

基于卷积神经网络和 SVM 的中国画情感分类

王 征¹, 李皓月¹, 许洪山¹, 孙美君²

(1. 大数据分析系统实验室(天津大学软件学院), 天津 300350)

(2. 天津大学计算机科学与技术学院, 天津 300350)

[摘要] 图像情感是指计算机识别数字图像所表达内容引起人的情感反应, 根据不同的情感反应, 可以对不同的图像进行分类。在信息量急剧增长的今天, 图像情感分类有助于图像的标注和检索, 蕴藏着很大的社会和商业价值。不同于西洋画的“以形写形”, 中国画有着自己明显的特征: 传统的国画不讲焦点透视, 不强调自然界对于物体的光色变化, 不拘泥于物体外表的肖似, 而多强调抒发作者的主观情趣。这比弥合一般的低层特征和人类情感高层语义之间的鸿沟的难度更大。基于卷积神经网络因为其具有结构简单、适应性强、训练参数少、连接点多等特点, 可以直接输入原始图像, 能够避免对图像进行复杂的前期预处理。相比传统图像特征提取方法, 卷积神经网络具有明显的优势。本文的目的是利用卷积神经网络发掘低层特征和情感语义之间的联系, 提取国画图像特征, 对得到的特征进行 PCA 降维、归一化等操作后, 利用支持向量机(SVM)分类器进行情感分类。

[关键词] 图像情感, 中国画, 卷积神经网络, 特征提取, 支持向量机

[中图分类号] K879.4, TP183 **[文献标志码]** A **[文章编号]** 1001-4616(2017)03-0074-06

Chinese Painting Emotion Classification Based on Convolution Neural Network and SVM

Wang Zheng¹, Li Haoyue¹, Xu Hongshan¹, Sun Meijun²

(1. Laboratory for Big Data Analysis and System(School of Software, Tianjin University), Tianjin 300350, China)

(2. School of Computer Science and Technology, Tianjin University, Tianjin 300350, China)

Abstract: Image emotions are human emotional responses caused by the contents of digital images. Computers are able to classify different images according to different human emotional responses. With the rapid growth of the amount of information, image emotion classification will contribute to the image annotation and search producing great social and commercial value. Chinese paintings have obvious characteristics; traditional Chinese paintings do not focus on the perspective, and do not emphasize the light color changes of objects in nature, and do not rigidly adhere to the appearance of objects. They more focus on the expression of authors' subjective consciousness making it harder to bridge the semantic gap between general low-level features and human emotions. The structure of convolutional neural network(CNN) is simple, yet its adaptability is strong. CNN also has less training parameters and more junctions, and are able to read images directly without preprocessing images complexly. It has a huge advantage over traditional image-processing method. This paper aims to explore the relationships between low-level features and emotional semantics by CNN, and extract the features of Chinese paintings and process the features by PCA and normalization. Finally we classify the features by SVM.

Key words: image emotion, Chinese painting, convolution neural network, feature extraction, support vector machine

国画是我们国家的艺术瑰宝, 传统国画讲求意象, 作者往往在画面中创造性地将客体进行艺术阐释^[1]。可以说, 中国画自古就是以情感为中心的艺术形式。优秀的国画作品是画家用真情实感描绘的精神产物, 画家不会仅仅简单地描摹客观事物, 更重要的是抒发自己的内心情感, 以达到物我结合、精神升华的状态^[2]。因为国画这种重在“写意”和“以情感表达为中心”的艺术特点, 我们在对国画的鉴赏、研究、保护工作中, 不可避免地需要考虑画中的情感因素, 理解高层的情感语义可以帮助我们鉴赏和研究国画。同时

收稿日期: 2017-03-18.

基金项目: 国家自然科学基金(61572351、61772360).

通讯联系人: 孙美君, 副教授, 研究方向: 图形学、图像及光谱数据处理. E-mail: sunmeijun@tju.edu.cn

也能丰富国画的组织、管理和检索方法,普及国画艺术的传播。

计算机从低层视觉数据中提取的特征信息与人类的情感解释之间一般存在较大的差异^[3-4],我们将这种差异称为情感鸿沟(Motional Gap)^[5-6]。我们希望计算机能够深层次地对图像进行理解,准确识别图像中蕴含的情感语义;同时在理解图像情感语义的基础上,根据所理解的数字图像情感语义的不同,对图像进行情感分类。近年来,图像的语义分类引起人们的广泛关注,在计算机视觉领域中非常活跃。图像的情感分类是其中一个研究方向,也是语义分类中层次最高的一个^[7]。

我们的目标是计算机能够通过适当的模型,描述人欣赏国画图像时所产生的情感反应,并据此进行归类。情感语义相对于认知层的语义来说,更加主观,人的认知水平、文化背景以及美学标准等的差异,都会对情感语义造成影响。由此,我们自己设计特征提取方法,设计感知分类算法的难度相当之高。而深度学习算法的优良特性,给弥合情感鸿沟,实现理解情感的认知提供了一条行之有效的途径。

深度学习实质是一种特征学习方法。原始数据通过一些简单、非线性的模型变换,成为更高层次的、更加抽象的表达。深度学习的各层都不是人工来设计的,而是通过一种通用的学习过程从数据中学到的^[8]。在分类任务中,高层在强化输入数据的区分能力的同时,削弱了不相关因素的影响。

在本文中,我们尽量选取具有代表性的国画画作作为样本,根据不同情感制作样本标签,每种情绪的样本都涵盖国画的不同领域,包括山水、花鸟鱼虫、人物。在样本集上进行训练,获得特征提取模型。在此基础上,我们对提取的特征进行一些预处理,比如利用 PCA 降维等。由于颜色因素在情感表达的独特作用,我们还会提取画作的颜色分布直方图,与神经网络提取的特征混合后,尝试支持向量机(SVM)的分类方法,以期得到更好的分类结果。本文第 1 节介绍情感分类的依据,第 2 节介绍特征提取和分类方法,第 3 节对实验及实验结果进行介绍。最后总结全文,并对未来值得关注的研究方向进行初步探讨。

1 国画中影响情感的特征

1.1 中国画中的色彩

归根到底,绘画是视觉的艺术。在生活中,色彩能给人最直观、最强烈的心理感受^[9-10]。同样的规律也适用于中国画的鉴赏。举例来讲,两幅都属于山水画范畴的画作,一幅色调明丽,另一幅色调较为浓重。由于色调的差异,人的视觉感觉也有明显差异。从情感空间角度来说,前者的情感表达可能更加正向。从颜色角度考虑情感,表 1^[11]比较全面地反映了颜色可能对应的情感类型集合。

表 1 颜色情感映射
Table 1 Color emotion mapping

| 颜色 | 情感说明 |
|-----|------------------------------------|
| 红色 | 兴奋、热情、强烈、好战、希望、奋进 |
| 黄色 | 欢快、欣喜、喧闹、活泼、希望、健康、奔放、开朗 |
| 蓝色 | 宁静、文静、冷静、理智、透明、沉稳、清爽 |
| 橙色 | 辉煌、富贵、温暖 |
| 绿色 | 永远、和平、年轻、新鲜、舒缓、清新、稳重、舒适、积极、幽雅、心旷神怡 |
| 紫色 | 积极、威严、尊贵、神奇、怪异 |
| 黑色 | 寂寞、神秘、严肃、含蓄、庄重、恐怖、悲哀、稳重、力度、死亡 |
| 白色 | 纯洁、洁白、光明、纯真、轻快、和谐、膨胀感、淡雅、干净、高尚 |
| 粉色 | 温柔、浪漫 |
| 灰色 | 简约、平和、浑厚、温存、稳重、朴素 |
| 银灰色 | 健康、积极、厚重 |
| 褐色 | 沉稳、醇厚、严密、深沉 |
| 青色 | 坚强、圆顺、冷清 |

1.2 中国画中的形状与纹理

颜色特征不是唯一影响画作语义特征的信息,纹理和形状也会对人的心理产生不可忽视的影响,图的内容构成的主要形状、线条、纹理不同,表达的情感也有所不同。一般来说,光滑给人以细腻感,粗糙给人以苍老感;规则的形状让人舒服,而不规则的形状则相反。表 2、表 3^[9]简单描述了形状和纹理对情感的映射关系。

表 2 形状情感映射

Table 2 Shape emotion mapping

| 形状 | 情感 | 形状 | 情感 |
|--------|-------|---------|-------|
| 整齐(规则) | 舒服 | 弧形 | 松弛 |
| 缺(不规则) | 厌恶、烦躁 | 曲线 | 动感 |
| 梭形 | 正统 | 不稳定 | 焦虑 |
| 方形 | 端正集中 | 稳(对称均匀) | 平静 |
| 圆形 | 圆润松弛感 | 三角形 | 机械、冷漠 |

1.3 中国画中的其他特征

除了比较重要的颜色特征和形状特征,从其他角度来说,例如图像的几何特征、灰度直方图、频带变化、图像理解的变化等都可能帮助我们计算分析图像的情感语义^[12]. 另外在中国画中,一些独特的意向往往对应特定的含义,比如梅兰竹菊一般代表高洁的品格,这需要鉴赏者具有相关的知识背景.

表 3 纹理情感映射

Table 3 Texture emotion mapping

| 纹理 | 情感 |
|----|-------|
| 光滑 | 细腻、放松 |
| 粗糙 | 苍老 |
| 柔软 | 温馨、温柔 |
| 坚硬 | 刚强 |

2 特征的提取与分类

2.1 颜色特征的提取

我们把颜色分布特征作为推断图像情感的一个重要依据. 我们采用颜色直方图这种被广泛采用的颜色特征. 颜色直方图描述的是不同色彩在图像中分别所占的比例大小,与色彩所处的空间位置无关.

2.2 其他特征的提取

我们利用卷积神经网络提取特征向量,通常提取高层网络特征,此时获取的特征已经是抽象特征的组合描述.

2.2.1 深度学习模型

我们需要训练得到一个用于提取特征向量的神经网络模型,神经网络模型是进行深度学习的核心所在,也是进行实验的基础. 不同网络模型具有不同特性,我们选取了不同的网络来提取特征. 本节选取 AlexNet 网络(见图 1)作为样例进行介绍.

神经网络由多个感知器互相连接而成,整个网络包含输入层、隐含层和输出层^[13-14]. 在多层网络结构模型中,卷积神经网络(CNN)是第一个能真正成功训练的多层网络,利用卷积和池化操作降低需要学习的参数数量. 不同于深度置信网络(DBN),卷积神经网络属于区分性训练算法^[15-17].

在 CNN 中,每一个图像块作为多层结构网络的可视层输入. 在卷积层中通过一组过滤器和非线性层变换,提取出图像局部特征. 为了减少需要训练的权值数量,引入了其他优化方法:权值共享^[18]. 经过卷积层得到的特征向量通常维数较大,冗余信息很多,易引起计算过拟合. 因而在这层的不同位置对特征进行汇聚统计的降采样操作,即池化(pooling). 通过组合多层卷积层和池化层形成的深度学习模型能够获取到对平移、缩放和旋转不变的观测样本的显著特征^[19-20].

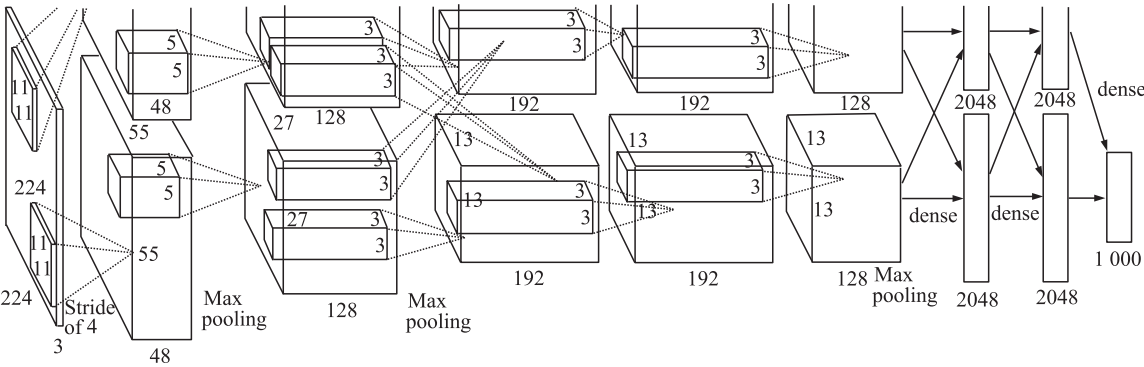


图 1 AlexNet 网络结构

Fig. 1 AlexNet network structure

AlexNet 共 8 层,前 5 层是卷积层,后 3 层为全连接层. 首层为卷积层,在 3 个通道上采用 11×11 的卷积模板、间隔为 4 像素的采样频率对图像进行卷积操作后,得到基本的卷积数据,然后进行一次 ReLU 和 Norm 变换,随后进行池化,得到规模为 $27 \times 27 \times 96$ 的结果作为输出传入下一层. 第二层仍是卷积层,处理过程和第一层类似,经过卷积、ReLU、Norm 和池化,输出结果规模为 $13 \times 13 \times 256$. 卷积层处理过程基本相同,第四、五层与前三层不同之处在于其处理过程仅是卷积和 ReLU,第四层的输出数据规模为 $13 \times 13 \times 324$. 第五层输出的数据规模变为 $6 \times 6 \times 256$. 经过五个卷积过程,第六层进入全连接层,全连接后达到 4 096 个节点. 第七层处理过程与第六层相似,fc7 对 fc6 进行 ReLU 后,进行 dropout 再进行全连接,本层节点数量仍为 4 096 个. 最后一层全连接层 fc8 是对上一个 fc7 进行 ReLU 以及 dropout 后再进行全连接的结果. 最后输出为融合 label 的 softmax loss^[21-23].

2.2.2 提取

图 2 对比了经过 AlexNet 卷积运算之后的特征图像. 得到的特征向量包含图像边缘、形状等特征信息. 之后每次卷积即一次抽象,最后得到特征的组合. 我们选择提取 fc7 层的特征向量,与颜色特征混合,处理后进行分类计算.

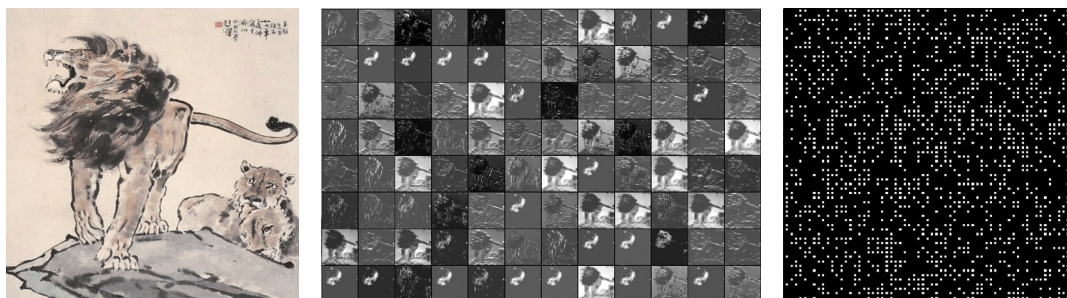


图 2 同一幅图的原始图像、卷积一次的特征图像和卷积多次的特征图像的对比

Fig. 2 The contrast among the original image, the feature image convoluted once and the feature image convoluted several times

2.3 特征的处理

主成分分析(Principal Component Analysis, PCA)可以从多元事物中解析出主要影响因素,简化复杂问题^[24-25]. 其目的是将高维数据投影到较低维空间,用原有变量的线性组合来表示图像的主要方面^[24].

数据归一化,就是要把需要处理的数据经过处理后(通过某种算法)限制在需要的一定范围内,使后续数据处理更加方便,并保证程序运行时收敛加快^[26].

利用卷积神经网络提取全连接层的特征向量,以 AlexNet 为例,特征向量为 4 096 维,对于 SVM 分类器,如此高维度的数据上找到合适的超平面分类相对难度比较大,因而使用 PCA 降低特征维度,与颜色特征分别进行归一化处理后混合,最终利用分类器进行分类.

2.4 SVM 分类器

在机器学习领域,支持向量机 SVM(Support Vector Machine)作为一个有监督的学习模型,通常用来进行模式识别、分类以及回归分析. 其适用于解决小样本、非线性和高维的模式识别问题,而且能够被推广和应用到机器学习问题中^[27-28].

3 实验及结果分析

我们的关注点主要在于特征提取. 不同的网络模型在同样的数据集上,提取的特征并不相同. 对比不同特征提取器和分类器结合的实验结果,得到相对更准确的分类模型. 当然还需要一些必要的操作,例如制作标签数据,标准化图像数据的大小,训练网络,对选取的特征做必要的操作以符合分类器的需要.

3.1 数据集

选取不同年代不同作者的画作,尽可能包括山水、花鸟、鱼虫等多个方面的题材,希望数据样本在小样数量下能具有更好的代表性. 同时对图像进行平移、翻转等变换,同 Dropout 正则化作用相似^[29],都是为了防止过拟合. 我们还计算国画图像数据的均值,这是因为图像本身具有统计不变性,即图像的一部分的统计特性和另一部分相同. 最后对数据进行标识,在本次实验中,我们将情感标签看做离散的变量进行操作

作,方便有监督的训练. 我们将图像的情感大致分为 6 类,图 3 给出了不同情感的代表作,表 4 对应其表达的不同情感标签.

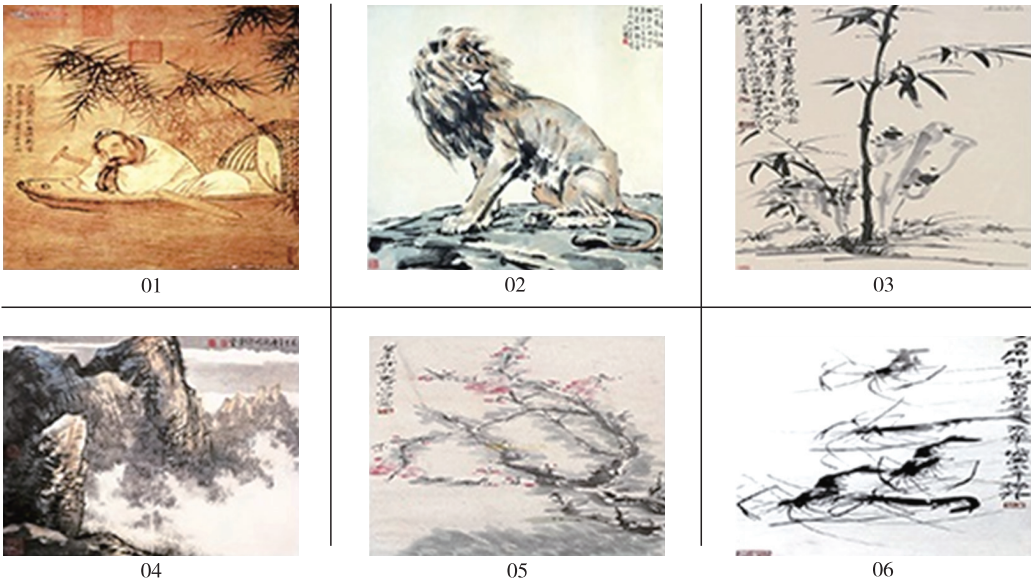


图 3 不同情感的代表作品

Fig. 3 Representation of different emotions

表 4 数据标签的标识

Table 4 The labels of the data

| 情感 | 编号 | 情感 | 编号 | 情感 | 编号 |
|-------|----|---------|----|---------|----|
| 孤寂 | 01 | 气节、百折不挠 | 03 | 清雅 | 05 |
| 焦虑、担忧 | 02 | 气势磅礴 | 04 | 恬淡、生活情趣 | 06 |

3.2 网络模型训练

此次实验选取了 AlexNet、GoogleNet 和残差神经网络(ResNet)3 种不同的模型训练,在有限的小数据集上,得到 3 种训练准确率不同的深度学习模型. 对于 AlexNet,我们提取 fc7 层的特征,其他两个网络都提取其 SoftMax 层^[30]之前的下采样层的特征. 表 5 是 3 种网络的训练准确率. 在本次实验的小数据集上, AlexNet 模型的准确率相对更高,达到了 0.92. 使用已得到的训练网络模型进行预测. 如图 4 所示.

表 5 神经网络训练结果

Table 5 Neural network training results

| 网络名称 | 训练结果 |
|-----------|------|
| AlexNet | 0.92 |
| GoogleNet | 0.87 |
| ResNet | 0.83 |

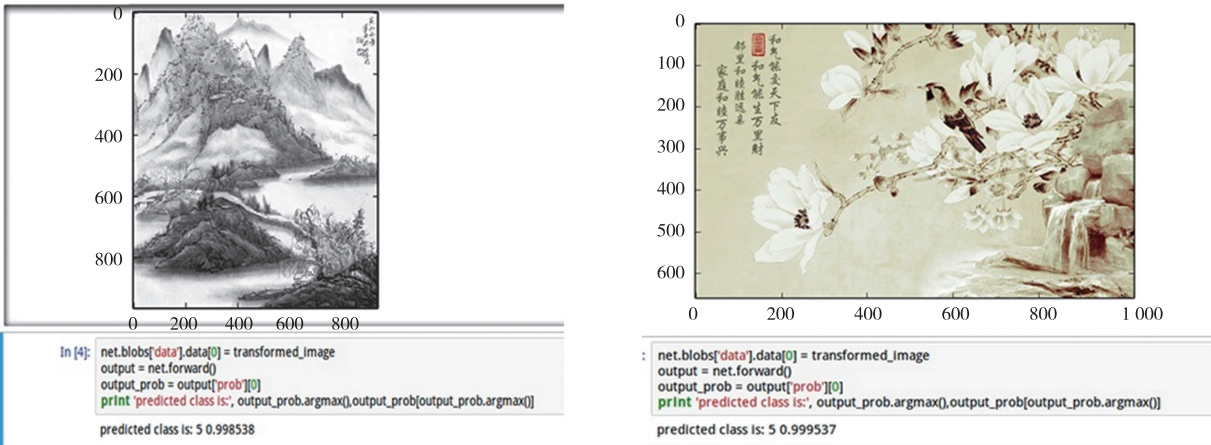


图 4 情感预测结果

Fig. 4 Emotional prediction results

3.3 准确率对比

对于提取的不同特征向量,都会经过 PCA 降维、颜色特征混合和归一化之后,利用 SVM 分类器分类,不同特征的分类结果也并不相同. 结果如表 6 所示.

AlexNet 训练结果准确率为 0.812 5,测试结果准确率为 0.666 7,仍是三者中最高的.

表 6 支持向量机训练结果

| Table 6 Support vector machine training results | | |
|---|---------|---------|
| 训练结果 | 训练集 | 测试集 |
| AlexNet | 0.812 5 | 0.666 7 |
| GoogleNet | 0.937 5 | 0.416 7 |
| ResNet | 0.750 0 | 0.416 7 |

4 结语

本文概述了国画情感特点、深度学习的概念以及卷积神经网络对于图像分类具有的优势,训练了不同的神经网络. 然后利用深度学习提取的特征,处理之后和支持向量机结合,构建分类模型.

目前深度学习正在飞速发展,在许多领域已经广泛应用,但在图像情感分类方面应用不多. 本文基于深度学习对中国画情感分类进行了一次尝试,但是仍有很多工作需要完成,比如如何将中国画的某些先验知识融合到模型卷积神经网络,以发掘出更好的情感特征. 相信不久的将来,机器学习能力会得到进一步加强,情感计算能力也会大幅提升.

[参考文献]

- [1] 柴刚. 解析国画作品中的意象世界[J]. 科技资讯,2009(20):197-198.
- [2] 雷莹. 浅析“情感”在国画创作中的艺术精神[J]. 大众文艺(学术版),2016(8):91.
- [3] WRIGHT J, MA Y, MAIRAL J, et al. Sparse representation for computer vision and pattern recognition[J]. Proceedings of the IEEE, 2010, 98(6):1 031-1 044.
- [4] SUN Y, FISHER R. Object-based visual attention for computer vision[J]. Artificial intelligence, 2003, 146(1):77-123.
- [5] DATTA R. Semantics and aesthetics inference for image search; statistical leaning approaches[D]. University Park: The Pennsylvania State University, 2009.
- [6] GUDIVADA V N. A geometry-based representation for efficient and effective retrieval of images by spatial similarity[J]. IEEE transactions of knowledge and data engineering, 1998, 10(3):504-512.
- [7] HANJALIC, ALAN. Video and image retrieval beyond the cognitive level: the needs and possibilities[C]//Storage and Retrieval for Media Databases, San Jose, 2001. Bellingham: Optical Engineering, 2001:130-140.
- [8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7 553):436-444.
- [9] HEMPHILL M. A note on adults' color-emotion associations[J]. Journal of genetic psychology, 1996, 157(3):275-280.
- [10] ELLIOT A J, MAIER M A, MOLLER A C, et al. Color and psychological functioning: the effect of red on performance attainment[J]. Journal of experimental psychology general, 2007, 136(1):154.
- [11] 陈俊杰. 图像情感语义分析技术[M]. 北京:电子工业出版社, 2011:13-15.
- [12] 余英林, 田菁, 蔡志峰. 图像视觉感知信息的初步研究[J]. 电子学报, 2001, 29(10):1 373-1 375.
- [13] WANG T, WU D J, COATES A, et al. End-to-end text recognition with convolutional neural networks[C]//International Conference on Pattern Recognition. Tsukuba Japan, 2012. New York: IEEE, 2012:3 304-3 308.
- [14] ABDEL H O, MOHAMED A R, JIANG H, et al. Convolutional neural networks for speech recognition[J]. IEEE/ACM transactions on audio speech and language processing, 2014, 22(10):1 533-1 545.
- [15] CECOTTI H, GRÄSER A. Convolutional neural networks for P300 detection with application to brain-computer interfaces[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(3):433-45.
- [16] NEBAUER C. Evaluation of convolutional neural networks for visual recognition[J]. IEEE transactions on neural networks, 1998, 9(4):685.
- [17] BAI S. Growing random forest on deep convolutional neural networks for scene categorization[M]. London: Pergamon Press, 2017:50-70.
- [18] SUN M, SONG Z, JIANG X, et al. Learning pooling for convolutional neural network[J]. Neurocomputing, 2017, 224:96-104.
- [19] YU S, JIA S, XU C. Convolutional neural networks for hyperspectral image classification[J]. Neurocomputing, 2017, 219:88-98.
- [20] KUMAR A, KIM J, LYNDON D, et al. An ensemble of fine-tuned convolutional neural networks for medical image classification[J]. IEEE journal of biomedical and health informatics, 2017, 21(99):1.

(下转第 86 页)

-
- [8] KIRALY A, CHRISTIDOU M, CHOVAN T, et al. Minimization of off-grade production in multi-site multi-product plants by solving multiple traveling salesman problem[J]. Journal of cleaner production, 2016, 111: 253–261.
- [9] 王勇臻, 陈燕, 于莹莹. 求解多旅行商问题的改进分组遗传算法[J]. 电子与信息学报, 2017, 39(1): 198–205.
- [10] SINGH A, BAGHEL A S. A new grouping genetic algorithm approach to the multiple traveling salesperson problem[J]. Soft computing, 2009, 13(1): 95–101.
- [11] YUAN S, SKINNER B, HUANG S, et al. A new crossover approach for solving the multiple travelling salesmen problem using genetic algorithms[J]. European journal of operational research, 2013, 228(1): 72–82.
- [12] LIU W, LI S, ZHAO F, et al. An ant colony optimization algorithm for the multiple traveling salesmen problem[C]//4th IEEE Conference on Industrial Electronics and Applications, 2009. ICIEA 2009. IEEE, 2009: 1 533–1 537.
- [13] NECULA R, BREABAN M, RASCHIP M. Performance evaluation of ant colony systems for the single-depot multiple traveling salesman problem[C]//International Conference on Hybrid Artificial Intelligence Systems. Cham: Springer, 2015: 257–268.
- [14] XUE M H, WANG T Z, MAO S. Double evolutionary artificial bee colony algorithm for multiple traveling salesman problem [C]//MATEC Web of Conferences. EDP Sciences, 2016: 44.
- [15] VENKATESH P, SINGH A. Two metaheuristic approaches for the multiple traveling salesperson problem[J]. Applied soft computing, 2015, 26: 74–89.

[责任编辑: 丁 蓉]

(上接第79页)

- [21] 乐毅, 王斌. 深度学习-Caffe之经典模型详解与实战[M]. 北京: 电子工业出版社, 2016: 107–111.
- [22] ZENG R, WU J, SHAO Z, et al. Quaternion softmax classifier[J]. Electronics letters, 2014, 50(25): 1 929–1 931.
- [23] CHEN X, YE Q, ZOU J, et al. Visual trajectory analysis via replicated softmax-based models[J]. Signal, image and video processing, 2014, 8(1): 183–190.
- [24] LEE J Y, KIM K. A feature-based approach to extracting machining features[J]. Computer-aided design, 1998, 30(13): 1 019–1 035.
- [25] CANDÈS E J, LI X, MA Y, et al. Robust principal component analysis? [J]. Journal of the Acm, 2009, 58(3): 11.
- [26] XIE X, ZHENG W S, LAI J, et al. Normalization of face illumination based on large- and small-scale features[J]. IEEE transactions on image processing a publication of the IEEE signal processing society, 2011, 20(7): 1 807–1 821.
- [27] QI Z, TIAN Y, SHI Y. Robust twin support vector machine for pattern classification[J]. Pattern recognition, 2013, 46(1): 305–316.
- [28] QI Z, TIAN Y, SHI Y. Structural twin support vector machine for classification[J]. Knowledge-based systems, 2013, 43(2): 74–81.
- [29] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of machine learning research, 2014, 15(1): 1 929–1 958.
- [30] JR S E. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals[J]. Journal of applied measurement, 2002, 3(2): 205–231.

[责任编辑: 丁 蓉]