Expert Systems With Applications 135 (2019) 90–101

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multitask painting categorization by deep multibranch neural network

Simone Bianco*, Davide Mazzini, Paolo Napoletano, Raimondo Schettini

Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Viale Sarca 336, Milan 20126, Italy

ARTICLE INFO

Article history:
Received 27 July 2018
Revised 10 April 2019
Accepted 25 May 2019
Available online 1 June 2019

2010 MSC:
00-01
99-00

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

ABSTRACT

We propose a novel deep multibranch and multitask neural network for artist, style, and genre painting categorization. The multibranch approach allows us to exploit at the same time the coarse layout of the painting and the fine-grained structures by using painting crops at different resolutions that are wisely extracted using a Spatial Transformer Network trained to identify the most discriminative subregions of paintings. The effectiveness of the proposed network is proved in experiments that are performed on a new dataset originally sourced from wikiart.org and hosted by Kaggle, and made suitable for artist, style and genre multitask learning. The dataset here proposed and made available for research is named *MultitaskPainting100k*, and is composed by 100K paintings, 1508 artists, 125 styles and 41 genres annotated by human experts. Among the different variants of the proposed network, the best method achieves accuracy levels of 56.5%, 57.2%, and 63.6% on the *MultitaskPainting100k* dataset for the tasks of artist, style and genre prediction respectively.

1. Introduction

Automatic categorization and retrieval of digital paintings is gaining increasing attention due to the large quantities of visual artistic data made available by art museums that have digitized or are digitizing their artworks (Carneiro, da Silva, Del Bue, & Costeira, 2012; Khan, Beigpour, Van de Weijer, & Felsberg, 2014; Mao, Cheung, & She, 2017; Mensink & Van Gemert, 2014). In this work, we deal with the problem of categorizing paintings by automatically predicting the artist who painted them (e.g. Monet, van Gogh, etc.), the pictorial styles (e.g. Impressionism, Baroque, etc.), and the genres (e.g. portrait, landscape, etc.) (Anwer, Khan, van de Weijer, & Laaksonen, 2016). These three tasks are very challenging due to the large amount of both inter- and intra-class variations: in fact there are different personal styles in the same art movement, and the same artist may have drawn in one or more different pictorial styles and genres. To have an idea of the difficulty of these tasks some examples taken from the dataset used in this work (i.e. *MultitaskPainting100k*) are reported in Fig. 1.

Artist classification consists in automatically associating the painting to its painter. In this task factors such as stroke patterns, the color palette used, the scene composition, and the subject depicted must be taken into account (Fichner-Rathus, 2011). Style classification consists in automatically assigning 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 (Widjaja, Leow, & Wu, 2003). Style classification is complicated by the fact that styles may not remain pure but could be influenced by others. Finally, genre classification consists in automatically categorizing a painting on the basis of the subject depicted.

The problems of automatic painter, style and genre categorization have been faced using different techniques. Some earlier approaches made use of traditional hand-crafted features (Carneiro et al., 2012; Khan et al., 2014) whereas more recent works rely on the use of deep neural networks for these tasks. Saleh and Elgammal (2016) investigated a comprehensive list of visual features and metric learning approaches to learn an optimized similarity measure between paintings, which is then used to predict painting style, genre and artist. Anwer et al. (2016) used a deformable part model in order to combine low-level details and an holistic representation of the whole painting. Inspired from the results obtained by deep networks as features extractors to solve different tasks (Bianco, Mazzini, Pau, & Schettini, 2015; Sharif Razavian, Azizpour, Sullivan, & Carlsson, 2014), Peng and Chen (2015b) used pretrained deep networks to deal with the small quantity of images available for painter and style categorization. Tan, Chan, Aguirre, and Tanaka (2016) made different experiments by training a network from scratch or fine-tuning an

Inter and intra-class variations
Metric learning approaches

* Corresponding author.
E-mail addresses: bianco@disco.unimib.it, simone.bianco@unimib.it (S. Bianco), mazzini@disco.unimib.it (D. Mazzini), napoletano@disco.unimib.it (P. Napoletano), schettini@disco.unimib.it (R. Schettini).

https://doi.org/10.1016/j.eswa.2019.05.036
0957-4174/

Fig. 1. Paintings from the dataset adopted in this work, i.e. *MultitaskPainting100k* dataset. Each row contains samples from a different artist. For each artist we show paintings with different genres and styles. Color coding is used to distinguish between genres and styles.

existing network for the tasks of style and painter recognition. Cetinic, Lipic, and Grgic (2018) investigated different fine-tuning setups and evaluated the effect of pretraining networks on tasks other than object recognition. Similarly, also Banerji and Sinha (2016) investigated the use of a pre-trained network. Hentschel, Wiradarma, and Sack (2016) performed interesting experiments about the quantity of data needed to fine-tune the network by Krizhevsky, Sutskever, and Hinton (2012) for the task of style classification. Chu and Wu (2016, 2018) investigated the use of deep intra-layer and inter-layer correlation features as style descriptors, showing their superiority with respect to CNN features coming from fully-connected layers. Puthenputhussery, Liu, and Liu (2016a,b) presented a novel set of image features that encode the local, color, spatial, relative intensity information and gradient orientation of the painting image for painting artist classification, style classification as well as artist and style influence analysis. Falomir, Museros, Sanz, and Gonzalez-Abril (2018) presented a system to categorize painting styles based on qualitative color descriptors and quantitative global features. Their approach has the main advantage of being easily explainable, being based on linguistical color palettes. Peng and Chen (2016) approached the problem of painter and style categorization together with other abstract tasks. Mao et al. (2017) with the aim to generate a better representation of visual arts, presented a unified framework to learn joint representations that can simultaneously capture content and style of visual arts. Huang, Zhong, and Xiao (2017) proposed a novel two-channel deep residual network to classify fine-art painting images, where the first channel is the RGB channel and the second one is the brush stroke information channel. Bianco, Mazzini, and Schettini (2017) proposed a novel deep multibranch neural network to automatically predict painting's artist and style, where the different branches processed the input image at different scales to jointly model the fine and coarse features of the painting.

All these works measure their performance mainly on three large scale datasets. The most used is the Painting-91 dataset (Khan et al., 2014), which consists of 4266 painting images from 91 different painters belonging to 13 different styles. This dataset is also the one used more consistently, since all the works adopting it use the same number of painters and styles. Another large scale dataset is the WikiArt-WikiPaintings, that consists of 100,000 high-art images (Karayev et al., 2014). The dataset was built for the task of style recognition and originally from these images only the styles with more than 1000 examples were selected, for a total of 25 styles and 85,000 images. Concerning the genre and artist recognition tasks, later works extracted from this dataset 10 different genres and from 19 to 23 artists. The largest and most recent dataset is the Art500k (Mao et al., 2017), which contains 554,198 images of visual arts mainly scraped from WikiArt, Web Gallery of Art, Rijks Museum, and Google Arts & Culture websites. From these images 1000 artists, 55 styles and 42 genres were extracted.

Three datasets in all the relevant studies

The average accuracy for the task of artist, style, and genre classification obtained by the state of the art approaches described, measured on the datasets respectively adopted are reported in Table 1. For each entry we also report as a subscript the number of classes considered in the experiments presented in each paper, that may be lower than the number of classes actually available in the original dataset.

This work builds on the results obtained in our previous work (Bianco et al., 2017) and significantly extends it, adding the following main contributions:

Multitask setup

- a novel deep neural network architecture, where different branches process the input image at different scales to jointly model fine and coarse features of the painting. The architecture is designed to simultaneously perform the classification of the author, the style and the genre in a multitask setup in order to both reduce the processing time and to induce a form of regularization;
- the use of a trainable crop strategy to feed the network branches with the most significant regions for painting categorization (Jaderberg, Simonyan, Zisserman et al., 2015);
- the use of hand-crafted features for painting categorization; the best performing features have been exploited in our model by applying feature injection;
- a new dataset created starting from a dataset originally collected for a public competition on painter verification, and made suitable for artist, style and genre multitask learning. The final dataset is named *MultitaskPainting100k*, and is composed of 100k paintings from 1508 artists, 125 styles and 41 genres;
- the evaluation of different strategies, with our best performing method achieving an accuracy level of 56.5%, 57.2%, and 63.6% on the tasks of artist, style and genre prediction respectively on the *MultitaskPainting100k* dataset considering all the 1508 artists, 125 styles and 41 genres.

Table 1
State of the art results for artist, style and genre categorization on the most used large scale painting datasets: Painting-91 (Khan et al., 2014), WikiArt-WikiPaintings (Karayev et al., 2014) and Art500k (Mao et al., 2017).

Model	Painting-91		WikiArt-WikiPaintings/Art500k			# Artists	# Styles	# Genres	Single/Multi-task	Handcrafted/ DeepFeatures/ DeepLearning	Single/Multi-Scale	Cropping Strategy	Ad-hoc CNN design
	Artist	Style	Artist	Style	Genre								
Khan et al. (2014)	53.1(91)	62.2(13)				LOW	LOW	–	●/○	●/○/○	●/●	○	–
Peng and Chen (2015a)	56.4(91)	69.2(13)				LOW	LOW	–	●/○	○/●/○	●/○	○	○
Anwer et al. (2016)	64.5(91)	78.4(13)				LOW	LOW	–	●/○	○/●/○	●/●	●	○
Chu and Wu (2016)	63.2(91)	73.6(13)		58.2(25)		LOW	LOW	–	●/○	○/●/○	●/○	○	○
Puthenputhussery et al. (2016a)	59.0(91)	67.4(13)				LOW	LOW	–	●/○	●/○/○	●/●	○	–
Puthenputhussery, Liu, and Liu (2016b)	65.8(91)	73.2(13)				LOW	LOW	–	●/○	●/○/○	●/●	○	–
Peng and Chen (2016)	57.3(91)	70.1(13)				LOW	LOW	–	●/○	○/●/●	●/○	○	–
Banerji and Sinha (2016)	45.0(91)	64.5(13)				LOW	LOW	–	●/○	○/●/○	●/○	○	○
Tan et al. (2016)			76.1(23)	54.5(27)	74.1(10)	LOW	LOW	LOW	●/○	○/●/●	●/○	●	○
Saleh and Elgammal (2016)			63.1(23)	46.0(27)	60.3(10)	LOW	LOW	LOW	●/○	●/●/○	●/○	○	○
Bianco et al. (2017)	78.5(91)	84.4(13)				LOW	LOW	–	●/○	○/○/●	●/●	●	●
Huang et al. (2017)			81.9(19)	50.1(25)	69.0(10)	LOW	LOW	LOW	●/○	○/●/○	●/○	○	○
Mao et al. (2017)			30.2(1000)	39.2(55)	39.2(42)	HIGH	MED	HIGH	○/●	○/●/○	●/○	○	●
Chu and Wu (2018)	64.3(91)	78.3(13)				LOW	LOW	–	●/○	○/●/○	●/●	○	○
Falomir et al. (2018)		67.3(3)				–	LOW	–	●/○	●/○/○	●/○	○	–
Cetinic et al. (2018)			81.9(23)	56.4(27)	77.6(10)	LOW	LOW	LOW	●/○	○/○/●	●/○	○	○
Our method			56.5(1508)	57.2(125)	63.6(41)	HIGH+	HIGH	HIGH	●/●	●/○/●	●/●	●	●

2. Deep multibranch neural network

Fig. 2 shows the scheme of our deep multibranch neural network. The proposed network has three main characteristics that we consider innovative: a multibranch multitask architecture with three different branches that process the input image at different scales. This allows to model fine and coarse features of the painting. Since discriminative fine details can be anywhere in the painting, we investigate two possible ways to extract small subregions from the original image within the *Regions Of Interest (ROIs) proposal* module: random crop selection and a trainable cropping strategy. Random crop selection leads to a simpler architecture since no additional parameters are required to perform ROI extraction. The network can be run different times with different random initializations and the predictions of different runs can be averaged. The trainable crop strategy allows with a single forward pass to select the regions that will possibly contain the most informative regions. This additional module is trained end-to-end with the whole network. The scene composition and the subject depicted are other important clues to recognize a particular author or a painting style. Such clues are more easily inferred by looking at an holistic representation of the whole painting. For this reason our model is composed by a third branch that elaborates the entire image.

The second main characteristic of our model is to perform simultaneously three different but related tasks: artist, style and genre prediction. This configuration results in a faster system both at training and at testing time since only one model needs to be trained/tested instead of three. Moreover this configuration achieved better results in terms of accuracy on the three tasks with respect to the three separate models, with each task acting as a sort of regularizer for the other ones.

The last main characteristic is the use of feature injection. Our model exploits the joint use of handcrafted and neural features to improve the classification accuracy. Neural and handcrafted features have different characteristics. Neural features are learned from data, that could result in a high discriminative power but also in a possible dataset overfit. Handcrafted features have been engineered for generic tasks and tend to be more general purpose while being less powerful on the specific task. Our model takes full advantage of both worlds to achieve a high recognition accuracy.

In the following sections each module of the scheme is discussed in more detail.

2.1. ROIs proposal

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 and fine level details. On the basis of these considerations we extract three subregions by following a multiresolution and multi-regions approach: two squared “small” subregions are extracted from the high-resolution image and one “large” subregion is extracted from the low-resolution image. We use only two scales since, in our preliminary experiments, the use of a higher number of scales brought a slight improvement compared to the exponential increase of computational burden.

Since paintings exhibit high variability in terms of aspect-ratios, the input images are resized such as the minimum side is 512 pixels and the aspect ratio is preserved. From the resulting images we extract two squared subregions of 224 by 224 pixels. Two possible ways to extract these two subregions are investigated: random crop selection, and a trainable cropping strategy based on a Spatial Transformer Network.

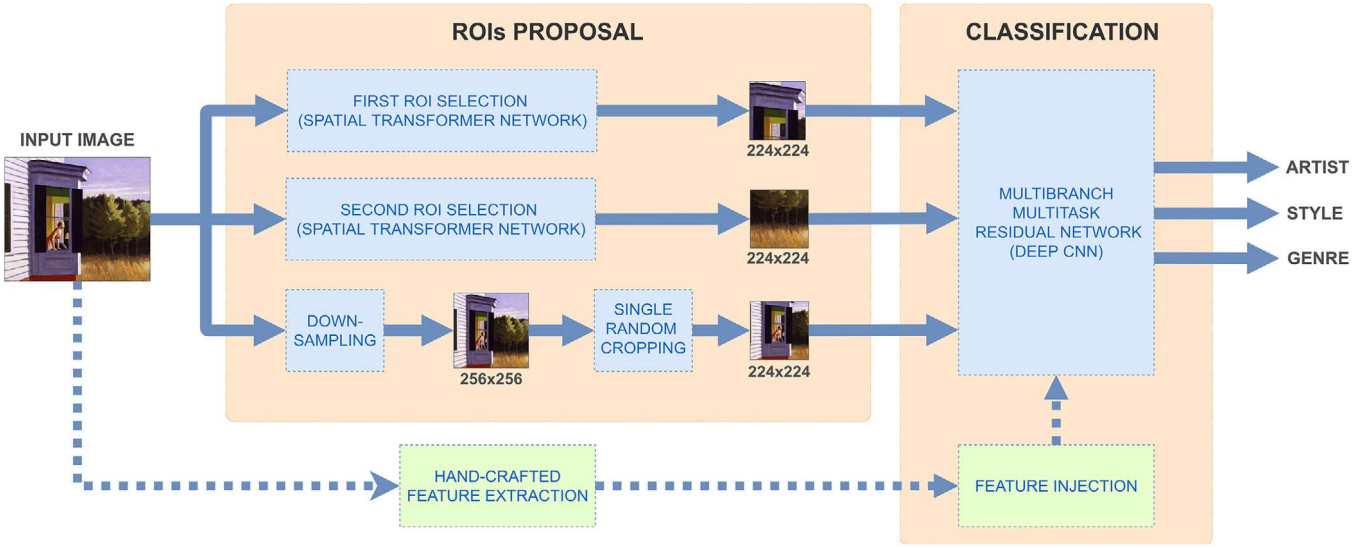

Fig. 2. Scheme of our deep multibranch neural network.

The third subregion of 224 by 224 pixels is randomly selected from the images downsampled so that the minimum side is 256 pixels.

All the subregions extracted are squared, independently from the original aspect ratio of the input image. This is done to improve the computational efficiency of the GPU memory. Images and regions sizes have been chosen as a trade-off between the resolution of fine details in smaller images and the computational burden of processing larger images.

2.1.1. Random subregion selection

The coordinates of the subregions inside the input image are randomly chosen with the only constraint that the selected subregions do not overlap. The rationale behind this choice is that the salient details can be anywhere inside the painting, and the extraction of subregions at no-overlapping random locations permits to increase the probability to get different painting details.

2.1.2. Trainable subregion selection

The subregions inside the input image are extracted by a trainable strategy that in the training phase learns how to extract the best subregions to maximize classification accuracy. The implemented strategy exploits a Spatial Transformer Network (STN), that was introduced by Jaderberg et al. (2015) to explicitly model the spatial manipulation of data within the network. The STN is composed by three modules. The first module is a *Localization Network*, that takes the input feature map $U \in \mathbb{R}^{HxWxC}$ where H, W, C represent the feature map width, height and channels respectively, and outputs the parameters θ of the transformation to be applied to the feature map, i.e. $\theta = f_{loc}(U)$. The second module is a *Parametrized Sampling Grid*, that takes as input the parameters from the Localization Network and produces a sampling grid. The third module is the *Bilinear Sampler*, which is a differentiable bilinear interpolation layer that takes as input the feature map and the sampling grid and performs the actual spatial warping.

In our implementation we used as Localization Network the ResNet-18 (He, Zhang, Ren, & Sun, 2016) with the same type of Residual Blocks used for the main network as described in Section 2.2. In order to maintain the geometric structure of the paintings, we also limited the type of transformations handled by the Sampling Grid layer allowing only translation and scale. We used two STNs, one for each of the first two branches of our network, that are jointly trained with the rest of the network for multitask painting categorization.

Table 2
Multibranch multitask deep neural network.

Output size	Layers		
	branch 1	branch 2	branch 3
	Conv7	Conv7	Conv7
	BatchNorm	BatchNorm	BatchNorm
	ReLU	ReLU	ReLU
112x112x64	MaxPool	MaxPool	MaxPool
56x56x256	3 × ResBlock	3 × ResBlock	3 × ResBlock
56x56x768	Concatenation (channel dimension)		
56x56x256	Join ResBlock, stride 1		
	ResBlock, stride 2		
28x28x512	2 × ResBlock		
	ResBlock, stride 2		
14x14x1024	5 × ResBlock		
	ResBlock, stride 2		
7x7x2048	3 × ResBlock		
1x1x2048	AvgPool		
Num. Classes	FC-1508	FC-125	FC-41

2.2. Classification: deep network architecture

A novel architecture based on Residual Blocks (He et al., 2016) that includes three branches and deals with the problem of artist, style and genre prediction at the same time is proposed. Table 2 shows the architecture of our network more in detail. Each branch processes the subregions coming from the *ROIs proposal* module separately until the processing flow is merged through the concatenation along the channel dimension of three $56 \times 56 \times 256$ feature maps to produce a $56 \times 56 \times 768$ feature map.

Both in the three branches and in the classification module our deep network makes use of Residual Blocks which have been shown to be an effective architectural choice to build very deep networks (He et al., 2016) 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 (He et al., 2016). 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 maintaining a low computational complexity and low memory use inside each block. Our Residual Block

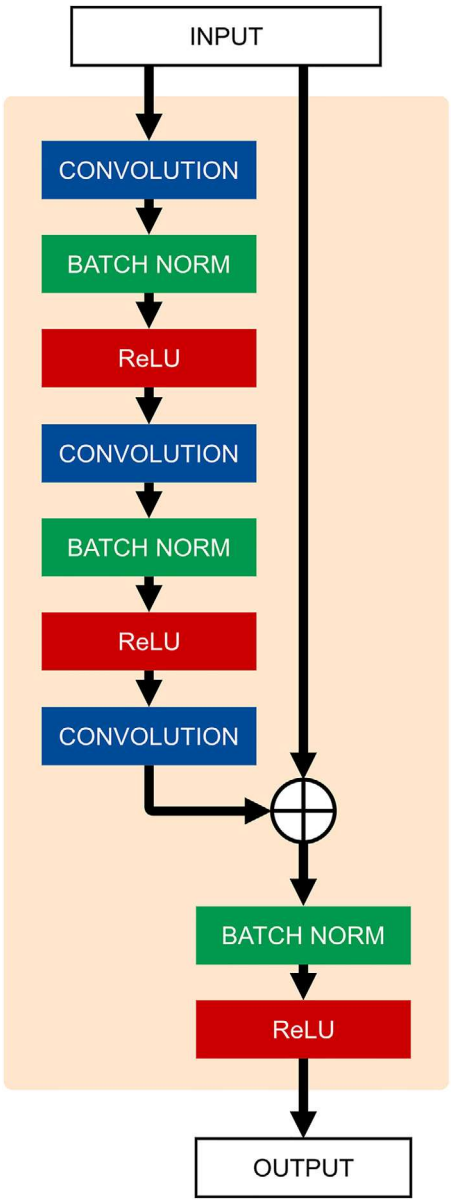

Fig. 3. The type of Residual Block used in our deep neural network.

structure is different from the one used by He et al. (2016): we moved the Batch Normalization layer (Ioffe & Szegedy, 2015) 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. 3. In our network each of the three branches is composed by three Residual Blocks plus four layers near the input which perform the first processing (Convolution + BatchNorm (Ioffe & Szegedy, 2015) + ReLU (Nair & Hinton, 2010)) and an initial downsampling (Max Pooling). The concatenation layer gathers the output of the three branches and stacks the output features along the channel dimension. The join layer is a particular Residual Block then converts the concatenated features 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.

The common part of the network is composed by 13 Residual Blocks plus a spatial Average Pooling layer. While the Residual Blocks in the three branches do not include any downsampling operator, this part of the network includes convolution operators with stride two to perform a spatial downsampling of the input. This leads to a gradual increasing of the receptive-field 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 the right number of classes depending on the task, respectively artist, style or genre prediction. Finally, the Fully-connected layer is followed by a Softmax layer that outputs the classes probabilities.

2.3. Classification: hand-crafted feature injection

We investigated the joint use of hand-crafted features along with learned neural features by adding them to the input of the last fully-connected layer of our network (Bianco, 2017).

Hand-crafted descriptors are features extracted using manually predefined algorithms based on the expert knowledge. These features can be global and local (Bianco et al., 2015; Napoletano, 2018). Global hand-crafted features describe an image as a whole in terms of colour, texture and shape distributions (Mirmehdi, 2008), while local hand-crafted descriptors, like Scale Invariant Feature Transform (SIFT) (Bianco et al., 2015; Lowe, 2004), provide a way to describe salient patches around properly chosen key points within the images.

The hand-crafted features evaluated in this paper are the following:

- 256-dimensional gray-scale histogram (Hist L) (Novak, Shafer et al., 1992);
- 768-dimensional RGB histograms (Hist RGB) (Pietikainen, Niemenen, Marszalec, & Ojala, 1996);
- 10-dimensional feature vector composed of normalized chromaticity moments, as defined in Paschos, Radev, and Prabakar (2003) (Chromaticity);
- 8-dimensional *Dual Tree Complex Wavelet Transform* features obtained considering, for each color channel, four scales, mean and standard deviation (DT-CWT and DT-CWT L) (Barilla & Spann, 2008; Bianconi, Harvey, Southam, & Fernández, 2011);
- 512-dimensional *Gist* features obtained considering eight orientations, four scales and 4 sub-windows for each channel (Gist RGB) (Oliva & Torralba, 2001);
- 32-dimensional *Gabor* features composed of, for each color channel, mean and standard deviation of six orientations extracted at four frequencies, and normalized to be rotation invariant (Gabor L and Gabor RGB) (Bianconi & Fernández, 2007; Bianconi et al., 2011);
- 243-dimensional *Local Binary Patterns* (LBP) feature vector for each channel. We consider LBP applied to gray images and to color images represented in RGB (Mäenpää & Pietikäinen, 2004). We select the LBP with a circular neighbourhood of radius 2 and 16 elements, and 18 uniform and no-rotation invariant patterns (LBP L and LBP RGB).
- 499-dimensional LBP L combined with the Local Color Contrast (LCC) descriptor, as described in Bianco, Cusano, Napoletano, and Schettini (2013); Cusano, Napoletano, and Schettini (2013, 2014, 2016).
- 144-dimensional Colour and Edge Directivity Descriptor (CEDD) features (Chatzichristofis & Boutalis, 2008). This descriptor uses a fuzzy version of the five digital filters proposed by the MPEG-7 Edge Histogram Descriptor (EHD), forming 6 texture areas. CEDD uses 2 fuzzy systems that map the colours of the image in a 24-color custom palette;
- 81-dimensional Histogram of Oriented Gradients feature vector (Junior, Delgado, Gonçalves, & Nunes, 2009). Nine

histograms with nine bins are concatenated to achieve the final feature vector (HoG);
- 1024-dimensional *Bag of Visual Words* (BoVW) of a 128-dimensional Scale Invariant Feature Transform (SIFT) calculated on the gray-scale image. The codebook of 1024 visual words is obtained by exploiting images from external sources (Yang & Newsam, 2010).

The gray-scale image L is defined as follows: L = 0.299R + 0.587G + 0.114B. All feature vectors have been l^2-normalized (i.e. they have been divided by their l^2-norm).

3. Artist, style and genre: the *MultitaskPainting100k* dataset

The dataset used for the evaluation of our multitask deep multibranch neural network has been obtained from the Painter by Numbers Kaggle competition.[1] The goal of the competition was to predict if a pair of images are artworks made by the same artist or not. The dataset contained 103,250 images of paintings obtained mainly from WikiArt.org, that is a publicly available provider of digital artworks. Additional paintings have been provided by artists specifically for the competition. Images are at different resolutions but in general not smaller than 512px per side. The dataset includes a set of metadata for each painting, such as the artist name, style and genre of the painting. Giorgio De Chirico and Salvador Dalí are some examples of artist names. Romanticism and impressionism are some examples of painting styles, while cityscape and landscape are some examples of painting genres.

While the competition provided a training/test split of the data to accomplish the task of predicting from a pair of images whether or not they are made by the same artist, we use this dataset here for another task: the prediction, given an image painting, of the artist name, style and genre. For this reason, the original split is not suitable for our task. To accomplish our task we select a subset of the original dataset such that there are at least 10 images in every class for each of the three tasks, i.e. author, style and genre classification. After this selection the dataset contains 99,816 images for a total of 1508 artists, 125 styles and 41 genres. We call this selection the *MultitaskPainting100k* dataset. The dataset is split in two parts: a random 70% belonging to the train set and the remaining 30% to the test set. We report in Fig. 1 a subset of the paintings in the *MultitaskPainting100k* dataset from three different artists (Pablo Picasso, Leonardo Da Vinci and Gustav Klimt) to let the reader getting the complexity of the recognition task. Each of the three selected artists has drawn, during his life, paintings with several styles and genres. This behavior is quite common among artists and this, along with the fact that the painting distribution is unbalanced across classes, makes recognition task quite challenging. Figs. 4(a)–(c) show the distributions of artists, genres and styles in terms of number of paintings for each class in the *MultitaskPainting100k* dataset. In the case of the artist distribution it is clear that about 70% of all artists have less than 100 paintings and about 50% of the artists have less than 50 paintings. In the case of genres and styles we observe a similar behavior: 50% of all the genres and styles have less than 1000 and 500 paintings respectively.

Figs. 5(a) and (b) show a sample of each genre and style class within the *MultitaskPainting100k* dataset. Images and annotations of the *MultitaskPainting100k* dataset together with our train-test split will be made available on our website.[2]

[1] https://www.kaggle.com/c/painter-by-numbers.
[2] http://www.ivl.disco.unimib.it/activities/paintings/.

4. Experiments

Our multitask deep multibrach neural network solutions is evaluated on the *MultitaskPainting100k* dataset. We compare our solution with that from Bianco et al. (2017) that, as already shown in Table 1, has demonstrated to perform much better than the state of the art on the Painting-91 dataset (Khan et al., 2014). The solution in Bianco et al. (2017) included two different networks, one for the prediction of the artist name and another for the prediction of the painting style. To make it possible the comparison with the network proposed in this paper, we train three networks on the *MultitaskPainting100k* dataset in order to accomplish the artist, style and genre prediction tasks separately. In addition we compare also with the method by Mao et al. (2017), that is the only work reporting results on a dataset having the closest number of classes to the proposed *MultitaskPainting100k* dataset. Given the difficulty to exactly replicate their approach, we report the results taken from their paper.

In all the experiments, to cope with the small amount of training data we exploited some data augmentation techniques:

- Color jitter. It consists in randomly modifying contrast, brightness and saturation of the input image independently.
- 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. (2012).
- Gaussian blur. It consists in applying a blur filter with fixed σ to random images chosen with probability 0.5.
- Geometric transforms. It includes small changes in scale and aspect-ratio of the input image.

All the models have been trained by backpropagation using Stochastic Gradient Descent with momentum. For our best network (Multibranch multitask with STN crop strategy and HOG features injection) we trained the model for 120 epochs with batch size 32 and learning rate 0.01. The learning rate is decreased with a fixed step policy: i.e. it is multiplied by 0.1 at 60 and 90 epochs. The learning rate policy has been estimated by looking at plateau in the loss curve. To train our models we used a single NVIDIA Titan X Pascal GPU with PyTorch 0.4 as deep learning framework.

4.1. Results

Table 3 reports the comparison between the network proposed in Bianco et al. (2017), the one proposed in Mao et al. (2017), and three variants of our proposal:

- the multibranch-multitask network coupled with a random crop selection strategy;
- the multibranch-multitask network coupled with the trained crop selection strategy (STN);
- the multibranch-multitask network with the injection of HOG features, and coupled with the trained crop selection strategy (STN).

The performance is measured on the *MultitaskPainting100k* dataset in terms of average classification accuracy, that is the mean of the accuracy obtained for each class.

Looking at the results reported in Table 3 it is quite clear that the joint multitask training over all the tasks gives a big boost on style accuracy at the expense of a small decrease in performance on genre. The same happens with the injection of HOG features. We suppose that artist and style are much more correlated tasks, thus the training can benefit more from a joint loss optimization. The use of Spatial Transformer Networks improves the performances on all tasks showing the contribution of the smart crop extraction strategy. Although a direct comparison with other state of the art method is not possible, we can indirectly compare

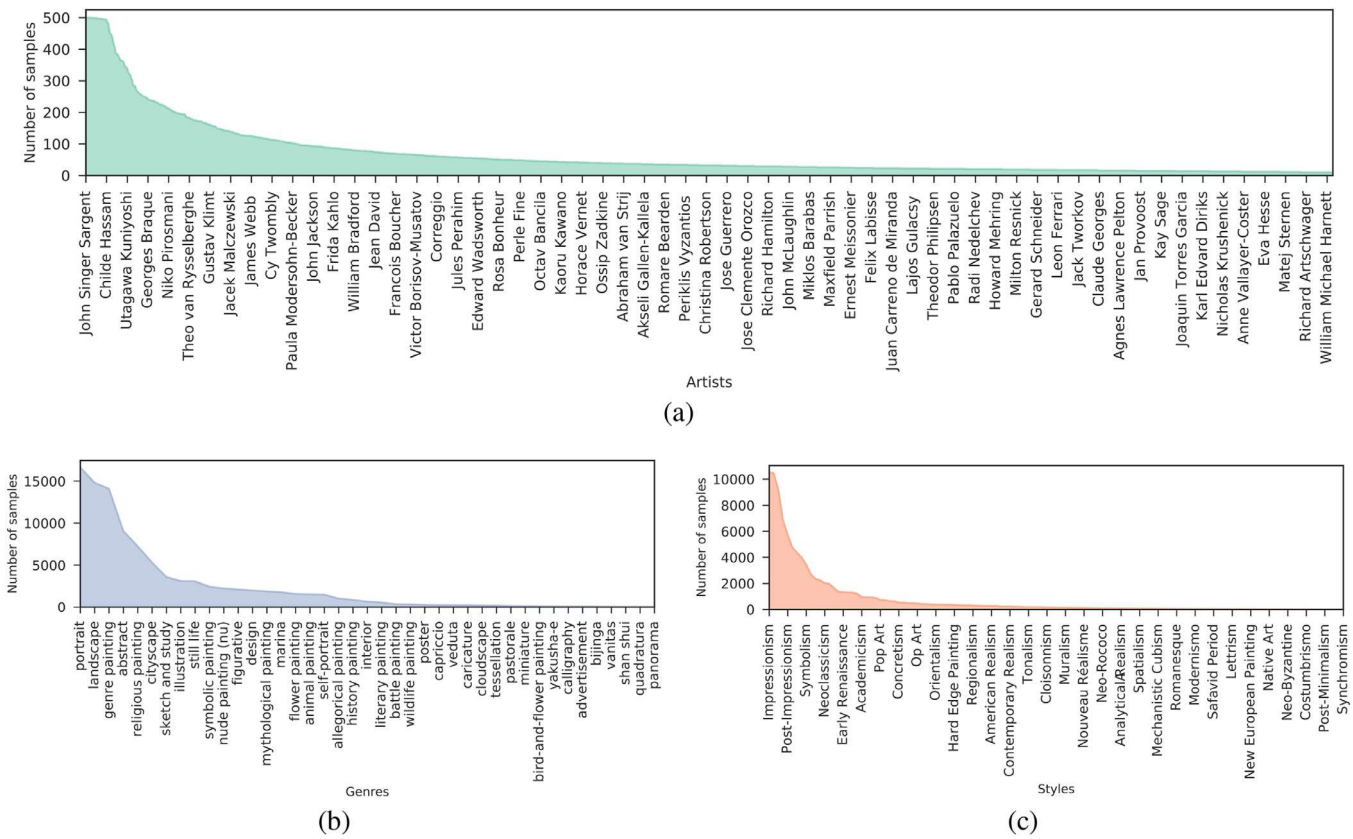

Fig. 4. Distributions of number of samples available for each of the 1508 artists (a), 41 genres (b) and 125 styles (c) within the *MultitaskPainting100k* dataset. The names of classes are partially shown for lack of space.

Table 3
Classification accuracy for the three different tasks on *MultitaskPainting100k* dataset. Different models exploiting multi-task, Spatial Transformer Networks (STN) and the injection of HOG features.

Model	Crop strategy	Feat. injection	*MultitaskPainting100k*			
			Artist	Style	Genre	Average
Mao et al. (2017)†			$30.2_{(1000)}$	$39.2_{(55)}$	$39.2_{(42)}$	36.2
Multibranch (Bianco et al., 2017)	random	–	$53.1_{(1508)}$	$51.5_{(125)}$	$\mathbf{64.3}_{(41)}$	56.3
Multibranch multitask	random	–	$53.3_{(1508)}$	$55.4_{(125)}$	$63.0_{(41)}$	57.2
Multibranch multitask	STN	–	$56.1_{(1508)}$	$57.0_{(125)}$	$64.1_{(41)}$	**59.1**
Multibranch multitask	STN	HOG	$\mathbf{56.5}_{(1508)}$	$\mathbf{57.2}_{(125)}$	$63.6_{(41)}$	**59.1**

†Evaluated on the Art500k dataset.

our solution with that of Mao et al. (2017), that uses a dataset with a common origin to the *MultitaskPainting100k* and has a similar number of classes. We can observe that for the artist task, even if we have 508 more classes, our results are 23.3% better; for the style task, even if we have 70 more classes, our results are 18.0% better; for the Genre task, where the results are more comparable since we have just one less class, our results are 24.4% better.

The proposed multi-branch approach obtains higher results thanks to the multi-resolution processing that enables at the same time to take into account fine details such stroke patterns or line styles and, at the same time, to have an holistic view on high level features. The multi-branch architecture works like an ensemble of methods, reaching higher accuracy than each branch individually. The multitask network has shown better results, in our experiments, with respect to the baseline especially for the task of artist and style predictions due to the high degree of correlation between the two tasks. Furthermore, in the multitask configuration each task acts as a sort of regularizer for the other ones, and this permits to have a higher generalization capability.

4.2. Evaluation of hand-crafted features

To assess the improvement that hand-crafted features could bring to our existing architecture we did some preliminary experiments. We trained a linear classifier on top of each hand-crafted feature in order to classify each of the three tasks: artist, genre and style. Fig. 6 shows the percentage of accuracy achieved by each hand-crafted feature and for each task. This experiment gives a first glance on the discriminative power of the considered features for our classification tasks. As expected the accuracy for the task of artist prediction is quite low. This is the most difficult of the three tasks due to the large set of classes (i.e. 1508). On style and genre prediction some descriptors show an accuracy over 4%. In particular the best features for style prediction are HOG and Gabor L, both grayscale descriptors, whereas for genre prediction genre classification the best descriptors are GIST color and chromaticity moments which relies both strongly on color information.

We made a second experiment in order to evaluate whenever an hand-crafted descriptor is able to correctly classify

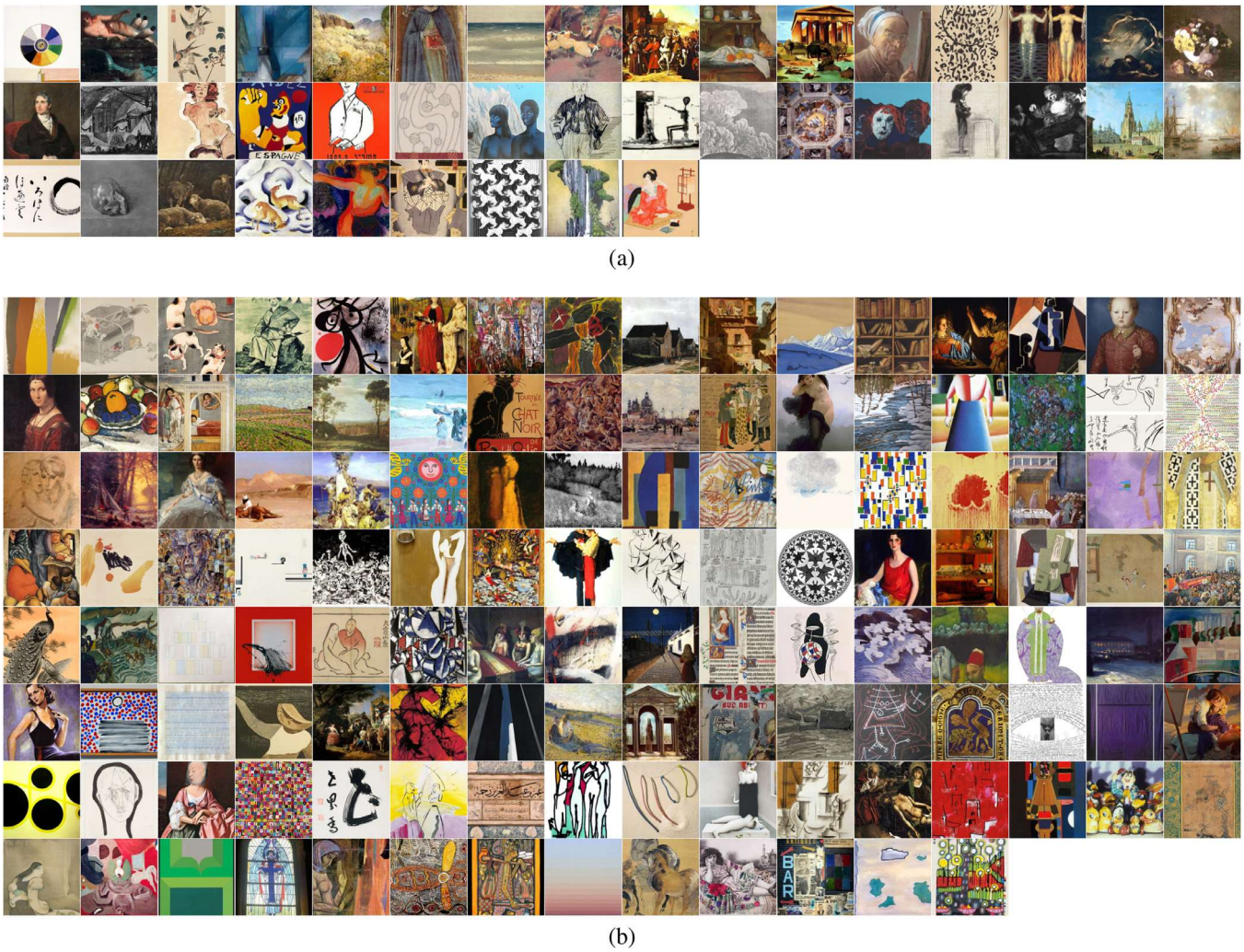

Fig. 5. Examples of each genre (a) and style (c) within the *MultitaskPainting100k* dataset.

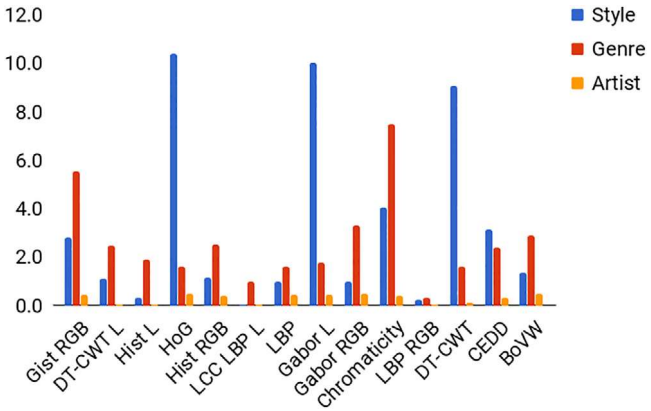

Fig. 6. Calssification accuracy (percentage). Hand-crafted features combined with a linear classifier to solve the three classification tasks.

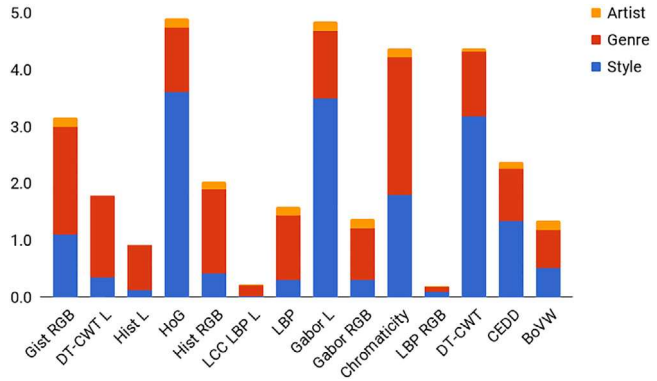

Fig. 7. Percentage of correctly classified examples by the hand-crafted features considered out of all misclassified examples by our multibranch multitask neural network with STN cropping strategy. Stacked bar chart for the three tasks together. HOG and Gabor L give the highest improvement.

examples that are misclassified by our deep architecture. We used the trained linear classifiers on top of hand-crafted features to classify only the misclassified examples of our neural network architecture. Fig. 7 shows a stacked bar graph. Each bar represents the cumulative contribution for all of the three tasks. From this graph are clearly visible the features that correctly classify the highest number of examples: HOG, Gabor L, Chromaticity Moments and DT-CWT.

These preliminary experiments help to highlight that among the hand-crafted descriptors, HOG is the most promising to be included in our classification pipeline. We fed the extracted features directly before the last fully-connected layer of our deep network.

Fig. 8. Similarity results for *Guernica* by Picasso, belonging to the Cubism style. Similarity results: top four painting retrieved using the artist features (first row) and the style features (second row).

Fig. 9. Similarity results for *Judith beheading Holofernes* by Caravaggio, belonging to the Baroque style. Similarity results: top four painting retrieved using the artist features (first row) and the style features (second row).

Fig. 10. Similarity results for a forged painting by van Meegeren, belonging to the Baroque style. Similarity results: top four painting retrieved using the artist features (first row) and the style features (second row).

More in detail, the hand-crafted features are computed on the input image. The neural network is used as a feature extractor by removing the last fully-connected layer and extracting neural features from the input image. Then hand-crafted features and neural features are then concatenated and fed to the classification layer that has been trained to output the final prediction. The network is thus trained end-to-end with two inputs: the first is the image itself and the second input is the descriptor computed on the input image (descriptors can be precomputed and stored on disk for efficiency reasons). The whole network is trained with backpropagation with the settings described in Section 4. Table 3 shows the accuracy achieved by the deep network combined with the HOG descriptor for the three classification tasks. In the case of artist and style classification, the use of HOG slightly improves the accuracy achieved by the deep network. In contrast, in the case of genre classification, the use of HOG features does not produce any improvements, so that the average accuracy over the three tasks is exactly the same with and without HOG features.

4.3. Similarity search

In the following we show some interesting results obtained when the proposed method is used for similarity search. Given one painting as input we extract three sets of features, one for each classification task faced (i.e. artist, style and genre). The features are the l^2-normalized activations of the last Fully-connected layer before the Softmax layer. In this way, each set of features can be used to compute the similarity in terms of artist, style and genre respectively. We report the similarity results for four different paintings. For each of them, we retrieve the four most similar paintings using the artist features, and the four most similar paintings using the style features. The reported examples show how even if the considered painting author would be incorrectly classified by our system from a purely top-1 accuracy point of view, the system could be used to find interesting influences among artists. In Fig. 8 the *Guernica* painting by Picasso is fed to the system. All the four most similar paintings retrieved by the system using the artist features do not belong to Picasso, but all of them belong to M.C. Escher, that in fact are much more similar to the input than any other painting from Picasso himself. This example shows the difficulty of the task of painter recognition, especially for those artist that have painted with many different styles across their artistic production as for example Picasso himself (have a look to the first row of Fig. 1 to see some examples). On the other hand, all the first four paintings retrieved using the style features belong to the Cubism style, that is the same style of the input painting.

A second example is reported in Fig. 9, where the painting *Judith beheading Holofernes* by Caravaggio is given as input to our system. Although the most similar painting retrieved with the artist features is not from the correct author, the second and the fourth ones are from Caravaggio. Furthermore, it is interesting to notice that the third most similar painting, although from a different author, depicts the same subject, i.e. Judith beheading Holofernes, with very similar body poses. For what concerns the style, all the first four paintings retrieved using the style features belong to the Baroque style, that is the same style of the input painting.

A third example is reported in Fig. 10, where a painting by H.A. van Meegeren (that does not belong to the *MultitaskPainting100k* dataset), is given as input to the system. H.A. van Meegeren is famous for having forged paintings of some of the world's most famous artists, including Frans Hals, Pieter de Hooch, Gerard ter Borch, and Johannes Vermeer. He so well replicated the styles and colors of the artists that the best art critics and experts of the time regarded his paintings as genuine and sometimes exquisite. Then, it makes completely sense that the all the four most similar paintings retrieved by our system using the artist features are from Johannes Vermeer. Furthermore, the retrieved paintings are similar to the input also from a compositional point of view, with a girl painted against a clear wall, close to a table and with the same light coming from a window in the upper left corner of the painting. Concerning artistic style, all the first four paintings retrieved using the style features belong to the Baroque style, that is the same style of the input painting.

A final example is reported in Fig. 11, where a painting from The Next Rembrandt (https://www.nextrembrandt.com/) is given as input to our system. The painting has been synthetically generated from data derived from 346 known paintings by Rembrandt, and was created from a deep, 18-month analysis of his work. A facial recognition algorithm learned Rembrandt's techniques, pixel data helped the computer mimic brush strokes, and an advanced 3D printer brought the painting to life using 13 layers of ink. The portrait consists of 148 million pixels and is based on 168,263 fragments from Rembrandt's portfolio. Interestingly, all the four most similar paintings retrieved by the system using the artist features are from Rembrandt, and all the first four paintings retrieved

Fig. 11. Similarity results for a painting by The Next Rembrandt, belonging to the Baroque style. Similarity results: top four painting retrieved using the artist features (first row) and the style features (second row).

using the style features belong to the Baroque style, that is the same style of the input painting.

4.4. Discussion

The experimental results obtained permit to analyze the strengths and weaknesses of the proposed method in comparison with the state of the art:

- the adopted multiscale analysis leads to accuracy improvements, as also proved by Bianco et al. (2017). Few papers in the state of the art perform multiscale analysis, see column "Single/Multi-scale" of Table 1. The way the multiscale analysis is implemented, i.e. with a multi-branch architecture, works like an ensemble of methods, reaching higher accuracy than each branch (i.e. scale) individually.
- Selection of a proper ROI from the input image (column "Cropping strategy" of Table 1) demonstrates accuracy improvements, see also Table 3. Proposed method selects the ROI on the basis of the Spatial Transform Network which adaptively selects the ROI that maximizes classification accuracy. Most of the methods in the state of the art do not adopt ROI selection strategies, while a few ones randomly select a ROI from the input image.
- The proposed CNN is a multi-task network, while the majority of the methods in the state of the art are single task (see column "Single/Multi-task" of Table 1). Multitask approach, as showed in Table 3, demonstrates accuracy improvements with respect to single-task approaches. The multitask formulation permits to use each task as a regularizer for the other ones, resulting in a model with higher generalization capability.
- Literature on painting categorization can be grouped in three main classes: methods that use handcrafted features, methods that use deep features combined with traditional classifiers (such as SVM) and end-to-end deep learning methods (see Table 3). Saleh and Elgammal (2016) experimented handcrafted and deep features separately thus concluding that each method has its own strengths and that the fusion of both features achieves the best performance. Our method exploits the strengths of both traditional handcrafted features and deep learning by using the feature injection scheme. Results in Table 3 confirm our intuition in the case of artist and style classification.
- Our proposed CNN has been designed specially for the three tasks (ad-hoc): artist, style and genre classification. It is a quite complex architecture that requires a long training with many images. Most of the approaches in the state of the art use less complex architectures or pre-trained ones. A few of them are based on ad-hoc architectures (see Table 1). However, given the availability of dataset with increasing cardinality such as the *MultitaskPainting100k* proposed in this work, the main weakness of the proposed method is also its strength, since having an ad-hoc designed deep neural network permits to fully exploit the data available.

5. Conclusions

In this work we tackled the problem of artist, style, and genre categorization of paintings. We proposed a new deep multibranch neural network to solve simultaneously all the three the problems in a multitask formulation. The branches of the proposed network are fed with crops at different resolutions in order to gather clues from low-level texture details and exploit at the same time the coarse layout of the painting. We proposed and compared two different cropping strategies: a random one, and one based on Spatial Transformer Networks. Furthermore, we have also experimented the injection in the proposed network of different hand-crafted features directly computed on the input images. The evaluation has been carried out on a new dataset originally sourced from wikiart.org and hosted by Kaggle, that we made suitable for artist, style and genre multitask learning. This dataset, that we named *MultitaskPainting100k*, is composed of 100K paintings divided into 1508 artists, 125 styles and 41 genres. We used *MultitaskPainting100k* to evaluate and compare the effectiveness of the different variants of the proposed approach. The best solution, which exploits the STN cropping strategy and the injection of HOG features, achieved accuracy levels of 56.5%, 57.2%, and 63.6% on the tasks of artist, style and genre prediction respectively. In order to facilitate a fair comparison with other methods in the state of the art, the *MultitaskPainting100k* dataset along with the training and test splits used are made available as well as a web demo that makes it possible to interactively experience the proposed method (http://www.ivl.disco.unimib.it/activities/paintings/).

An expert system is a system that is able to emulate and hopefully outperform the decision-making ability of a human expert (Jackson (1998)). Although the performance of our system are far from being perfect, its present implementation is often superior to human performance for non-trained users such as the average visitor of a museum or the average high school student. The designed system not only classifies the image in terms of author, style, and genre, but also provides similar paintings with respect to these three aspects makes the system results more understandable for the users. We can imagine that this system, if used by many users that provide some form of feedback or correction to the system output, could make it possible to further improve the recognition performance. From the point of view of neural network based expert and intelligent systems (Sahin, Tolun, and Hassanpour (2012) the designed multibranch, multiscale, multitask network and the hand-crafted feature injections represent an approach that could be inspiring in other challenging application domains. Another contribution of this work to the expert and intelligent systems community is the collection of a large painting dataset that is made available to the research community, the definition of a clear training and testing protocol for making experimental results by other authors directly comparable.

The reported results, although far from being perfect, support the suitability of the proposed CNN for painting recognition and classification. We can for example imagine the application of the proposed network as a core building block for mobile apps that are able to recognize, retrieve suitable information and similar paintings, in the framework of museums and art galleries. It is well known that image quality and size/resolution play a fundamental role in the level of recognition accuracy that can be reached. Moreover for some painting the texture is not well characterized using a single front-view image. To this end one possible research direction could be the use of multi-view or video acquisitions of the painting. Another issue that would deserve more attention is related to the color imaging conditions such as lighting, camera

filters characterization and color space used. In particular, the sRGB color space gamut is probably too small to faithfully represent the painting colors. Moreover digital cameras process the acquired painting just like typical natural images modifying image contrast and color balance accordingly. We can imagine a pre-classification tool that recognizing the painting as the main subject of the photo, would save the image in a unprocessed image format or in a much larger color space. Also, image compression that is typical in consumer digital photos, that is lossy, may alter the fine-grained image textures of some paintings. Obviously the capability of detecting painter signature or specific patterns would greatly improve the recognition performance. Concerning the classification task, we probably could improve the results by clustering the paintings of some artists into their different artistic periods (think for example of Picasso or Mondrian). One of the biggest issues in image classification that we have not been able to solve is the long tail distribution, i.e. the very limited number of paintings to be used for training for some artists. We could imagine as a future work a sort of data augmentation by image synthesis and style transfer. Although these techniques have been used in different scenarios, their application for data augmentation in challenging image classification task such as painting classification and recognition is an open issue that should be deeply investigated. Another point we would like to investigate in the future is the possibility of extending our methods to 3D cultural artifacts. As a future work the system could be also used to investigate the influence and similarity of a new artist that is not in the database with respect to the large archive already indexed.

Declaration of Competing Interest

All the authors declare to not have any conflict of interest.

Credit authorship contribution statement

Simone Bianco: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **Davide Mazzini:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **Paolo Napoletano:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **Raimondo Schettini:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing.

Acknowledgment

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

References

Anwer, R. M., Khan, F. S., van de Weijer, J., & Laaksonen, J. (2016). 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.

Banerji, S., & Sinha, A. (2016). Painting classification using a pre-trained convolutional neural network. In *International conference on computer vision, graphics, and image processing* (pp. 168–179). Springer.

Barilla, M., & Spann, M. (2008). Colour-based texture image classification using the complex wavelet transform. In *Electrical engineering, computing science and automatic control, 2008. CCE 2008. 5th international conference on* (pp. 358–363).

Bianco, S. (2017). Large age-gap face verification by feature injection in deep networks. *Pattern Recognition Letters, 90*, 36–42.

Bianco, S., Cusano, C., Napoletano, P., & Schettini, R. (2013). On the robustness of color texture descriptors across illuminants. In *International conference on image analysis and processing* (pp. 652–662). Springer.

Bianco, S., Mazzini, D., Pau, D., & Schettini, R. (2015). Local detectors and compact descriptors for visual search: a quantitative comparison. *Digital Signal Processing, 44*, 1–13.

Bianco, S., Mazzini, D., & Schettini, R. (2017). Deep multibranch neural network for painting categorization. In *International conference on image analysis and processing* (pp. 414–423). Springer.

Bianconi, F., & Fernández, A. (2007). Evaluation of the effects of gabor filter parameters on texture classification. *Pattern Recognition, 40*(12), 3325–3335.

Bianconi, F., Harvey, R., Southam, P., & Fernández, A. (2011). Theoretical and experimental comparison of different approaches for color texture classification. *Journal of Electronic Imaging, 20*(4).

Carneiro, G., da Silva, N. P., Del Bue, A., & Costeira, J. P. (2012). Artistic image classification: An analysis on the printart database. In *European conference on computer vision* (pp. 143–157). Springer.

Cetinic, E., Lipic, T., & Grgic, S. (2018). Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications, 114*, 107–118.

Chatzichristofis, S. A., & Boutalis, Y. S. (2008). Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *International conference on computer vision systems* (pp. 312–322). Springer.

Chu, W.-T., & Wu, Y.-L. (2016). Deep correlation features for image style classification. In *Proceedings of the 2016 ACM on multimedia conference* (pp. 402–406). ACM.

Chu, W.-T., & Wu, Y.-L. (2018). Image style classification based on learnt deep correlation features. *IEEE Transactions on Multimedia*.

Cusano, C., Napoletano, P., & Schettini, R. (2013). Illuminant invariant descriptors for color texture classification. In *Computational color imaging*. In *Lecture Notes in Computer Science: 7786* (pp. 239–249).

Cusano, C., Napoletano, P., & Schettini, R. (2014). Combining local binary patterns and local color contrast for texture classification under varying illumination. *JOSA A, 31*(7), 1453–1461.

Cusano, C., Napoletano, P., & Schettini, R. (2016). Combining multiple features for color texture classification. *Journal of Electronic Imaging, 25*(6). 061410–061410

Falomir, Z., Museros, L., Sanz, I., & Gonzalez-Abril, L. (2018). Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qart-learn). *Expert Systems with Applications, 97*, 83–94.

Fichner-Rathus, L. (2011). *Foundations of art and design: An enhanced media edition*. Cengage Learning.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hentschel, C., Wiradarma, T. P., & Sack, H. (2016). Fine tuning cnns with scarce training data adapting imagenet to art epoch classification. In *Image processing (ICIP), 2016 IEEE international conference on* (pp. 3693–3697). IEEE.

Huang, X., Zhong, S.-h., & Xiao, Z. (2017). Fine-art painting classification via two-channel deep residual network. In *Pacific RIM conference on multimedia* (pp. 79–88). Springer.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd international conference on machine learning* (pp. 448–456).

Jackson, P. (1998). *Introduction to expert systems*. Addison-Wesley Longman Publishing Co., Inc..

Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. In *Advances in neural information processing systems* (pp. 2017–2025).

Junior, O. L., Delgado, D., Gonçalves, V., & Nunes, U. (2009). Trainable classifier-fusion schemes: an application to pedestrian detection. *Intelligent transportation systems*.

Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., & Winnemoeller, H. (2014). Recognizing image style. In *BMVC 2014 - proceedings of the british machine vision conference 2014* (pp. 1–20).

Khan, F. S., Beigpour, S., Van de Weijer, J., & Felsberg, M. (2014). Painting-91: a large scale database for computational painting categorization. *Machine Vision and Applications, 25*(6), 1385–1397.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision, 60*(2), 91–110.

Mäenpää, T., & Pietikäinen, M. (2004). Classification with color and texture: jointly or separately? *Pattern Recognition, 37*(8), 1629–1640.

Mao, H., Cheung, M., & She, J. (2017). Deepart: Learning joint representations of visual arts. In *Proceedings of the 2017 ACM on multimedia conference* (pp. 1183–1191). ACM.

Mensink, T., & Van Gemert, J. (2014). The rijksmuseum challenge: Museum-centered visual recognition. In *Proceedings of international conference on multimedia retrieval* (p. 451). ACM.

Mirmehdi, M. (2008). *Handbook of texture analysis*. Imperial College Press.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814).

Napoletano, P. (2018). Visual descriptors for content-based retrieval of remote-sensing images. *International Journal of Remote Sensing, 39*(5), 1343–1376.

Novak, C. L., Shafer, S., et al. (1992). Anatomy of a color histogram. In *Computer vision and pattern recognition, 1992. proceedings CVPR'92., 1992 IEEE computer society conference on* (pp. 599–605). IEEE.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.

Paschos, G., Radev, I., & Prabakar, N. (2003). Image content-based retrieval using chromaticity moments. *IEEE Transactions on Knowledge and Data Engineering, 15*(5), 1069–1072.

Peng, K.-C., & Chen, T. (2015a). Cross-layer features in convolutional neural networks for generic classification tasks. In *Image processing (ICIP), 2015 ieee international conference on* (pp. 3057–3061). IEEE.

Peng, K.-C., & Chen, T. (2015b). 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.

Peng, K.-C., & Chen, T. (2016). Toward correlating and solving abstract tasks using convolutional neural networks. In *Applications of computer vision (WACV), 2016 IEEE winter conference on* (pp. 1–9). IEEE.

Pietikainen, M., Nieminen, S., Marszalec, E., & Ojala, T. (1996). Accurate color discrimination with classification based on feature distributions. In *Proceedings of the 13th international conference on pattern recognition: 3* (pp. 833–838).

Puthenputhussery, A., Liu, Q., & Liu, C. (2016a). Color multi-fusion fisher vector feature for fine art painting categorization and influence analysis. In *Applications of computer vision (WACV), 2016 IEEE winter conference on* (pp. 1–9). IEEE.

Puthenputhussery, A., Liu, Q., & Liu, C. (2016b). Sparse representation based complete kernel marginal fisher analysis framework for computational art painting categorization. In *European conference on computer vision* (pp. 612–627). Springer.

Sahin, S., Tolun, M. R., & Hassanpour, R. (2012). Hybrid expert systems: A survey of current approaches and applications. *Expert Systems with Applications, 39*(4), 4609–4617.

Saleh, B., & Elgammal, A. (2016). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *International Journal for Digital Art History, 2*(Oct), 71–93.

Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). 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).

Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2016). Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *Image processing (ICIP), 2016 IEEE international conference on* (pp. 3703–3707). IEEE.

Widjaja, I., Leow, W. K., & Wu, F.-C. (2003). Identifying painters from color profiles of skin patches in painting images. In *Image processing, 2003. ICIP 2003. proceedings. 2003 international conference on: 1* (pp. I–845). IEEE.

Yang, Y., & Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. In *Proc. of the int'l conf. on advances in geographic information systems* (pp. 270–279).