

基于改进 VGG16 网络模型的花卉分类^①



侯向宁¹, 刘华春¹, 侯宛贞²

¹(成都理工大学工程技术学院, 乐山 614000)

²(西华大学 计算机与软件工程学院, 成都 610039)

通信作者: 侯向宁, E-mail: houxiangning_2005@163.com

摘要: 为进一步提高花卉分类的准确率, 在对现有的 VGG16 网络模型进行研究的基础上, 提出一种基于视觉注意力机制的网络模型. 将 SE 视觉注意力模块嵌入到 VGG16 网络模型中, 实现了对花卉显著性区域特征的提取; 为有效防止梯度爆炸及梯度消失, 加快网络的训练和收敛的速度, 在各卷积层后加入 BN 层; 采用多损失函数融合的方式对新模型进行训练. 新模型能有效提取花卉的花蕊、花瓣等显著性区域, 放大了花卉的类间距离, 缩小了类内距离, 加快了网络的收敛, 进一步提高了花卉分类的准确率. 实验结果表明, 新模型在 Oxford-102 数据集上的分类准确率比未引入注意力前有较大提高, 与参考文献相比, 分类准确率也有较大的提高.

关键词: VGG16; 注意力机制; SE 模块; 损失函数; 深度学习; 图像分类

引用格式: 侯向宁, 刘华春, 侯宛贞. 基于改进 VGG16 网络模型的花卉分类. 计算机系统应用, 2022, 31(7): 172-178. <http://www.c-s-a.org.cn/1003-3254/8582.html>

Flower Classification Based on Improved VGG16 Network Model

HOU Xiang-Ning¹, LIU Hua-Chun¹, HOU Wan-Zhen²

¹(The Engineering & Technical College of Chengdu University of Technology, Leshan 614000, China)

²(School of Computer and Software Engineering, Xihua University, Chengdu 610039, China)

Abstract: To further improve the accuracy of flower classification, this study proposes a network model based on visual attention mechanism after the research on the VGG16 network model. Squeeze-and-excitation (SE) attention is embedded in the VGG16 network model to extract salient region features of flowers. BN layer is added following the convolutional layer to effectively prevent gradient explosion and gradient disappearance and to speed up the training and convergence of the network. Multi-loss function fusion is utilized to train the new model. The new model can effectively extract salient regions such as stamens and petals. It can enlarge the distance between and within classes and accelerate the convergence, further improving the classification accuracy. Experimental results show that the new model advances in the classification accuracy of the Oxford-102 dataset after the introduction of the attention mechanism and outstrips the current reference.

Key words: VGG16; attention mechanism; squeeze-and-excitation (SE) module; loss function; deep learning; image classification

步入全面小康社会, 人们在享受物质生活的同时, 也在不断追求高品质的精神生活, 比如对美的欣赏及追求. 花卉观赏可使人心情愉悦, 给人恬美的视觉体验; 然而人们对花卉的分类却较为迷惑, 究其原因主要有两点: 首先是花卉具有较高的类内相似性, 即不同种类的花卉之间存

在着难以分辨的相似的颜色及形状等; 其次是花卉具有较大的类内差异性, 即同类花卉之间受气候和光照等因素的影响, 存在着颜色、形状等方面的不同之处.

我国幅员辽阔, 已知的花卉种类约 3 万种. 在植物研究领域, 基于人工提取特征的传统花卉分类方式分

① 基金项目: 四川省教育厅重点项目 (18ZA0077); 成都理工大学工程技术学院基金 (C122020006)

收稿时间: 2021-10-12; 修改时间: 2021-11-08; 采用时间: 2021-11-19; csa 在线出版时间: 2022-05-30

类准确率低且耗费人力;而具有自动提取特征的基于深度学习的花卉分类方法逐渐受到人们的关注.文献[1]设计了一个8层的卷积神经网络,并在Oxford-102花卉数据集上进行了测试,由于模型的深度较浅,分类效果并不理想.文献[2]通过迁移学习,对Inception-v3网络模型在花卉数据集上进行微调,对分类准确率有少量的提升.文献[3]将AlexNet和VGG16模型的特征进行串联,利用mRMR算法选择有效特征,最后采用SVM分类器进行分类,该方法稍显繁琐且分类准确率不高.文献[4]在VGG16模型的基础上将多层深度卷积特征进行融合,并在Oxford-102花卉数据集进行了测试,但分类的准确率并不理想.文献[5]对LeNet-5模型进行了适当改造,取得了较好的分类效果,但其对花卉图像进行了灰度预处理,并未考虑到颜色对识别精度的影响.文献[6]采用多层特征融合及提取感兴趣区域的方法对花卉进行分类,模型较为复杂.

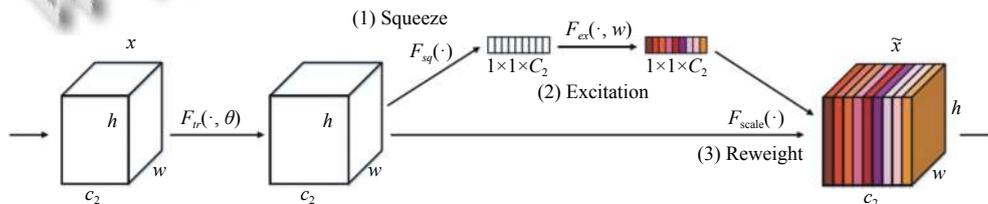


图1 SE模块

在对特征进行重标定之前先进行一个标准的卷积操作 F_{tr} , 如式(1)所示:

$$u_c = v_c * X = \sum_{s=1}^{c'} v_c^s * x^s \quad (1)$$

其中, 输入特征 $X \in R^{H \times W \times C'}$, 输出特征 $U \in R^{H \times W \times C}$, 卷积核 $V = \{v_1, v_2, \dots, v_c\}$, 其中 v_c 代表第 c 个卷积核. $*$ 代表卷积操作, v_c^s 表示第 s 个通道的卷积核, x^s 表示第 s 个通道的输入特征, u_c 表示第 c 个卷积核所对应的输出特征.

(1) Squeeze^[9] 即 F_{sq} 操作, 该操作较为简单, 主要是对特征图进行一个全局平均池化, 从而获取各通道的全局特征, 如式(2)所示:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

F_{sq} 操作将 $H \times W \times C$ 的输入转换成 $1 \times 1 \times C$ 的输出, 也就是将每个二维的特征通道变成一个实数, 这个实数

文献[7]将SE视觉注意力模块引入VGG16网络模型,用于对行人进行检测,取得了较好的检测效果.本文在VGG16-BN模型的基础上,引入视觉注意力机制,将SE视觉注意力模块嵌入VGG16-BN网络模型,对花卉图片的显著性区域进行特征提取,充分利用VGG16-BN模型的宽度和深度优势,以进一步提高花卉分类的准确率.

1 相关工作

1.1 SENet网络

SENet是最后一届ImageNet分类任务的冠军.SENet^[8]的本质是采用通道注意力机制,通过深度学习的方式自动获取图像各个特征通道的权重,以增强有用特征并抑制无用特征.SENet的核心模块是squeeze-and-excitation(SE),主要分3个步骤对特征进行重标定,如图1所示.

某种程度上具有全局的感受野,它表征着在特征通道上响应的全局分布.

(2) Excitation^[10] 即 F_{ex} 操作,通过两个全连接层先降维后升维对squeeze操作的结果进行非线性变换,来为每个特征通道生成权值,该权值表示特征通道之间的相关性,如式(3)所示:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

其中, $W_1 z$ 是第一个全连接层操作, W_1 的维度是 $C/r \times C$, z 的维度是 $1 \times 1 \times C$, 因此 $W_1 z$ 的输出维度是 $1 \times 1 \times C/r$, 即通过 r (维度的缩放因子) 进行了降维, 然后采用ReLU激活. 第2个全连接层是将上个全连接层的输出乘以 W_2 , 其中 W_2 的维度是 $C \times C/r$, 因此最终输出的维度为 $1 \times 1 \times C$, 即同squeeze操作输出的维度相同, 最后再经过Sigmoid函数用于获取各通道归一化后的权重, 得到维度为 $1 \times 1 \times C$ 的 s , s 用来表示通过之前两个全连接层的非线性变换学习到的第 C 个特征图的权值. Excitation这种先降维后升维的操作一方面降

低模型复杂度,使网络具有更好的非线性,另一方面更好的拟合通道间复杂的相关性,提升了模型泛化能力。

(3) $\text{Reweight}^{[11]}$ 即 F_{scale} 操作,将上一步 excitation 操作得到的权值 s 通过乘法逐通道加权到原始的特征上,完成在通道维度上的对原始特征的重标定,如式(4)所示:

$$\tilde{x}_c = F_{\text{scale}}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

总之,SE模块通过自动学习,以获取每个特征通

道的重要程度,然后根据各个通道的重要程度,一方面去提升有用的特征,另一方面抑制对当前任务不相关或作用不大的特征。

1.2 VGG16网络

VGGNet^[12] 荣膺 2014 年 ImageNet 图像分类第 2 名的好成绩,其中 VGG16 是 VGGNet 中分类性能最好的网络之一,其网络结构如图 2 所示。

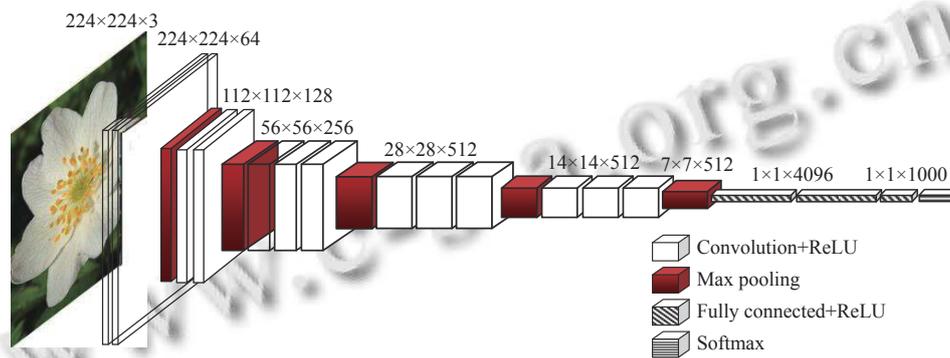


图2 VGG16网络结构

(1) VGG16网络可分为6段,即5段卷积加1段全连接,其中5段卷积包含13个卷积层,1段全连接指网络最后的3个全连接层,因此VGG16网络总共有13+3=16层。

(2) 5段卷积用以提取低、中、高各层的图像特征,每一段有2或3个卷积层。为了增加网络的非线性、防止梯度消失、减少过拟合以及提高网络训练的速度,各卷积层后均采用ReLU激活函数。为利于捕捉细节变化,获得更好的非线性效果并减少参数数量,每个卷积层均采用3×3的卷积核,使得网络结构更加简洁,在必要时3×3卷积核的堆叠还可以替代5×5、7×7等较大的卷积核。

(3) 5段卷积的尾部均接有一个最大池化层,该池化层采用2×2的池化核,能够减小卷积层参数误差造成估计值均值的偏移,更容易捕捉图像和梯度的变化,有利于保留纹理等细节信息。

(4) VGG16网络的最后一段是3个全连接层,全连接层中的每一个节点都与上一层每个节点连接,把前一层的输出特征综合起来,起到分类器的作用。

总之,VGG16网络的深度为16层,这种较深的网络通过逐层的抽象,能够不断学习由低到高各层的特征,具有更强的非线性表达能力,能表达更为丰富的特

征,拟合更为复杂的输入特征。另外,VGG16网络最开始采用64个3×3卷积核,随着网络的加深,卷积核数量逐渐从64,增加到128、256、512,因此使其具有较大的网络宽度,宽度的增加能使网络各层学习到更为丰富的颜色、纹理等特征。

2 改进后的VGG16网络

2.1 SE-VGGConv模块

SE模块的最大特点在于其内部采用常见的池化及全连接层,因此具有很强的通用性,可以方便的嵌入到其他常见的网络模型中。在VGG网络模型的卷积层之后加入SE视觉注意力单元,如图3所示。

如前所述,在VGG网络的卷积层后,首先经过一个GAP全局平均池化层,即图1中的squeeze操作,用于获取通道级的全局特征。然后进入第一个FC层进行降维操作,用ReLU函数激活后进入第2个FC层进行升维操作,Sigmoid函数用于获取各通道归一化后的权重。 $\text{Scale}^{[8,13]}$ (即reweight操作)将归一化后的权重加权到每个原始通道的特征之上,实现了在通道维度上的对原始特征的重标定。

2.2 SE-VGG16-BN网络模型

VGG16因其具有较好的深度及宽度,在图像分类

的应用具有一定的优势,但对具有类间相似性高,类内差异性大以及存在复杂背景干扰的花卉分类,其准确率还有待提高.因此,在VGG16的基础上引入BN层及SE视觉注意力单元,可以充分提取花卉分类任务中类间相似性高、类内差异较大的敏感特征,从而提高花卉分类的准确率.

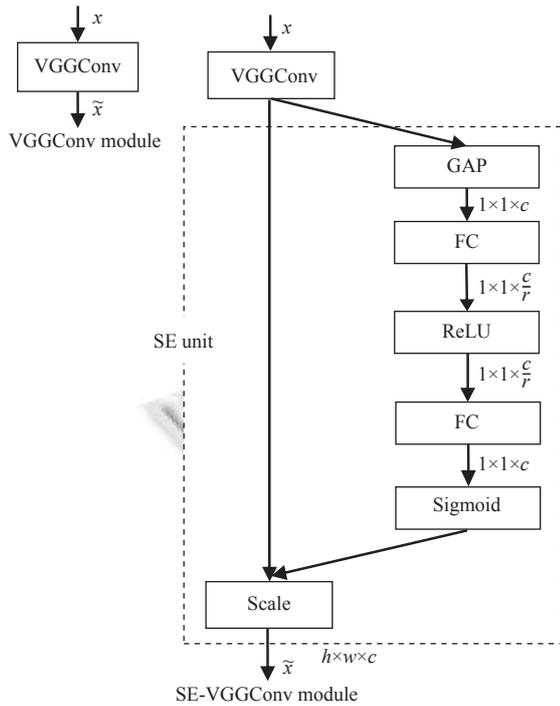


图3 SE-VGGConv 模块

在VGG16加入BN层及SE视觉注意力单元后的网络结构如图4所示.

图4对VGG16网络做了如下改进:前5段卷积的每个卷积层中均加入SE视觉注意力单元、BN层和ReLU激活函数.其中SE单元用于学习各通道的重要程度,从而增强有用特征并抑制无用特征.BN层(batch normalization)的作用是加快网络的训练和收敛的速度,防止梯度爆炸及梯度消失,使模型会变得更加稳定;ReLU激活函数^[14]能增强网络的非线性、防止梯度消失、减少过拟合并提高网络训练的速度.为防止过拟合,第6段的两个FC层后面均加入dropout.Softmax用于最终的分类,由于本文采用的Oxford-102花卉数据集有102种花卉,因此输出的是原始图片对应于102类花卉的概率.

2.3 多损失函数融合

为扩大花卉的类间距离,减少类内距离,以进一步

提高花卉分类的准确率,采用多损失函数融合的方式,具体如下:

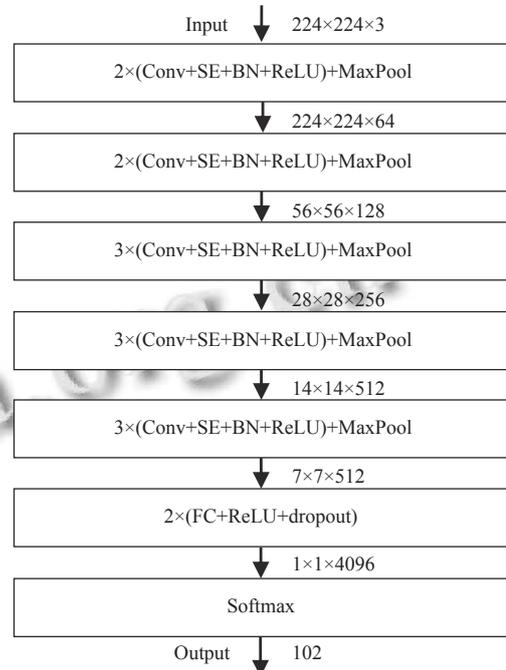


图4 SE-VGG16-BN 网络结构

交叉熵损失函数(cross-entropy cost function)经常用于分类任务中,起着控制模型的总体走势的作用,交叉熵损失函数的定义如下:

$$C = - \sum_{x=1}^n [y \log \hat{y} + (1-y) \log (1-\hat{y})] \quad (5)$$

其中, n 是批处理样本数, x 为输入, y 为标签值, \hat{y} 表示实际输出.

中心损失函数(center loss)优势是可以学习类内距离更小的特征,从而减小类内的差异,并能在一定程度上增大类间差异性,从而提高分类的准确率,中心损失函数的定义如下:

$$L_c = \frac{1}{2} \sum_{i=1}^n \|x_i - C_{y_i}\|^2 \quad (6)$$

其中, n 是批处理样本数, x_i 表示 y_i 类别的第 i 个特征, C_{y_i} 表示 i 类特征的中心值.

将交叉熵损失函数和中心损失函数进行融合,并将其推广至多层神经网络,假设输出神经元的期望值是 $y = \{y_1, y_2, y_3, \dots\}$, $\hat{y}^L = \{\hat{y}_1^L, \hat{y}_2^L, \hat{y}_3^L, \dots\}$,则融合后的计算公式如下:

$$L = C + \lambda L_c$$

$$= - \sum_{x=1}^n [y \log \hat{y}_x^l + (1-y) \log (1 - \hat{y}_x^l)] + \frac{\lambda}{2} \sum_{i=1}^n \|x_i - C_{y_i}\|_2^2 \quad (7)$$

其中,融合系数 λ 的取值范围是0-1,通过多损失函数的融合,放大了花卉的类间距离,缩小了类内距离,加快了网络的收敛速度,进一步提高了分类的效率和准确率。

3 实验结果与分析

本实验的软硬件环境如下:操作系统为Ubuntu 16.04.1 LTS,显卡为NVIDIA GeForce GTX 1080Ti,CPU是Intel Core i7-9700K,32 GB内存。采用PyTorch 1.6.0深度学习框架,Python 3.6编程环境以及Pycharm开发工具。

3.1 数据集

选用牛津大学公开的Oxford-102花卉数据集,该数据集包括102类花卉,每类由40-258张图片组成,共计8189张花卉图片。由于Oxford-102花卉数据集的数据量较小,为了防止过拟合,提高网络模型的性能,对该数据集进行了随机旋转、翻转、平移、裁剪、放缩等操作,通过数据增强,将Oxford-102数据集扩充至

8189+8189×5=49134张。

3.2 模型训练及参数设置

(1) 为提高训练的效果,加快网络模型的收敛,对两个数据集的花卉图片按照保持长宽比的方式归一化,归一化后的尺寸为224×224×3。

(2) 将数据增强后的每类花卉图片数的70%划分为训练集,剩余30%作为测试集。

(3) 训练时保留VGG16经ImageNet预训练产生的用于特征提取的参数,SE单元模块中用于放缩参数 r 设置为文献[8]的作者所推荐的16,其余参数均使用正态分布随机值进行初始化。

(4) 采用随机梯度下降法来优化模型,batchsize设置为32,epoch设为3000,学习率设为0.001,动量因子设为0.9,权重衰减设为0.0005。

(5) 为了防止过拟合,SE-VGG16网络模型第6段的两个全连接层的dropout设置为0.5。

(6) 多损失函数融合公式中 λ 参数的值设置为0.5。

3.3 实验对比与分析

(1) 数据可视化

VGG16网络卷积层前36个通道特征图可视化如图5所示。

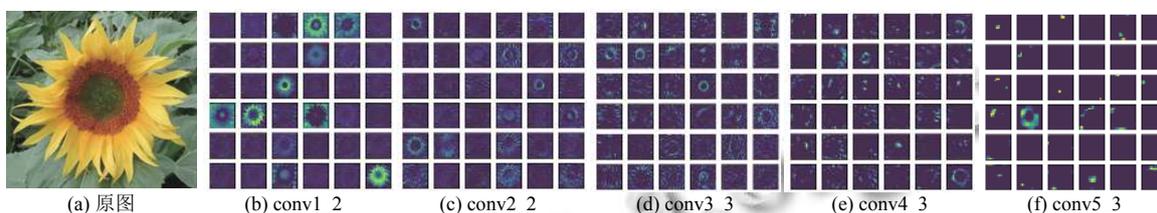


图5 原图及特征图

图5中网络底层conv1_2主要提取的是花卉的颜色及边缘特征,网络中间层conv3_3主要提取的是花卉的简单纹理特征,网络高层conv5_3主要提取的是花瓣、花蕊等细微的抽象特征。

本文与其他网络的可视化类激活图的效果对比如图6所示。

从图6中可以看出,与其他模型相比,本文模型的可视化类激活图所覆盖的范围较大,温度也较高,表明加入SE视觉注意力单元后,网络能够有效提取花卉的花蕊、花瓣等显著性区域。

(2) 分类准确率对比

本文与常见网络模型及引用文献中的网络模型,

在Oxford-102花卉数据集上做了比较,准确率对比结果如表1所示。

本文的模型比所引文献中在Oxford-102花卉数据集上平均高出近6.2个百分点,本文的模型比未添加SE视觉注意力单元的VGG16高出近15.85个百分点。这一方面缘于本文对数据进行了增强,提高了模型的泛化能力和鲁棒性,另一方面本文模型中嵌入了SE视觉注意力单元,能够有效学习花瓣、花蕊等部分的细节的信息,从而提高了模型分类的准确率。

(3) 损失函数效果对比

多损失函数融合在Oxford-102花卉数据集上,与常用的交叉熵及中心损失函数对分类准确率影响的对

比, 实验的结果如表 2 所示.

对比的结果显示, 采用多损失函数融合的方法比单独使用交叉熵及中心损失函数在 Oxford-102 花卉数

据集上的分类准确率分别高出 0.85 及 1.11 个百分点. 结果表明通过多损失函数的融合, 放大了花卉的类间距离, 缩小了类内距离, 进一步提高了分类的准确率.

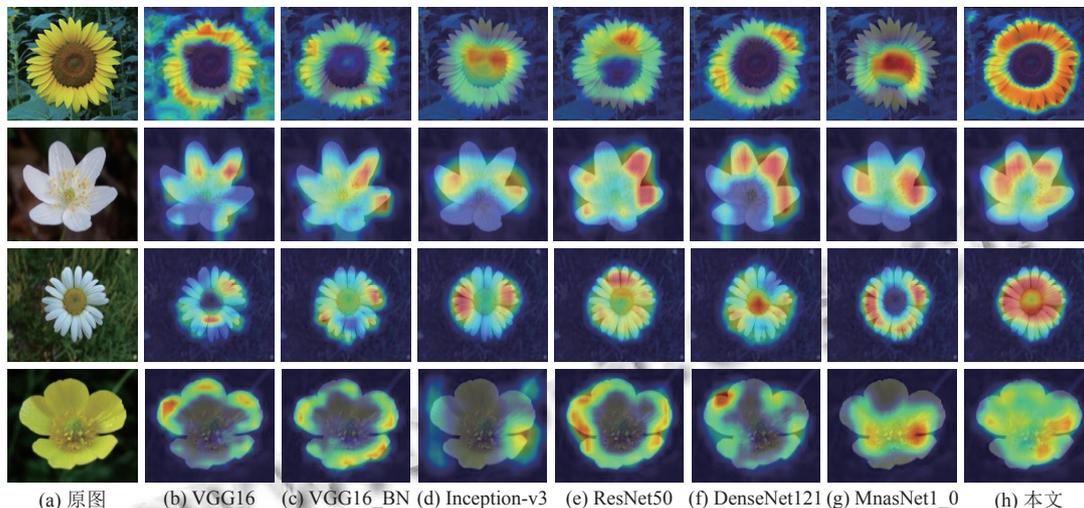


图 6 可视化类激活图

表 1 Oxford-102 上不同网络模型分类准确率对比 (%)

模型结构	分类准确率
VGG16	80.91
文献[1]	84.02
文献[4]	85.55
文献[6]	93.57
文献[2]	94.00
文献[3]	95.70
本文	96.76

表 2 Oxford-102 上各损失函数分类准确率的对比 (%)

损失函数	分类准确率
交叉熵损失函数	95.91
中心损失函数	95.65
多损失函数融合	96.76

(4) 融合系数效果对比

多损失函数融合公式中融合系数 λ 取值对花卉分类准确率影响的对比, 实验的结果如表 3 所示.

表 3 Oxford-102 上融合系数 λ 对分类准确率的影响

λ	分类准确率 (%)
0.2	95.37
0.3	96.61
0.4	96.85
0.5	96.76
0.6	96.16
0.7	95.53
0.8	95.12

实验结果显示随着融合系数 λ 的增长, 在 Oxford-102 花卉数据集上, 分类的准确率逐渐提高, 当 $\lambda=0.5$ 时, 准确率达到最高, 之后又逐渐下降. 实验表明交叉熵损失函数在总体上控制着模型的走势, 中心损失函数能够捕捉并学习类内距离更小的特征, 两者的适度融合, 可以有效提高花卉分类的准确率.

(5) BN 层效果对比

当 $\lambda=0.5$ 时, 未加入 BN 层及加入 BN 层后的分类准确率效果对比分别如图 7、图 8 所示.

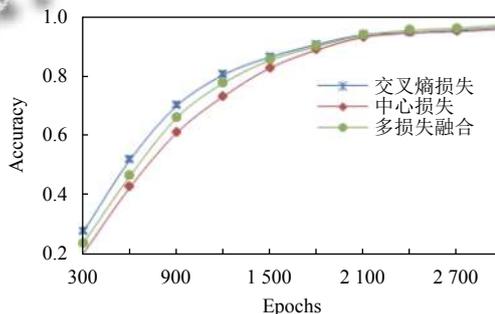


图 7 未加入 BN 层的分类准确率效果

实验结果显示, 未加入 BN 层的曲线收敛较慢, 在 2400 个 epoch 时才逐渐趋于稳定, 而加入 BN 层的模型曲线收敛的较快, 在 1200 个 epoch 时已经基本趋于稳定. 实验表明, 在模型中加入 BN 层后, 能够利用小批量上的均值和标准差, 使网络中间层参数服从于相

同的分布,加快了网络的训练和收敛的速度,防止梯度爆炸及梯度消失,使模型变得更加稳定。

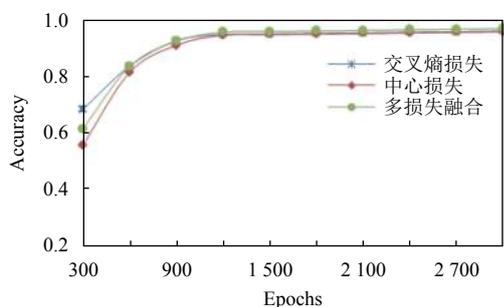


图8 加入BN层后的分类准确率效果

4 结论与展望

本文在VGG16网络模型的基础上,引入视觉注意力机制.通过将SE视觉注意力模块嵌入到VGG16网络模型各卷积层之后,以提取花卉显著性区域的图像特征;各卷积层之后分别加入BN层,有效防止了梯度爆炸及梯度消失,加快了网络的训练和收敛的速度;采用多损失函数融合的方式对新模型进行训练,放大了花卉的类间距离,缩小了类内距离,进一步提高了花卉分类的准确率.实验结果表明,新模型在Oxford 102数据集上的准确率比未引入注意力前有较大的提升。

参考文献

- Liu YY, Tang F, Zhou DW, *et al.* Flower classification via convolutional neural network. IEEE International Conference on Functional-structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA). Qingdao: IEEE, 2016. 110–116.
- Xia XL, Xu C, Nan B. Inception-v3 for flower classification. 2017 2nd International Conference on Image, Vision and Computing (ICIVC). Chengdu: IEEE, 2017. 783–787.
- Cbuk M, Budak U, Guo YH, *et al.* Efficient deep features

selections and classification for flower species recognition. Measurement, 2019, 137: 7–13. [doi: 10.1016/j.measurement.2019.01.041]

- 尹红,符祥,曾接贤,等.选择性卷积特征融合的花卉图像分类.中国图象图形学报,2019,24(5):762–772.[doi:10.11834/jig.180426]
- 吴丽娜,王林山.改进的LeNet-5模型在花卉识别中的应用.计算机工程与设计,2020,41(3):850–855.
- 杨旺功,淮永建.多层特征融合及感兴趣区域的花卉图像分类.哈尔滨工程大学学报,2021,42(4):588–594.
- 李克文,李新宇.基于SENet改进的Faster R-CNN行人检测模型.计算机系统应用,2020,29(4):266–271.[doi:10.15888/j.cnki.csa.007321]
- Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023. [doi: 10.1109/TPAMI.2019.2913372]
- 孟庆宽,张漫,杨晓霞,等.基于轻量卷积结合特征信息融合的玉米幼苗与杂草识别.农业机械学报,2020,51(12):238–245,303.[doi:10.6041/j.issn.1000-1298.2020.12.026]
- Roy SK, Dubey SR, Chatterjee S, *et al.* FuSENet: Fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. IET Image Processing, 2020, 14(8): 1653–1661. [doi: 10.1049/iet-ipr.2019.1462]
- 李昊玥,陈桂芬,裴傲.基于改进Mask R-CNN的奶牛个体识别方法研究.华南农业大学学报,2020,41(6):161–168.[doi:10.7671/j.issn.1001-411X.202003030]
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2015.
- 严春满,王铖.卷积神经网络模型发展及应用.计算机科学与探索,2021,15(1):27–46.[doi:10.3778/j.issn.1673-9418.2008016]
- Liang XL, Xu J. Biased ReLU neural networks. Neurocomputing, 2021, 423: 71–79. [doi: 10.1016/j.neucom.2020.09.050]

(校对责编:孙君艳)