

# 基于迁移学习的图像分类在诗词中的应用研究

武苏雯,赵慧杰,刘鑫,王佳豪

(中原工学院 计算机学院,河南 郑州 451191)

**摘要:**中国传统诗词中蕴含着丰富的文化内涵。为了从海量的诗词库中搜索出最符合图像意境的诗词,实现解析图像内容、提取图像特征关键词,结合项目需求,提出一种基于迁移学习的多 EfficientNet 融合网络的图像分类算法。收集、整理了基础诗词库,创建了项目专有的诗词意象图像数据集 NID(nature image dataset),其中共有 9 大类;将在 ImageNet 图像数据集上训练好的 EfficientNet 模型迁移到 NID 中,对 NID 进行特征提取和图像标签匹配度的权值计算,结合每种图像类别训练得到的不同模型权重,融合 9 种模型权重,部署为一个多 EfficientNet 融合网络模型;最后对比了多种深度学习模型在 NID 上的表现性能。实验结果表明,多 EfficientNet 融合网络模型能够较为准确地解析图像,得到具有区分性的分类特征,并对 NID 的分类效果明显,收敛速度更快,精确率更高,符合项目中对诗词搜索的要求。

**关键词:**迁移学习;图像分类;Nature Image Dataset/NID;特征提取;多 EfficientNet 融合网络

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2021)07-0215-06

doi:10.3969/j.issn.1673-629X.2021.07.036

## Research on Application of Image Classification Based on Transfer Learning in Poetry

WU Su-wen, ZHAO Hui-jie, LIU Xin, WANG Jia-hao

(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 451191, China)

**Abstract:** There are rich cultural connotations in traditional Chinese poetry. In order to search for the most suitable poetry from the massive poetry dictionary, analyze the content of the image, extract the key words of the image characteristics, combined with the project requirements, an EfficientNet fusion network image classification algorithm based on transfer learning is proposed. The basic poetry library is collected and sorted out, and the project's proprietary poetry image dataset NID (nature image dataset) is created, which has nine categories. The EfficientNet model, which is trained on the ImageNet image dataset, is migrated to NID to perform feature extraction and image label matching weight calculation. Combined with the different model weights obtained from each image category training, the nine model weights are merged and deployed into a multi-EfficientNet fusion network model. Finally, the performance of various deep learning models on NID is compared. The experiment shows that the multi-EfficientNet fusion network model can analyze images more accurately, obtain distinguishing classification features, and have obvious classification effects on NID, with faster convergence speed and higher accuracy, which meets the requirements for poetry search in the project.

**Key words:** transfer learning; image classification; nature image dataset/NID; feature extraction; multi-EfficientNet fusion network

### 0 引言

图像分类是将不同类别的图片分开的图像处理方法。它在工业、农业、医学等领域都起着重要的作用。图像识别在诗词搜索中的应用研究是根据图像中的主要特征,例如花、鸟、山来进行图像分类,得到关键词,通过关键词在已经创建好的数据库中搜索诗词,得到搜索结果。当前,许多诗词应用系统已投入市场,例如西窗烛、古诗词典和古诗词分类赏析等应用软件,这类

软件为用户提供了一个搜索诗词的平台,便于用户快速获取到诗词信息。考虑到现有的搜索方式较为单一,且许多用户联想到诗词的情景各有不同,提出一种新型应用系统,通过对用户输入的图像进行特征提取,将特征提取后的分类结果进行整合,从已有的诗词库中自动匹配,输出与图像意境匹配最佳的诗词推荐给用户。

目前,图像与诗词匹配的算法日新月异,国内外致

收稿日期:2020-09-30

修回日期:2021-01-30

基金项目:国家自然科学基金资助项目(61673404);教育部产学合作协同育人(201901093055)

作者简介:武苏雯(1995-),女,硕士研究生,CCF 会员(C5317G),研究方向为人工智能与图像处理;赵慧杰,副教授,博士,研究方向为计算智能、图像语义解析、物联网移动应用。

力于这一领域的研究也取得了一定的研究成果。文献[1]通过学习深度耦合的视觉诗意嵌入模型与基于RNN的联合学习生成器结合进行多对抗训练,实现跨模态相关性和诗意的相关性,提出了一种由图像自动生成英文诗歌的模型。文献[2]通过改进的VGG16模型,预测输入图形的中文预测结果,然后根据RNN的预测结果创作一首诗歌。文献[3]将深度卷积网络迁移应用到小规模花卉数据集上,在进行模型微调后可以取得较好的识别结果。文献[4]通过使用EfficientNet和MixNet构建了一个专家系统,能够准确、快速识别水果的种类。文献[5]提出了基于自然性和视觉特征通道的场景分类方法,用Gist模型在Oliva等<sup>[6]</sup>提出的图像库依据其语义类别进行分类,得到了良好的分类效果。文献[7]通过使用改进的InceptionV3网络以及改进的激活函数,实现了大规模花卉图像数据集的分类。

虽然深度学习在单一类别图像的特征提取和识别上有一定的应用,但在多种类、多场景的图像分类实验中泛化能力较差。考虑到复杂自然场景、物种类别的分类,该文整合了9类63种不同的图像,选用一种基于迁移学习的多EfficientNet融合网络模型,将训练好的9种不同模型部署到应用系统中,实现图像分类在

诗词搜索中的应用研究。

## 1 相关工作

### 1.1 图像数据集

由于项目的特殊需求,本实验专门构建了意象图像数据NID(nature image dataset),图像数据集NID是通过从标准数据集(例如Flower102、ImageNet、CIFAR-10、SUN397-15场景数据集、scene15category等)、百度图片库中选取的。删除无关、重复的图片以及非JPG格式的图像数据后,得到了60种共70 206张有效图片。由于物种类别的多样性,结合项目中诗词意象的特点,人为将类型相近的物种归为一类,便于识别分类。共分为九种不同的类别,对每一种类别的物种分别进行识别,具体的图像数据集如表1所示。其中有45类的图像数据在1 000张以上,最多的一种包含1 800张图像,有15类的图像数据低于1 000张,最少的一类为478张。由于图像的分类精度会受到数据样本数量的影响,文中采用数据增强<sup>[8]</sup>的方法来进行数据扩充,主要包括随机裁减、随机旋转、亮度调节等操作。其中设置了训练集和测试集,分别占数据集的80%和20%。

表 1 NID 数据集

种类	种类	数量/种
兽类	熊、骆驼、猫、牛、鹿、狗、大象、狐狸、马、狮子、猴子、鼠、猪、老虎、狼	15
鸟类	鸡、鹤、乌鸦、鸭、鹰、鹅、黄鹌、喜鹊、孔雀、麻雀、燕子	11
花类	菊花、桂花、荷花、桃花、牡丹、玫瑰、向日葵	7
昆虫	蝴蝶、毛毛虫、蝉、龙、蜻蜓、蛇	6
景物类	建筑、草原、月景、山川、大海	6
树类	竹子、银杏、松树、柳树	4
水生动物	青蛙、金鱼、鲨鱼、乌龟、鲸鱼	5
天气	雨天、雾、晴天、雪	4
人	男人、女人	2
共计	—	60

### 1.2 EfficientNet 模型

EfficientNet 卷积神经网络<sup>[9]</sup>是通过使用复合系数网络的深度(卷积层的数量)、宽度(卷积层的通道数)、分辨率(输入特征的尺寸)中的所有维度进行均匀缩放的网络结构。因为单独缩放其中一个维度,模型会随着模型的增大,精度变小,在整个神经网络的缩放过程中要综合均衡所有的维度才是关键,得到协调的进度和效率。在EfficientNet模型中,其使用一组固定的缩放系数统一缩放网络深度、宽度和分辨率。若想使用 $2N$ 倍的计算资源,可将网络深度扩大 $\alpha^N$ 倍、宽度扩大 $\beta^N$ 倍、图像尺寸扩大 $\gamma^N$ 倍,这里的 $\alpha, \beta, \gamma$ 都是由原来的小模型上做微小的网格得到的最优解, $N$

则是表明资源的可用范围,调节模型的可用资源。图1为EfficientNet的设计思路,从三个方面同时扩充网络的特性。

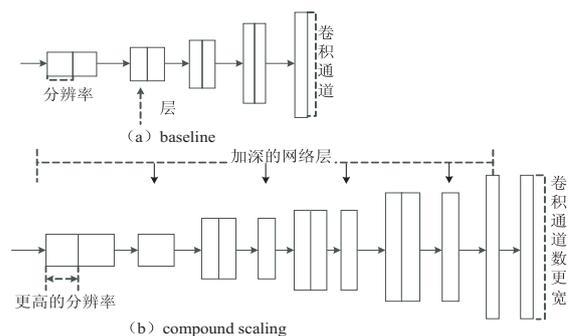


图 1 模型的扩展过程

神经网络架构采用搜索的方式,在限定存储空间和目标浮点运算次数的条件下,搜索到了一系列 EfficientNet B0 - B6 模型。文中采用轻量级的 EfficientNet-B0 网络作为主干网络,具体参数如表 2

所示。其中考虑到在 stem stage 层的小模型对前期底层特征的依赖,采取的是步长为 2 的卷积操作,可以有效地避免信息丢失的问题。

表 2 轻量级 EfficientNet-B0 主干网络参数

卷积阶段类型	参数	卷积核大小	通道数	分辨率 (像素×像素)
Stem Stage	Conv3×3,通道数为 32,步长为 2	—	—	224×224
MBConv Stage1	MBConv1	3	16	112×112
MBConv Stage2	[MBConv6]×2	3	24	112×112
MBConv Stage3	[MBConv6]×2	5	40	56×56
MBConv Stage4	[MBConv6]×2	3	80	28×28
MBConv Stage5	[MBConv6]×2	5	112	28×28
MBConv Stage7	[MBConv6]×2	5	192	14×14
MBConv Stage8	MBConv6	3	320	7×7
Pooling Stage	Conv1×1,池化,全连接	—	—	7×7

### 1.3 模型构建

文中使用的 EfficientNet-B0 基本网络作为主干网络,主要由 1 个 Stem 和 16 个大 Blocks 堆叠构成,其中 16 个大 Blocks 可以分为 1、2、2、3、3、4、1 个 Block。Block 的通用结构如图 2 所示,其总体的设计思路是 Inverted residuals 结构和残差结构,在 3×3 或者 5×5 网络结构前利用 1×1 卷积升维,在 3×3 或者 5×5 网络结构后增加了一个关于通道的注意力机制<sup>[10]</sup>,最后利用 1×1 卷积降维后增加一个大残差<sup>[11]</sup>边。

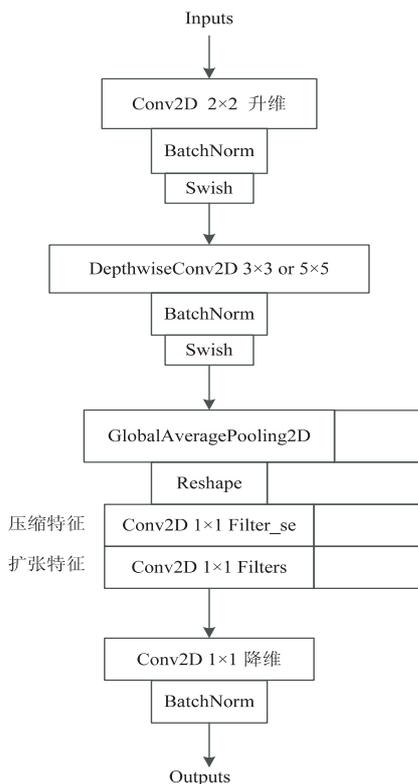


图 2 Block 的通用结构

### 1.4 迁移学习及微调

为了能加快模型的学习速度,本实验使用迁移学习。迁移学习<sup>[12-15]</sup>是将训练好的深度神经网络模型应用到另一个新的任务中,在这个过程中模型解冻、参数微调,训练得出适合新任务的模型。将已经在 ImageNet 数据集上训练好的 EfficientNet-b0 参数作为特征提取模型的初始参数,从而继承模型从 ImageNet 数据集上学习到的特征提取能力,再在此基础上进行微调。ImageNet 数据集的数据量十分庞大,可以赋予模型极强的特征提取能力以及泛化能力。该方法不仅可以节省大量的训练时间和计算资源,更可以在一定程度上削弱过拟合现象,使模型拥有比随机初始化参数更好的表现。

## 2 多 EfficientNet 模型的融合应用

### 2.1 基于迁移学习的 EfficientNet 图像分类算法

由于 NID 的图像数据量目前还不够大,难以将网络模型训练到理想的分类效果,而 ImageNet 数据集的图像数量高达 100 多万张, EfficientNet 模型先在 ImageNet 数据集上完成训练,对训练模型可以得到良好的识别率。因此文中采用基于迁移学习的 EfficientNet 图像分类算法来完成 NID 的分类任务。首先将 ImageNet 数据集在 EfficientNet 预训练模型上进行训练,在训练过程中调整模型超参数,使预训练模型能够较好地提取出图像的低层特征,例如边缘特征、纹理特征,使模型具有更佳的通用性,得到最佳识别结果,并记录相应的权重文件。再将 NID 其中一类图像随机选择 80% 用于训练,20% 用于测试,训练集可以再次对模型超参数进行微调,充分训练和调整模型的

语义层面的参数信息,提取出具有区分性的图像特征,最终得到最适合该类图像数据集的基于迁移学习的

EfficientNet 图像分类模型。具体图像分类顺序如图 3 所示。

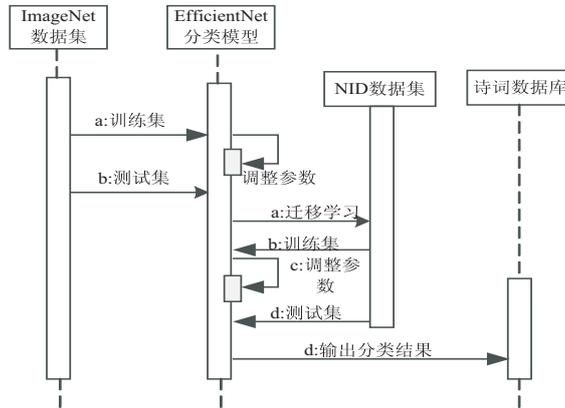


图 3 基于迁移学习的 EfficientNet 图像分类算法顺序

### 2.2 融合模型的图像分类算法

NID 中 9 大类图像分别通过基于迁移学习的 EfficientNet 图像分类算法得到 9 种不同参数的 EfficientNet 模型,其中包括兽类、鸟类、花类、昆虫、景物类、树类、水生动物、天气、人等。将这 9 种 EfficientNet 模型部署为一种融合模型,用于系统中图像处理的应用,图像识别与诗词搜索的系统框架如图

4 所示。其中的图像处理过程可以分为两个过程,第一个过程是图像裁剪、预处理,第二个过程是将预处理后的图像进行通道融合,选取相应的分类模型进行识别,将分类之后的输出结果传送到诗词数据库中进行下一步的搜索处理,图像分类过程的基本流程如图 4 所示。

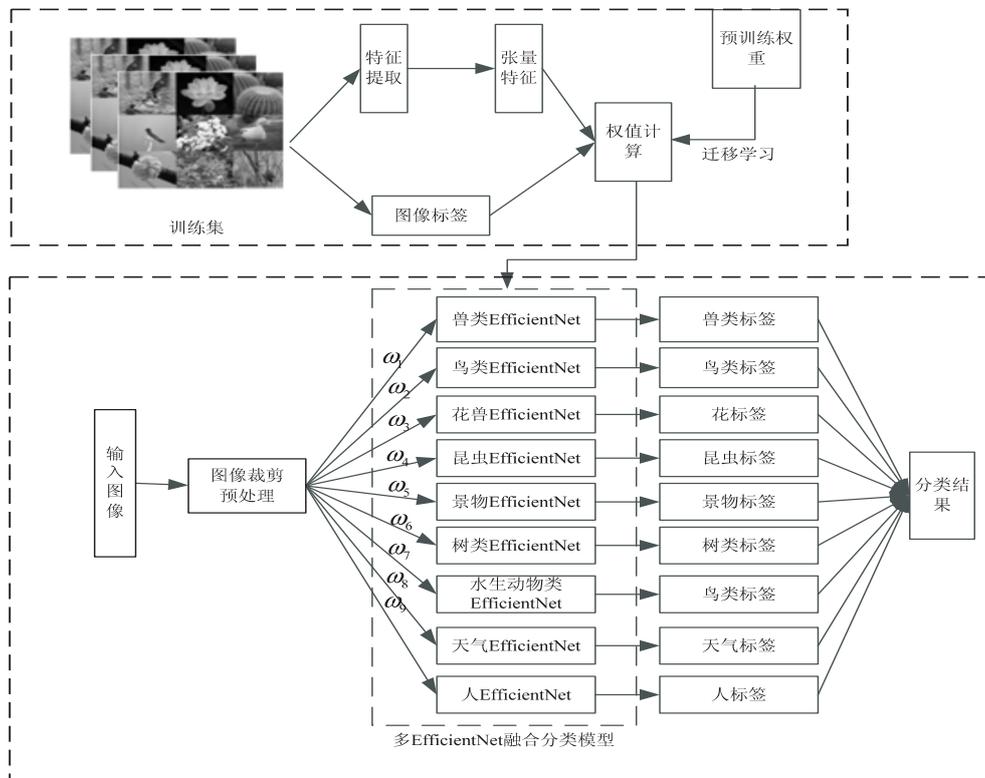


图 4 图像分类过程的基本流程

## 3 实验结果与分析

### 3.1 实验设置

#### 3.1.1 环境设置

本实验均在 kaggle 平台上进行,使用 python 编程

语言结合深度学习模型 TensorFlow 的高阶 API; tf.keras 实现神经网络的搭建、训练和测试。环境配置如下:操作系统为 Windows10;GPU 为 Nvidia Tesla P100-PCIE-16 GB 1.328 5 GHz;CPU Frequency 为 2.3 GHz;RAM 为 14 GB;Disk 为 5.2 GB。

3.1.2 参数设置

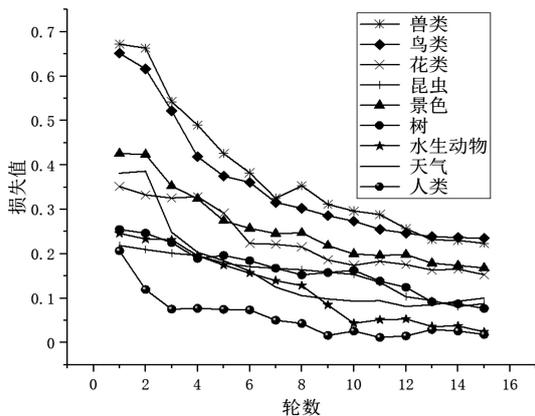
该文使用 TensorFlow 的深度学习库 Keras 构建基于迁移学习的图像分类模型,通过 EfficientNet 模型迁移学习,设置学习率 0.001,批尺寸 128,采用 AMSGrad 优化器及交叉熵损失函数,实现了对 NID 的特征提取和分类。

AMSGrad 算法针对 Adam 算法收敛性证明的缺陷,对 Adam 的更新规则做出了一些改进,在某些应用场景下拥有超过 Adam 算法的收敛速度。

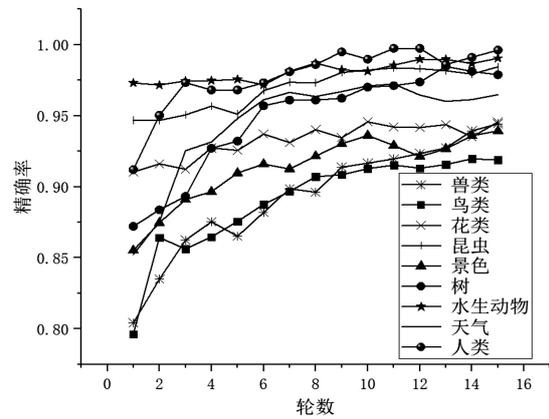
3.2 结果对比分析

3.2.1 多 EfficientNet 融合模型结果分析

实验过程中分别在 NID 上对 9 种模型进行训练,实验训练总轮数为 15 轮,在训练过程中计算出每一轮的平均损失值,所有的损失值都随着训练轮数的增加而减小,收敛速度较快,并未出现过拟合现象。根据记录下的每一种模型精度绘制出的变化图可以看出,模型精度也在随训练轮数的增加趋于稳定,其中每一类图像的分类精度如表 3 所示,9 种模型的损失值和精确率变化情况如图 5 所示。



(a) 9 种模型的损失值变化曲线



(b) 9 种模型的精确率变化曲线

图 5 评估参数变化趋势对比

表 3 NID 的分类结果

指标	兽类	鸟类	花类	昆虫	风景	树类	水生动物	天气	人类	平均精确率
损失值	0.239 6	0.235 1	0.153 1	0.087 9	0.168 2	0.77	0.024 5	0.100 5	0.062 1	—
精确率/%	94.43	91.88	94.55	98.44	93.93	97.88	99.05	96.48	99.62	96.25

由表 3 可知,文中方法图像分类结果均在 90% 以上,精确率最高的一类为水生动物,达到了 99.62%,相对来说精确率较低的一类为鸟类,精确率为 91.88%。

结合图 5 来看,基于迁移学习的 EfficientNet 的图像分类算法使 NID 的各类别图像的平均精确率达到了 96.25%。这足以说明在 NID 图像分类研究方面的有效性。文中采用的迁移学习思想提取了深度特征避免了模型过拟合,同时结合低层特征来充分表达图像信息,使得文中分类算法更稳定,鲁棒性更强。

3.2.2 不同模型实验结果分析

为了验证 EfficientNet 对于 NID 的分类效果优于其他模型,将 EfficientNet 与 Xception、Inception-V3、VGG16、ResNet50 四种模型作比较,在构建的花类物种数据集上进行对比实验,结果如表 4 所示。可以看出,文中算法的精确率分别较 Xception、Inception-V3、VGG16、ResNet50 模型高 5.11%、4.31%、13.32%、8.2%,且参数量为 11 574 063,相比其他四种算法的参数量大大减少,降低了运算复杂度,减少了运算时间,符合项目中对图像分类的要求。

表 4 EfficientNet 模型与其他模型的对比

方法	花类精确率/%	损失值	参数量
Xception	89.44	0.15	22 910 480
Inception-V3	90.24	0.20	23 851 784
VGG16	81.23	0.41	138 357 544
ResNet50	86.35	0.26	25 636 712
EfficientNet	94.55	0.15	11 574 063

#### 4 结束语

文中构建了一个大型的诗词图像数据集 NID,在此基础上提出了一种基于迁移学习的多 EfficientNet 融合的图像分类算法模型,并取得了良好的分类效果,各类图像的平均识别率为 96.25%,收敛速度快,且训练集和验证集之间并未出现过拟合现象。对比其他四个模型,实验结果表明,文中采用的模型识别结果稳定,且在 NID 上的识别率最高。

本研究仅对数据集中包含的 60 种意象景物进行了分类识别实验,要建立有效的、用于诗词搜索的图像识别系统,就要有更强大的训练数据作为基础,并且在提升模型的识别率,同时持续补充诗词库,以保证搜索到更全面、更精确的诗词。因此,在后期的工作中将补充图像数据研究种类、丰富诗词库,并增加图像识别模型识别的特征目标,以便获取更多的图像数据信息,以建立更准确的诗词搜索系统。

#### 参考文献:

- [1] LIU Bei, FU Jianlong, KATO M P, et al. Beyond narrative description: generating poetry from images by multi-adversarial training [C]//Proceedings of the 26th ACM international conference on multimedia. Seoul, Republic of Korea: Association for Computing Machinery, 2018: 783-791.
- [2] WANG Xiaoyu, ZHONG Xian, LI Lin. Generating Chinese classical poems based on images [C]//Proceedings of the international multiconference of engineers and computer scientists. Hong Kong: [s. n.], 2018.
- [3] 曹晓杰, 么 娆, 严雨灵. 应用迁移学习的卷积神经网络花卉图像识别 [J]. 计算机应用与软件, 2020, 37(8): 142-148.
- [4] DUONG L T, NGUYEN P T, SIPIO C D, et al. Automated fruit recognition using EfficientNet and MixNet [J]. Computers and Electronics in Agriculture, 2020, 171: 105326.
- [5] 赵宏伟, 李明昭, 刘 静, 等. 基于自然性和视觉特征通道的场景分类 [J]. 吉林大学学报: 工学版, 2019, 49(5): 1668-1675.
- [6] OLIVA A, SCHYNS P G. Diagnostic colors mediate scene recognition [J]. Cognitive Psychology, 2000, 41(2): 176-210.
- [7] 吴 迪, 侯凌燕, 刘秀磊, 等. 一种改进的深度学习的花卉图像分类 [J]. 河南大学学报: 自然科学版, 2019, 49(2): 192-203.
- [8] 杨旺功, 准永建. 基于深度卷积生成对抗网络的花朵图像增强与分类 [J]. 计算机科学, 2020, 47(6): 176-179.
- [9] TAN MINGXING, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [C]//Proceedings of the 36th international conference on machine learning. California: PMLR, 2019: 6105-6114.
- [10] 雷鹏程, 刘 丛, 唐坚刚, 等. 分层特征融合注意力网络图像超分辨率重建 [J]. 中国图象图形学报, 2020, 25(9): 1773-1786.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770-778.
- [12] 吴 建, 许 镜, 丁 韬. 基于集成迁移学习的细粒度图像分类算法 [J]. 重庆邮电大学学报: 自然科学版, 2020, 32(3): 452-458.
- [13] 谢小红, 陆建波, 李文韬, 等. 基于迁移学习的服装图像分类模型研究 [J]. 计算机应用与软件, 2020, 37(9): 88-93.
- [14] 贾宇霞, 樊帅昌, 易晓梅. 基于显著性增强和迁移学习的鱼类识别研究 [J]. 渔业现代化, 2020, 47(1): 38-46.
- [15] 石祥滨, 房雪键, 张德园, 等. 基于深度学习混合模型迁移学习的图像分类 [J]. 系统仿真学报, 2016, 28(1): 167-173.