Deep Multibranch Neural Network for Painting Categorization

Simone Bianco(✉), Davide Mazzini, and Raimondo Schettini

Dipartimento di Informatica, Sistemistica e Comunicazione,
Università degli Studi di Milano-Bicocca, viale Sarca 336, 20126 Milan, Italy
{simone.bianco,davide.mazzini,raimondo.schettini}@disco.unimib.it

Abstract. Coarse features, such as scene composition and subject together with fine details, such as strokes and line styles, are useful clues for painter and style categorization. In this work, to automatically predict painting's artist and style, we propose a novel deep multibranch neural network, where the different branches process the input image at different scales to jointly model the fine and coarse features of the painting. Experiments for both artist and style classification tasks are performed on the challenging Painting-91 dataset, that includes 91 different painters and 13 diverse painting styles. Our method outperforms the best method in the state of the art by 14.0% and 9.6% on artist and style classification respectively.

Keywords: Painting categorization · Painting style classification · Painter recognition · Deep convolutional neural network · Multiresolution

1 Introduction

Research on digital analysis of paintings is gaining increasing attention due to the large quantities of visual artistic data [4,10,12], made available from art museums digitizing their collection for cultural heritage, and the need of automatic tools to organize and manage them. In this work, we approach the problem of categorizing a painting by automatically predicting its artist and style given solely the digital version of the painting itself [1]. Both these tasks are very challenging due to the large amount both inter- and intra-class variations, e.g. the different personal styles in the same art movement, or the same artist adhering to different schools in different periods in his/her production. Artist classification consists in automatically associate the painting to its painter. In this task factors such as stroke patterns, the color palette used, the scene composition, and the subject must be taken into account. Style classification consists in automatically categorize a painting into the school or art movement it belongs to. Art theorists define an artistic style as the combination of iconographic, technical and compositional features that give to a work its character [20]. Style categorization is complicated by the fact that styles may not remain pure but could be influenced by others.

S. Battiato et al. (Eds.): ICIAP 2017, Part I, LNCS 10484, pp. 414–423, 2017.
https://doi.org/10.1007/978-3-319-68560-1_37

1.1 Contribution

We propose a multiresolution approach to solve the tasks of artist and style categorization. A particular random-crop strategy permits to gather clues from low-level texture details and, at the same time, exploit the coarse layout of the painting. The classification process is carried on by a specifically-taylored multibranch neural network.

Experiments are performed on the challenging painting-91 dataset [10]. On both artist and style classification tasks our approach improves the mean classification accuracy by 14.3% and 10.2% respectively, compared to the previous state-of-the-art models.

1.2 Prior Works

The problem of painter or style categorization has been faced using different tecniques. Some existing approaches make use of traditional handcrafted features [4,10] whereas more recent works relay on the use of deep networks [1,14,15,18,19]. Zhao et al. [21] used a pretrained neural network in a two-step bootstrap approach to categorize ancient illustration from the British Library. Peng and Chen [15] use a multiresolution approach to exploit both small details and the overall image structure. A more sophisticated technique is used by [1] where the use of a deformable part model is adopted in order to combine low-level details and an holistic representation of the whole painting. Deep CNNs have been widely used as features extractors to solve different tasks [3,16], Peng and Chen [15] and Anwer et al. [1] relay on pretrained deep CNNs to deal with the small quantity of images of the Painting-91 dataset. Tan et al. [18] made different experiments by training a network from scratch or finetuning an existing network for the task of style and painter recognition. They adopted a network structure similar to the one used by Krizhevsky et al. [11]. Hentschel et al. [8] performed interesting experiments about the quantity of data needed to fine-tune the network by Krizhevsky et al. [11] for the task of style classification.

2 Our Approach

The scene composition and the subject depicted are important clues to recognize a particular author or a painting style. These elements need to be extracted from the whole painting. At the same time finer details, such as stroke patterns or the line styles, are also very good clues. Obviously a powerful discriminative model should consider both the coarse level and fine details. On the basis of these considerations we decided to adopt a multiresolution approach: first, a predefined number of squared "small" crops are extracted from the high-resolution image. Then, the image is downsampled and another "large" crop is extracted from the low-resolution image (see Sect. 2.1). All the crops are then fed to the branches of a deep neural network that extracts the corresponding features. The outputs of the branches are collected by a join layer and fed to a deep neural network that carries on the categorization process.

2.1 Input Preprocessing

The first preprocessing step consists in normalizing the input image by subtracting the mean and dividing by the standard deviation of the pixel distribution of whole training set. This contrast normalization preprocessing is known to improve CNNs accuracy in different domains [2] by limiting the variability of the input range. The second step consists in a particular cropping strategy. Crops are taken at multiple resolutions to capture both fine details and coarse structures. Since paintings exhibit high variability in terms of aspect-ratios, the input image is resized such as the minimum side is 512 pixels and the aspect ratio is preserved. From the resulting image we extract two squared random crops of 227 pixels side. Then the image is further downsampled, using an average pooling layer, such as the minimum side is 256 pixels and another squared crop of 227 by 227 pixels is extracted. All the crops are squared, independently from the original aspect ratio of the input image. This is done to improve the computational efficiency allocating GPU memory blocks only once. Images and crops sizes has been choosen as a tradeoff to exploit fine details and to limit the computational burden accordingly to the size and quality of the original images. The coordinates of the crops inside the input image are randomly chosen with the only constraint that crops coming from the same scale do not overlap. The rational behind this choice is that the salient details can be anywhere inside the painting, and the extraction of crops at random locations permits the implementation of a consensus strategy by simply processing the same input image several times. The consensus strategy consists in averaging the output of the last fully-connected layer for the multiple passes of the same image trough the network, resulting in a feature vector that is then fed to the softmax layer to get the final prediction.

2.2 Deep Network Structure

We propose a novel network whose structure is shown in Fig. 1. It is composed of five modules: three branches to extract the low level structures of the painting crops, a join module to gather the output of the three branches and a classification module to make the prediction. Each branch is trained with crops from a specific scale, thus becoming specialized in processing texture patterns at that specific resolution. We decided to use only two scales since, in our preliminary experiments, the use of higher scales brought a slight improvement compared to the exponential increase of computational burden.

In the three branches and in the classification model our deep network makes use of Residual Blocks which have been shown to be an effective architectural choice to build very deep networks [7] and tackle the problem of vanishing gradients by using shortcut connections. In particular, we used “bottleneck” Residual Blocks, which allow the network architecture to be even deeper [7]. Each skip connection has four times the number of channels with respect to the internal elements of the block. This permits a large troughput of information among layers while mantaining a low computational complexity and low memory use

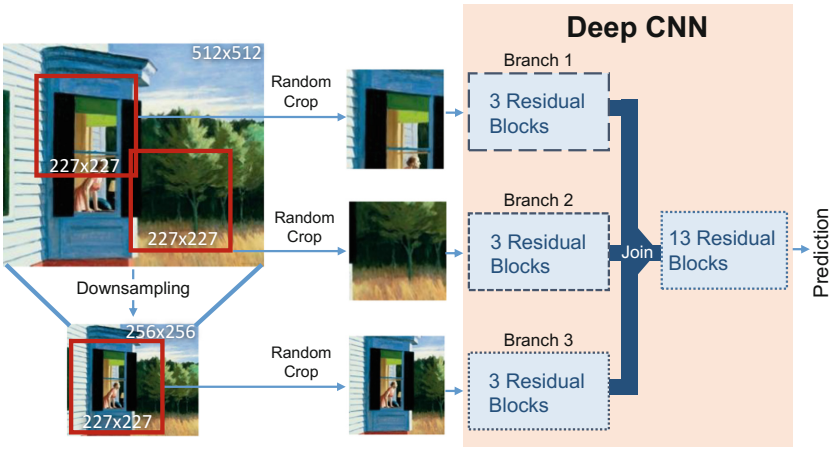

Fig. 1. Scheme of our deep multibranch neural network

insde each block. Our Residual Block structure is different from the one used by He et al. [7]: we moved the Batch Normalization layer [9] after the sum with the skip connection because, in our experiments, the resulting configuration has shown better performances.

The Residual Block we used is shown in Fig. 2. In our network (see Fig. 1) each of the three branches is composed by three Residual Blocks plus four layers near the input which perform the first processing (Convolution + BatchNorm [9] + ReLU [13]) and an initial downsampling (Max Pooling). The join module is a particular Residual Block which gathers the output of the three branches. It stacks the output features and then converts them to a smaller-dimensional feature space by compressing information along the channel dimension. The reason behind this operation is to make the computations feasible in the following layers by reducing the channel dimension of the output by a factor of three.

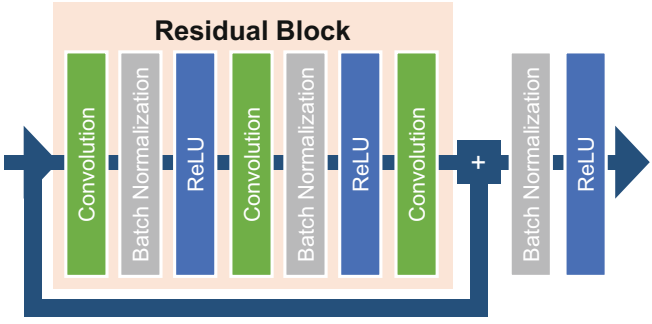

Fig. 2. The type of residual block used in our deep neural network

The classification module is composed by 13 Residual Blocks plus a Spatial Average Pooling layer, a Fully-connected layer and a Softmax layer that outputs the classes probabilities. While the Residual Blocks in the three branches do not include any downsampling operator, the classification module uses convolution operators with stride two to perform a spatial downsampling of the input. Every five blocks the input is spatially reduced by a factor of two. At the same time the number of channels is increased by the same amount. This leads to a gradual increasing of the receptive-fields of the network in the deeper layers and also favors more abstract representations of the input. In the final part of the classification module a fully-connected layer maps the output to 13 or 91 classes depending on the task, respectively artist or style categorization.

3 Experiments

3.1 Dataset

We evaluate our recognition pipeline on the challenging Painting91 dataset [10] for both artist and style classification tasks. The dataset consists of 4266 paintings of 91 painters. As train and test split we used those provided by the authors which are in both cases nearly 50%. For the task of artist recognition, the whole dataset is used whereas for the task of style recognition only 2338 groundtruth are provided.

3.2 Training

Our training procedure was carried on in two phases. We first pretrained our deep network on the Kaggle dataset Painterbynumbers.[1] This dataset is intended for a similar task, i.e. painter verification, but it is much bigger. It contains more than 1500 authors and a training set of 79433 images. Then we finetuned it two times (one for each of the two tasks) on the Painting91 dataset, substituting the last fully connected layer with a new one that matched the number of classes needed for each task.

To cope with the small amount of training data we exploited some data augmentation techniques:

- Color Jitter. It consists in randomly modifying constrast, brightness and saturation of the input image indipendently.
- Lighting noise. It is a pixelwise transform based on the eigenvalues of the RGB pixel distribution of the dataset. It has been introduced by Krizhevsky et al. [11].
- Gaussian Blur. It consists in applying a blur filter with fixed σ to random images choosen with probability 0.5.

[1] https://www.kaggle.com/. We took part to the Painterbynumbers competition and ended among the top positions. Our method, that is disclosed here, achieves an accuracy of 53.8% on validation set for the task of artist classification.

– Geometric transforms. It includes small changes in scale and aspect-ratio of the input image.

As explained in the Subsect. 2.1 our network exploits random crops. Therefore if the same input is processed several times by the same network, the final prediction vectors can be averaged before being fed to the last softmax layer. In Table 1 we report the performance in terms of accuracy at different number of passes. Results are averaged over ten independent runs. The biggest improvement is obtained by exploiting two passes with respect to the single one. The best performance are obtained using four passes.

Table 1. Accuracy vs number of passes trough the network. Each value represents the average of 10 runs.

Passes	1	2	4	8
Artist	77.5	78.1	**78.5**	78.3
Style	83.6	84.1	**84.4**	84.3

3.3 Results

In Table 2 we report the performances of our method with respect to the state-of-the-art on the Paintings-91 dataset. Concerning our method, we report the average accuracy over ten independent runs together with the minimum and maximum values. Considering our average performance, our method outperforms the best method in literature by 14.0% and 9.6% on the task of artist and style categorization respectively.

Table 2. Comparison with the state of the art. Average classification rates on the Paintings-91 dataset for the tasks of Artist and Style recognition. Our values are obtained as the maximum of 10 runs.

Method	Artist	Style
VGG-16 FC [17]	51.7	67.2
MF [10]	53.1	62.2
CL-CNN [14]	56.4	69.2
MS-MCNN [15]	58.1	71.0
MOP [6]	59.7	68.8
Holistic [5]	61.8	70.1
Holistic + Part Based [1]	64.5	74.8
Ours (worst performance among 10 runs)	77.9	83.8
Ours (average performance among 10 runs)	78.5	84.4
Ours (best performance among 10 runs)	**78.8**	**85.0**

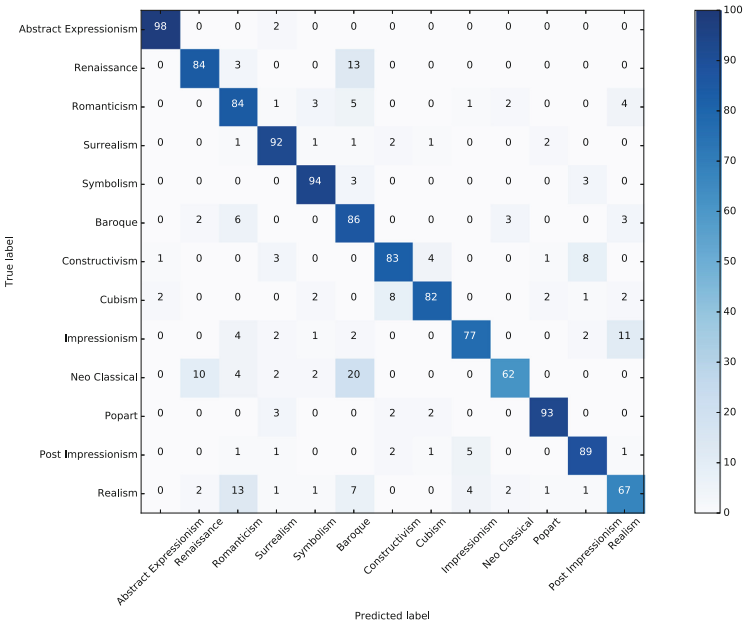

Fig. 3. Confusion matrix for the task of style recognition. The highest error rates are between Neo-Classical paintings, Baroque and Renaissance.

Figure 3 shows the confusion matrix for the style recognition task. The highest classification errors are between the Neo-Classical, Baroque and Renaissance classes. This seems to agree with styles' contaminations and influences as studied by art historians. For example Caravaggio paintings are classified as Baroque in Paintings-91 groundtruth. Actually he lived at the end of the Renaissance era, having a great influence on future Baroque painters.

Figure 4 shows the confusion matrix for the task of artist recognition. The highest error rates are between Memling and Van Eyck (27%), and Zurbaran and Vermeer (30%). Memling and Van Eyck are contemporaneous and both belonging to the Dutch and Flemish Renaissance, while Zurbaran and Vermeer are coeval painters, both belonging to the Baroque movement. To be able to actually discriminate between the last two painters, the network should be aware that Vermeer paintings are usually about indoor every-day life scenes whereas Zurbaran mostly painted religious subjects.

Figure 5 shows in the top row the highest scored errors. To better denote the complexity of the task, we also reported the highest scored and correctly classified example for the corresponding painter. Most confusions are between coeval painters. Even for an untrained human it could be difficult to predict the correct artist for a new unseen painting.

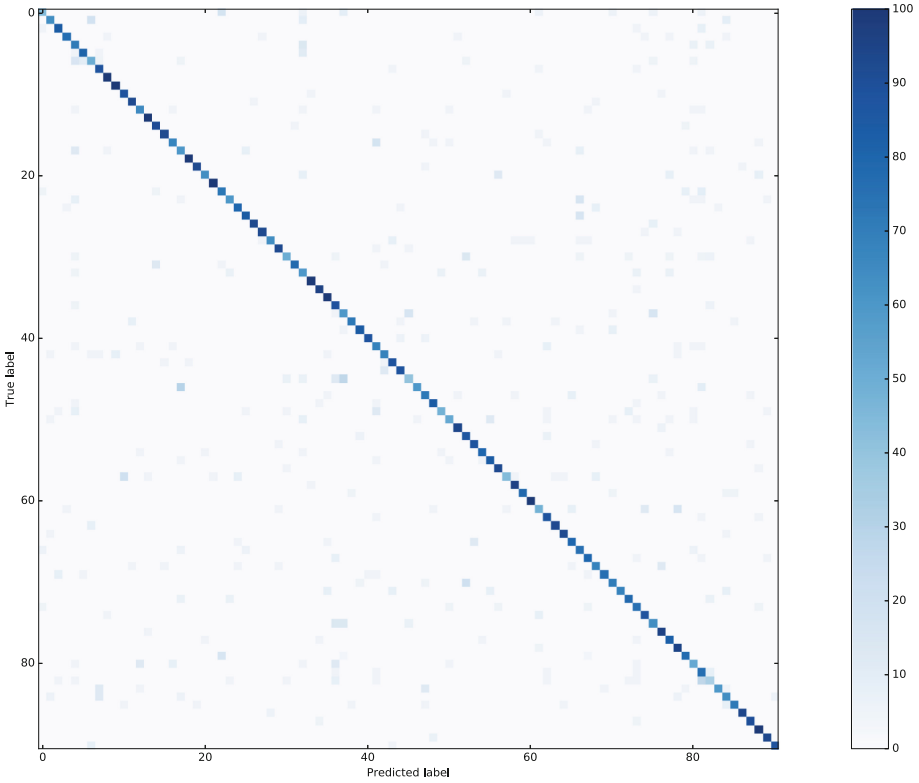

Fig. 4. Confusion matrix for the task of artist recognition. The highest error rates are between Zurbaran and Vermeer, Memling and Van Eyck. These painters are coeval and belongs to the same artistic movement.

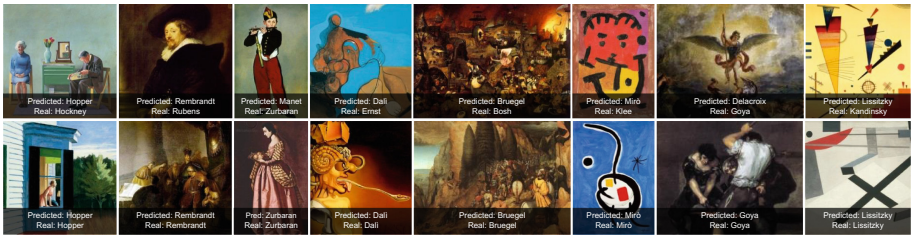

Fig. 5. Top row: highest scored errors for the task of painters classification. Bottom row: for each of the predicted painters, we report the correctly classified example with the highest score.

4 Conclusions

We proposed a novel approach to accomplish the task of painter and style recognition on the challenging Painting91 dataset. Our particular crop strategy permits to exploit multiple cues at different scales. Both fine details and coarse structures are considered during the classification process. The crops are fed to a multibranch deep neural network which merge the information at multiple scales and different spatial locations and performs the final prediction. Since the classification process is not fully deterministic we reported the results as average performance and best performance among ten runs. Our approach clearly outperforms state-of-the-art methods on Paintings-91 dataset by a large margin.

References

1. Anwer, R.M., Khan, F.S., van de Weijer, J., Laaksonen, J.: Combining holistic and part-based deep representations for computational painting categorization. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, pp. 339–342. ACM (2016)
2. Bianco, S., Buzzelli, M., Mazzini, D., Schettini, R.: Deep learning for logo recognition. Neurocomputing (2017). http://dx.doi.org/10.1016/j.neucom.2017.03.051
3. Bianco, S., Mazzini, D., Pau, D., Schettini, R.: Local detectors and compact descriptors for visual search: a quantitative comparison. Digit. Sig. Proc. **44**, 1–13 (2015)
4. Carneiro, G., da Silva, N.P., Del Bue, A., Costeira, J.P.: Artistic image classification: an analysis on the PRINTART database. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7575, pp. 143–157. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33765-9_11
5. Cimpoi, M., Maji, S., Vedaldi, A.: Deep filter banks for texture recognition and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3828–3836 (2015)
6. Gong, Y., Wang, L., Guo, R., Lazebnik, S.: Multi-scale orderless pooling of deep convolutional activation features. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 392–407. Springer, Cham (2014). doi:10.1007/978-3-319-10584-0_26
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Hentschel, C., Wiradarma, T.P., Sack, H.: Fine tuning CNNS with scarce training data–adapting imagenet to art epoch classification. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3693–3697. IEEE (2016)
9. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning, pp. 448–456 (2015)
10. 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)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

12. Mensink, T., Van Gemert, J.: The rijksmuseum challenge: museum-centered visual recognition. In: Proceedings of International Conference on Multimedia Retrieval, p. 451. ACM (2014)
13. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 807–814 (2010)
14. Peng, K.C., Chen, T.: Cross-layer features in convolutional neural networks for generic classification tasks. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 3057–3061. IEEE (2015)
15. Peng, K.C., Chen, T.: A framework of extracting multi-scale features using multiple convolutional neural networks. In: 2015 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2015)
16. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR) (2015)
18. 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: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3703–3707. IEEE (2016)
19. Westlake, N., Cai, H., Hall, P.: Detecting people in artwork with CNNs. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9913, pp. 825–841. Springer, Cham (2016). doi:10.1007/978-3-319-46604-0_57
20. Widjaja, I., Leow, W.K., Wu, F.C.: Identifying painters from color profiles of skin patches in painting images. In: Proceedings of 2003 International Conference on Image Processing, ICIP 2003, vol. 1, pp. I–845. IEEE (2003)
21. Zhao, L., Wang, K., Do, B.: Sherlocknet: exploring 400 years of western book illustrations with convolutional neural networks. Technical report, Stanford University (2016)