

文章编号: 2095-4980(2020)01-0136-06

高仿真光敏印章盖印印文的自动识别

张 倩¹, 韩星周²

(1.中国人民公安大学 刑事科学技术学院, 北京 102623; 2.公安部物证鉴定中心, 北京 100038)

摘 要: 为了实现高仿真光敏印章印文的自动识别, 探究训练样本量、网络模型对识别准确率的影响, 通过扫描打印伪造法、拓印设计伪造法制备2枚高仿光敏印章, 盖印3 000枚印文作为训练样本, 30枚印文作为测试样本, 利用卷积神经网络4种模型实现高仿真光敏印章印文的鉴别。4种网络模型均能得到100%的识别准确率。仿真实验结果表明, 针对高仿真光敏印章印文识别任务, 卷积神经网络能作为一种可行的方法为检验提供辅助参考; 综合分析4种网络模型, Resnet50是最优选择。

关键词: 高仿真光敏印章; 盖印印文; 自动识别; 卷积神经网络

中图分类号: TN915.05; TP391.41 **文献标志码:** A **doi:** 10.11805/TKYDA2019163

Automatic recognition of high-simulation photosensitive seal stamping

ZHANG Qian¹, HAN Xingzhou²

(1.Forensic Science College, People's Public Security University of China, Beijing 102623, China; 2.Institute of Forensic Science, Ministry of Public Security, Beijing 100038, China)

Abstract: To realize the automatic recognition of high simulation photosensitive seals and explore the influence of training sample size and network model on the recognition accuracy, two high imitation photosensitive seals were prepared by scanning, printing and rubbing design forgery. 3 000 seals were stamped as training samples and 30 seals were stamped as test samples. Four models of convolutional neural network were utilized to identify the high simulation photosensitive seals. According to the results, all the four network models can get 100% recognition accuracy. The convolutional neural network can be used as a feasible method to provide auxiliary reference for the test. By comprehensive analyzing the four network models, Resnet50 is the best choice for the task of high-simulation photosensitive seal printing.

Keywords: high simulation photosensitive seal; stamp impression; auto recognition; convolutional neural network

印文自动识别是一种模式识别方法, 在银行系统的验印操作中已得到成功应用^[1-3], 提高了工作效率, 增强了操作的科学性。与一般印文识别不同, 印文的真伪识别不仅需要区分真伪印文的差异性, 还需要包容真印文在盖印过程中形成的非本质差异特征, 是真伪印文自动识别的一大难点。近年来利用高仿真印章伪造文件进行诈骗活动的案件屡见不鲜, 对于真伪印文的识别研究刻不容缓。

伪造印章主要有 3 种类型: 伪造原子印章、伪造光敏印章、伪造激光雕刻印章。其中高仿真光敏印章在伪造与制作过程中相对简单, 且采购制章设备容易, 成本较低, 是目前伪造盖印印文文件的主要方法^[4]。智能科技的发展使得机器视觉技术和深度学习方法成为学术研究热点, 卷积神经网络作为深度学习方法之一在众多领域得以研究^[5-8]。

本文基于卷积神经网络的开发和应用研究成果, 提出一种利用卷积神经网络对高仿真光敏印章盖印印文的自动识别方法, 从识别准确率、泛化能力、消耗代价等方面综合比较、分析、探究训练样本量和网络模型对自动识别的影响。

收稿日期: 2019-05-29; 修回日期: 2019-07-03

基金项目: 财政部基本科研业务基金资助项目(2018JB022)

作者简介: 张 倩(1995-), 女, 在读硕士研究生, 主要研究方向为文件检验。email:490793861@qq.com

1 高仿真光敏印章印文检验

1.1 高仿真光敏印章的概念

高仿真光敏印章是指基于真实印章，制备在印章的规格特征、章面图文布局特征、细节特征等各方面高度相似的伪造印章。目前市场上出现较多、使用较广的伪造光敏印章有扫描打印伪造印章和拓印设计伪造印章，二者都基于对真实印文的扫描：扫描打印伪造印章使用扫描的印文得到印文模板，修饰后直接进行印章刻制；拓印设计伪造印章通过电脑印章排版系统扫描印文为母版，设计伪造印文模板进行印章刻制^[9]。

扫描打印伪造的光敏印章逼真度受真实印文质量、文件上印文背景干扰、扫描仪分辨率、制章机类型等因素影响，形成的盖印印文的细节特征具有差异，如印文中的露白特征、印文边框清晰整齐度、防伪标志变化、打印附加特征等。拓印设计伪造的光敏印章逼真度还受电脑印章排版系统中不同字体库的影响，因此在印文的图文布局特征上存在差异。这些制章过程和盖印过程形成的差异特征是进行高仿真光敏印章盖印印文鉴别的理论依据，充分了解印章类型，理解特征变化的规律是完成真伪鉴别的基础。

1.2 高仿真光敏印章印文检验的方法

高仿真光敏印章印文的检验是依据真伪印文的本质差异特征，运用同一认定的理论和方法进行比对检验、综合分析从而得出真伪判断的结论。传统高仿真光敏印章印文检验主要通过对检材和样本的人工比对，利用显微镜、文检仪等光学仪器，首先判断印文的形成方式，再通过图像处理软件进行印文规格特征、细节特征比对检验，必要时还需通过薄层色谱法、X 射线荧光光谱法等进行油墨成分的检验等。传统的检验方法对鉴定专家的专业技能和经验具有较强的依赖性，且伪造印章的技术手段不断提高，易伪难鉴的情况是印文检验需要应对的一大挑战。

随着计算机设备不断更新，基于各种分类算法和机器学习理论，利用计算机辅助人工检验方法应运而生，如参考印文重合率阈值范围^[10]、基于印文特征指标的神经网络方法^[11]、基于印文纹理特征的神经网络方法^[12]、基于几何统计特征的 BP 神经网络方法^[13]、基于支持向量机多分类器融合印文鉴别的方法^[14]、基于量化边缘差异特征的自动识别真伪印文的方法^[15]、基于模糊集理论的方法等。准确、高效的印文自动识别方法能降低印文检材的质量要求，削弱检验方法的专家依赖性，提高正确识别率、强化鉴定结论的科学性。在模式识别方法的不断优化下，进一步减少自动识别中人工处理步骤。基于迁移学习理论，将卷积神经网络成功用于高相似度的图像分类和识别任务中^[16]，使利用该方法解决印文自动识别问题成为可能。王茜等^[17]利用 Lenet 卷积神经网络对不同种类印章的印文进行自动识别，获得 100% 的分类准确率，成功利用 Caffe 框架实现卷积神经网络对印章章型分类识别任务。在众多的深度学习开源框架中，优异的性能和强大的泛化能力使得卷积神经网络模型的应用领域不断拓展^[18-20]，实现其高仿真光敏印章印文的自动识别中的应用具备科学性与可行性。

2 实验设计

2.1 实验原理与模型结构

卷积神经网络是一种针对图像识别问题设计的神经网络。与一般的神经网络不同，卷积神经网络中包含多个“卷积”和“池化”，用来提取特征的组合操作。组合操作的本质原理是对图像像素的加权矩阵运算，并通过反向传播，不断修正运算过程中的权值来提升预测分类结果的质量。整个印文识别过程主要有 3 个阶段：实现网络模型的运行、网络模型的训练、测试验证。基于迁移学习的理论，实验选取 Pytorch 为框架实现开源卷积神经网络模型的运行，从 Python 库中直接导入预训练的网络模型，以当前识别任务对网络模型参数做微调。

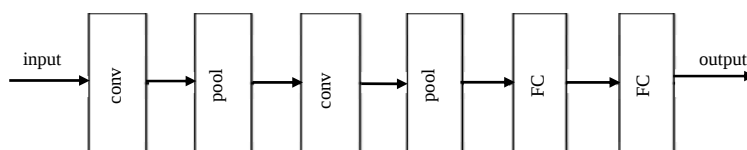


Fig.1 Lenet5 network structure

图 1 Lenet5 网络结构图

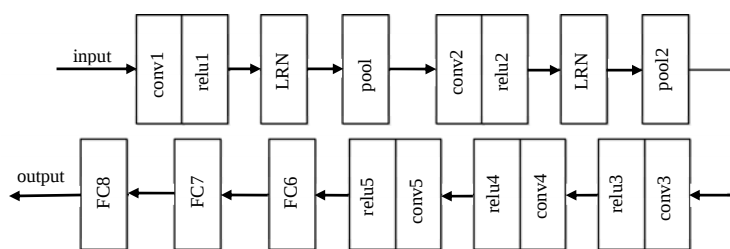


Fig.2 Alexnet network structure

图 2 Alexnet 网络结构图

卷积神经网络模型的一般结构主要包括：卷积层(convolution, conv)、池化层(pool)、全连接层(Fully connected, FC)，有些网络模型还包括“残差模块”(building block)，不同网络模型的层级之间会添加局部响应归一化(Local Response Normalization, LRN)、批量标准归一化(Batch Normalization, Batch Norm)、Relu 激活函数等处理。近年来深度学习领域在图像分类方向发展迅速，涌现了大量优秀的卷积神经网络模型并在图像分类任务中得以应用。这些卷积神经网络模型大都是以 Lenet, Alexnet, VGGnet, Resnet 模型的构架、理念为基础和灵感衍生而来。因而实验选取开发时间由早至晚的 4 种经典网络模型：Lenet5, Alexnet, VGG16, Resnet50，模型结构由简单到复杂，如图 1~图 4 所示。

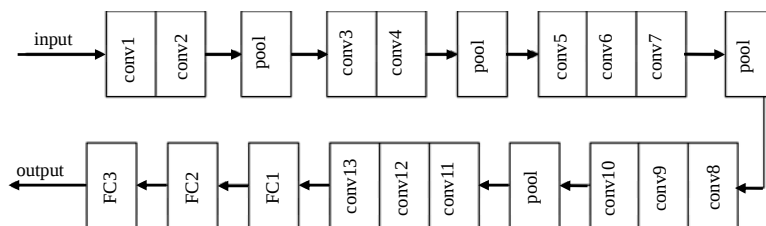


Fig.3 VGG16 network structure

图 3 VGG16 网络结构图

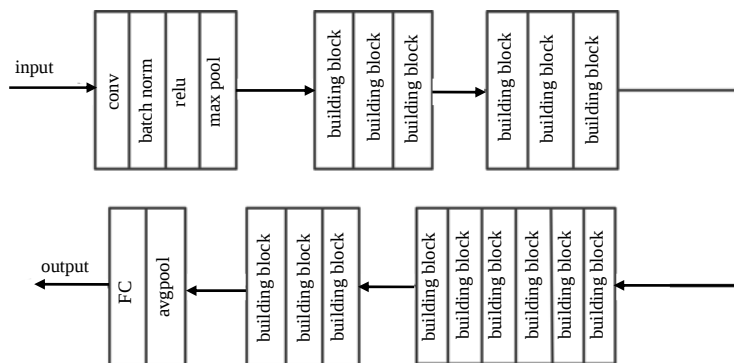


Fig.4 Resnet50 network structure

图 4 Resnet50 网络结构图

2.2 样本设计

实验选取以“海市省区县解财务发票文件实验室专用章”为印面图文内容制备光敏印章，再运用扫描打印伪造、拓印设计伪造方法分别制备 2 种高仿真光敏印章。将此 3 种印章以均匀适中的盖印压力在 A4 复印纸上每类印章分别盖印 1 000 枚印文，扫描得到印文图像作为训练样本数据，每类印章分别盖印 10 枚印文，扫描得到印文图像作为测试样本数据。

运用 Lenet5, Alexnet, VGG16, Resnet50 4 种卷积神经网络，以每类印文 100, 200, 300, ..., 1 000 张逐级递增的训练样本量共设计 40 次训练，以测试样本数据进行性能验证，将训练、测试的识别准确率、训练的交叉熵(loss)、消耗时间进行综合比对，分析训练样本量、网络模型的改变对识别准确率的影响。

3 实验结果

3.1 训练样本量对识别准确率的影响

完成 Lenet5 网络模型的 10 次不同训练样本量的训练和测试，训练、测试的准确率和 loss 值汇总如图 5 所示。训练识别准确率和测试识别准确率变化趋势一致，随着训练样本量的增多，准确率得到 30% 的提升，且 loss 值不断减小，证明训练后的网络模型成功收敛且拟合数据能力不断增强。以 3 000 张训练网络时，最高准确率可达 90.8%。

完成 Alexnet 网络模型的 10 次不同训练样本量的训练和测试，训练、测试的准确率和 loss 值汇总如图 6 所示，在 300 张训练样本量的条件下，loss 值较大，说明网络模型收敛欠佳，学习特征不足以充分训练构架中的权重参数，导致网络欠拟合，准确率仅为 34%。但随着样本量的增多，准确率得到 53% 的提升，以 2 700 张训练网络时，最高准确率可达 96.23%。

完成 VGG16, Resnet50 网络模型的 10 次不同训练样本量的训练和测试，训练、测试的准确率和 loss 值汇总见图 7~图 8，训练、测试准确率较高且稳定，当训练样本量增至 600 张以上时，测试准确率能稳定在 100% 左右，当以 300 张样本训练时，Resnet50 模型的测试准确率为 86.67%，VGG16 模型的测试准确率为 90%，VGG16 提升了约 4% 准确率归于网络模型中待训练的权重参数较少，对训练样本量的需求相对较少。

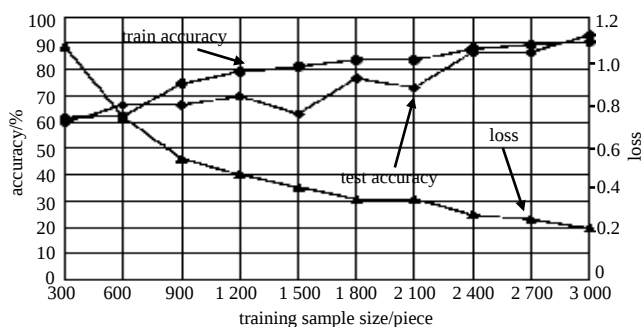


Fig.5 Accuracy rate, loss value change of Lenet5 model

图 5 Lenet5 模型准确率、loss 值变化折线图

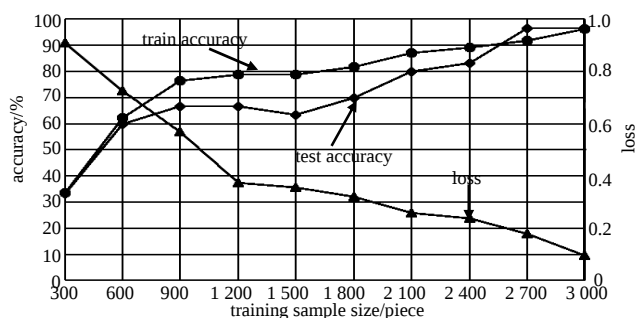


Fig.6 Accuracy rate, loss value change of Alexnet model
图 6 Alexnet 模型准确率、loss 值变化折线图

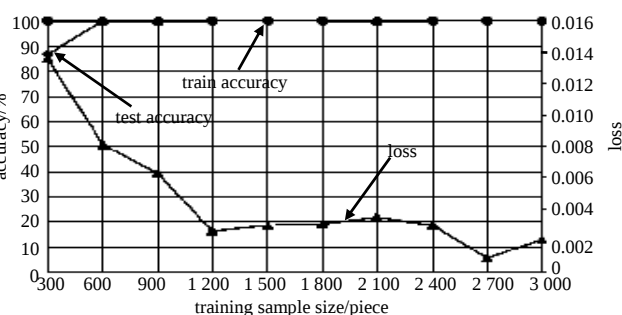


Fig.7 Accuracy rate, loss value change of VGG16 model
图 7 VGG16 模型准确率、loss 值变化折线图

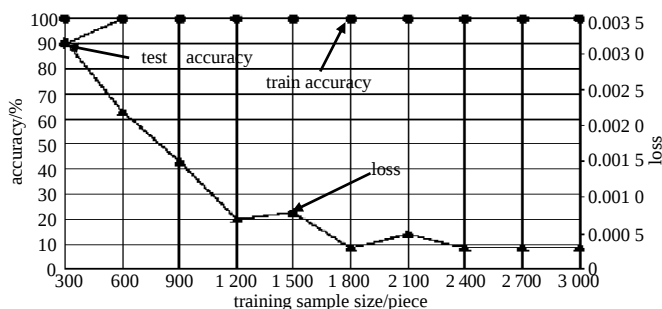


Fig.8 Accuracy rate, loss value change of Resnet50 model
图 8 Resnet50 模型准确率、loss 值变化折线图

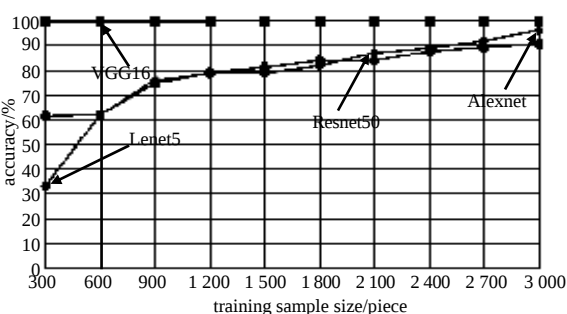


Fig.9 Training accuracy rate changes of four network models
图 9 4 种网络模型训练准确率变化图

3.2 网络模型对识别准确率的影响

综合比较 4 种网络模型的训练、测试准确率，结果如图 9~10 所示，VGG16 和 Resnet50 网络模型的优势突出，即便在训练样本量较少的条件下，网络模型强大的特征学习能力也能得到近 90% 的识别准确率；而 Alexnet, Lenet5 由于网络层级相对较少，结构简单，识别准确率波动较大，在样本量较少的条件下，Alexnet 识别性能相对较差。训练样本量至少为 2 400 张，才能保证 4 种网络模型得到 80% 以上的测试准确率。从消耗时间方面比较，Lenet5, Alexnet, Resnet50 消耗时间均在 10 min 内，如图 11 所示 VGG16 消耗时间受样本量的变化影响最大，相对于其他 3 种网络模型，所耗时间至少提升了约 70%。

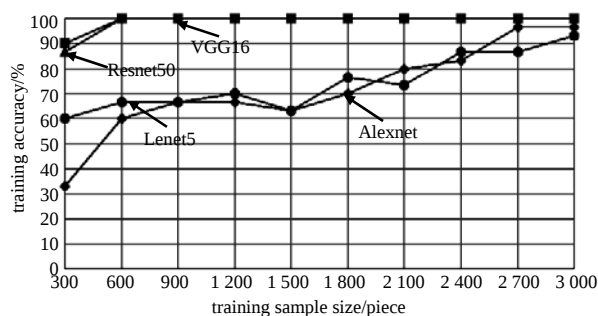


Fig.10 Test accuracy changes of four network models
图 10 4 种网络模型测试准确率变化图

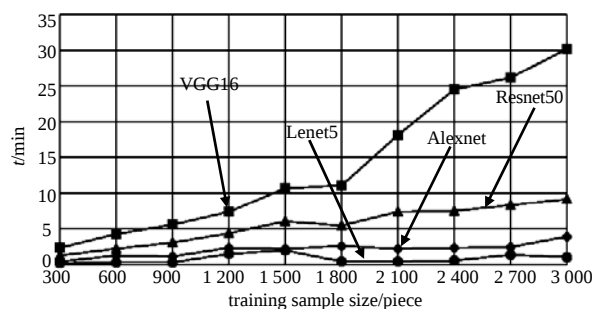


Fig.11 Time consumption changes of four network models
图 11 4 种网络模型消耗时间变化图

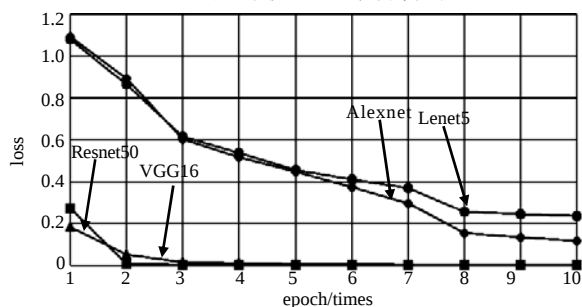


Fig.12 Loss value changes of four network models (1 000 training samples)
图 12 4 种网络模型 loss 值变化折线图 (1 000 张训练样本)

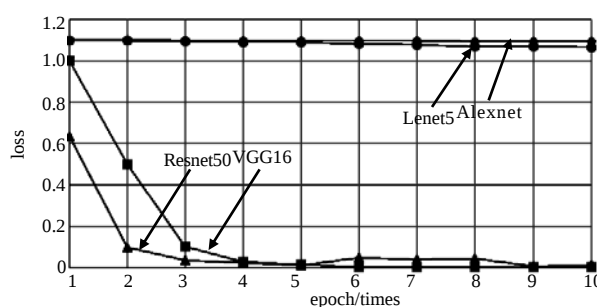


Fig.13 Loss value change of four network models (100 training samples)
图 13 4 种网络模型 loss 值变化折线图 (100 张训练样本)

优秀的网络模型不仅需要取得较高的识别准确率同时还应具备强大的泛化能力,从训练得到的 loss 值变化趋势可见,在训练样本量增多的情况下(见图 12),4 种网络模型 loss 均不断减少,但在训练样本量较少的情况下(见图 13),学习回合(epoch)增多,但 Alexnet 和 Lenet5 模型的 loss 值并没有减少,表明网络模型的“学以致用”能力欠佳。

4 结论

基于 Pytorch 实现 4 种卷积神经网络模型完成高仿真光敏印章印文的自动识别任务,通过实验对比分析得出以下结论。

1) 在高仿真光敏印章印文自动识别任务中,4 种经典卷积神经网络模型均能取得 100%的识别准确率,证明卷积神经网络能作为一种成功的方法为检验鉴定提供参考依据。

2) 充足的训练样本量是优秀模型得到高识别率的重要因素,为了取得 90%以上的识别准确率,Lenet5 网络模型至少需 3 000 张训练样本, Alexnet 网络模型至少需 2 700 张训练样本, VGG16 和 Resnet50 网络模型至少需 600 张训练样本。

3) 综合比较、分析得出:4 种网络模型 Lenet5,Alexnet,VGG16,Resnet50 泛化能力由弱到强,运行 VGG16 网络的内存开销最大,收敛速度最快; Resnet50,Alexnet,Lenet5 依次减小。

4) 综合考虑消耗时间、制作样本量代价、较高的识别准确率等方面, Resnet50 更适用于高仿真伪光敏印章印文的自动识别任务。

实验首次证明了卷积神经网络在真伪印章印文自动识别中的可行性,但在司法鉴定实际工作中,对于真伪印章印文的鉴别不能单纯用差异特征来下定论,差异特征是否为本质性是进行鉴别的关键,且需要根据案件情况和样本性质进行综合判断特征价值。因而印章印文的自动识别即使能取得 100%的识别准确率,也只能是检验鉴定的辅助依据。在实际案件中,样本量往往较少且高质量得不到保证,因而探究针对少量、质量差的样本条件如何提高识别准确率与稳定性是今后的研究方向。

参考文献:

- [1] 廖闻剑,成渝. 银行印鉴的模糊识别方法[J]. 南京航空航天大学学报, 1998(6):663-667. (LIAO Wenjian, CHENG Yu. Fuzzy identification method for bank seals[J]. Journal of Nanjing University of Aeronautics & Astronautics, 1998(6):663-667.)
- [2] 艾武,彭磊,李玲玲,等. 用于支票印鉴鉴别的图像预处理及配准算法[J]. 计算机工程与应用, 2003,39(3):34-37. (AI Wu,PENG Lei,LI Lingling,et al. Preprocessing and registration algorithm for check seal verification[J]. Computer Engineering and Applications, 2003,39(3):34-37.) doi:10.3321/j.issn:1002-8331.2003.03.011.
- [3] 张艳. 票据自动处理系统中若干关键技术研究[D]. 南京:南京理工大学, 2008. (ZHANG Yan. Research on some key technologies in automatic bill processing system[D]. Nanjing,China:Nanjing University of Science and Technology, 2008.) doi:10.7666/d.y1542675.
- [4] 邓绍秋,方郁,谢滨键. 高仿真法伪造文件的识别及检验[J]. 湖北警官学院学报, 2015,28(10):140-142. (DENG Shaoqiu, FANG Fang,XIE Binjian. Identification and verification of forged documents with high simulation method[J]. Journal of Hubei University of Police, 2015,28(10):140-142.) doi:10.3969/j.issn.1673-2391.2015.10.036.
- [5] 贾世杰,杨东坡,刘金环. 基于卷积神经网络的商品图像精细分类[J]. 山东科技大学学报(自然科学版), 2014,33(6):91-96. (JIA Shijie,YANG Dongpo,LIU Jinhuan. Product image fine-grained classification based on convolutional neural network[J]. Journal of Shandong University of Science and Technology(Natural Science), 2014,33(6):91-96.) doi:10.3969/j.issn.1672-3767.2014.06.013.
- [6] DAN C,MEIER U,SCHMIDHUBER J. Multi-column deep neural networks for image classification[C]// Conference on Computer Vision and Pattern Recognition. [S.l.]:IEEE, 2012. doi:10.1109/CVPR.2012.6248110.
- [7] CIRESAN D C,MEIER U,GAMBARDELLA L M,et al. Convolutional neural network committees for handwritten character classification[C]// International Conference on Document Analysis & Recognition. [S.l.]:IEEE, 2011:1135-1139. doi:10.1109/ICDAR.2011.229.
- [8] HINTON G,DENG L,YU D,et al. Deep neural networks for acoustic modeling in speech recognition:the shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012,29(6):82-97.
- [9] 周光磊,王长亮,杨旭,等. 高仿真光敏印章的伪造方法与检验鉴定研究[J]. 中国司法鉴定, 2018,100(5):76-79. (ZHOU Guanglei,WANG Changliang,YANG Xu,et al. Research on the methods for simulating photo-sensitive seals and its Forensic Identification[J]. Chinese Journal of Forensic Sciences, 2018,100(5):76-79.) doi:CNKI:SUN:SFJD.0.2018-05-009.

- [10] 张乾,郭磊,李金璐. 利用重合率鉴别光敏印文的实验研究[J]. 湖北警官学院学报, 2013(2):169–171. (ZHANG Qian, GUO Lei, LI Jinlu. Experimental study on identifying photosensitive seals by using coincidence ratio[J]. Journal of Hubei University of Police, 2013(2):169–171.) doi:10.3969/j.issn.1673–2391.2013.02.053.
- [11] 刘铁根,何瑾,邓集杰,等. 基于嵌入式系统结构的印鉴鉴别系统[J]. 光学学报, 2009,29(11):3043–3049. (LIU Tiegeng, HE Jin, DENG Jijie, et al. A seal imprint identification system based on embedded structure[J]. Acta Optica Sinica, 2009,29(11):3043–3049.) doi:10.3788/AOS20092911.3043.
- [12] 常军民,陈国定. 一种基于神经网络判断的印鉴自动鉴别新方法[J]. 浙江工业大学学报, 2005,33(3):268–271. (CHANG Junmin, CHEN Guoding. A new automatic seal imprint verification method based on neural networks[J]. Journal of Zhejiang University of Technology, 2005,33(3):268–271.) doi:10.3969/j.issn.1006–4303.2005.03.009.
- [13] 陈运文,王逸飞. 基于几何统计特征的印鉴自动识别算法[J]. 计算机工程与应用, 2005,41(27):4–6. (CHEN Yunwen, WANG Yifei. Seal identification using geometrical statistic features[J]. Computer Engineering and Applications, 2005,41(27):4–6.) doi:10.3321/j.issn:1002–8331.2005.27.002.
- [14] 廖闻剑,成瑜. 基于多分类器决策融合的印鉴真伪鉴别方法[J]. 南京航空航天大学学报, 2002,34(4):368–371. (LIAO Wenjian, CHENG Yu. Multi-classifier decision fusion based seal imprint verification approach[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2002,34(4):368–371.) doi:10.3969/j.issn.1005–2615.2002.04.014.
- [15] 何瑾. 金融印鉴真伪识别新方法研究[D]. 天津:天津大学, 2010. (HE Jin. Research on new method of identification of financial seals[D]. Tianjin, China: Tianjin University, 2010.) doi:10.7666/d.y1874588.
- [16] 姜文超,刘海波,杨宇杰,等. 一种融合小波变换与卷积神经网络的高相似度图像识别与分类算法[J]. 计算机工程与科学, 2018,40(9):124–130. (JIANG Wenchao, LIU Haibo, YANG Yujie, et al. A high similar image recognition and classification algorithm fusing wavelet transform and convolution neural network[J]. Computer Engineering and Science, 2018,40(9):124–130.) doi:CNKI:SUN:JSJK.0.2018-09-018.
- [17] 王茜,张海仙. 深度学习框架 Caffe 在图像分类中的应用[J]. 现代计算机, 2016(5):72–75. (WANG Xi, ZHANG Haixian. Application of deep learning framework Caffe in image classification[J]. Modern Computer, 2016(5):72–75.) doi:10.3969/j.issn.1007–1423.2016.05.016.
- [18] SUN M, ZHANG D, REN J, et al. Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-wash paintings[C]// International Conference on Image Processing(ICIP). [S.l.]:IEEE, 2015. doi:10.1109/ICIP.2015.7350874.
- [19] ABDEL-HAMID 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):1533–1545. doi:10.1109/TASLP.2014.2339736.
- [20] LI S, LIU Z Q, CHAN A B. Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network[C]// Conference on Computer Vision & Pattern Recognition Workshops. [S.l.]:IEEE, 2014. doi:10.1109/CVPRW.2014.78.