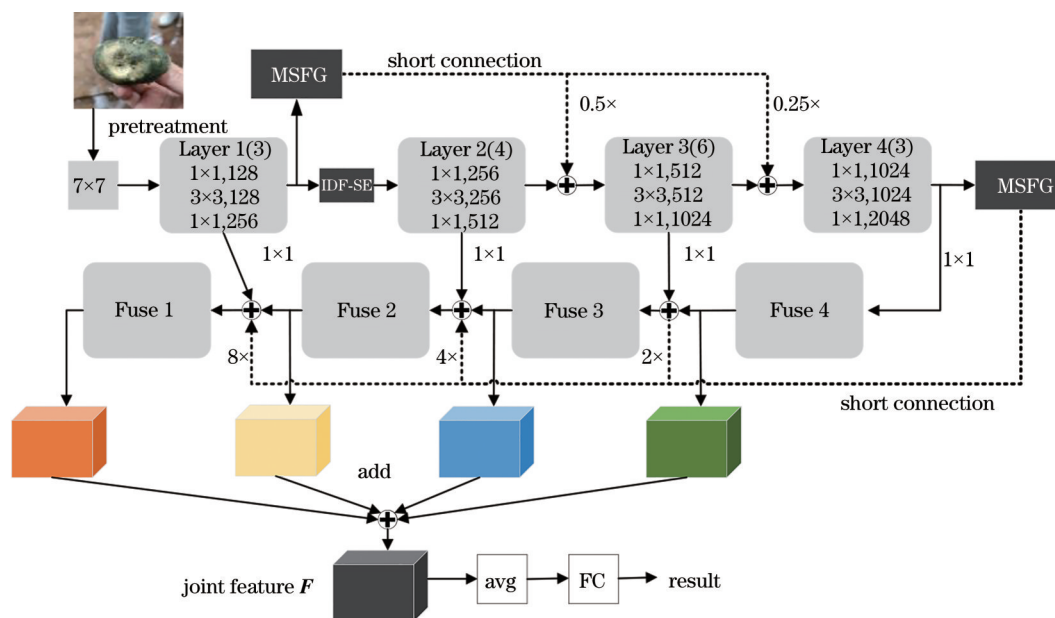


图 4 改进之后的整体网络结构

Fig. 4 Overall network structure after the improvement

别具有重要的作用,模型在 Layer 1 得到的底层特征图后面添加 MSFG 结构,利用多感受野来获取更多的全局信息;再将 MSFG 得到的特征图分别作为输入送到 ResNeXt50 网络的 Layer 3 和 Layer 4 中,引导模型学习底层特征;接着,在 Layer 1 后面添加 IDF-SE 模块,让网络更加注重有效信息的获取,尽可能多地保留特征信息;其次,在 ResNeXt50 网络的基础上构建特征金字塔结构,图中的特征图 Fuse 1、Fuse 2、Fuse 3、Fuse 4 就是特征金字塔的自顶向下结构。由于卷积网络的高级语义特征可以定位野生菌主体所在位置,于是还在 ResNeXt50 网络模型的 Layer 4 之后继续添加 MSFG 结构,获取具有不同感受野的多尺度特征图,以便更好地定位野生菌主体所在位置;再利用短连接将 MSFG 输出的特征图分别添加到 Fuse 1、Fuse 2 和 Fuse 3 中,这样做的目的就是为将 MSFG 处理得到的多尺度高级语义特征融合到其他层的特征图中,以此来定位野生菌主体的位置,引导网络学习关键信息,减少背景信息的干扰,实现特征之间的高度融合;然后对特征金字塔融合之后得到的四个特征图进行相加,得到通道数为 256 的联合特征 F ,将其依次通过最大平均池化、全连接层;最后利用 softmax 函数进行概率归一化,采用梯度下降法对网络的参数进行学习和调整,不断减小预测值和标签之间的差距,来训练最终的模型。

3 实验及分析

3.1 实验环境及参数设置

实验环境基于操作系统为 Ubuntu20.04 的服务器,该服务器的 CPU 为 i9 7900, GPU 为两块 Geforce GTX 3090。深度学习框架选择 pytorch 1.6,编程环境选择 python 3.8,CUDA 驱动版本选择 11.1。损失函数选择交叉熵损失函数,实验初始学习率设置为 0.01,momentum 设置为 0.9,在训练过程中选择线性学习率衰减法调整学习率,每训练 10 次,学习率有 0.1 的缩减。在梯度下降过程中,使用 SGD 算法进行优化。实验中 batchsize 的大小设置为 64,训练的 epoch 个数设置为 200,IDF-SE 中的温度系数 T 设置为 0.3。

3.2 数据集构建

在国家自然科学基金支撑下,通过实地拍摄构建了一个高质量、多角度且普适性强的野生菌数据集,有效解决了利用互联网上的图片来构建数据集时图片清晰度不高,图片质量差,野生菌细节丢失,不利于细粒度识别等问题。本文采集的数据集共有常见菌类图片 6881 张,图片分辨率为 3000×4000 ,图片清晰度较高。数据集一共有 21 种常见的野生菌,每一个种类的图片数量为 200~400 张。这 21 种野生菌分别是黄牛肝、黑牛肝、红牛肝、鸡油菌、谷熟菌、灵芝、奶浆菌、皮条菌、荞面菌、青头菌、珊瑚菌、松露、松茸、铜绿菌、虎掌菌、平菇、

竹荪、虫草花、鸡枞菌、干巴菌及羊肚菌。同时,此数据集不仅是由单一视角的野生菌图片构成的,而且是一个多视角的野生菌数据集,每一个种类的野生菌图片都是从多个角度进行采集的,例如从上

到下、从左到右、从下而上等角度拍摄,从而采集到野生菌的菌盖、菌柄和菌褶等不同部位的图片,能够实现从多角度识别野生菌种类。该数据集部分种类的多视角图像如图 5 所示。



图 5 部分数据集图片示例

Fig. 5 Examples of some pictures of the dataset

图 5 是部分野生菌的三个视角图,可以发现,不同野生菌种类之间的差异性较小。同时该数据集中有外形相似的无毒野生菌和有毒野生菌,图 5 中的前三种野生菌分别是黑牛肝菌、红牛肝菌以及黄牛肝菌,其中红牛肝菌和黄牛肝菌便是常见的有毒野生菌,而黑牛肝菌则是无毒的。

为了提升模型的鲁棒性和泛化能力,采用数据增强策略对数据集进行扩充,这种方法在 AlexNet^[25]中就已经被应用,还能有效地解决过拟合现象。主要采取的数据增强方法有旋转、裁剪、放大、镜像对称、添加白噪声,其中旋转角度使用了 15°, 30°, 45°, 90°, 150°, 扩充后数据集一共有 13762 张图片,本文的数据集按照 8:1:1 的比例依次划分为训练集、验证集、测试集。训练集用来训练模型,验证集用来选择模型,测试集用来检测模型的优劣及泛化能力。

3.3 实验结果分析

改进后的模型在训练集和验证集上训练了 200 个 epoch,模型训练的损失曲线如图 6 所示。

模型在训练集和验证集上的损失收敛效果好,整个训练过程呈现下降趋势,没有出现较大的振荡。选择在验证集上准确率最高的模型作为最终模型,再对这一模型在测试集上进行测试,测试的准确率为 96.47%。实验结果的混淆矩阵如图 7 所示。

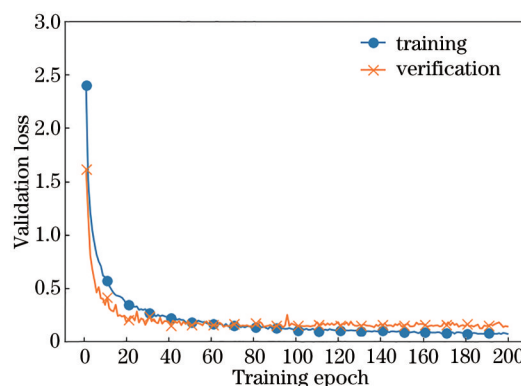


图 6 训练和验证的损失变化

Fig. 6 Loss value changes for training and verification

通过混淆矩阵可以发现,每个种类的大部分野生菌图片都能被正确识别,只有极少部分的野生菌图片被误识别。其中,所提模型对虫草花、谷熟菌、红牛肝菌、平菇、荞面菌、皮条菌及铜绿菌的所有测试集样本具有较好的细粒度识别率。

为了进一步验证所提方法的有效性,在测试集中又单独选择了 8 种(如图 5 所示)较难区分的图片测试集来进行进一步验证,选择精确率和召回率作为模型评价指标,各类别实验结果如表 1 所示。

通过观察表 1 可以发现,所提模型对以上种类的野生菌的精确率都大于或等于 95.5%,召回率都大于或等于 95.2%,其中对红牛肝菌的精确率和召回率最高,分别是 97.6% 和 98.3%,对与红牛肝菌

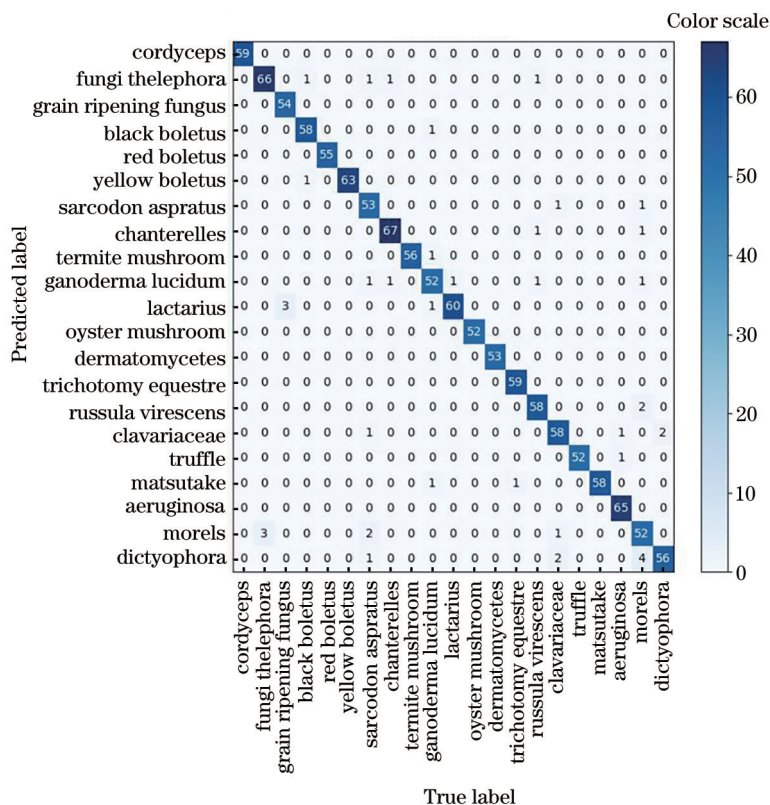


图 7 混淆矩阵

Fig. 7 Confusion matrix

表 1 所提模型对不同野生菌的识别结果

Table 1 Recognition results of the proposed model for different mushroom

Specie	Precision /%	Recall /%
Black boletus	97.3	96.7
Red boletus	97.6	98.3
Yellow boletus	97.4	96.8
Aeruginosa	96.5	97.1
Lactarius	97.2	96.3
Chanterelles	96.6	98.3
Grain ripening fungus	95.5	95.2
Trichotomy equestre	96.4	95.7

相似的黑牛肝菌和黄牛肝菌的精确率也都在 97% 以上。实验结果表明,所提模型对相似性较强的野生菌种类有较高的识别率。

为了更加直观地对比模型改进前后的效果和改进方法的可信度,使用卷积神经网络可视化方法 Grad-CAM^[26]对改进模型进行热力图的可视化。通过观察热力图,可以判断哪部分图片区域有利于最终的分类。可视化结果如图 8 所示。

图 8 是基于未改进的 ResNeXt50 和改进的 ResNeXt50 模型对随机选择的三张背景复杂图片的

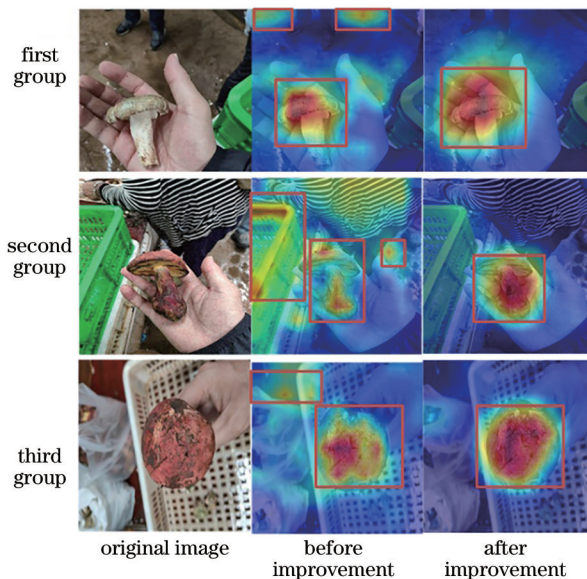


图 8 热力图对比

Fig. 8 Comparison of heat maps

热力图结果。可以发现,模型改进之前的热力图关注区域较多,即图中矩形框中的区域,包括复杂背景和部分野生菌区域;而模型改进之后得到的热力图效果有了明显的改善,网络减少了对背景的关注,能够准确定位到野生菌所在位置。通过以上三

组实验结果对比可以发现,改进之后的模型能够更好地找到图像中关键特征的位置,减少对背景区域的关注,降低背景干扰,提取关键特征,从而减少复杂背景信息对识别精度的影响。

为了验证改进之后的网络是否优于其他网络,将其与 VGG19、DenseNet121、Inception_v3、ResNet50、ResNeXt50 以及 ShuffleNet_v2 这 6 个网络进行对比,同时对这 6 个网络的模型参数进行调整和优化,每个模型的训练次数为 200 个 epoch。各模型在验证集上的损失曲线如图 9 所示。

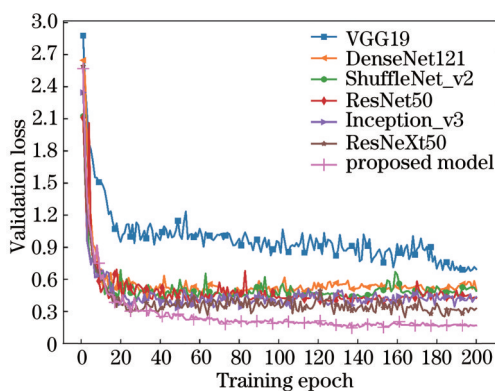


图 9 各模型损失曲线对比

Fig. 9 Comparison of loss curves of various models

由图 9 可以直观地发现:所提模型在验证集上的损失最小; DenseNet121、Inception_V3、ResNet50、ShuffleNet_v2 和 ResNeXt50 损失值较低,但这 5 个网络模型的损失值都高于所提模型,效果最差的是 VGG19,损失值最大。实验结果表明,所提模型在验证集上的效果优于 VGG19、DenseNet121、Inception_v3、ResNet50、ResNeXt50 和 ShuffleNet_v2。

为了更加准确地评价所提模型,分别利用准确

率、精确率、召回率以及 F1 值来进一步对模型性能进行评估,由于本实验是多分类问题,采用精确率、召回率以及 F1 的平均值来进行对比。不同模型在测试集上的各指标如表 2 所示。

表 2 不同模型在数据集上的识别结果

Table 2 Recognition results of different models on the data set

Model	Accuracy /%	Precision /%	Recall /%	F1 /%
VGG19	88.37	88.99	89.37	89.18
DenseNet121	91.34	91.26	90.89	91.07
Inception_V3	93.23	93.35	93.44	93.39
ResNet50	93.17	93.13	93.41	93.27
ShuffleNet_v2	92.22	88.89	92.85	90.83
ResNeXt50	93.83	93.97	93.94	93.95
Proposed model	96.47	96.58	96.56	96.57

由表 2 可以看出,在所使用的模型中,所提模型在该数据集上具有良好的表现,较改进之前的模型 ResNeXt50 相比,所提改进的 ResNeXt50 在准确率、宏精度、宏召回率和宏 F1 上分别提升了 2.64 个百分点、2.61 个百分点、2.62 个百分点和 2.62 个百分点。实验结果表明,改进之后的模型优于原模型 ResNeXt50。另外,与 VGG19、DenseNet121、Inception_v3、ResNet50 和 ShuffleNet_v2 相比,所提改进 ResNeXt50 的准确率分别提升了 8.10 个百分点、5.13 个百分点、3.24 个百分点、3.30 个百分点和 4.25 个百分点。综上所述,针对野生菌细粒度识别问题,所提改进 ResNeXt50 在这些模型中识别效果最佳。

为了更好地对比改进过程中不同模块对模型性能的增益,对不同模块进行了实验对比,实验结果如表 3 所示。

表 3 不同模块对网络模型的增益

Table 3 Gain of different modules to the network model

Model	Accuracy /%	Precision /%	Recall /%	F1 /%
ResNeXt50	93.83	93.97	93.94	93.95
Backbone+ SE	94.02	93.98	92.85	93.41
Backbone+ IDF-SE	94.65	94.61	94.58	94.59
Backbone+ MSFG	95.91	96.02	94.01	94.99
Backbone+ MSFG+ IDF-SE	96.47	96.58	96.56	96.57

其中 Backbone 是指基于 ResNeXt50 网络构建的特征金字塔网络。通过观察表 3 可以发现:在准确率指标上,模型添加 IDF-SE 注意力机制比模型添加 SE 注意力机制提升 0.63 个百分点,说明所提

的 IDF-SE 注意力机制能够有效地提升识别精度;在主干网络上添加 MSFG 模块能够让准确率提升 2.08 个百分点,实验结果说明,所提 MSFG 结构能够有效地提升细粒度识别精度。

4 结 论

近年来国内野生菌中毒事件频发,为了使用深度学习算法帮助人们识别野生菌种类,针对传统网络模型对背景复杂、外形差异较小的野生菌图片的细粒度识别率低的问题,提出了一种改进的 ResNeXt50。改进的 ResNeXt50 利用 MSFG 模块进行多尺度特征引导,接着使用 IDF-SE 注意力机制来增强网络对有用特征的学习,再构建特征金字塔进行不同层级特征的融合,最后使用联合特征来进行识别。从实验准确率结果可以看出,所提改进 ResNeXt50 能够在一定程度上提升细粒度图像识别精度,较 ResNeXt50 在准确率上提升了 2.64 个百分点,并且优于其他 5 种模型。从热力图结果可以看出,所提改进 ResNeXt50 明显地减少了背景信息的干扰,让网络重点关注野生菌主体所在位置。今后,课题组将继续拍摄更多的野生菌图像来扩充野生菌图像数据库,不断完善模型结构,优化模型参数,进一步提升网络模型的精度,以便在云服务器上部署深度学习模型,开发具有野生菌识别功能的移动端 App 应用,为普通人群鉴定野生菌的种类提供有力的支持,从而在一定程度上减少野生菌中毒事件的发生。

参 考 文 献

- [1] 万蓉, 刘志涛, 万青青, 等. 2011—2017 年云南省野生菌中毒情况分析[J]. 卫生软科学, 2019, 33(10): 84-86, 97.
Wan R, Liu Z T, Wan Q Q, et al. Analysis on wild mushroom poisoning in Yunnan Province from 2011 to 2017[J]. Soft Science of Health, 2019, 33(10): 84-86, 97.
- [2] 李林静, 李高阳, 谢秋涛. 毒蘑菇毒素的分类与识别研究进展[J]. 中国食品卫生杂志, 2013, 25(4): 383-387.
Li L J, Li G Y, Xie Q T. Research progress on poisonous mushroom toxins classification and recognition [J]. Chinese Journal of Food Hygiene, 2013, 25(4): 383-387.
- [3] Grandstaff D E, Terry D O, Jr. Rare earth element composition of Paleogene vertebrate fossils from Toadstool Geologic Park, Nebraska, USA[J]. Applied Geochemistry, 2009, 24(4): 733-745.
- [4] Zhao J, Cao M, Zhang J, et al. Pathological effects of the mushroom toxin α -amanitin on BALB/c mice [J]. Peptides, 2006, 27(12): 3047-3052.
- [5] 桂明英, 何容, 郭永红, 等. 基于形态特征和 ITS 序列对新疆芦苇根蘑菇的分类鉴定[J]. 食用菌, 2014, 36(4): 14-16.
Gui M Y, He R, Guo Y H, et al. Based on morphological characteristics and ITS classification and identification of Xinjiang Reed Root mushrooms by sequence[J]. Edible Fungi, 2014, 36(4): 14-16.
- [6] 曾令奎. 毒蘑菇中毒的识别和预防[J]. 中国林副特产, 2008(1): 97-98.
Zeng L K. Recognition and prevention of poisonous mushroom poisoning[J]. Forest by-Product and Speciality in China, 2008(1): 97-98.
- [7] 刘洋, 王涛, 左月明. 基于支持向量机的野生蘑菇近红外识别模型[J]. 食品与机械, 2016, 32(4): 92-94, 112.
Liu Y, Wang T, Zuo Y M. Study on recognition model for wild mushroom based on support vector machine of near infrared spectral diagnosis[J]. Food & Machinery, 2016, 32(4): 92-94, 112.
- [8] 刘斌, 张振东, 张婷婷. 基于贝叶斯分类的毒蘑菇识别[J]. 软件导刊, 2015, 14(11): 60-62.
Liu B, Zhang Z D, Zhang T T. Recognition of poisonous mushrooms based on Bayesian classification [J]. Software Guide, 2015, 14(11): 60-62.
- [9] 闫志峰. 基于 Stacking 算法的多分类器融合在毒蘑菇识别中的应用[D]. 太谷: 山西农业大学, 2019.
Yan Z F. Application of multi-classifier fusion based on Stacking algorithm in identification of poisonous mushrooms[D]. Taigu: Shanxi Agricultural University, 2019.
- [10] 沈若兰, 黄英来, 温馨, 等. 基于 Xception 与 ResNet50 模型的蘑菇分类方法[J]. 黑河学院学报, 2020, 11(7): 181-184.
Shen R L, Huang Y L, Wen X, et al. Mushroom classification based on Xception and ResNet50 models[J]. Journal of Heihe University, 2020, 11(7): 181-184.
- [11] 肖杰文, 赵铖博, 李欣洁, 等. 基于深度学习的蘑菇图像分类研究[J]. 软件工程, 2020, 23(7): 21-26.
Xiao J W, Zhao C B, Li X J, et al. Research on mushroom image classification based on deep learning [J]. Software Engineering, 2020, 23(7): 21-26.
- [12] 陈德刚, 艾孜尔古丽, 尹鹏博, 等. 基于改进 Xception 迁移学习的野生菌种类识别研究[J]. 激光与光电子学进展, 2021, 58(8): 0810023.
Chen D G, Aizierguli, Yin P B, et al. Research on identification of wild mushroom species based on

- improved Xception transfer learning[J]. *Laser & Optoelectronics Progress*, 2021, 58(8): 0810023.
- [13] 张洁, 赵红东, 李宇海, 等. 复杂背景下车型识别分类器[J]. *激光与光电子学进展*, 2019, 56(4): 041501. Zhang J, Zhao H D, Li Y H, et al. Classifier for recognition of fine-grained vehicle models under complex background[J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 041501.
- [14] 冯春, 赵南京, 殷高方, 等. 基于光谱相似性分析的水体致病菌种类识别方法[J]. *光学学报*, 2020, 40(3): 0330002. Feng C, Zhao N J, Yin G F, et al. Recognition of waterborne pathogens based on spectral similarity analysis[J]. *Acta Optica Sinica*, 2020, 40(3): 0330002.
- [15] 袁培森, 申成吉, 徐焕良. 基于迁移学习和双线性 CNN 的细粒度菌菇表型识别[J]. *农业机械学报*, 2021, 52(7): 151-158. Yuan P S, Shen C J, Xu H L. Fine-grained mushroom phenotype recognition based on transfer learning and bilinear CNN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(7): 151-158.
- [16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [17] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 1552-1560.
- [18] Liu J J, Hou Q B, Cheng M M, et al. A simple pooling-based design for real-time salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3912-3921.
- [19] Liu Y, Zhang X Y, Bian J W, et al. SAMNet: stereoscopically attentive multi-scale network for lightweight salient object detection[J]. *IEEE Transactions on Image Processing*, 2021, 30: 3804-3814.
- [20] Stergiou A, Poppe R, Kalliatakis G. Refining activation downsampling with SoftPool[EB/OL]. (2021-01-02) [2021-04-02]. <https://arxiv.org/abs/2101.00440>.
- [21] 李宇昕, 杨帆, 刘钊, 等. 基于改进残差网络的道口车辆分类方法[J]. *激光与光电子学进展*, 2021, 58(4): 0415009. Li Y X, Yang F, Liu Z, et al. Classification method of crossing vehicle based on improved residual network[J]. *Laser & Optoelectronics Progress*, 2021, 58(4): 0415009.
- [22] 徐志京, 王东. 基于双路循环生成对抗网络的多姿态人脸识别方法[J]. *光学学报*, 2020, 40(19): 1910002. Xu Z J, Wang D. Multi-pose face recognition with two-cycle generative adversarial network[J]. *Acta Optica Sinica*, 2020, 40(19): 1910002.
- [23] 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.
- [24] 彭俊利, 谷雨, 张震, 等. 融合单词贡献度与 Word2Vec 词向量的文档表示[J]. *计算机工程*, 2021, 47(4): 62-67. Peng J L, Gu Y, Zhang Z, et al. Document representation fused with term contribution and Word2Vec word vector[J]. *Computer Engineering*, 2021, 47(4): 62-67.
- [25] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [26] 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.