

# 基于特征融合与注意力机制的野生菌细粒度分类

钱嘉鑫, 余鹏飞\*, 李海燕, 李红松

云南大学信息学院, 云南 昆明 650500

**摘要** 鉴别野生菌的种类是预防误食有毒野生菌的重要途径。因此,为了提高野生菌细粒度识别分类的准确率,对卷积模块的注意力机制模块(CBAM)进行改进,提出了一种并行相加卷积模块的注意力机制模块PA\_CBAM,将CBAM中的通道注意力模块和空间注意力模块从原来的串行连接改为并行连接并相加,解决了2种注意力模块因串行连接带来的互相干扰问题。另外,参考特征金字塔的思想改进ResNet50,其Top-1和Top-5准确率达到86.03%和97.19%,较原来提升0.86和0.73个百分点;其添加PA\_CBAM后在Top-1和Top-5准确率达到88.52%、97.58%,较CBAM模块提高了3.03和0.69个百分点。此外,为了将模型移植到移动端,结合迁移学习,提出了MobileNet\_v2+PA\_CBAM的识别方法,准确率达到94.87%,较之前提升0.66个百分点。研究表明:提出的注意力机制模块PA\_CBAM在野生菌细粒度识别研究中具有更好的识别效果,具有一定的泛用性,并且MobileNet\_v2+PA\_CBAM训练后模型大小仅为27.8 MB,识别图片的平均耗时仅为1.3 ms,是在移动端部署野生菌识别的理想模型。

**关键词** 图像识别; 细粒度分类; 特征融合; 注意力机制; 迁移学习

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP212774

## Fine-Grained Classification of Wild Mushrooms Based on Feature Fusion and Attention Mechanism

Qian Jiabin, Yu Pengfei\*, Li Haiyan, Li Hongsong

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

**Abstract** Identifying the species of wild mushrooms is important to prevent mistaking the toxic type of mushrooms for non-toxic ones. Therefore, to improve the accuracy of the fine-grained classification of wild mushrooms, a parallel addition convolutional block attention module (PA\_CBAM), which is improved from the convolutional block attention module (CBAM), is proposed. PA\_CBAM changes the connections of the channel and spatial attention modules from serial to parallel and adds their results together. Consequently, the interference caused by cascading these attention modules is solved. In addition, the proposed method improves the performance of ResNet50 by referring to the concept of a feature pyramid, whose accuracies of the Top-1 and Top-5 are 86.03% and 97.19%, which are 0.86 and 0.73 percentage points higher than those of the original method, respectively. Furthermore, the Top-1 and Top-5 reach 88.52% and 97.58% using PA\_CBAM, which are 3.03 and 0.69 percentage points higher, respectively. Moreover, to adapt the model for mobile terminals, combined with migration learning, the MobileNet\_v2+PA\_CBAM recognition method is proposed, obtaining an accuracy of 94.87%, which is 0.66 percentage points higher than that previously obtained. The results show that PA\_CBAM has a better recognition and generalization effect in the fine-grained classification of wild mushrooms. Meanwhile, the size of MobileNet\_v2+PA\_CBAM is only 27.8 MB, and the recognition time required for a picture is only 1.3 ms, which is an ideal model for deploying wild mushrooms classification on mobile devices.

**Key words** image recognition; fine-grained classification; feature fusion; attention mechanism; transfer learning

## 1 引言

据中国科学院昆明植物研究所统计,我国已知的

野生食用菌有800多种,云南占3/4<sup>[1]</sup>;由于野生菌种类繁杂,普通人缺乏专业知识、鉴别能力有限,每年都有食用野生菌中毒的伤亡情况发生。2011—2019年

收稿日期: 2021-10-21; 修回日期: 2021-12-03; 录用日期: 2021-12-21; 网络首发日期: 2022-01-12

基金项目: 国家自然科学基金(62066046)

通信作者: \*pfyu@ynu.edu.cn

云南省共报告食物中毒事件 5025 起,中毒 36247 人,死亡 445 人,野生菌是引起云南省食物中毒的主要致病因素<sup>[2]</sup>。传统的野生菌识别方法主要包括民间的图文对比和经验判断法,这两种方法十分主观,容易出错。其他如:文献[3]中提到的化学实验分析法,实验过程较为复杂,只对已有数据的野生菌毒素有较高识别率、未知毒素的识别具有一定局限性;动物检验法和真菌分类学鉴定法虽然具有较高识别率,但是其实验周期过长,所需设备材料多。这些方法识别过程复杂,需具备一定专业知识,难以在实际生活中广泛应用。就现状而言,传统途径解决野生菌的中毒问题十分困难。如果能借助神经网络模型来对野生菌种类进行识别,将有利于促进问题的解决。

目前通过深度学习对野生菌进行识别的研究工作较少。文献[4]采用基于贝叶斯分类模型的毒蘑菇识别方法,通过对已知有毒蘑菇特征的学习,对毒蘑菇识别准确率达到 98% 以上,但是其需要人工标注数据,制作成本较高难以扩充。文献[5]采用 ShuffleNetV2 模型进行蘑菇分类,Top-1 准确率仅为 55.18%, Top-5 准确率为 93.55%, 准确率不高,实验数据较少,只有 7 种野生菌以及 1675 张图片,且图片通过爬虫工具获取,图像质量难以保证。文献[6]采用迁移学习<sup>[7]</sup>,基于 Xception 与 ResNet50 模型,使用 Kaggle<sup>[8]</sup>中的一个包含 6714 张图片共计 9 类野生菌的数据集, ResNet50 模型精度达 93.46%, Xception 模型精度达 95.10%, 但是数据种类与总量较少,且来源是国外的野生菌,与我国本土的野生菌有一定差别。文献[9-10]通过网上采集数据,分别建立拥有 23 种和 33 种野生菌的数据集,准确率也达到了 92.17% 和 96.32%, 和上述方法相比在数据量和精度上都有提升,但是其图片都来自互联网,主要通过旋转的方式扩充数量,且图片清晰度不高,容易造成细节丢失。文献[11]通过自行采集数据,建立了 21 种共计 6881 张高清图片数据集,以 ResNeXt50 为基础提出多尺度特征融合,得到的模型准确率为 96.47%。

本文从细粒度<sup>[12-14]</sup>的角度对野生菌分类进行研究,参考特征金字塔网络<sup>[15]</sup>多尺度融合特征<sup>[16]</sup>的思想,改进了 ResNet50 的网络结构,使其能够关注到低级和高级语义中更多的细节特征;其次,在模型中采用了基于视觉注意力机制<sup>[17]</sup>的细粒度图像分类方法,减少模型对背景等干扰信息的学习,对卷积模块的注意力机制模块(Convolutional block attention module, CBAM)<sup>[18]</sup>进行了改进,得到并行相加卷积模块的注意力机制模块(Parallel addition convolutional block attention module, PA\_CBAM),使其避免了串行结构带来的不同注意力模块之间的干扰。最后,结合迁移学习,测试并选用轻量化模型 MobileNet\_v2 结合 PA\_CBAM,以探寻野生菌识别在移动端应用的理想模型。实验识别率表明,改进后的 ResNet50 和 PA\_CBAM 能够有效提

升细粒度识别精度,识别率得到明显提升。MobileNet\_v2 加入 PA\_CBAM 后也有明显提升,证明 PA\_CBAM 具有一定的泛用性。

## 2 ResNet50 与 CBAM 的改进

### 2.1 改进 ResNet50

野生菌子类繁多,且类间差异参差不齐,同子类下衍生不同亚种,在颜色和形态上都有不同程度变化,这要求网络模型具备更强的细节特征学习能力。残差网络(Residual neural network, ResNet)<sup>[19]</sup>采用了残差块结构,如图 1 所示。该结构的引入使得网络层之间能够进行短连接,让网络不会出现梯度爆炸或梯度消失的问题,在正向和反向传导信息时十分畅通,缓解了模型梯度爆炸和消失的问题,使得网络层数加深,能够学习到更多特征信息,同时不增加额外的参数,也不增加计算复杂度,在此基础上进行野生菌识别。

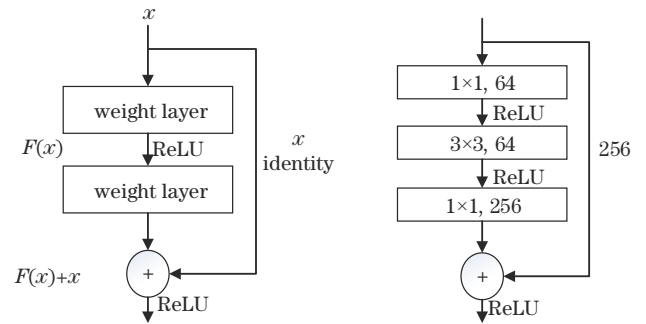


图 1 残差块结构图

Fig. 1 Structure diagram of residual block

参考多层特征融合的思想,对 ResNet50 网络结构进行了改进,其结构如图 2 所示。

图 2 中,框外的部分是 ResNet50 原有网络结构, layer1 至 layer4 分别由 3、4、6、3 个残差块组成;框内部是对 layer1 至 layer4 的改进,用公式分析过程表示为

$$\left\{ \begin{array}{l} L_1(256, 56, 56) = \text{layer1}(\text{input}_{(64, 56, 56)}) \\ L_2(512, 28, 28) = \text{layer2}(L_1(256, 56, 56)) \\ S_1(512, 28, 28) = \text{shortcut1}(L_1(256, 56, 56)) \\ L_3(1024, 14, 14) = \text{layer3}(L_2(512, 28, 28) + S_1(512, 28, 28)) \\ S_2(1024, 14, 14) = \text{shortcut2}(L_1(256, 56, 56)) \\ S_3(1024, 14, 14) = \text{shortcut3}(L_2(512, 28, 28)) \\ L_4(2048, 7, 7) = \text{layer4}(L_3(1024, 14, 14) + S_2(1024, 14, 14) + S_3(1024, 14, 14)) \end{array} \right. \quad (1)$$

shortcut 将上层输出改变通道数和特征尺寸后与下层特征进行融合匹配,输入到后面的卷积层。操作为输入一个  $(C, H, W)$  的特征图,其中  $C$  为通道数,  $H$  为高度,  $W$  为宽度,通过一个 Conv  $1 \times 1$  的卷积改变其通道数,批量归一化后激活输入到一个 Conv  $3 \times 3$  卷积改变其尺寸大小,再次批量归一化后激活输出。首先,使用 shortcut 让 layer1 的输出  $L_1$  满足 layer3 的输入条件,即将形状由  $(256, 56, 56)$  转变为  $(512, 28, 28)$  得

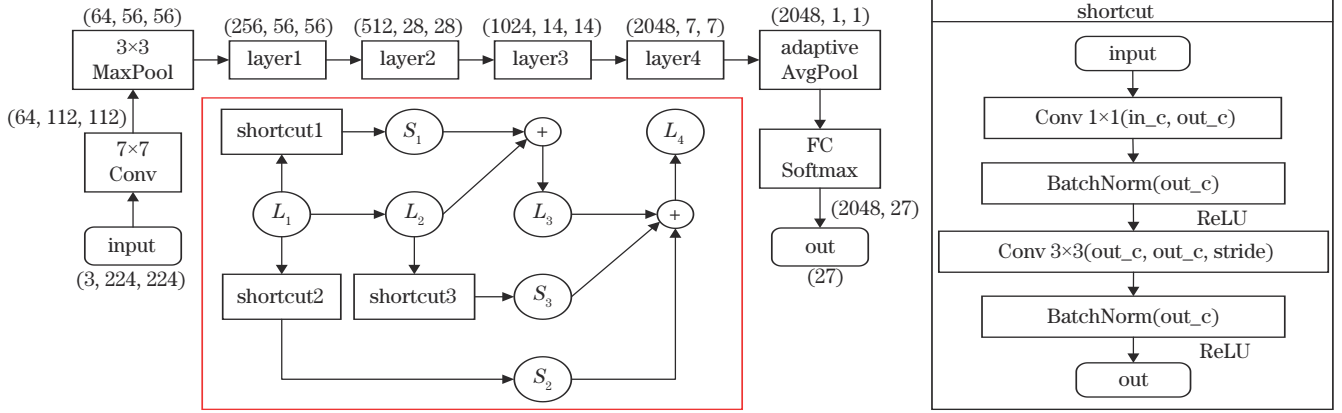


图2 ResNet50改进结构图及shortcut结构

Fig. 2 Structure diagram of improved ResNet50 and shortcut

到  $S_1$ , 与  $L_2$  融合特征后输入到 layer3; 接着使  $L_1$  满足 layer4 的输入条件, 将 (256, 56, 56) 变为 (1024, 14, 14) 得到  $S_2$ ; 让 layer2 的输出  $L_2$  满足 layer4 的输入条件, 将 (512, 28, 28) 的  $L_2$  变为 (1024, 14, 14) 得到  $S_3$ ; 然后  $S_2$ 、 $S_3$  和 layer3 的输出  $L_3$  融合特征后输入到 layer4。将拥有低级语义特征的  $L_1$  和  $L_2$  经过 shortcut 处理后与后面拥有更高级语义特征的  $L_3$  和  $L_4$  进行融合, 促使网络学习到更多特征信息。

## 2.2 改进 CBAM

计算机视觉中的注意力机制是让模型学会关注图像中的重点信息, 忽略无关信息, 其已经被应用于多个领域<sup>[20-22]</sup>, 即让模型学会关注野生菌, 忽略背景信息等干扰。卷积神经网络中的注意力分为硬注意力和软注意力<sup>[23]</sup>。硬注意力是一个随机的预测过程, 更强调动态变化, 同时其不可微, 训练通常需要通过增强学习来完成。软注意力是可微的, 可分为空间域、通道域和混合域, 可通过计算梯度、利用反向传播获得, 是本文关注的重点。

空间注意力在图像上表现为对特征映射上不同位置的关注程度不同, 在网络中表现为对每个通道中的图片特征做同等处理, 忽略了通道域中的信息, 局限在原始图片特征提取阶段, 应用在网络深层的可解释性不强。通道注意力在图像上表现为对不同的图像通道的关注程度不同, 在网络中表现为对每个通道内的信息直接全局平均池化, 忽略了通道内的局部信息。因此, 结合两种注意力机制, 得出了使用混合域的卷积模块的注意力机制模块 CBAM。

CBAM 作为一种即插即用的模块, 将其插入到残差块中使用。CBAM 是一种串行结构, 如图 3 所示, 将卷积层输出的特征映射  $F$  作为输入, 按顺序先进入通道注意力模块得到  $F_1$ , 再通过空间注意力模块得到  $F_2$ , 公式为

$$\begin{cases} F_1 = M_c(F) \otimes F \\ F_2 = M_s(F_1) \otimes F_1 \end{cases} \quad (2)$$

式中:  $M_c(F)$  为通道模块权重;  $M_s(F_1)$  为空间模块权

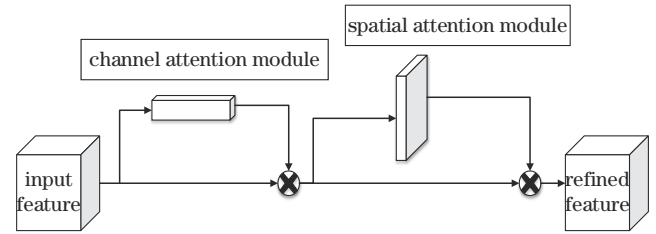


图3 CBAM 模块结构图

Fig. 3 Structure diagram of CBAM module

重;  $\otimes$  为卷积运算。

因 CBAM 为串行结构, 空间注意力和通道注意力无论先后顺序, 后面的模块所学的内容都会被前面的模块处理过的内容影响<sup>[24]</sup>。在细粒度分类问题中, 这种串行干扰会使模型效果变得不稳定, 无法保证准确率的有效提升。

因此, 对 CBAM 进行改进, 将其由串行改为并行相加, 由此得到改进的 CBAM 模块——PA\_CBAM, 让两种注意力机制都能直接从原始输入特征映射中学习各自所需的内容而不互相影响。PA\_CBAM 将输入的特征图  $F$  输入到通道注意力模块和空间注意力模块两个并行的模块中, 得到对应的权值, 然后将特征图  $F$  分别与两个权值加权得到  $F_1$  和  $F_2$ , 最后相加得到输出特征图  $F_{out}$ 。过程公式为

$$\begin{cases} F_1 = M_c(F) \otimes F \\ F_2 = M_s(F) \otimes F \\ F_{out} = F_1 + F_2 \end{cases} \quad (3)$$

调整后的结构如图 4 所示。

## 2.3 野生菌分类模型结构

基于 ResNet50 主干网络进行改进, 结合改进的 CBAM 模块, 改进之后的整体网络如图 5 所示, 其中  $L_1$  至  $L_4$  中的残差块+PA\_CBAM 如图 6 所示。

首先从训练集读取某个批次的野生菌图片进行预处理操作, 将图片调整为  $224 \times 224$  并进行标准化处理, 然后转化为张量, 通过图 2 中改进的 ResNet50 网络中  $7 \times 7$  大小卷积核的卷积层、 $L_1$  至  $L_4$  模块来提取特

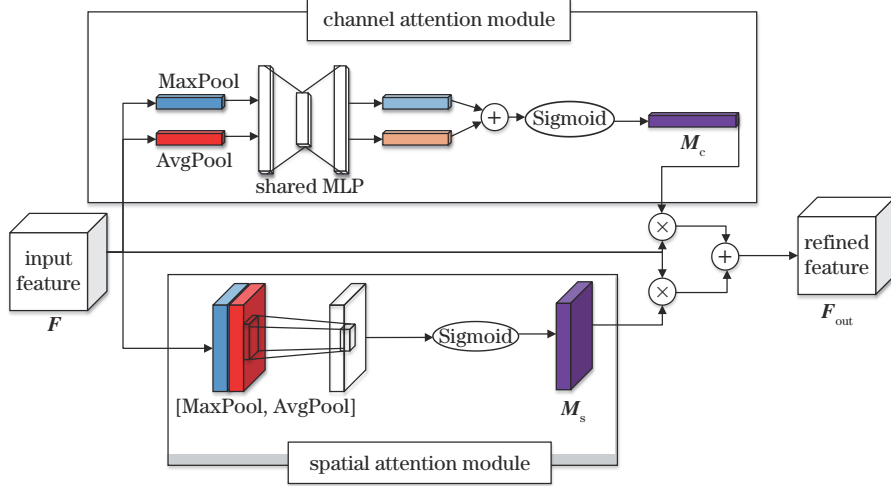


图 4 PA\_CBAM 模块结构图

Fig. 4 Structure diagram of PA\_CBAM module

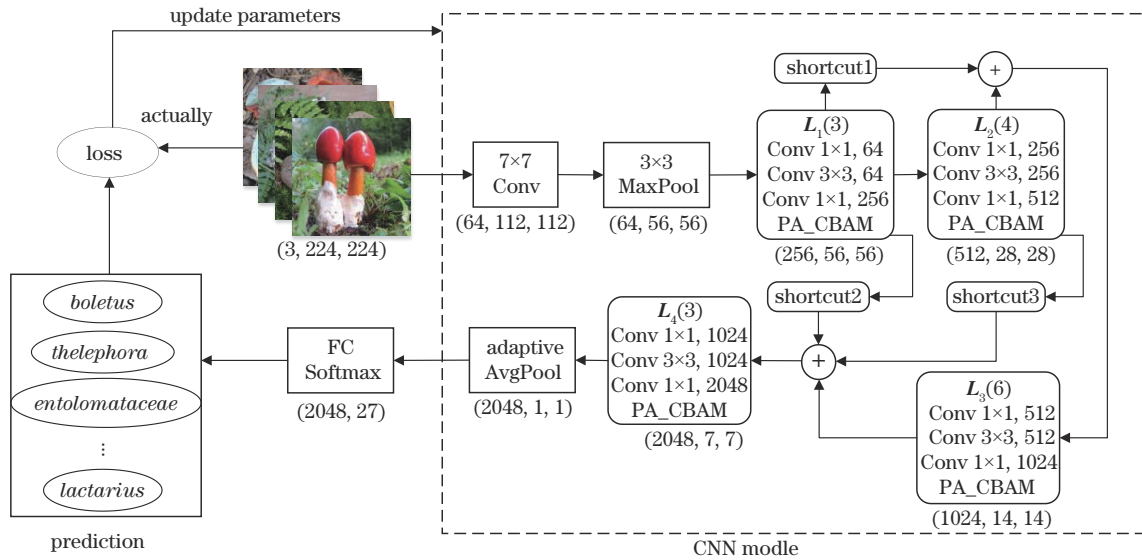


图 5 训练流程图及残差块+PA\_CBAM

Fig. 5 Training process and residual block+ PA\_CBAM

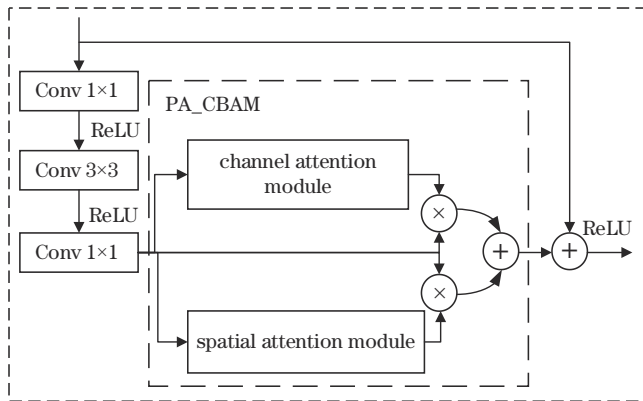


图 6 残差块+PA\_CBAM

Fig. 6 Residual block+ PA\_CBAM

征,这 4 个模块包含的残差块数量分别是 3、4、6 和 3。野生菌图片经过网络的第一层进行特征提取后得到如

颜色和纹理等较多的低级语义特征图,对野生菌识别这类细粒度分类任务来说这些底层的特征往往具有重要的作用,因此模型在图中  $L_1$  至  $L_4$  的每个残差块中都添加了 PA\_CBAM,利用改进后的 CBAM 获取更多的特征信息,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。最后,通过自适应平均池化、全连接层的处理,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更新,减小预测和标签之间的损失,完成训练模型。

### 3 实验及分析

#### 3.1 实验数据

数据集主要由国家自然科学基金项目“基于多视角学习的野生菌种类识别技术研究”提供资金支持进行现场拍摄,获取高质量多角度适用性更强的野生菌

图片(分辨率为  $3000 \times 2000$ ),结合部分网络爬虫工具筛选的图片构成。目前数据集包含 27 种野生菌,包括虫草菌、鹅膏菌、粉褶菌、干巴菌、谷熟菌、红菇、猴头菌、虎掌菌、鸡油菌、鸡枞菌、灵芝、蘑菇、奶浆菌、牛肝菌、平菇、皮条菌、青头菌、荞面菌、乳菇、珊瑚菌、湿伞菌、丝膜菌、松露、松茸、铜绿菌、羊肚菌和竹荪,每个种

类有 400 到 600 张图片。数据集所用野生菌图像共 13581 幅,按照 8:2 将其分为训练集和验证集,其中训练集包含 10865 张图片,验证集包含 2716 张图片。数据集部分图像如图 7 所示,分别是青头菌、灵芝、谷熟菌、干巴菌、牛肝菌、鹅膏菌、红菇和松茸。



图 7 野生菌数据集图像示例

Fig. 7 Image examples from wild mushrooms dataset

为进一步提高模型的泛化能力,避免训练过程中出现过拟合现象,每个训练周期前都会对训练集进行随机的数据增强,主要包括缩放、剪裁、旋转、亮度变化以及 mixup<sup>[25]</sup>,以达到每次训练所用图片都不同的效果,但对验证集不做数据增强操作。最后对所有图片进行归一化操作以提高模型的学习速度。mixup 的具体操作为

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda) x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda) y_j,\end{aligned}\quad (4)$$

式中:  $(x_i, y_i)$ 、 $(x_j, y_j)$  为训练数据中随机抽取的两个特征目标向量。对每一张图片的  $x_i$ , 在数据集中选取另一张图片的  $x_j$ , 选定一个权重参数  $\lambda$ ,  $\lambda \in (0, 1)$ , 然后将两张图片及其标签分别进行加权平均, 由此得到新图片以及对应的标签。

### 3.2 实验环境、参数设置及评估

实验采用的服务器配置为 NVIDIA 3090 24 G 的显卡两块, 128 G 的内存, Intel®Core™i7-10700 K CPU@3.80 GHz×16 的 CPU, 操作系统为 Ubuntu20.04, CUDA 11.1 的并行计算框架, 使用 Anaconda 搭建环境, Jupyter Notebook 作为开发平台, 编程环境为 Python 3.8, 使用了 PyTorch 和 Fastai<sup>[26-27]</sup> 作为框架。

设置模型训练 epoch 统一为 300 次, 根据显卡显存大小, 设置每个批次读取图片数量的 batch\_size=128,

优化函数为 Adam, 初始学习率为  $1 \times 10^{-4}$ , 损失函数为交叉熵(Cross entropy)损失函数。在不影响训练精度前提下, 将训练数据精度从 32 位改为 16 位, 以加速训练, 并减少内存消耗。

选取 Top-1 和 Top-5 准确率(Accuracy)、损失值(Loss)、召回率(Recall)、精确度(Precision)、F1 score、模型预测一张图片的平均时间(ms)、模型规模与模型参数数量作为评价标准。其中准确率、召回率、精确度和 F1 score 的计算公式为

$$A = \frac{f_{TP} + f_{TN}}{f_{TP} + f_{TN} + f_{FP} + f_{FN}}, \quad (5)$$

$$P = \frac{f_{TP}}{f_{TP} + f_{FP}}, \quad (6)$$

$$R = \frac{f_{TP}}{f_{TP} + f_{FN}}, \quad (7)$$

$$F_1 = 2 \frac{P \cdot R}{P + R}, \quad (8)$$

式中:  $A$  为准确率;  $P$  为精确度;  $R$  为召回率;  $f_{TP}$  (True positive) 为某类预测结果为正, 实际结果为正的数量;  $f_{TN}$  (True negative) 为某类预测结果为负, 实际结果为负的数量;  $f_{FP}$  (False positive) 为某类预测结果为正, 实际结果为负的数量;  $f_{FN}$  (False negative) 为某类预测结果为负, 实际结果为正的数量;  $F_1$  为 F1 score, 精确度

与召回率的加权平均。

为防止实验数据中各类数据量差异影响指标可靠性,分别对精确率、召回率和 F1 score 进行加权平均,其公式为

$$\frac{\sum_{i=0}^m \left( \begin{matrix} P_i \\ R_i \times n_i \\ F_{1i} \end{matrix} \right)}{N}, \quad (9)$$

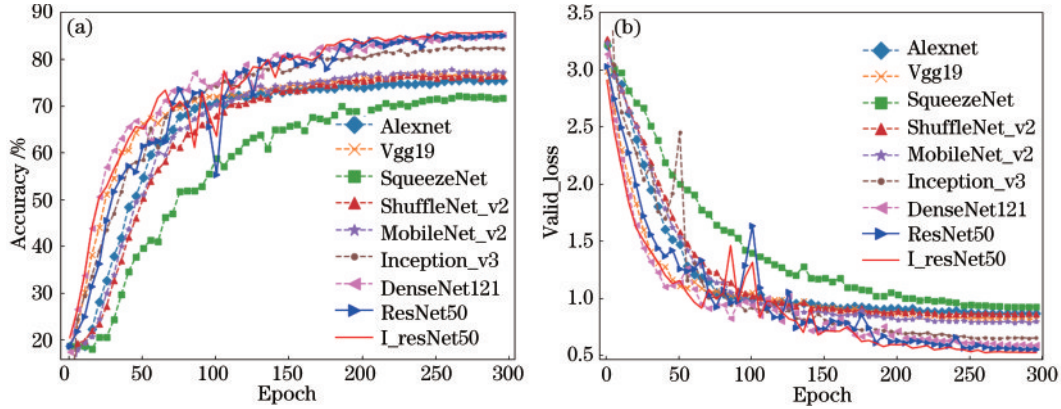


图 8 不同模型在验证集上的实验过程。(a)准确率收敛曲线;(b)损失收敛曲线

Fig. 8 Experimental process of different models on the validation set. (a) Accuracy convergence curve; (b) loss convergence curve

模型在验证集上的准确率稳步增长、损失收敛效果好,训练过程呈下降趋势,未出现较大的震荡,选择在验证集上准确率最高的模型作为最终模型,识别准确率结果如表 1 所示。

表 1 不同模型的识别效果  
Table 1 Recognition results of different models

Model	Accuracy / %	Precision / %	Recall / %	F1 score / %
AlexNet	75.43	76.58	75.53	75.78
Vgg19	76.55	76.87	76.65	76.74
SqueezeNet	71.68	71.81	71.71	71.44
ShuffleNet_v2	76.37	76.37	76.37	76.21
MobileNet_v2	77.59	77.66	77.60	77.47
Inception_v3	82.43	82.66	82.43	82.38
DenseNet121	84.81	85.39	85.10	85.08
ResNet50	85.17	85.63	85.13	85.40
I_ResNet50	86.03	86.00	85.97	86.24

可以看出,在对比的 8 个模型中,ResNet50 具有最好的识别效果,识别准确为 85.17%。改进后的模型 I\_ResNet50 优于 ResNet50,与之相比在准确率、精确度、召回率和 F1\_score 4 个评价标准上分别提升了 0.86、0.37、0.84 和 0.84 个百分点。改进后的特征提取网络通过多层特征融合,加强了模型对图片特征语义信息的学习,实现了更好的分类效果,更适合本文进行野生菌识别的研究。

结合注意力模块,对 ResNet50 与 I\_ResNet50 进行

式中: $i$ 为某类的数字代码,从 0 开始到最后一类 $m$ ; $n_i$ 为某类的数量; $N$ 为所有类的数量。

### 3.3 实验结果及分析

通过与 AlexNet、Vgg19、SqueezeNet、ShuffleNet\_v2、MobileNet\_v2、Inception\_v3、DenseNet121 和 ResNet50 这 8 个网络模型进行对比实验,对比模型改进前后的效果以及改进方法的可信度。模型训练的损失和准确率变化如图 8 所示。

更细致的对比。文献[24]将通道注意力、空间注意力和混合注意力模块 CBAM 加入到 ResNet50 残差块中最后一个卷积层之后。最后将 PA\_CBAM 与 ResNet50 和 I\_ResNet50 结合进行对比实验。实验过程的 Top1 准确率和验证集损失变化曲线如图 9 所示(为了方便观察,epoch 截取 100~300,主要进行 CBAM 和 PA\_CBAM 的对比),详细的实验结果如表 2 所示。

加入注意力模块后,模型的准确率和损失在训练过程都出现了不同程度的震荡,其中改进后的 PA\_CBAM 模块的震荡最为明显,不过在第 200 次训练之后也趋于平稳收敛。

可以看出,I\_ResNet50 的各项精度标准都得到了提升,但加入的多尺度特征融合结构使其模型规模增加了 1 倍。

在数据集中,ResNet50 分别结合通道注意力、空间注意力和 CBAM 模块进行训练后,模型的准确率等各项评价指标并没有得到提升,Top1 准确率反而分别下降了 0.32、1.46 和 0.43 个百分点,I\_ResNet50 结合 CBAM 后 Top1 准确率同样下降了 0.54 个百分点,表明野生菌细粒度识别任务中,这几种注意力模块没能发挥其有效性,且加入 CBAM 的准确率比加入通道注意力的准确率低,进一步说明 CBAM 中顺序靠后的通道注意力模块因为学习到了前面空间注意力模块的特征而受到了干扰,由此产生了负作用,文献[24]中也通过实验验证了串行连接的注意力模块会导致模型学习效果不好、性能下降等问题。

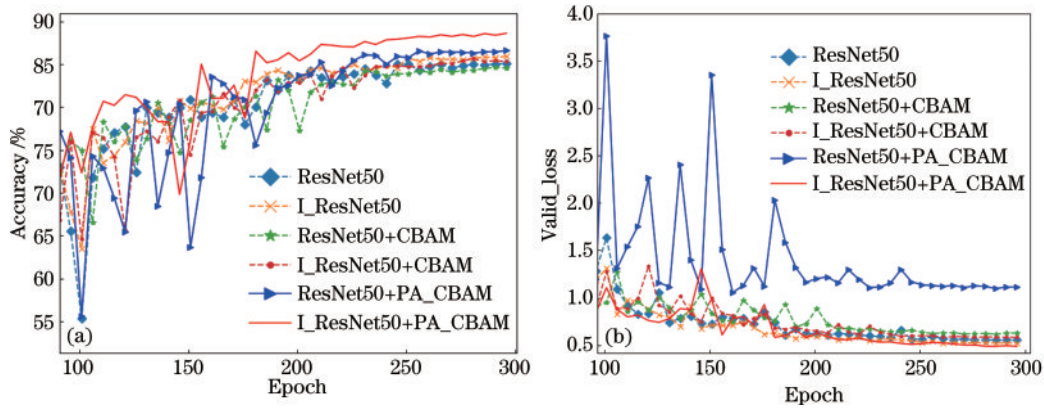


图9 对比实验在验证集上的实验过程。(a)准确率收敛曲线;(b)损失收敛曲线

Fig. 9 Experimental process of the comparison experiment on the validation set. (a) Accuracy convergence curve; (b) loss convergence curve

表2 对比实验的结果

Table 2 Comparative experiment results

Model	Accuracy of Top1 / %	Accuracy of Top5 / %	Loss	Recall / %	Precision / %
ResNet50	85.17	96.46	0.5552	85.13	85.15
I_ResNet50	<b>86.03</b>	<b>97.19</b>	<b>0.5212</b>	<b>85.97</b>	<b>86.00</b>
ResNet50+channel	84.85	95.81	0.6178	84.85	84.85
ResNet50+spatial	83.71	<b>96.89</b>	<b>0.5643</b>	83.62	83.81
ResNet50+CBAM	84.74	96.39	0.6214	84.55	84.68
I_ResNet50+CBAM	<b>85.49</b>	<b>96.89</b>	0.5797	<b>85.49</b>	<b>85.62</b>
ResNet50+PA_CBAM	87.66	97.11	0.5335	87.14	87.47
I_ResNet50+PA_CBAM	<b>88.52</b>	<b>97.58</b>	<b>0.4927</b>	<b>88.53</b>	<b>88.71</b>

Model	F1 score / %	Predicting time / ms	Size / MB	Parameter number / $10^7$	Training time / h
ResNet50	85.40	<b>1.557</b>	<b>270.06</b>	<b>2.332</b>	12.41
I_ResNet50	<b>86.24</b>	1.629	525.81	4.572	<b>11.83</b>
ResNet50+channel	84.48	1.592	286.50	2.356	4.50
ResNet50+spatial	83.60	<b>1.410</b>	<b>283.20</b>	<b>2.344</b>	<b>4.00</b>
ResNet50+CBAM	84.48	2.189	298.92	2.607	5.25
I_ResNet50+CBAM	<b>85.37</b>	3.496	554.66	4.824	9.67
ResNet50+PA_CBAM	87.71	<b>2.259</b>	<b>298.92</b>	<b>2.807</b>	<b>5.42</b>
I_ResNet50+PA_CBAM	<b>88.48</b>	3.365	554.66	4.824	9.33

而改进后的 PA\_CBAM 表现较优,由于改变了原来的 CBAM 串行连接模式,采用并行相加的结构,解决了串行所带来的干扰,让 ResNet50 和 I\_ResNet50 两个模型的各项精度评价指标都有所提升。其中,Top1 准确率和 Top5 准确率分别达到了 87.66%、97.11% 和 88.52%、97.58%,比原来提升了 2.49、0.65 和 2.49、0.39 个百分点。这两个模型与 PA\_CBAM 结合后参数和模型规模有所增加,预测时间分别从 1.557 ms 和 1.629 ms 增加到 2.259 ms 和 3.365 ms,分别增加了 0.702 ms 和 1.736 ms,对识别预测性能的影响可忽略不计,证明了 PA\_CBAM 在野生菌细粒度识别研究中的有效性。

对比所有模型的训练时间发现,虽然 ResNet50 在模型规模和参数量上是最小的,但是其所用的训练时

间却是最多的,耗时 12.41 h,而 I\_ResNet50 在模型规模和参数量翻倍的情况下,训练消耗了 11.83 h。将 ResNet50 和 I\_ResNet50 与 CBAM 结合后,训练时间下降至 5.25 h 和 9.67 h,与 PA\_CBAM 结合后为 5.42 h 和 9.33 h。所以,训练时间大幅下降的原因是引入了注意力机制后,模型更多地关注到了图片中的重要信息,减少了图片背景等无关信息对模型的干扰,再加上残差网络本身跳跃连接的特点,后向传播计算中存在很多梯度为零的参数,减少了许多不必要的计算,加速了模型训练,因此在模型规模和参数量都翻倍增加的情况下,模型的训练时间反而减少了,同时也说明模型存在冗余。

为了更加直观地看到注意力模块改进前后的识别效果,使用 Grad-CAM<sup>[28]</sup> 提供可视化,画出激活热力

图。随机选取了部分野生菌图像,分别通过 I\_ResNet50+CBAM 和 I\_ResNet50+PA\_CBAM 模型,得到热力图(图 10)。由图 10 可知,添加 PA\_CBAM 的 I\_ResNet50 比添加 CBAM 更能准确地选中并标记到目标所在位置。

为进一步验证 PA\_CBAM 的泛用性,将其加入到 AlexNet 和 Vgg19 最后一个卷积层之后进行实验。另外,如此大的模型不利于在实际中应用与部署,因此根据表 1 中 SqueezeNet、ShuffleNet\_v2 和 MobileNet\_v2

3 个经典轻量化模型的识别效果,选取 MobileNet\_v2 结合 PA\_CBAM 进行对比实验,验证 PA\_CBAM 在轻量化模型上的效果以及实际应用的可能性。采用轻量化模型 MobileNet\_v2,结合迁移学习,为避免将注意力模块 PA\_CBAM 插入到 Inverted\_res\_block 后模型规模翻倍,在训练过程中对预训练模型参数造成严重破坏,选择将其插入到 MobileNet\_v2 的最后一个卷积层之后。实验过程中准确率及损失变化曲线如图 11 所示,实验结果如表 3 所示。

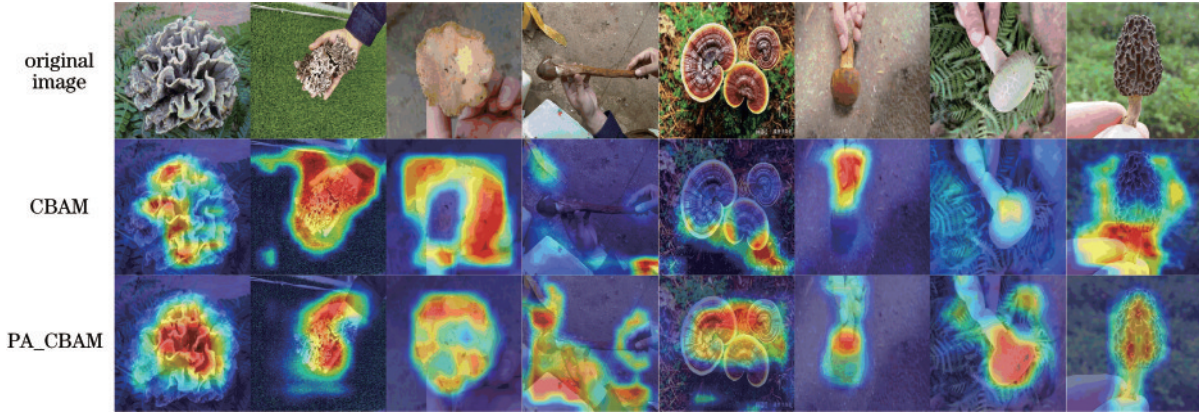


图 10 热力图效果对比

Fig. 10 Diagram of thermodynamic effect comparison

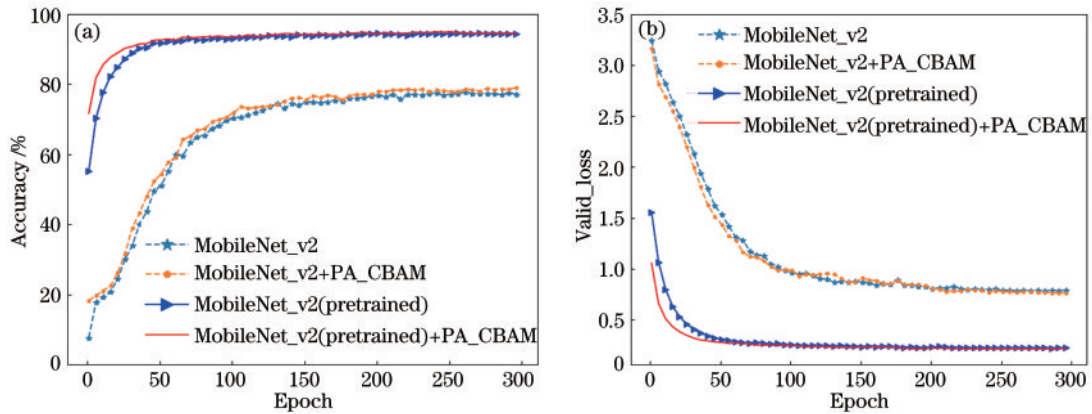


图 11 MobileNet\_v2 结合 PA\_CBAM 在验证集上的实验过程。(a) 准确率收敛曲线; (b) 损失收敛曲线

Fig. 11 Experimental process of MobileNet\_v2 combined with PA\_CBAM on the validation set. (a) Accuracy convergence curve; (b) loss convergence curve

表 3 PA\_CBAM 泛用性验证实验结果

Table 3 PA\_CBAM versatility verification experiment results

Model	Accuracy / %	Predicting time /ms	Size /MB
AlexNet + PA_CBAM	76.91	-	-
Vgg19 + PA_CBAM	-	-	-
MobileNet_v2 + PA_CBAM	79.18	1.27	27.7
MobileNet_v2(pretrained)	94.21	1.65	27.8
MobileNet_v2(pretrained) + PA_CBAM	94.87	1.30	27.8

PA\_CBAM 与 MobileNet\_v2 结合后,在训练过程中并未出现图 8 中的震荡,整个过程比较平稳,且模型

性能得到了稳步的提升。

由表 3 可知,PA\_CBAM 结合 AlexNet 后,准确率



与表 1 相比提升了 1.48 个百分点;而 Vgg19 经过多次尝试,在加入 PA\_CBAM 后都无法进行正常训练。MobileNet\_v2 的模型大小仅为 27.8 MB,相比 ResNet50 这类模型更为轻巧,是在移动端实现野生菌细粒度识别的理想模型。加入注意力模块 PA\_CBAM 后,对比表 1 中未经预训练的 MobileNet\_v2 的准确率提升了 1.59 个百分点;预训练后的模型准确率达到 94.21%,作为轻量化模型已经拥有一个不错的识别效果,加入 PA\_CBAM 后准确率达到 94.87%,提升了 0.66 个百分点,并且模型的预测耗时也仅为 1.3 ms,预测速度十分可靠,其在野生菌细粒度分类任务上具有较好的应用效果。实验结果表明,PA\_CBAM 对模型准确率的提升具有稳定性和有效性,且具有一定的泛用性。

## 4 结 论

针对野生菌细粒度识别问题,使用野生菌数据集,对野生菌识别分类问题展开研究,参考多层特征融合的思想,对 ResNet50 的网络结构进行了改进,接着对注意力机制模块 CBAM 进行改进,提出了一种并行相加的 PA\_CBAM 注意力模块,解决了 CBAM 中空间模块和通道模块串行连接造成的两种注意力互相干扰的问题,增强了网络对有效特征的学习。从实验结果来看,改进后的模型在一定程度提升了对野生菌细粒度图像的识别精度,优于其他 8 种对比模型,较 ResNet50 在准确率上提升了 0.86 个百分点,结合 PA\_CBAM 后准确率提升 2.49 个百分点。从热力图中可明显看出,PA\_CBAM 大大减少了背景干扰,使目标定位更加准确,让模型能更好地学习到有效的特征信息。考虑日常移动场景使用,结合迁移学习,使用 MobileNet\_v2,加入 PA\_CBAM 后准确率达到 94.21%,提升了 0.66 个百分点。实验结果表明,PA\_CBAM 具有一定的有效性和泛用性。今后将继续采集更多种类的野生菌图片补充数据集,并进行更加细致的分类,继续寻找和完善网络模型的结构,采用云服务器或本地移动端计算识别的方式开发应用,为大众提供可靠的技术支持,减少误食有毒野生菌事件的发生。

## 参 考 文 献

- [1] 杨祝良. 浅论云南野生蕈菌资源及其利用[J]. 自然资源学报, 2002, 17(4): 463-469.  
Yang Z L. On wild mushroom resources and their utilization in Yunnan Province, Southwest China[J]. Journal of Natural Resources, 2002, 17(4): 463-469.
- [2] 万蓉, 赵江, 万青青, 等. 2011—2019 年云南省食物中毒流行特征分析及预防措施探讨[J]. 食品安全质量检测学报, 2021, 12(4): 1620-1624.  
Wan R, Zhao J, Wan Q Q, et al. Epidemiological characteristics and preventive measures of food poisoning in Yunnan Province from 2011 to 2019[J]. Journal of Food Safety & Quality, 2021, 12(4): 1620-1624.
- [3] 王鹏惊. 对毒蘑菇毒素的分类与识别探讨[J]. 科技与创新, 2018(11): 61-62.  
Wang P J. Discussion on classification and identification of poisonous mushroom toxins[J]. Science and Technology & Innovation, 2018(11): 61-62.
- [4] 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. Poisonous mushroom recognition based on Bayesian classification[J]. Software Guide, 2015, 14(11): 60-62.
- [5] 肖杰文, 赵敏博, 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.
- [6] 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.
- [7] Torrey L, Shavlik J. Transfer learning[M]//Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. Hershey: IGI Global, 2010: 242-264.
- [8] CatoDogo. Mushrooms classification-Common genus's im-ages[EB/OL]. (2019-03-03) [2020-12-02]. <https://www.kaggle.com/maysee/mushrooms-classification-common-genuss-images/metadata>.
- [9] 樊帅昌, 易晓梅, 李剑, 等. 基于深度残差网络与迁移学习的毒蕈图像识别[J]. 传感技术学报, 2020, 33(1): 74-83.  
Fan S C, Yi X M, Li J, et al. Toadstool image recognition based on deep residual network and transfer learning[J]. Chinese Journal of Sensors and Actuators, 2020, 33(1): 74-83.
- [10] 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.
- [11] 张志刚, 余鹏飞, 李海燕, 等. 基于多尺度特征引导的细粒度野生菌图像识别[J]. 激光与光电子学进展, 2022, 59(12): 1210016.  
Zhang Z G, Yu P F, Li H Y, et al. Fine-grained image recognition of wild mushroom based on multiscale feature guide[J]. Laser & Optoelectronics Progress, 2022, 59(12): 1210016.
- [12] Wah C, Branson S, Welinder P, et al. The caltech-ucsd birds-200-2011 dataset[EB/OL]. [2020-12-06]. [http://www.vision.caltech.edu/datasets/cub\\_200\\_2011/](http://www.vision.caltech.edu/datasets/cub_200_2011/).
- [13] Maji S, Rahtu E, Kannala J, et al. Fine-grained visual classification of aircraft[EB/OL]. (2013-06-21)[2021-05-04]. <https://arxiv.org/abs/1306.5151>.

- [14] Krause J, Stark M, Jia D, et al. 3D object representations for fine-grained categorization[C]//2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 554-561.
- [15] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [16] 李思瑶, 刘宇红, 张荣芬. 多尺度特征融合的细粒度图像分类[J]. 激光与光电子学进展, 2020, 57(12): 121002.  
Li S Y, Liu Y H, Zhang R F. Fine-grained image classification based on multi-scale feature fusion[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121002.
- [17] Fu J L, Zheng H L, Mei T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4476-4484.
- [18] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[EB/OL]. (2018-07-17)[2021-04-07]. <https://arxiv.org/abs/1807.06521>.
- [19] 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.
- [20] 曹城硕, 袁杰. 基于YOLO-Mask算法的口罩佩戴检测方法[J]. 激光与光电子学进展, 2021, 58(8): 0810019.  
Cao C S, Yuan J. Mask-wearing detection method based on YOLO-mask[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810019.
- [21] 鲍海龙, 万敏, 刘忠祥, 等. 基于区域自我注意力的实时语义分割网络[J]. 激光与光电子学进展, 2021, 58(8): 0810018.  
Bao H L, Wan M, Liu Z X, et al. Real-time semantic segmentation network based on regional self-attention[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810018.
- [22] 陈子涵, 吴浩博, 裴浩东, 等. 基于自注意力深度网络的图像超分辨率重建方法[J]. 激光与光电子学进展, 2021, 58(4): 0410013.  
Chen Z H, Wu H B, Pei H D, et al. Image super-resolution reconstruction method based on self-attention deep network[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0410013.
- [23] Xu K, Ba J, Kiros R, et al. Show, attend and tell: neural image caption generation with visual attention[C]//Proceedings of the 32th International Conference on Machine Learning 2015, Lille, France. PMLR, 2015, 37: 2048-2057.
- [24] 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 M H, Wu Z X, Zhou Z G. 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.
- [25] Zhang H Y, Cisse M, Dauphin Y N, et al. mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25)[2021-05-08]. <https://arxiv.org/abs/1710.09412>.
- [26] Howard J, Gugger S. Fastai: a layered API for deep learning[J]. Information, 2020, 11(2): 108.
- [27] Howard J, Gugger S. Deep learning for coders with fastai and PyTorch[M]. Sebastopol: O'Reilly Media, 2020.
- [28] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 618-626.