



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 图形学报

2020
05
VOL.25

ISSN1006-8961
CN11-3758/TB



智能疫
“小珈”
助力



第25卷第5期（总第289期）

2020年5月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 顾行发

编辑出版《中国图象图形学报》编辑出版委员会

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@radi.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China

Zip code 100190

E-mail jig@radi.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 60.00元



多相图像分割Vese-Chan模型连续最大流方法(第0926页)



保持区域平滑的3维动画形状高效编辑(第1019页)



树形结构卷积神经网络优化的城区遥感图像语义分割(第1043页)

学者观点

多源空一谱遥感图像融合方法进展与挑战

肖亮, 刘鹏飞, 李恒 0851

综述

中国图像工程: 2019

章毓晋 0864

图像处理和编码

相同编码参数HEVC视频重压缩检测

潘鹏飞, 姚晔, 王慧 0879

密集连接卷积网络图像去模糊

吴迪, 赵洪田, 郑世宝 0890

密集网络图像哈希检索

王亚鸽, 康晓东, 郭军, 李博, 张华丽, 刘汉卿 0900

图像分析和识别

域自适应城市场景语义分割

张桂梅, 潘国峰, 刘建新 0913

多相图像分割Vese-Chan模型连续最大流方法

王洁, 潘振宽, 魏伟波, 徐子森 0926

结合注意力机制和多属性分类的行人再识别

郑鑫, 林兰, 叶茂, 王丽, 贺春林 0936

多分辨率特征注意力融合行人再识别

沈庆, 田畅, 王家宝, 焦珊珊, 杜麟 0946

全卷积网络电线识别方法

刘嘉玮, 李元祥, 龚政, 刘心刚, 周拥军 0956

双核压缩激活神经网络艺术图像分类

杨秀芹, 张华熊 0967

图像理解和计算机视觉

高效检测复杂场景的快速金字塔网络SPNet

李鑫泽, 张轩雄, 陈胜 0977

结合自底向上注意力机制和记忆网络的视觉问答模型

闫茹玉, 刘学亮 0993

统筹图像变换与缝合线生成的无参数影像拼接

高炯笠, 吴军, 刘祺昌, 徐刚 1007

计算机图形学

保持区域平滑的3维动画形状高效编辑

冼楚华, 杨煜, 黄俊贤, 李桂清 1019

医学图像处理

肌骨超声图像特征检测及拼接

颜焕欢, 张培镇, 王伊依, 李璇, 阳维, 王青 1032

遥感图像处理

树形结构卷积神经网络优化的城区遥感图像语义分割

胡伟, 高博川, 黄振航, 李瑞瑞 1043

融合累积变异比和集成超限学习机的高光谱图像分类

尹玉萍, 魏林, 刘万军 1053

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Continuous max-flow method for multiphase segmentation Vese-Chan model(P0926)



Efficient 3D animation shape manipulation with region smoothness preservation(P1019)



Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network (P1043)

Scholar View

Progress and challenges in the fusion of multisource spatial-spectral remote sensing images

Xiao Liang, Liu Pengfei, Li Heng 0851

Review

Image engineering in China: 2019

Zhang Yujin 0864

Image Processing and Coding

Detection of double compression for HEVC videos with the same coding parameters

Pan Pengfei, Yao Ye, Wang Hui 0879

Motion deblurring method based on DenseNets

Wu Di, Zhao Hongtian, Zheng Shibao 0890

Image Hash retrieval with DenseNet

Wang Yage, Kang Xiaodong, Guo Jun, Li bo, Zhang Huali, Liu Hanqing 0900

Image Analysis and Recognition

Domain adaptation for semantic segmentation based on adaption learning rate

Zhang Guimei, Pan Guofeng, Liu Jianxin 0913

Continuous max-flow method for multiphase segmentation Vese-Chan model

Wang Jie, Pan Zhenkuan, Wei Weibo, Xu Zisen 0926

Improving person re-identification by attention and multi-attributes

Zheng Xin, Lin Lan, Ye Mao, Wang Li, He Chunlin 0936

Multi-resolution feature attention fusion method for person re-identification

Shen Qing, Tian Chang, Wang Jiabao, Jiao Shanshan, Du Lin 0946

Power line recognition method via fully convolutional network

Liu Jiawei, Li Yuanxiang, Gong Zheng, Liu Xingang, Zhou Yongjun 0956

Art image classification with double kernel squeeze-and-excitation neural network

Yang Xiuqin, Zhang Huaxiong 0967

Image Understanding and Computer Vision

Snap Pyramid Network: Real-time complex scene detecting system

Li Xinze, Zhang Xuanxiong, Chen Sheng 0977

Visual question answering model based on bottom-up attention and memory network

Yan Ruyu, Liu Xueliang 0993

Stitching of parametric-free images based on coordinated image transformation and seam-line generation

Gao Jiongli, Wu Jun, Liu Qichang, Xu Gang 1007

Computer Graphics

Efficient 3D animation shape manipulation with region smoothness preservation

Xian Chuhua, Yang Yu, Huang Junxian, Li Guiqing 1019

Medical Image Processing

Feature detection algorithm of musculoskeletal ultrasound image and its application of image stitching

Yan Huanhuan, Zhang Peizhen, Wang Yinong, Li Xuan, Yang Wei, Wang Qing 1032

Remote Sensing Image Processing

Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network

Hu Wei, Gao Bochuan, Huang Zhenhang, Li Ruirui 1043

Ensemble extreme learning machine with cumulative variation quotient for hyperspectral image classification

Yin Yuping, Wei Lin, Liu Wanjun 1053

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2020)05-0967-10

论文引用格式: Yang X Q and Zhang H X. 2020. Art image classification with double kernel squeeze-and-excitation neural network. Journal of Image and Graphics, 25(05): 0967-0976 (杨秀芹, 张华熊. 2020. 双核压缩激活神经网络艺术图像分类. 中国图象图形学报, 25(05): 0967-0976) [DOI: 10.11834/jig.190245]

双核压缩激活神经网络艺术图像分类

杨秀芹, 张华熊

浙江理工大学信息学院, 杭州 310018

摘要: 目的 为了充分提取版画、中国画、油画、水彩画和水粉画等艺术图像的整体风格和局部细节特征, 实现计算机自动分类检索艺术图像的需求, 提出通过双核压缩激活模块(double kernel squeeze-and-excitation, DKSE)和深度可分离卷积搭建卷积神经网络对艺术图像进行分类。方法 根据 SKNet(selective kernel networks)自适应调节感受野提取图像整体与细节特征的结构特点和 SENet(squeeze-and-excitation networks)增强通道特征的特点构建 DKSE 模块, 利用 DKSE 模块分支上的卷积核提取输入图像的整体特征与局部细节特征; 将分支上的特征图进行特征融合, 并对融合后的特征图进行特征压缩和激活处理; 将处理后的特征加权映射到不同分支的特征图上并进行特征融合; 通过 DKSE 模块与深度可分离卷积搭建卷积神经网络对艺术图像进行分类。结果 使用本文网络模型对有无数据增强(5 类艺术图像数据增强后共 25 634 幅)处理的数据分类, 数据增强后的分类准确率比未增强处理的准确率高 9.21%。将本文方法与其他网络模型和传统分类方法相比, 本文方法的分类准确率达到 86.55%, 比传统分类方法高 26.35%。当 DKSE 模块分支上的卷积核为 1×1 和 5×5 , 且放在本文网络模型第 3 个深度可分离卷积后, 分类准确率达到 87.58%。结论 DKSE 模块可以有效提高模型分类性能, 充分提取艺术图像的整体与局部细节特征, 比传统网络模型具有更好的分类准确率。

关键词: 艺术图像分类; 深度可分离卷积; 卷积神经网络; 整体特征; 局部细节特征

Art image classification with double kernel squeeze-and-excitation neural network

Yang Xiuqin, Zhang Huaxiong

School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

Abstract: **Objective** The development of online digital media technology has promoted the sharing and spreading of natural art images. However, given the increasing number of art images, effective classification and retrieval are urgent problems that need to be solved. In the face of massive art image data, problems may occur in traditional manual feature extraction methods, such as tagging errors and subjective tagging. Moreover, the professional requirements of classifiers are relatively high. Convolutional neural networks (CNNs) are widely used in image classification because of its automatic feature extraction characteristics. Most of these network models are used for feature extraction in key areas of photographed images. However, natural art images are different from photographed images. Specifically, the distribution of overall style features and local detail features is evidently uniform. Selective kernel networks (SKNet) can adaptively adjust their receptive field size according to the input image to select multi-scale spatial information. However, the softmax gating mechanism in the mod-

收稿日期: 2019-06-03; 修回日期: 2019-10-16; 预印本日期: 2019-10-23

基金项目: 浙江省服装个性化定制协同创新中心项目(浙教高科[2016]63号); 中国纺织工业联合会应用基础研究项目(J201802)

Supported by: Collaborative Innovation Center for Garment Personalized Customization of Zhejiang, China (63-2016)

ule only strengthens the dependence between the channels of the feature map after the convolution operation of the receptive field with large response to stimulus. It also ignores the role of local detail features. Squeeze-and-excitation networks (SENet) can enhance the features in different channels but cannot extract the overall features and local detail features of the input. To fully extract and enhance the overall style features and local detail features of art images and realize the automatic classification and retrieval of art images, we combine the characteristics of SKNet and SENet to build a block called double kernel squeeze-and-excitation (DKSE) module. DKSE blocks and depthwise separable convolutions are mainly used to construct a CNN to classify art images. **Method** SKNet can capture the overall features and local detail features with different scales. According to the multi-scale structural characteristics of SKNet, we build the DKSE module with two branches. Each branch has a different convolutional kernel to extract the overall features and the local detail features and fuse the feature maps obtained by convolution operation. Then, according to the idea of compression and excitation in SENet, the fusion feature map spatial information is compressed into the channel descriptor by global average pooling (GAP). After the GAP operation, 1×1 convolutional kernel is used to compress and activate the feature map. The weight of normalization between (0, 1) is obtained through sigmoid gating mechanism. The weight is rescaled to the feature map of the different branches. The final output of the block is obtained by fusing the rescaled feature maps. Thus, more representative art image characteristics are extracted. In this study, we choose engraving, Chinese painting, oil painting, opaque watercolor painting and watercolor painting for classification. To enhance the data of artistic images, the images with high resolution of art images are artificially extracted and cut randomly into 299×299 pixels. The modules with rich style information are then selected. After the data augmentation, a total of 25 634 images under the five kinds of art images are obtained. The CNN is constructed by multiple DKSE modules and depthwise separable convolutions to classify the five kinds of art images. In all the experiments, 80% of the art images under each kind are randomly selected as the training sets, while the remaining 20% of the datasets are used as verification sets. Our CNN is implemented in a Keras frame. The input images are resized to 299×299 pixels for training. The Adam optimizer is used in the experiments. The initial value of learning rate is 0.001, a mini-batch size of 32 is observed, and the total of training epochs is 120. In the training process, the training sets rotate randomly from 0° to 20° , and the horizontal or vertical direction is randomly shifted between 0% and 10% and flipped randomly to enhance the generalization ability of the proposed CNN. The learning rate is decreased by a factor of 10 if the accuracy of training sets does not improve after three training cycles. **Result** Our network model is used to classify the data with or without data enhancement processing. The accuracy of art image classification after data augmentation is 9.21% higher than that of unenhanced processing. Compared with other network models and traditional art image classification methods, the classification accuracy of our method is 86.55%, more than 26.35% higher than that of traditional art image classification methods. Compared with Inception-V4 networks, the number of parameters is approximately 33% of the number of Inception-V4 parameters, and the time spent is approximately 25%. In this study, we place the proposed DKSE module in three different positions of the network and then verify the influence of DKSE on the classification results. When the module is placed at the third depthwise separable convolution of the network model, the reduction ratio is set to 4 and the convolution kernel sizes on the branches are 1×1 and 5×5 . Moreover, the classification accuracy is 87.58%, which is 1.58% higher than that of the other eight state-of-the-art network models. The classification accuracy of the reduction ratio of 4 is superior to the reduction ratio set to 16. We use gradient-weighted class activation mapping (Grad-CAM) algorithm with our network model, ours + SK model and ours + SE model, to visualize the overall features and local detail features of each kind of art images. Experimental results show that compared with the other two network models, our network model can fully extract the overall features and local detail features of art images. **Conclusion** Experimental results show that the proposed DKSE module can effectively improve the classification performance of the network model and fully extract the overall features and local detail features of the art images. The network model in this study has better classification accuracy than do the other CNN models.

Key words: art image classification; depthwise separable convolution; convolutional neural network (CNN); overall feature; local detail feature

0 引言

自然艺术绘画是人们追求生活品质和精神品质的体现,能够表达语言和文字无法表达的情感,包括版画、中国画、油画、水粉画、水彩画等,如图1所示。版画是艺术家通过构思创造、雕刻制版和印刷得到的艺术作品,以丰满密集和萧疏简淡的构图方式表现主题风格,简洁、单纯、颜色数目有限、亮度和纯度较低、颜色间的过渡较快。中国画使用毛笔蘸水、墨或彩在绢或纸上作画,讲究气韵、应物象形和骨法用笔,其中骨法表现用笔的艺术性,包含笔力、力感、表现结构等(亓玉权,2009)。中国画不注重画面背景的渲染,画中多留有空白,且以墨代色,使墨色产生丰富而细微的色度变化,也就是“墨分五色”,运笔疾缓及笔触长短大小的不同,使笔墨技巧千变万化,明暗调子丰富多彩。油画不同于中国画,注重写实,画家通过冷暖色调对比、颜色明暗强度对比、颜料厚薄层次对比进行光感的创造。油画具有不透明性,颜色丰富饱和度高。水彩画是以水为媒介通过调和

透明颜料作画的一种绘画形式,颜料的透明性使画面通透明澈,水的流动性增强画面淋漓酣畅、自然洒脱的意趣。水粉画是用水调和粉质颜料作画的一种绘画形式,颜色透明性介于水彩画和油画之间。水粉画在湿的时候颜色饱和度很高,干后颜色失去光泽,饱和度下降。油画、水彩画和水粉画都可以表现出颜色丰富多彩、纹理粗犷、色块明显的特性,但是油画中明暗色调的交互以及对光影的运用使画面整体的空间感和立体感更加强烈,细节上纹理丰富,颜料反复堆叠形成的线条增强油画的语义信息,透明度最低;水粉画透明度次之,水粉干后,画面出现明显粉状色块,立体感不强,色彩饱和度低于油画;水彩画透明度高,能表达出清新明快的质感,画面上水渍湿痕是区别它们的关键特征之一。中国画与水彩画在审美表达、意向气韵方面有异曲同工之处,但是国画整体上多用墨画制,线条丰富,细节处以点、线、面的形式描述对象的形貌、骨法、光暗及情态神韵,画面整体颜色深沉;水彩画整体颜色丰富鲜明,明快润泽,亮度大,细节处不注重以线作画,多借助“水”与“色”的相互作用表现画面的情态神韵。



(a) 版画



(b) 中国画



(c) 油画



(d) 水粉画



(e) 水彩画

图1 5类艺术图像

Fig.1 Five types of artistic images

((a) engravings; (b) Chinese paintings; (c) oil paintings; (d) opaque watercolor paintings; (e) watercolor paintings)

目前,艺术图像管理的研究工作主要是根据题材与表现手法对画作分类、根据创作风格对画家分类以及艺术画像真假鉴别等,基于多类艺术图像风

格特点的分类研究较少。不同画家的运笔力道不同,画中线条的粗细可以展现画家的绘画风格。Li和Wang(2004)提出用2维多分辨率隐马尔可夫混

合模型对图像的大部分区域进行分析,捕获关键区域笔画特征,对画家分类和比较。国画多以墨代色, Sheng 和 Jiang(2013)使用灰度直方图提取水墨画整体风格特征,并用 Sobel 边缘检测方法获得笔画风格丰富的局部细节特征,使用信息熵融合算法实现画家分类。Shen(2009)将图像分成 4×4 的子块,对各个子块分别提取颜色、纹理、形状等整体特征和局部纹理特征并输入基于径向基函数(radial basis function, RBF)的神经网络进行训练,然后对 RBF 神经网络输出的向量进行汉明距离计算以实现对输入的图像进行所属画家分类。Sun 等人(2016)采用蒙特卡洛凸壳特征选择模型和支持向量机对传统中国画进行特征描述。王征等人(2013)根据国画中纹理、颜色、形状等特征,采用传统方法进行监督式异构稀疏特征提取,但是特征只有 96 维,不足以描述国画的整体特征。Jiang 等人(2006)根据国画中工笔画与写意画在表现手法上的区别,利用图像低层的纹理、形状、颜色、边缘特征,对中国传统工笔画和写意画进行分类。高峰等人(2017)利用尺度不变特征变换(scale invariant feature transform, SIFT)特征检测子和边缘检测得到图像关键区域,通过关键区域特征和邻域内部差异性的描述,采用级联分类器分析得出工笔画和写意画在表现手法上的不同。但是,艺术展现的特征往往以一种有机的形式组合而成,组合技巧很难概括,不同种类艺术图像在颜色、纹理、线条等特征方面存在相似性,使用传统方法提取艺术图像的纹理、颜色、线条等特征并不能充分区分每类艺术图像的风格特点。

卷积神经网络(convolutional neural networks, CNNs)在提取图像整体特征以及局部细节特征上取得了很好的成果。InceptionV4 模块(Szegedy 等, 2017)分支中的不同卷积核提取输入图像的整体与细节特征,但这些模块大多考虑空间维度信息,没有考虑通道间的依赖性,对提取的特征没有进一步增强处理。SENet(squeeze-and-excitation networks)(Hu 等, 2018)中的 SE(squeeze-and-excitation)模块融合空间维度和通道间的依赖关系,强化提取的有用特征,抑制无关的特征信息。SKNet(selective kernel networks)(Li 等, 2019)中的 SK(selective kernel)模块使网络可以根据输入信息的多个尺度自适应调节感受野大小,提取图像的细节与整体特征。SE 与 SK 模块如图 2 所示。本文根据 SE 和 SK 模块的功

能特点,构建双核压缩激活模块(double kernel squeeze-and-excitation, DKSE)。DKSE 模块分支上不同大小的卷积核提取图像的整体与细节特征,并进行融合、压缩和激活,然后将提取的风格特征与各个分支上的特征图进行加权映射,最后将映射后的不同特征图进行特征融合,从而强化不同卷积核提取出的特征信息。本文用 DKSE 网络模块和深度可分离卷积搭建卷积神经网络,对 5 类艺术图像进行分类。

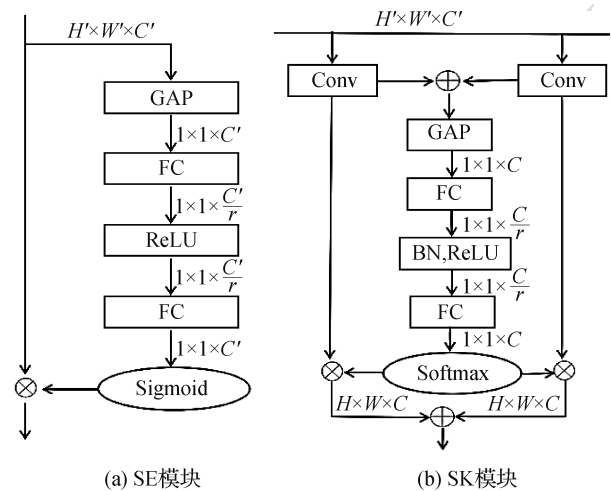


图 2 SE 模块与 SK 模块

Fig. 2 SE module and SK module ((a) SE; (b) SK)

1 基于卷积神经网络的艺术图像分类

1.1 DKSE 模块

本文双核压缩激活模块(DKSE)结合了 SE 模块和 SK 模块的特点,能够更好地增强提取的艺术图像整体风格特征与局部细节特征,由 split、squeeze、excitation 和 scale 4 个子模块组成,如图 3 所示,表达式为

$$V = \sum_{i=1}^N U_i \cdot F_{ex}(F_{sq}(F_{gp}(U))) \quad (1)$$

式中, U 是 DKSE 模块分支上融合后的特征图, $F_{gp}(\cdot)$ 表示全局平均池化操作(global average pooling, GAP), $F_{sq}(\cdot)$ 表示通道压缩处理, $F_{ex}(\cdot)$ 为通道特征激活操作, N 为 DKSE 模块分支数,本文取 $N=2$ 。

1) 子模块 split。对一个中间特征图 $X \in \mathbf{R}^{H' \times W' \times C'}$, 使用两个不同大小的卷积核卷积映射,即

$$F_1: X \rightarrow U_1 \in \mathbf{R}^{H \times W \times C}$$

$$F_2: X \rightarrow U_2 \in \mathbf{R}^{H \times W \times C}$$

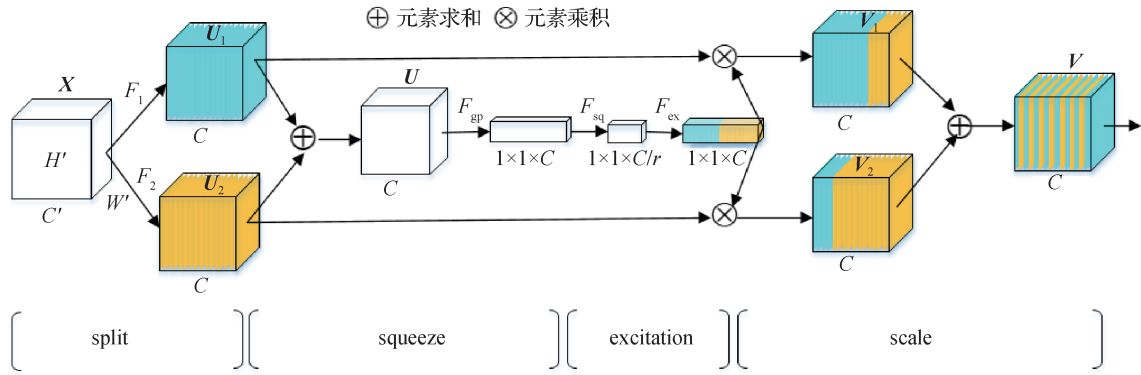


图3 DKSE 模块

Fig. 3 DKSE module

式中, H' 、 W' 和 C' 分别表示特征图 X 的高、宽和通道数。 F_1 和 F_2 是分别经过卷积核卷积映射、批标准化(batch normalization, BN) (Ioffe 和 Szegedy, 2015) 和 ReLU 激励函数(Nair 和 Hinton, 2010)处理的映射过程。 H 、 W 和 C 分别表示经过 F_1 和 F_2 操作后特征图的高、宽和通道数。卷积过滤器 $w = [w_1, w_2, \dots, w_c, \dots, w_{c'}]$, 其中 w_c 表示第 c 个过滤器的参数。在第 i 个分支上, 每个过滤器对中间特征图 X 的卷积映射公式为

$$u'_{ic} = w_{ic} \times X = \sum_{k=1}^{c'} w_{ic}^k \times X^k + b_{ic}$$

$$u_{ic} = \delta(\mathcal{N}(u'_{ic})), i = 1, \dots, N \quad (2)$$

式中, c' 表示过滤器和特征图的通道数, b_{ic} 表示偏置, $\mathcal{N}(\cdot)$ 表示批标准化处理, $\delta(X') = \max(0, X')$ 是 ReLU 函数操作, $X' = \mathcal{N}(u'_{ic})$, $U_i = [u_{i1}, u_{i2}, \dots, u_{ic}]$ 。

2) 子模块 squeeze。经过 split 操作, 得到两个新的特征图 U_1 和 U_2 , 通过对应元素求和, 融合两个特征图的特征信息, 即

$$U = U_1 + U_2 \quad (3)$$

融合后的特征图 U 结合了 U_1 和 U_2 的特征信息, 使用全局平均池化对每个特征通道中的全局空间特征节点进行池化处理。全局平均池化将特征图 U 空间信息压缩到 C 个通道描述符中, 生成描述特征通道信息的统计量 $S \in \mathbf{R}^C$ 。统计量 S 中第 c 个元素通过压缩 U 空间信息计算得到, 具体为

$$S_c = F_{sq}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (4)$$

3) 子模块 excitation。为增强从每类艺术图像

提取的风格特征, 减弱作用不大的特征信息, 通过 1×1 卷积操作将全局平均池化后的特征图通道维度降低到原通道的 $1/r$, r 为下降率, 然后经过 BN 处理和 ReLU 函数激活后, 再通过 1×1 卷积操作将通道数升到原先的数目, 最后通过一个 Sigmoid 的门机制获得 $(0, 1)$ 之间归一化的权重, 得到每类艺术图像的风格特征信息

$$Z = F_{ex}(S) = \sigma(W_2 F_{sq}(S)) = \sigma(W_2 \delta(\mathcal{N}(W_1 S))) \quad (5)$$

式中, $\sigma(X'') = (1 + e^{-X''})^{-1}$ 是 Sigmoid 激活函数, $X'' = W_2 \delta(\mathcal{N}(W_1 S))$, $W_1 \in \mathbf{R}^{\frac{C}{r} \times C}$, $W_2 \in \mathbf{R}^{C \times \frac{C}{r}}$, r 表示特征图通道降维比例。

4) 子模块 scale。通过 squeeze 和 excitation 操作得到的权重 Z 是经过主要特征筛选、抑制次要特征的图像特征信息, 分别与 U_1 和 U_2 进行加权求和, 则

$$V_c = u_{1c} \cdot Z_c + u_{2c} \cdot Z_c \quad (6)$$

式中, $V_c \in \mathbf{R}^{H \times W}$ 表示特征图 V 的第 c 个通道的特征信息, $V = [V_1, V_2, \dots, V_c]$ 。

1.2 卷积神经网络结构

参考 MobileNetV1 网络 (Howard 等, 2017) 结构, 使用深度可分离卷积及 DKSE 模块搭建卷积神经网络, 第 1 层使用空洞卷积 (Yu 和 Koltun, 2016) 提取原始艺术图像的特征。空洞卷积与普通卷积相比, 在保持计算量不增加的情况下, 具有更大的网络感受野, 可以保持更多内部数据结构和原图信息。深度可分离卷积由深度卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution) 组成。DKSE 模块嵌入深度可分离卷积中操作公式为

$$Y(X) = [Y_1 \cdot Y_2 \cdot Y_3](X) \quad (7)$$

式中, $Y_1: \mathbf{R}^{H' \times W' \times C'} \rightarrow \mathbf{R}^{\frac{H'}{s} \times \frac{W'}{s} \times C'}$ 为深度卷积操作, $Y_2: \mathbf{R}^{\frac{H'}{s} \times \frac{W'}{s} \times C'} \rightarrow \mathbf{R}^{\frac{H'}{s} \times \frac{W'}{s} \times tC'}$ 为逐点卷积操作, $Y_3: \mathbf{R}^{\frac{H'}{s} \times \frac{W'}{s} \times tC'} \rightarrow \mathbf{R}^{H \times W \times C}$ 为 DKSE 模块操作, s 表示深度卷积对特征图维度进行降维处理, t 表示逐点卷积对特征图通道数的处理。每次深度卷积和逐点卷积操作后, 再进行 BN 和 ReLU 激励函数处理。

表 1 是在 DKSE 分支上取 3×3 和 5×5 进行普通卷积的操作情况, 为了减少训练中的过拟合现象, 在逐点卷积操作中使用 L2 正则化方法, 并在最后进行 GAP 处理的前后进行 dropout (Krizhevsky 等, 2012) 处理。表 1 中, dilated_conv 表示空洞卷积, $d = 2$ 表示扩张率为 2, depthwise_conv 为深度卷积, dep_separable_conv 表示深度可分离卷积, $s1$ 表示步长为 1 像素, 依次类推。

表 1 艺术图像分类卷积神经网络模型
Table 1 Convolutional neural network
model of art images classification

ID	输入尺寸	卷积类型	卷积核大小
1	$299 \times 299 \times 3$	dilated_conv/s1	$3 \times 3, 32, d = 2$
2	$299 \times 299 \times 32$	depthwise_conv/s3	3×3
3	$99 \times 99 \times 32$	dep_separable_conv/s1	$3 \times 3, 64$
4	$99 \times 99 \times 64$	dep_separable_conv/s2	$3 \times 3, 128$
5	$50 \times 50 \times 128$	dep_separable_conv/s1	$3 \times 3, 128$
6	$50 \times 50 \times 128$	DKSE_block/s1	$3 \times 3, 5 \times 5, 128, r = 4$
7	$50 \times 50 \times 128$	dep_separable_conv/s2	$3 \times 3, 256$
8	$25 \times 25 \times 256$	dep_separable_conv/s1	$3 \times 3, 256$
9	$25 \times 25 \times 256$	DKSE_block/s1	$3 \times 3, 5 \times 5, 256, r = 4$
10	$25 \times 25 \times 256$	dep_separable_conv/s2	$3 \times 3, 512$
11	$13 \times 13 \times 512$	dep_separable_conv/s1	$3 \times 3, 512$
12	$13 \times 13 \times 512$	DKSE_block/s1	$3 \times 3, 5 \times 5, 512, r = 4$
13	$13 \times 13 \times 512$	dep_separable_conv/s2	$3 \times 3, 1\ 024$
14	$7 \times 7 \times 1\ 024$	dep_separable_conv/s1	$3 \times 3, 1\ 024$
15	$7 \times 7 \times 1\ 024$	global average pool	7×7
16	$1 \times 1 \times 1\ 024$	fully connected	$1\ 024 \times 5$
17	$1 \times 1 \times 5$	softmax	-

2 实验结果与分析

2.1 实验环境

实验采用 Adam (Kingma 和 Ba, 2017) 作为优化器, 学习率初始值为 0.001, 训练周期为 120 次。训练中将训练集随机旋转 $0^\circ \sim 20^\circ$, 且在水平和垂直方向分别随机翻转和偏移 $0 \sim 10\%$, 以增强网络的泛化能力。网络训练准确率在 3 次训练没有提升的情况下, 学习率下降到原来的 10%。实验中随机选取图像库的 80% 为训练集, 剩下 20% 为验证集。实验硬件为: 1 块 NVIDIA TeslaP100-12G GPU, 2 块 Intel Xeon E5-2620 V4 CPU, 4 条 DDR4 2 400 MHz 16 GB 内存, 深度学习框架采用 Keras。

2.2 艺术图像库

本文使用的艺术图像从 Artlib 世界艺术鉴赏库和大艺网等网站获取, 其中版画 3 393 张、中国画 3 449 幅、油画 3 400 幅、水粉画 3 378 幅和水彩画 3 390 幅, 各类艺术图像的风格信息分布比较均匀。将高分辨率且风格信息丰富的图像裁切成若干个 299×299 像素大小的图像, 并人工从中筛选风格信息丰富的图像, 对数据进行增强处理。数据增强后, 版画 5 116 幅、中国画 5 151 幅、油画 5 117 幅、水彩画 5 122 幅和水粉画 5 128 幅。使用表 1 的网络模型对原始数据和数据增强后的数据进行实验, 结果如表 2 所示, 可以看出, 数据增强后的分类准确率提升了 9.21%。

表 2 数据增强前后的分类结果
Table 2 Classification results before and
after data augmentation

图像数据处理	准确率 /%
无数据增强	77.34
有数据增强	86.55

注: 加粗字体为最优结果。

2.3 实验结果分析

2.3.1 不同分类方法结果比较

1) 与传统网络模型的比较。为了验证本文网络模型对艺术图像分类的效果, 将本文数据分别输入传统网络模型和本文网络模型进行训练和验证, 结果如表 3 所示。

表3 不同网络模型的分类结果比较
Table 3 Comparison of classification results of different network models

网络模型	参数/M	时间/min	准确率/%
ResNet-50	23.5	6 040	85.85
Xception	20.8	4 320	85.26
InceptionV4	41.7	6 550	85.87
MobileNetV1	3.2	1 655	86.00
MobileNetV2	2.2	1 370	77.87
ShuffleNet	9.2	558	80.39
本文 + SE ($r = 16$)	2.2	1 356	85.34
本文 + SK ($r = 16$)	8.4	2 368	84.95
本文 ($r = 16$)	13.8	2 140	86.20
本文 ($r = 4$)	14	1 660	86.55

注:加粗和斜体分别为最优和次优准确率结果。

从表3可以看出,与 InceptionV4、ResNet-50(He等,2016)和 Xception(Chollet,2017)网络模型相比,本文网络模型轻量、训练用时短、准确率最优。但是 DKSE 网络模块中包含并行的卷积操作,使本文网络模型参数比 ShuffleNet(Zhang等,2018)、MobileNetV1 和 MobileNetV2(Sandler等,2018)高。本文方法 + SE 和本文方法 + SK 是用 SE 和 SK 模块替换表1中 DKSE 模块的网络模型,模块中的下降率 r 值均取 16。实验结果显示,本文网络模型准确率比本文方法 + SE 和本文方法 + SK 分别高 0.86% 和 1.25%。当 DKSE 模块的 r 值取 4 时,分类准确率比 r 取 16 时高。

2)与传统方法的比较。用 Shen(2009)以及 Sheng 和 Jiang(2013)的传统整体特征与局部特征提取方法对本文数据分类,结果如表4所示。

表4 本文方法与传统方法的结果对比
Table 4 Comparison between the results of traditional methods and ours

方法	准确率/%
Shen(2009)	66.78
Sheng 和 Jiang(2013)	60.2
本文($r = 4$)	86.55

注:加粗字体为最优结果。

从表4可以看出,传统手工提取的基于颜色、形

状等底层整体特征与局部特征不能充分区分各类艺术图像的风格特征,而本文方法可以较好地提取艺术图像的整体特征与局部细节特征,提高了分类准确率。

2.3.2 DKSE 模块作用

1)位置。为了查看 DKSE 模块在网络模型不同位置对艺术图像分类的影响,将其单独放在本文网络模型编号为 6、9、12 的位置,分支上分别取 3×3 , 5×5 的卷积核,下降率 r 值取 4,结果如表5所示。

表5 DKSE 模块在不同位置时的准确率

Table 5 Accuracy of DKSE module at different positions

位置编号	参数/M	时间/min	准确率/%	运算量/GFLOPs
6	2.7	1 800	87.24	0.75
9	4.4	2 043	87.14	1.77
12	11.1	1 860	86.46	3.37

注:加粗字体为最优准确率结果。

从表5可以看出,DKSE 模块单独放在网络模型编号为 6 的位置时,艺术图像的分类准确率最高,计算量消耗最少。

2)下降率 r 和分支卷积核大小。下降率 r 和分支卷积核大小是 DKSE 模块中对计算资源和实验准确率进行控制的一组重要参数。取 DKSE 模块单独放在编号位置为 6 的网络模型,对 DKSE 的 r 值和分支卷积核大小进行实验和分析,结果如表6所示。可以看出,当 DKSE 模块分支上卷积核大小固定时,

表6 下降率与卷积核大小分类结果比较

Table 6 Comparison of classification results between reduction ratio and convolution kernel size

下降率	1×1	3×3	5×5	参数/M	时间/min	准确率/%
$r = 4$	✓	✓		2.3	1 367	87.26
	✓		✓	2.6	1 680	87.58
		✓	✓	2.7	1 800	87.24
	✓	✓	✓	2.7	2 370	87.35
$r = 16$	✓	✓		2.3	1 230	86.35
	✓		✓	2.6	1 220	86.85
		✓	✓	2.7	1 440	86.15
	✓	✓	✓	2.7	1 770	86.42

注:加粗字体分别为 $r = 4$ 和 $r = 16$ 的最优准确率结果。

r 值取 4 的分类结果比 r 值取 16 时高;当 r 值固定时,2 分支的 DKSE 模块比 3 分支的训练用时短、参数少;当 DKSE 模块分支上卷积核分别取 1×1 和 5×5 时,实验结果准确率最高。

3) 空洞卷积。为了比较空洞卷积核对艺术图像特征提取的效果,在 DKSE 模块上取不同大小的空洞卷积核对本文数据进行实验,结果如表 7 所示,其中 K3 表示普通的 3×3 卷积核,K5 表示空洞卷积扩张率为 2 的 3×3 卷积核、感受野为 5×5 ,K7 表示扩张率为 3 的 3×3 卷积核、感受野为 7×7 。

表 7 DKSE 模块分支取不同空洞卷积核的分类结果
Table 7 Classification results of different dilated convolution kernels in DKSE module

下降率	K3	K5	K7	参数/M	时间/min	准确率/%
$r = 4$	✓	✓		2.4	1 530	86.07
	✓		✓	2.4	1 657	86.00
		✓	✓	2.4	1 560	84.50
	✓	✓	✓	2.6	2 220	86.40

注:加粗字体为最优准确率结果。

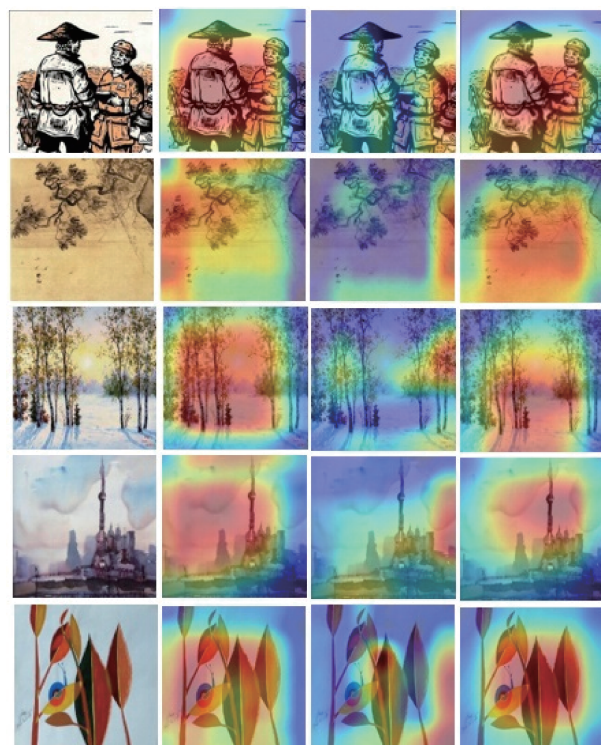
实验结果显示,空洞卷积的参数比具有相同感受野的普通卷积参数少,但是分类准确率没有普通卷积高。

2.3.3 特征可视化

将分支上卷积核为 1×1 和 5×5 且 $r = 4$ 的 DKSE 模块单独放在网络模型编号为 6 的位置,称为 DKSE-MobileNet。为便于观察网络模型学习的关键区域特征,取“本文方法 + SE”、“本文方法 + SK”和 DKSE-MobileNet 网络模型使用 Grad-CAM (gradient-weighted class activation mapping) 算法 (Selvaraju 等,2017) 对提取的艺术图像特征区域进行可视化显示,如图 4 所示。图 4(a) 为艺术图像,图 4(b)~(d) 为特征热力图,暖色越深表示类别判断越依赖该区域信息。从图 4 可以看出,艺术图像的风格特征在整体分布中比较均匀,与“本文方法 + SE”和“本文方法 + SK”网络模型相比,本文的网络模型提取艺术图像的整体特征更全面,局部细节特征更突出。

2.3.4 分类预测结果分析

为了直观展示本文方法的分类性能,使用 DKSE-MobileNet 网络对图 5 中的图像进行分类验



(a) 原画 (b) DKSE-MobileNet (c) 本文方法+SK (d) 本文方法+SE

图 4 特征分布热力图

Fig. 4 Characteristic distribution thermal diagram

((a) original paintings; (b) DKSE-MobileNet thermal diagrams; (c) ours + SK; (d) ours + SE)

证,图 5 中的蓝框表示油画错分成水粉画,绿框表示油画错分成水彩画,红框表示水粉画错分成油画。从图 5 可以看出,油画、水粉画、水彩画无论在绘画技法、颜色布局、纹理、冷暖色调的运用或光感的设计上都存在一定的相似性。错分成水彩画、水粉画的油画和错分成油画的水粉画,在其所在类别应具有的风格特征并不是很明显。

2.3.5 分类性能比较

使用查全率、查准率和 F-Measure 值对 DKSE-MobileNet 进行性能评估,结果如表 8 所示。国画的查准率与 F-Measure 值最高,版画的查全率最高。版画独特的雕刻技法和线条特征是区别于其他画种的关键特征。中国画独特的绘画技巧及墨彩的运用,使国画不同于西方画种。

ROC (receiver operating characteristic) 曲线及 AUC (area under curve) 值是一组评价分类器性能的指标,ROC 曲线下的面积值越大,模型分类性能越好。使用 ROC 及 AUC 衡量 DKSE-MobileNet 分类性

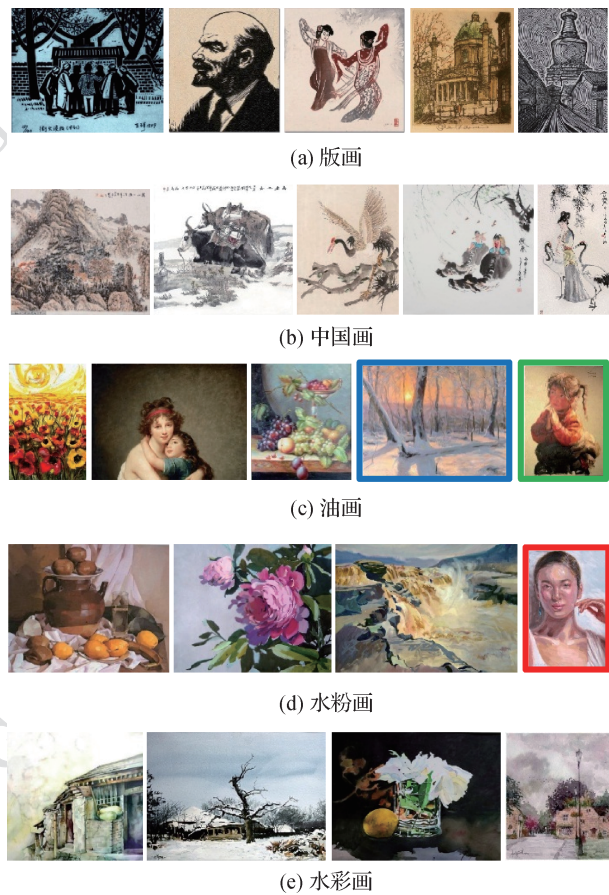


图 5 5 类艺术图像分类结果

Fig. 5 Classification results of five types of artistic images ((a) engravings; (b) Chinese paintings; (c) oil paintings; (d) opaque watercolor paintings; (e) watercolor paintings)

能,结果如图 6 所示,可以看出,版画的 AUC 值达到 0.99,分类器对版画的分类结果最好,国画次之。因为油画的色块、线条、作画技法等特征与其他画种在整体特征与局部细节存在相似处,油画的 AUC 值最小。

表 8 分类性能
Table 8 Classification performance

种类	查准率	查全率	F-Measure
版画	87.25	91.03	88.83
中国画	93.08	87.32	90.11
油画	88.15	84.44	86.26
水彩画	87.22	86.44	86.83
水粉画	82.00	87.77	84.35

注:加粗字体为每列最优结果。

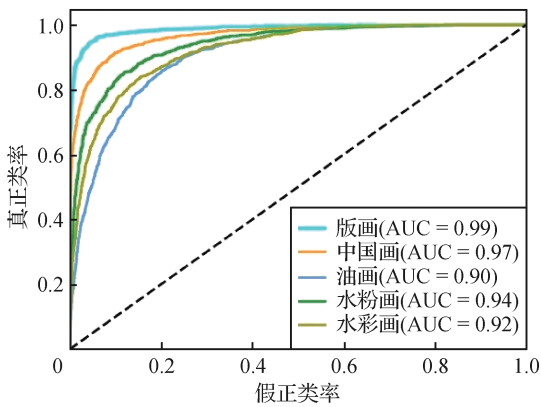


图 6 5 类艺术图像 ROC 曲线图

Fig. 6 Roc curve of five types of artistic images

3 结 论

本文主要提出了 DKSE 模块。DKSE 模块使用双分支卷积核提取图像整体特征与局部细节特征,将提取的特征进行融合,并用两次 1×1 卷积操作和 Sigmoid 门机制提取每类艺术图像的关键特征,抑制作用不大的特征。用 DKSE 模块结合轻量级的深度可分离卷积构建卷积神经网络,对版画、中国画、油画、水彩画和水粉画等艺术图像进行分类,更好地实现了对艺术图像的特征提取和分类。本文方法较好地解决了现有多类艺术图像分类研究不足的问题,取得了较现有网络模型和传统分类方法更好的艺术图像的分类效果。但是 DKSE 模块中的并行双卷积操作,使网络模型中添加多个 DKSE 模块,网络模型参数量增加明显。在后续的工作中将优化艺术图像分类网络模型,扩大艺术图像样本库,进一步提高艺术图像分类准确率和网络模型分类的效率。

参考文献 (References)

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Gao F, Nie J, Huang L, Duan L Y and Li X M. 2017. Traditional Chinese painting classification based on painting techniques. Chinese Journal of Computers, 40(12): 2871-2882 (高峰, 聂婕, 黄磊, 段凌宇, 李晓明. 2017. 基于表现手法的国画分类方法研究. 计算机学报, 40(12): 2871-2882) [DOI: 10.11897/SP. J. 1016.2017.02871]

- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Howard A G, Zhu M L, Chen B and Kalenichenko D. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2019-05-25]. <https://arxiv.org/pdf/1704.04861.pdf>
- Hu J, Shen L, Albanie S, Sun G and Wu E H. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 7132-7141
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2019-05-25]. <https://arxiv.org/pdf/1502.03167.pdf>
- Jiang S Q, Huang Q M, Ye Q X and Gao W. 2006. An effective method to detect and categorize digitized traditional Chinese paintings. Pattern Recognition Letters, 27(7): 734-746 [DOI: 10.1016/j.patrec.2005.10.017]
- Kingma D P and Ba J L. 2017. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2019-05-25]. <https://arxiv.org/pdf/1412.6980.pdf>
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: ACM: 1097-1105
- Li J and Wang J Z. 2004. Studying digital imagery of ancient paintings by mixtures of stochastic models. IEEE Transactions on Image Processing, 13(3): 340-353 [DOI: 10.1109/TIP.2003.821349]
- Li X, Wang W H, Hu X L and Yang J. 2019. Selective kernel networks [EB/OL]. (2019-03-18) [2019-05-25]. <https://arxiv.org/pdf/1903.06586.pdf>
- Nair V and Hinton G E. 2010. Rectified linear units improve restricted Boltzmann machines//Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, Israel: Omnipress: 807-814
- Qi Y Q. 2009. The creation and appreciation of Chinese painting——Xie He's "The six laws". Science and Technology Information, (14): 231 (元玉权. 2009. 中国画的创作和鉴赏——谢赫的“六法论”. 科技资讯, (14): 231) [DOI: 10.3969/j.issn.1672-3791.2009.14.192]
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- Shen J L. 2009. Stochastic modeling western paintings for effective classification. Pattern Recognition, 42(2): 293-301 [DOI: 10.1016/j.patcog.2008.04.016]
- Sheng J C and Jiang J M. 2013. Style-based classification of Chinese ink and wash paintings. Optical Engineering, 52(9): #093101 [DOI: 10.1117/1.oe.52.9.093101]
- Sun M J, Zhang D, Wang Z, Ren J C and Jin J S. 2016. Monte Carlo convex hull model for classification of traditional Chinese paintings. Neurocomputing, 171: 788-797 [DOI: 10.1016/j.neucom.2015.08.013]
- Szegedy C, Ioffe S, Vanhoucke V and Alemi A A. 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning//Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17). San Francisco: AIAA: 4278-4284
- Wang Z, Sun M J, Han Y H and Zhang D. 2013. Supervised heterogeneous sparse feature selection for Chinese paintings classification. Journal of Computer-Aided Design and Computer Graphics, 25(12): 1848-1855 (王征, 孙美君, 韩亚洪, 张冬. 2013. 监督式异构稀疏特征选择的国画分类和预测. 计算机辅助设计与图形学学报, 25(12): 1848-1855)
- Yu F and Koltun V. 2016. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-04-30) [2019-05-25]. <https://arxiv.org/pdf/1511.07122.pdf>
- Zhang X Y, Zhou X Y, Lin M X and Sun J. 2018. ShuffleNet: an extremely efficient convolutional neural network for mobile devices//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 6848-6856 [DOI: 10.1109/CVPR.2018.00716]

作者简介



杨秀芹, 1994年生, 女, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 2238849807@qq.com



张华熊, 通信作者, 男, 教授, 主要研究方向为智能信息处理。

E-mail: zhxhz@zstu.edu.cn