Fine-Art Painting Classification via Two-Channel Deep Residual Network

Xingsheng Huang, Sheng-hua Zhong(✉), and Zhijiao Xiao

College of Computer Science and Software Engineering, Shenzhen University,
Shenzhen 518000, People's Republic of China
huangxingsheng2016@email.szu.edu.cn,
{csshzhong,cindyxzj}@szu.edu.cn

Abstract. Automatic fine-art painting classification is an important task to assist the analysis of fine-art paintings. In this paper, we propose a novel two-channel deep residual network to classify fine-art painting images. In detail, we take the advantage of the ImageNet to pre-train the deep residual network. Our two channels include the RGB channel and the brush stroke information channel. The gray-level co-occurrence matrix is used to detect the brush stroke information, which has never been considered in the task of fine-art painting classification. Experiments demonstrate that the proposed model achieves better classification performance than other models. Moreover, each stage of our model is effective for the image classification.

Keywords: Image classification · Fine-art painting classification
Gray-level co-occurrence · Deep residual network

1 Introduction

In the history of world civilization, fine-art painting plays a very important role. Fine-art painting fully expresses the state of mind and social culture of mankind in different times. Nowadays, smart mobile devices have penetrated into every detail of people's daily life, which leads to the rapid development of digital collection of fine-art paintings. Hence, vast digital collections have been made available across the Internet and museums. With a large number of digital works collection, it is very important to automatically process and analyze the fine-art paintings. Moreover, automatic fine-art painting classification is an important task to assist the analysis of fine-art paintings, such as forging detection [1], object retrieval [2, 3], archiving and retrieval of works of fine-art [4, 5] and so on.

Since Krizhevsky and Hinton successfully applied the CNN model for image classification, there has been a significant shift away from shallow image descriptors towards deep features [6]. In the classification of natural images task, Ren et al. has achieved great success [7]. However, for the classification of fine-art paintings, CNN's performance is somewhat unsatisfactory. One of the main reasons is that the number of samples for fine-art painting classification is limited. For example, Painting-91, which is the largest number of fine-art painting dataset [8], only has 4266 images. Therefore,

B. Zeng et al. (Eds.): PCM 2017, LNCS 10736, pp. 79–88, 2018.
https://doi.org/10.1007/978-3-319-77383-4_8

considering the very limited training data, CNN is difficult to effectively extract features and achieve good performance.

Evidences from previous work show that CNN's success is, in the field of computer vision, relied on the availability of large-scale datasets with labels [9–14]. For example, for the classification of ImageNet, Krizhevsky proposed CNN model to effectively solve the problem of over-fitting [6]. One important reason is that the consistency of the up to 144 million parameters of the CNN model and the millions samples of ImageNet dataset. In view of the limited number of fine-art painting samples, Hentschel et al. proposed a fine-tuning method to solve this problem [15]. That is, a CNN model is firstly pre-trained on a large-scale dataset such as ImageNet, and then it is fine-tuned with the target dataset. Thus, the fine-tuning can, in the case of the limited sample of fine-art painting datasets, help us to construct an effective learning model based on the pre-trained CNN. Thus, in this paper, our proposed model also uses ImageNet dataset to pre-train our model.

With the stage of fine-tuning, CNN can solve the problem of insufficient data in the classification task of fine-art paintings. As we known, driven by the increases of depth, the notorious problem of vanishing/exploding gradients could hamper convergence of the deep networks. He et al. partially solved this problem by introducing a deep residual learning framework. Hence, our proposed model is based on deep residual neural networks [16–18].

In the task of fine-art painting classification, although some researchers tried to use some existing deep learning model or construct some new deep learning models, but all these models did not take into account the essential characteristics of fine-art paintings. Brush stroke is an important and powerful tool to understand the fine-art painting [19]. Unfortunately, this important character has never been considered in the classification of fine-art painting. Thus, in our work, we try to use the gray-level co-occurrence matrix (GLCM) to represent this information and it is set as the input of brush stroke information channel. In this paper, we propose a novel two-channel deep residual network for the classification task of fine-art paintings. This model is consisted of two channels, RGB channel and brush stroke information channel. This model firstly pre-trains on ImageNet dataset, and then it is fine-tuned with the fine-art painting dataset.

The rest of this paper is organized as followings. The second part briefly introduces the related work for fine-art painting classification. The third part introduces the architecture of our proposed model. The fourth part introduces the experimental setting and provides the experimental results. Finally, the conclusions are drawn in section five.

2 Related Work

Recently, CNN is widely used in the classification of fine-art painting images. Some researchers have suggested that CNN can be used as a feature extractor. Elgammal et al. investigated the effects of different features coupled with different types of metrics to perform the fine-art painting classification task. Although the CNN was employed, it was simply used as a feature extractor only [20]. As we described before, the number of samples in the fine-art painting datasets is very limited. Thus, some researchers try to combine the pre-training and fine-tuning stages to extract the effective image features from fine-art painting images. Tan et al. show that combine the pre-training and the fine-tuning stages can improve the performance of the deep learning model for the classification task of fine-art paintings [21]. Hentschel et al. pre-trained the deep representations on the ImageNet dataset and used fine-tuning for fine-art painting image datasets to evaluate the learned models [15].

More researchers have proposed novel models by reconstructing the structure of CNN to improve the performance of the classification task with fine-art painting images. Peng proposed cross-layer CNN is formed by cascading a number of modified CNN [22, 23]. Each modified CNN in the cross-layer is as same as Krizhevsky's CNN except that the convolution layer is removed. Tan replaced the last layer of CNN with a SVM classifier instead of a softmax layer [21].

3 Fine-Art Painting via Two-Channel Deep Residual Network

3.1 Two-Channel Deep Residual Network Architecture

In this part, we will introduce the details about the proposed model: fine-art painting via two-channel deep residual network (FPTD). The structure of the proposed FPTD is shown in Fig. 1. In RGB channel, the original RGB image of each fine-art painting image is input into the deep residual network (ResNet). In brush stroke information channel, the GLCM image is used to extract the brush stroke information and is input into the ResNet. The output of each channel is a 2048-dimensional vector, and 2048 is the number of kernel of the last convolution layer. Then, they are combined as a 4096-dimensional feature. This feature is input to the SVM classifier. We use LIBSVM Toolbox to implement SVM classifier and use the gaussian kernel and the grid

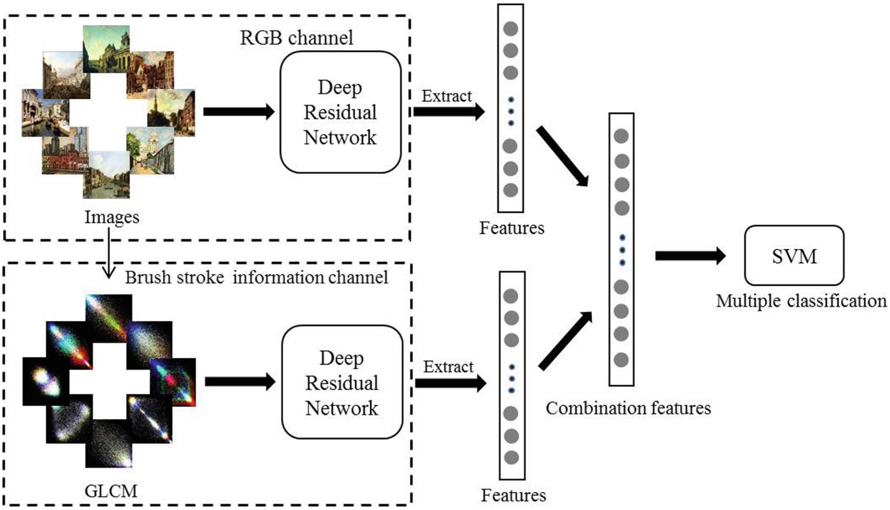

Fig. 1. Two-channel framework for fine-art paintings classification

optimization to find the optimal value of C in the parameter space [2−10: 1000] with a step of one [20]. To overcome the limitation of the number of samples, our model firstly pre-trains on ImageNet dataset, and then it is fine-tuned with the fine-art painting dataset.

3.2 RGB Channel

The RGB channel uses the original fine-art painting image as the input to learn the model. The output is a 2048-dimensional vector, and 2048 is the number of kernel of the last convolution layer. In this paper, we use two versions of ResNet, including 14 layers and 50 layers, The ResNet structure of each channel is shown in Table 1. To the setting of building blocks, the number of blocks stacked, and the down-sampling stages, we follow the previous work of He et al. [16].

Table 1. The architecture of ResNet in our proposed model

Layer name	Output size	14-layer ResNet	50-layer ResNet
Conv1	112 × 112	7 × 7, 64, stride 2	
Conv2_x	56 × 56	3 × 3max pool, stride 2	
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28 × 28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	14 × 14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5_x	7 × 7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1 × 1	Average pool, 1000-dimensional fc, softmax	

3.3 Brush Stroke Information Channel

The brush stroke is a fundamental part of fine-art paintings, and it can also be an important tool to analyze or classify fine-art paintings. Because the brush stroke is also known as the texture information of painting, we use gray-level co-occurrence matrix (GLCM) to describe this kind of information in fine-art painting images and it is utilized as the input of the brush stroke information channel.

Q is an operator that defines the relative position of two pixels relative to each other and consider an image I of size $M \times N$ with L possible gray levels. $\mathbf{G}$ is a matrix whose element g is the number of times the pixel pair with gray levels i and j appear at the position specified by Q in I, where $1 \leq i, j \leq L$. In this paper, Q is defined as one pixel immediately to the right. Hence, $\mathbf{G}$ could be defined as Eq. 1.

$$\mathbf{G} = \left(g_{ij}\right)_{L \times L} \tag{1}$$

$$g_{ij} = \left|\left\{(x, y) | \mathbf{I}(x, y) = i,\ \mathbf{I}(x, y+1) = j, 8 \cdot m \leq x \leq 8 \cdot m + 7, 8 \cdot n \leq y \leq 8 \cdot n + 7\right\}\right| \tag{2}$$

$$m = 0, 1, 2, \cdots, \left\lfloor \frac{M}{8} \right\rfloor + 1 \tag{3}$$

$$m = 0, 1, 2, \cdots, \left\lfloor \frac{N}{8} \right\rfloor + 1 \tag{4}$$

In our work, we obtain the gray-level co-occurrence matrix $\mathbf{G}$ for each color channel (R, G, and B). And then we combine them as a 3D matrix, which is referred as GLCM image.

Figure 2 shows four sample images with different styles and their corresponding gray-level co-occurrence matrix images. Figures 2(a) and (c) are sample images of "neoclassicism". Figures 2(e) and (g) are sample images of "northern renaissance". Figures 2(b) and (d) are the corresponding GLCM image of Figs. 2(a) and (c), respectively. Figures 2(f) and (h) are the corresponding GLCM image of Figs. 2(e) and (g), respectively. We can find these two styles are not similar in style and in vision, and their GLCM image are different.

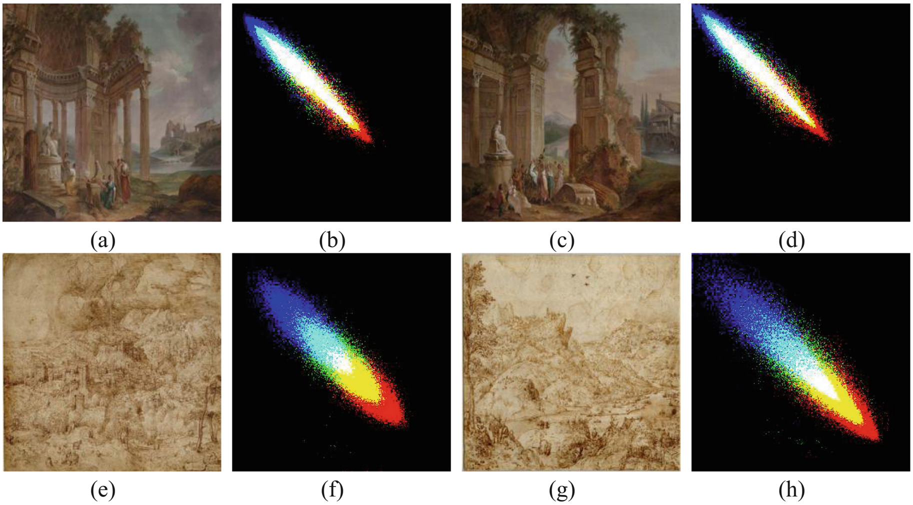

Fig. 2. (a) and (c) are sample images, which style is "neoclassicism". (b) and (d) are the extracted GLCM image of (a) and (c), respectively. (e) and (g) are sample images, which style is "northern renaissance". (f) and (h) are the extracted GLCM image of (e) and (g), respectively.

This GLCM image is used as the input of brush stroke information channel. The output of the brush stroke information channel is also a 2048-dimensional vector, and 2048 is the number of kernel of the last convolution layer. The structure of ResNet in brush stroke information channel and RGB channel is exactly the same, as shown in Table 1.

4 Experiment

In this section, we first introduce the experimental setting in Sect. 4.1. In Sect. 4.2, we evaluate the proposed method for the classification of fine-art painting on *style*, *genre* and *artist* datasets.

4.1 Experimental Setting

We conduct the experiments on three datasets to validate the performance of our method. The style dataset, the genre dataset and the artist dataset are downloaded from the WikiArt.org – Encyclopedia of fine-art painting website. The paintings in the website as well as the annotations are contributed by a community of experts [15].

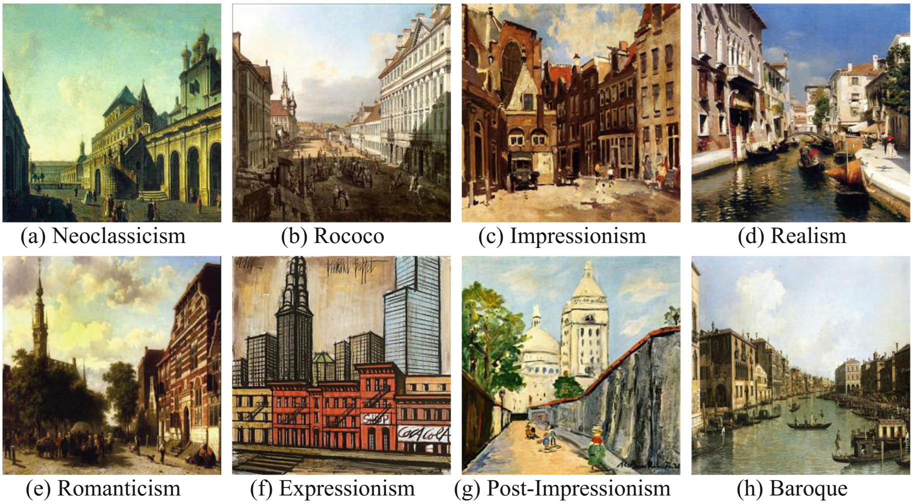
(a) Neoclassicism
(b) Rococo
(c) Impressionism
(d) Realism
(e) Romanticism
(f) Expressionism
(g) Post-Impressionism
(h) Baroque

Fig. 3. The sample images in the *style* dataset with different style labels

The *style* dataset is collected of a total 30825 images, including 25 styles, each style has 1233 pictures. The *genre* dataset has 28,760 images containing 10 genres and 2876 graphs for each genre. The *artist* dataset has 9766 images, including 19 artists, each has 514 images. We resize all images to 256 × 256, and 60% of images in each dataset are used for training the model, and the remaining 40% images are used for test.

As shown in Fig. 3, we can find that, although the foreground objects of all these images are “buildings”, they actually belong to different style categories. It brings a lot of difficulty for the fine-art painting classification.

In our experiments, AlexNet and ResNet were trained using the stochastic gradient descent (SGD) with a batch size of 256 images. By following the setting of the AlexNet [6], their learning rate ε for the training epoch p with respect to the current epoch i is set to be

$$\varepsilon_i = 10^{-1-4\times\frac{i-1}{p-1}} \tag{5}$$

where p is a positive integer to ensure that the model is convergent. In our experiments, p is set to be 180. At that time, all the learning models are already converged.

4.2 Experimental Result

4.2.1 Fine-Tuning Is Useful to Improve Classification Accuracy

In Table 2, we provide the classification accuracies on three datasets. We compare AlexNet and ResNet with pre-training and without pre-training for the classification task of fine-art painting. We provide two versions of ResNet, 14 layers and 50 layers. Moreover, all the deep learning models here only include RGB channel.

Table 2. The comparisons of the classification performance on *style*, *genre* and *artist* datasets using different network structures with or without pre-training

Dataset	Network	Without pre-training		With pre-training	
		Top-1 error rates (%)	Top-5 error rates (%)	Top-1 error rates (%)	Top-5 error rates (%)
Style	AlexNet	69.23	31.76	56.71	19.12
	50-layer ResNet	67.16	27.08	**49.91**	**11.96**
	14-layer ResNet	**62.28**	**21.88**	51.5	13.22
Genre	AlexNet	51.18	10.38	34.95	4.15
	50-layer ResNet	51.61	9.69	**31.04**	**3.04**
	14-layer ResNet	**48.65**	**7.78**	32.91	3.43
Artist	AlexNet	53.74	19.28	27.34	5.6
	50-layer ResNet	57.82	19.33	**18.13**	**2.75**
	14-layer ResNet	**44.29**	**11.42**	19.61	2.93

From Table 2, we can find the performance of ResNet is better than AlexNet. This is because ResNet has the advantage that solves the problem of vanishing/exploding gradients. To each case, the models with pre-training achieve better performance than the models without it. It evidences that the pre-training is helpful to learn an effective model. Here, we can also find if we do not use ImageNet to pre-train the model, 14-layer ResNet could obtain smaller error rate. But this case does not happen in the case of the models with pre-training. That is because 50-layer ResNet gains better learning effect than 14-layer ResNet in the stage of pre-training [16].

4.2.2 Brush Stroke Information Is More Helpful to Improve the Classification Accuracy

In the previous section, we have verified that the pre-training stage could improve the performance for different classification tasks. In this section, we try to validate the effectiveness of the brush stroke information. We compare the proposed method FPTD with the model which has only one channel (RGB channel). Here, we provide the results of two versions of ResNet, 14 layers and 50 layers.

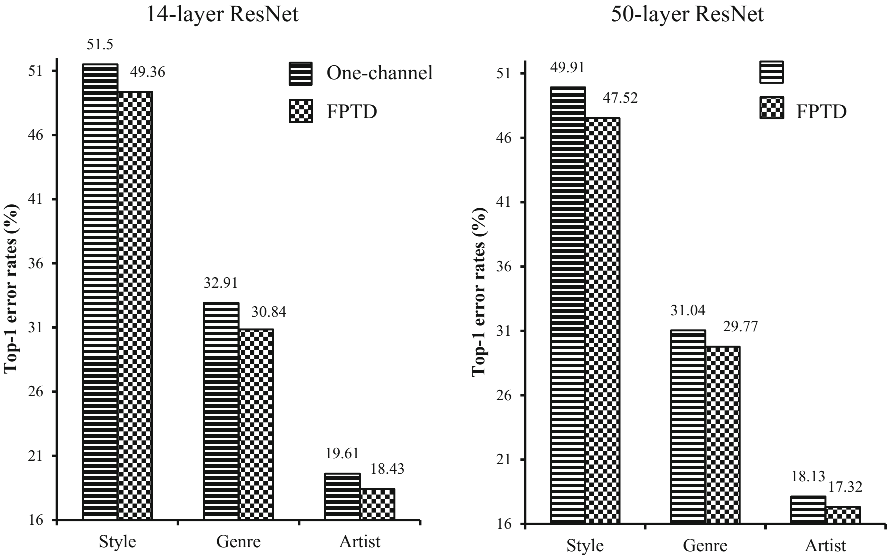

Fig. 4. The classification error rates comparisons of FPTD and one-channel model.

From Fig. 4, we can find the top-1 error rates of the proposed two-channel model FPTD are obviously less than the model with only one channel in each datatset. Moreover, the accuracy of the 50-layer ResNet is also better than the 14-layer ResNet. All the previous results demonstrate the importance of the brush stroke for the classification of fine-art paintings. Moreover, each stage of our model is effective for the fine-art painting classification.

5 Conclusion

Brush stroke is an important and powerful tool to understand the fine-art painting. Unfortunately, this important character has never been considered in the classification of fine-art painting. In this paper, we propose a novel model for fine-art painting classification via two-channel deep residual network, including RBG channel and brush stroke information channel. In detail, we take the advantage of the ImageNet to pre-train the deep residual network. The gray-level co-occurrence matrix is used to detect the brush stroke information, as the input of the brush stroke information channel. In order to validate the performance of our model, we run two experiments. In the first experiment, we find the pre-training is helpful to learn an effective model. We also find that the performance of ResNet is better than AlexNet, and the accuracy of 50-layer ResNet is better than 14-layer ResNet. In the second experiment, we find that the classification accuracy of our proposed two-channel model is obviously better than the model with only one channel. In future, we will try to integrate more characters of fine-art painting images into our model to improve the classification performance.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 61502311), the Natural Science Foundation of Guangdong Province (No. 2016A03 0310053), the Science and Technology Innovation Commission of Shenzhen under Grant (No. JCYJ20150324141711640, JCYJ20160422151736824), the Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), the Shenzhen high-level overseas talents program, and the Tencent “Rhinoceros Birds” - Scientific Research Foundation for Young Teachers of Shenzhen University (2015, 2016).

References

1. Polatkan, G., Jafarpour, S., Brasoveanu, A., Hughes, S., Daubechies, I.: Detection of forgery in paintings using supervised learning. In: IEEE International Conference on Image Processing (ICIP), pp. 2921–2924 (2009)
2. Crowley, E.J., Zisserman, A.: In search of art. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8925, pp. 54–70. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16178-5_4
3. Crowley, E., Zisserman, A.: The state of the art: object retrieval in paintings using discriminative regions. In: British Machine Vision Conference (BMVC) (2014)
4. Mensink, T., Van Gemert, J.: The rijksmuseum challenge: museum centered visual recognition. In: Proceedings of International Conference on Multimedia Retrieval (ICMR), p. 451 (2014)
5. Gatys, L.A., Ecker A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
6. Krizhevsky, A., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105 (2012)
7. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems (NIPS), pp. 91–99 (2015)
8. Khan, F.S., Beigpour, S., Van-De-Weijer, J., Felsberg, M.: Painting-91: a large scale database for computational painting categorization. Mach. Vis. Appl. **25**(6), 1385–1397 (2014)
9. Zhong, S., Liu, Y., Hua, K.: Field effect deep networks for image recognition with incomplete data. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **12**(4), 52 (2016)
10. Zhong, S., Liu, Y., Li, B., Long, J.: Query-oriented unsupervised multi-document summarization via deep learning. Expert Syst. Appl. (ESWA) **42**(21), 8146–8155 (2015)

11. Zhong, S., Liu, Y., Liu, Y.: Bilinear deep learning for image classification. In: Proceedings of 19th ACM International Conference on Multimedia (ACMMM) (2011)
12. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
13. Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P.: Deep-plant: plant identification with convolutional neural networks. In: IEEE International Conference on Image Processing (ICIP), pp. 452–456 (2015)
14. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_23
15. Hentschel, C., Wiradarma, T.P., Sack, H.: Fine tuning CNNs with scarce training data-adapting ImageNet to art epoch classification. In: IEEE International Conference on Image Processing (ICIP) (2016)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
17. Zheng, M., Zhong, S., Wu, S., Jiang, J.: Steganographer detection via deep residual network. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) (2017)
18. Wu, S., Zhong, S., Liu, Y.: Deep residual learning for image steganalysis. In: Multimedia Tools and Applications (MTAP) (2017)
19. Li, L.: The relationship between the brush strokes and the image, the color, the emotion. J. Yangtze Univ. (Socia1 Sciences), **34**(9), 181–182 (2011)
20. Saleh, B., Elgammal, A.: Large-scale classification of fine-art paintings: learning the right metric on the right feature. arXiv preprint arXiv:1505.00855 (2015)
21. Tan, W.R., Chan, C.S., Aguirre, H.E., Tanaka, K.: Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification. In: IEEE International Conference on Image Processing (ICIP) (2016)
22. Peng, K.C., Chen, T.: Cross-layer features in convolutional neural networks for generic classification tasks. In: IEEE International Conference on Image Processing (ICIP) (2015)
23. Peng, K.C., Chen, T.: A framework of extracting multi-scale features using multiple convolutional neural networks. In: International Conference on Multimedia and Expo (ICME) (2015)