Painting-91: A large scale database for computational painting categorization

4 authors:

Fahad Khan
Linköping University
58 PUBLICATIONS   **2,102** CITATIONS

SEE PROFILE

Shida Beigpour
Max Planck Institute for Informatics
12 PUBLICATIONS   **120** CITATIONS

SEE PROFILE

Joost van de Weijer
Autonomous University of Barcelona
121 PUBLICATIONS   **4,837** CITATIONS

SEE PROFILE

Michael Felsberg
Linköping University
193 PUBLICATIONS   **3,894** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Visual tracking evaluation View project

Project DIPLECS View project

ORIGINAL PAPER

Painting-91: a large scale database for computational painting categorization

**Fahad Shahbaz Khan · Shida Beigpour ·
Joost van de Weijer · Michael Felsberg**

Abstract Computer analysis of visual art, especially paintings, is an interesting cross-disciplinary research domain. Most of the research in the analysis of paintings involve medium to small range datasets with own specific settings. Interestingly, significant progress has been made in the field of object and scene recognition lately. A key factor in this success is the introduction and availability of benchmark datasets for evaluation. Surprisingly, such a benchmark setup is still missing in the area of computational painting categorization. In this work, we propose a novel large scale dataset of digital paintings. The dataset consists of paintings from 91 different painters. We further show three applications of our dataset namely: artist categorization, style classification and saliency detection. We investigate how local and global features popular in image classification perform for the tasks of artist and style categorization. For both categorization tasks, our experimental results suggest that combining multiple features significantly improves the final performance. We show that state-of-the-art computer vision methods can correctly classify 50 % of unseen paintings to its painter in a large dataset and correctly attribute its artistic style in over 60 % of the cases. Additionally, we explore the task of saliency detection on paintings and show experimental findings using state-of-the-art saliency estimation algorithms.

F. S. Khan (✉) · M. Felsberg
Computer Vision Laboratory, Linköping University,
Linköping, Sweden
e-mail: fahad@cvc.uab.es

S. Beigpour
Norwegian Colour and Visual Computing Laboratory,
Gjovik University College, Gjøvik, Norway

J. van de Weijer
Computer Vision Centre Barcelona, Universitat Autonoma
de Barcelona, Catalonia, Spain

1 Introduction

Visual art in any of its forms reflects our propensity to create and render the world around us. It is well known that the true beauty lies in the eye of the beholder. Be it a Picasso, a Van Gogh or a Monet masterpiece, paintings are enjoyed by everyone [37]. Analyzing visual art, especially paintings, is a highly complicated cognitive task [12] since it involves processes in different visual areas of the human brain. Significant amount of research has been done on how our brain responses to visual art forms [4,12,37]. In recent times, several works [3,41,42,44,45] have aimed at investigating different aspects of computational painting categorization.

The advent of internet has provided a dominant platform for photo sharing. In today's digital age, a large amount of art images are available on the internet. This significant amount of art images in digitized databases on the Internet are difficult to manage manually. It is therefore imperative to look into automatic techniques to manage these large art databases by classifying paintings into different sub-categories. An art image classification system will allow to automatically classify the genre, artist and other details of a new painting image which has many potential applications for tourism, crime investigations and museum industries. In this paper, we look into the problem of computational painting categorization.

Most of the research done in the field of computational painting categorization involves medium to small datasets. Whereas the success of image classification to large extent is due to the availability of large scale challenging datasets such as the popular PASCAL VOC series [10] and SUN dataset [49]. The public availability of these large scale

Fig. 1 Example of paintings from our dataset. Each image is provided with two labels: artist and style. The task is to automatically categorize an image to its artist and style

Artist: Pollock
Style: Abst. Expressionism

Artist: Ingres
Style: Neo classicial

Artist: Constable
Style: Romanticism

Artist: Rubens
Style: Baroque

Artist: Picasso
Style: Cubbism

Artist: Kandinsky
Style: Constructivism

datasets, the large number of images with a wide variety of categories and standard evaluation protocols, are some of the crucial factors for their success. To the best of our knowledge, such a dataset still does not exist for analyzing digital paintings. Therefore, we propose a novel dataset for digital paintings consisting of 4,266 images from 91 different painters. In addition, we label the images with the artistic style to which the painting belongs. Figure 1 shows some example images with annotations provided in our dataset. Recently Carneiro et al. [3], investigate the problem of digital painting processing and propose a novel dataset. The dataset deals with the difficult problem of visual theme categorization and pose estimation. However, it differs from ours in a number of aspects. Firstly, the dataset only contains monochromatic images. Secondly, the dataset lacks labels for artist and style categorization. Thirdly, no eye fixation data are provided for the dataset.

The problem of computational painting categorization can be sub-divided into two tasks namely: artist classification and style classification. The task of artist categorization involves classifying a painting to its respective painter. Whereas style categorization deals with the problem of categorizing artist by school of art such as Cubbism, Baroque and Impressionism. Both tasks are highly challenging since there exist large variations in style even within the images of the same painter. For example, the work of Wassily Kandinsky is known to have many different styles. The range of his work includes paintings influenced by the old Russian icon style, landscape paintings with strong color and his abstract paintings with geometric shapes. Other than the inherent difficulty within the problem, the images available on the internet are acquired

under different conditions. These factors make the problem of automatic painting categorization very challenging. In this paper, we focus on both artist and style categorization problems.

Significant amount of progress has been made in recent years in the field of object and scene recognition [9,10,23, 24,38,50]. Most of this success is attributed toward creating efficient visual features and sophisticated learning schemes. Local features with the bag-of-words approach have shown to provide excellent performance for object and scene recognition. The bag-of-words approach works by vector quantizing local features into a visual vocabulary. A histogram is then constructed over these visual words which is then input to a classifier. A variety of local features such as shape, color, texture, etc. are used for classification. In this work, we investigate both local and global features, popular in image classification, for computational painting categorization.

Saliency estimation in natural images is a well-studied domain in computer vision [14,17,29,46]. Visual saliency detection involves capturing the most distinctive regions in an image. Human fixations are used to compare the performance of saliency algorithms. Generally, human eye movement and fixations are estimated using eye-tracking data. How humans see art especially paintings and the accompanied processing in the brain is still an open research problem [36]. The task is challenging since perceiving paintings is a highly variable personal experience. Recent works have shown that estimating saliency in digital paintings have many potential applications such as non-photorealistic rendering and style transfer [5]. To also address this aspect, we perform an eye-tracking experiment on a subset of 182 paintings. As a pos-

sible application of this data, we evaluate existing saliency methods on painting images.

Concluding, we make the following contributions:

– We introduce a new large scale dataset of 4,266 painting images from 91 different painters.
– We show how the state-of-the-art visual features used for image classification perform for the task of artist and style categorization.
– We perform an eye-tracking experiments to obtain human fixation data on a selection of 182 paintings.
– We analyze several computational saliency methods and compare their performance to human fixations.

The paper is organized as follows. In Sect. 2, we discuss related work. In Sect. 3, we introduce our dataset. The analysis on the task of artist and style categorization with experimental results are provided in Sect. 4. In Sect. 5, we provide an overview of saliency algorithms. Section 6 finishes with a discussion and concluding remarks.

2 Related work

Recently several works have aimed at analyzing visual art especially paintings using computer vision techniques [3, 18,19,39,41,42,44,45]. The work of Sablatnig et al. [39] explore the structural signature using the brush strokes in portrait miniatures. A multiple feature- based framework is proposed by Shen [44] for classification of western painting image collections. Shamir and Tarakhovsky [42] propose an approach for automatically categorizing paintings by their artistic movements. Zujovic et al. [51] present an approach based on features focusing on salient aspects of an image to classify paintings into genres. The work of Shamir et al. [40] is based on automatically grouping artists which belong to the same school of art. Nine artists representing three different schools of art are used in their study. Similarly, this paper also aims at exploiting computer vision techniques for the analysis of digital paintings.

Computational painting categorization has many interesting applications. One such application is automatic classification of paintings with respect to their painters. A low level multiple feature-based approach is proposed by Jiang et al. [18] to categorize traditional Chinese paintings. In their approach texture, color and edge size histograms are used as low level features. Two types of feature learning schemes are employed. At the first level, C4.5-based decision trees are used for pre-classification. As a final classifier, an SVM-based framework is employed. A discriminative kernel approach based on training data selection using AdaBoost is proposed by Siddiquie et al. [45]. Recently, Carneiro et al. [3], propose a new dataset of monochromatic artistic images for

artistic image understanding. This suggests that automatic digital painting classification is still an open research problem. In this paper, we investigate the state-of-the-art object recognition methods for the task of painting and style classification.

The advent of large scale image classification datasets have contributed to the research field in an enormous fashion. One such example is the popular PASCAL VOC benchmark series [10]. Several key factors such as the public release of the datasets, large number of images and classes, and common evaluation criteria contribute to the success of these datasets. This is apparent from the significant improvement in classification performance of different approaches over recent years [10]. Other than classification, research in the field of object detection has also progressed at a rapid pace due to the emergence of PASCAL VOC datasets. Furthermore, the emergence of ImageNet dataset [8] has transformed the problem of image classification from several categories to several hundred categories. Other than image classification, a large scale scene classification dataset is proposed by Xiao et al. [49]. Murray et al. [32] propose a large scale database for aesthetic visual analysis. This suggests that the significant improvement in performance of computer vision applications is directly related to the availability of large scale benchmark datasets. To the best of our knowledge, such a large scale benchmark dataset for the analysis of digital paintings is still missing.

2.1 Relation to existing datasets

We provide a comparison of our dataset to existing databases containing painting images. We further outline the characteristics of all these datasets.

PrintART dataset [3]**:** The artistic dataset consists of 988 monochromatic images. The dataset contains 27 theme classes. Unlike other datasets used for artist and style classification, this dataset contains local and global annotations for visual themes. Furthermore, pose annotations of images are also provided.

Artistic genre dataset [51]**:** This dataset comprises of 353 different paintings categorized into 5 artistic genres namely: abstract impressionism, cubism, impressionism, pop art and realism. The images have been collected from the internet.

Artistic style dataset [40]**:** The artistic style dataset consists of 513 paintings of nine different painters. The artists included in the study are as follows: Monet, Pollock, Van Gogh, Kadnisky, Dali, Ernst and deChirico.

Painting genre dataset [45]**:** The painting genre dataset is challenging and consists of 498 painting images. The styles used in their study are as follows: abstract expressionist, baroque, cubist, graffiti, impressionist and renaissance. The images have been collected from the internet.

Table 1 Comparison of our dataset with other datasets in the literature

This paper	Artist labels Y	Style labels Y	Theme labels N	Eye-tracking data Y
PrintART dataset	N	N	Y	N
Artistic genre dataset	N	Y	N	N
Artistic style dataset	N	Y	N	N
Painting genre dataset	N	Y	N	N
Artist and genre dataset	Y	Y	N	N
Western paintings dataset	Y	N	N	N

Note that the most distinguishing feature of our work is the introduction of a large scale public dataset with both artist and style annotation. Additionally, we also provide eye-tracking data for paintings

Artist and genre dataset [42]**:** The dataset introduced in the work of [42] comprises of 994 paintings representing 34 painters. The images have been collected from the internet and used to perform both artist and style classification.

Artist identification dataset [19]**:** 101 paintings are used to analyze the brushwork and identification of Van Gogh's paintings. The paintings have been acquired from van Gogh and Kroller-Muller museums. Out of 101 paintings, 82 have been mostly attributed to van Gogh. Whereas 6 paintings have been denoted to be non-van Gogh and 13 paintings are open to debate by experts.

Western paintings dataset [44]**:** In this work, the dataset is collected from the internet and consists of 1,080 images from 25 different painters. Most of the artists in this dataset are active painters from sixteenth to eighteenth century.

Table 1 provides a comparison of characteristics of current datasets with ours. A key criteria for a benchmark dataset is to be sufficiently larger than the existing ones while providing consistent annotations and evaluation protocols for future analysis. In summary, the dataset proposed in this paper is introduced in the spirit of encouraging research in the direction of computational painting categorization. Inspired from the recent success and contribution of benchmark datasets in the field of image classification and visual aesthetics, we believe the extensive analysis provided in this paper will take this research line one step further.

3 Dataset construction

Our main contribution in this work is the creation of a new and extensive dataset of art images.[1] To the best of our knowledge, this is the first attempt to create a large scale dataset containing painting images of various artists. The images

are collected from the internet. The dataset consists of 4,266 images of 91 artists. The artists in our dataset come from different eras. Figure 2 shows example images of different artist categories from our dataset. There are variable number of images per artists ranging from 31 (Frida Kahlo) to 56 (Sandro Botticelli). The large number of images and artist cateogries make the problem of computational painting categorization extremely challenging. Table 2 shows the list of painters in our dataset.

3.1 Annotations

In our dataset, we provide two types of annotations namely: labels for artist and style.

Artist labels: Each image in the dataset is associated with one class label. The class label denotes the artist of the respective painting. There are 91 classes in the dataset. Table 2 shows the artist categories in our dataset.

Style labels: We further perform style classification to analyze the relationships between different painting styles. The style classification here refers to the school of arts to which an artist belongs to. Art historians often categorize different painters with respect to a certain style. Several recent works have aimed at investigating the problem of style classification [1,45,51]. In our study, we have categorized 50 painters into 13 different painting styles. We have taken care not to categorize painters which clearly belong to more than one style and have only categorized painters for which the large majority of their paintings, considered in our dataset, were correctly described by the style label.

The styles used in this paper are as follows: abstract expressionism, baroque, constructivism, cubbism, impressionism, neo-classical, popart, postimpressionism, realism, renaissance, romanticism, surrealism and symbolism.

3.2 Eye tracking: Where do people look at in a painting?

Here, we present an experiment done on a subset of our dataset which covers 182 images (2 image per painter). 10 subjects were asked to freely view the images in the way that each painting was viewed by three different users. We follow a similar protocol as in Judd et al. [20] to design and execute the experiment. We use the SMI table-mounted eye-tracking device. The iView X™system is a dark pupil eye-tracking system that uses infrared illumination and computer-based image processing to register movements of the eyes in real-time. We used a 5 point calibration system. In our experiment, users were presented with each image for 6 s and before presenting each image the observer fixated on a cross in the center of the gray screen for half a second. We discarded the first fixation of each scan path to avoid the initial fixation on the center. Figure 3 shows some examples of the results obtained

---

[1] The dataset is available at http://www.cat.uab.cat/~joost/painting91.html.

Fig. 2 Example images from the dataset. The images are collected from the internet thereby making the analysis of paintings a challenging task

using the eye tracker processed by BeGaze™2.1 software.[2] We will make the fixation data from the eye tracker publicly available along with the fixation map which is averaged over the subject who viewed each paining.

4 Computational painting categorization

In this section, we show two main applications of our dataset namely: artist and style categorization. For artist categorization, the goal is to associate a painting image with a painter who painted it. Style categorization involves classifying paintings with respect to a certain style of art. We use both the global and local (bag-of-words) feature-based image representations for the two classification tasks. We also evaluated the bio-inspired visual object recognition framework [33] for the painting categorization task.

4.1 Global features

Global image representations work by describing an image as a whole with a single vector. A variety of global features are used to construct an image-wide histogram representation. Here, we provide a brief introduction of the global features used in our evaluation.

LBP [34]: Local binary patterns (LBP) is the most commonly used texture descriptor for image description. The descriptor has shown to obtain state-of-the-art results for texture classification. Here, we use it to extract texture content from painting images. In this work, we use a rotation invariant uniform pattern-based representation of LBP with a pixel neighborhood of 20 and radius size of 2. The dimensionality of final histogram is 383. In our experiments, we also try only the rotation invariant extension without uniform patterns of LBP. However, inferior results were obtained.

Color LBP: To obtain a joint color-texture description, we also use a color LBP approach. In this case, we computed the LBP codes on the R, G and B color channels. The final descriptor is a concatenation of the texture codes from the individual color channels. As a result, a 1,149 dimensional histogram for each image is used for image description.

GIST [35]**:** The GIST is a holistic descriptor that captures the spatial structure of an image. As a preprocessing, we resize the image into 128×128 size. The output of each Gabor-like filter is averaged on a 4×4 grid. The final dimensionality of the descriptor is 320.

Color GIST: As in LBP, we also compute the GIST features on the R, G, B color channels. The final descriptor is the concatenation of three channels with a dimensionality of 960.

PHOG [2]**:** This descriptor captures the local shape of an image along with its spatial layout. We use the standard settings with 20 orientation bins where the orientations are in the range [0, 360]. The final dimensionality of the histogram is 1,700.

Color PHOG: Here, again, we compute the PHOG descriptor on the R,G,B channels making the dimensionality of final descriptor to 5,100.

4.2 Local features

We use the popular bag-of-words framework for local features. The bag-of-words-based image representation is the most successful approach for object and scene recognition [6,25,27,47]. The technique works by first detecting local features either using interest points or by dense sampling strategies. The next stage involves describing the detected regions in an image. Many features such as color, texture and shape have been employed to describe visual information. The feature extraction is followed by vocabulary construction where the local features are quantized into a visual codebook. Generally, a visual vocabulary is constructed using K-means Algorithm. The final stage involves assigning each descriptor to a single visual word in the codebook. As a result of this, a histogram is constructed by counting the number of occurrences of each visual word in an image. The resulting histogram is then input to a machine learning algorithm for classification. We employ the bag-of-

---

[2] The software at: www.smivision.com/en/gaze-and-eye-tracking-systems/products/begaze-analysis-software.html.

Table 2 A list of painters with the number of images, respectively, in our dataset

1:ALBRECHT DURER (43)	24:FRANCISCO DE GOYA (49)	47:JACQUES LOUIS DAVID (50)	70:PETER PAUL RUBENS (50)
2:AMEDEO MODIGLIANI (50)	25:FRANCISCO DE ZURBARAN (50)	48:JAMES ENSOR (50)	71:PICASSO (50)
3:ANDREA MANTEGNA (47)	26:FRANS HALS (50)	49:JAMES MCNEILL WHISTLER (50)	72:PIERRE AUGUSTE RENOIR (50)
4:ANDY WARHOL (32)	27:FRANZ MARC (50)	50:JAN VAN EYCK (40)	73:PIETER BRUEGEL THE ELDER (45)
5:ARSHILLE GORKY (33)	28:FREDERIC EDWIN CHURCH (50)	51:JAN VERMEER (35)	74:PIET MONDRIAN (38)
6:CAMILLE COROT (50)	29:FRIDA KAHLO (31)	52:JASPER JOHNS (50)	75:RAPHAEL (39)
7:CARAVAGGIO (41)	30:GENTILESCHI ARTEMISIA (50)	53:JEAN ANTOINE WATTEAU (50)	76:REMBRANDT VAN RIJN (47)
8:CASPAR DAVID FRIEDRICH (49)	31:GEORGES BRAQUE (50)	54:JEAN AUGUSTE DOMINIQUE INGRES (50)	77:RENE MAGRITTE (50)
9:CLAUDE LORRAIN (50)	32:GEORGES DE LA TOUR (50)	55:JEAN FRANCOIS MILLET (50)	78:ROGER VAN DER WEYDEN (39)
10:CLAUDE MONET (36)	33:GEORGES SEURAT (38)	56:JEAN MICHEL BASQUIAT (43)	79:ROY LICHTENSTEIN (50)
11:DANTE GABRIEL ROSSETTI (50)	34:GEORGIA O'KEEFE (50)	57:JOACHIM PATINIR (40)	80:SALVADOR DALI (46)
12:DAVID HOCKNEY (50)	35:GERHARD RICHTER (50)	58:JOAN MIRO (50)	81:SANDRO BOTTICELLI (56)
13:DIEGO VELAZQUEZ (39)	36:GIORGIO DE CHIRICO (50)	59:JOHN CONSTABLE (50)	82:THEODORE GERICAULT (50)
14:EDGAR DEGAS (51)	37:GIORGIONE (47)	60:JOSEPH MALLORD WILLIAM TURNER (50)	83:TINTORETTO (50)
15:EDOUARD MANET (45)	38:GIOTTO DI BONDONE (48)	61:KAZIMIR MALEVICH (50)	84:TITIAN (51)
16:EDVARD MUNCH (49)	39:GUSTAVE COURBET (40)	62:LUCIO FONTANA (50)	85:UMBERTO BOCCIONI (50)
17:EDWARD HOPPER (50)	40:GUSTAVE MOREAU (50)	63:MARC CHAGALL (44)	86:VINCENT VAN GOGH (42)
18:EGON SCHIELE (50)	41:GUSTAV KLIMT (36)	64:MARK ROTHKO (50)	87:WASSILY KANDINSKY (50)
19:EL LISSITZKY (50)	42:HANS HOLBEIN THE YOUNGER (43)	65:MAX ERNST (50)	88:WILLEM DE KOONING (43)
20:EUGENE DELACROIX (48)	43:HANS MEMLING (50)	66:NICOLAS POUSSIN (50)	89:WILLIAM BLAKE (45)
21:FERNAND LEGER (50)	44:HENRI MATISSE (50)	67:PAUL CEZANNE (50)	90:WILLIAM HOGARTH (39)
22:FRA ANGELICO (47)	45:HIERONYMUS BOSCH (50)	68:PAUL GAUGUIN (51)	91:WINSLOW HOMER (50)
23:FRANCIS BACON (50)	46:JACKSON POLLOCK (41)	69:PAUL KLEE (50)	

The large amount of painters in our dataset pose a challenging problem of automatic classification

Fig. 3 Example of eye-tracking experiment results. Each row from *top to down* represents: the original image, scanning path for three different users marked in different *colors*, the heat-map emphasizing the highest fixation times
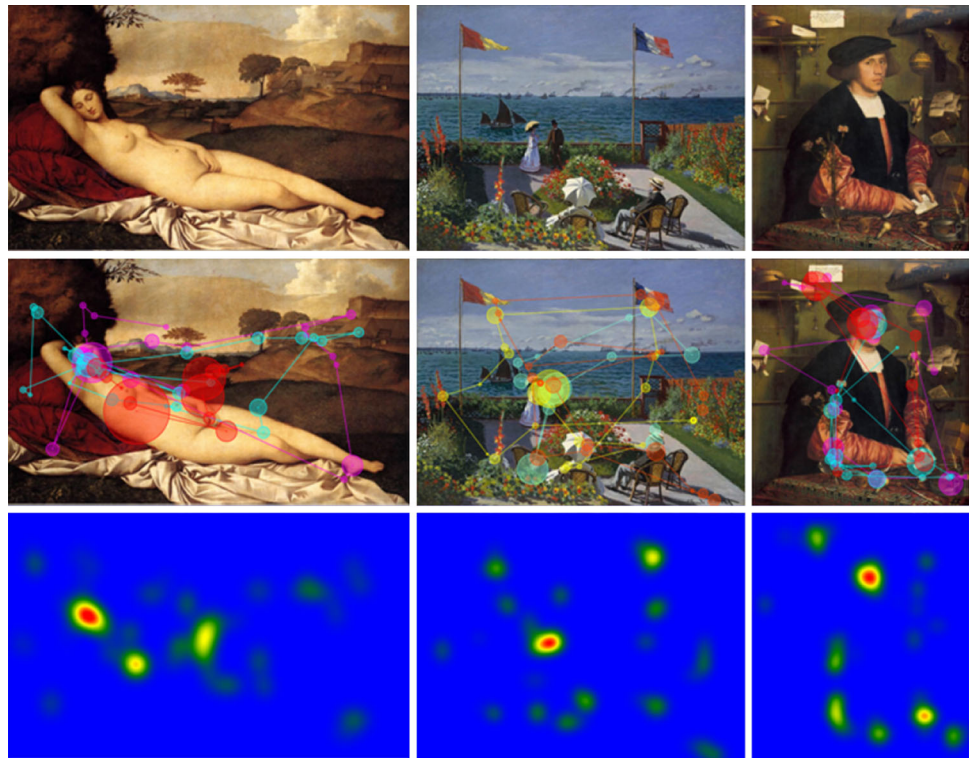
words framework for the task of painting categorization and evaluate a variety of visual features commonly used in object and scene recognition.

4.2.1 Feature detection

Dense sampling scheme is often advantageous since all regions in an image provide information for the categorization task. The technique works by scanning the image with either single or multiple scales at fixed locations forming a grid of rectangular windows. In this work, we use a dense single scale grid with a spacing of 5 pixels.

4.2.2 Feature extraction

The feature extraction stage involves describing the detected regions of an image. In this paper, we investigate a variety of visual features for image description.

Color names [48]**:** We use color names since they have shown to be superior to its counterparts for object recognition, object detection, action recognition, visual tracking and texture classification [7,21,22,25,26].

Complete LBP [13]**:** We also use a texture descriptor within the bag-of-words framework. In complete LBP, a region in an image is represented by its center pixel and a local difference sign-magnitude transform.

SIFT [30]**:** To capture the appearance of an image, we use the popular SIFT descriptor. The SIFT descriptor has shown to provide superior performance for object recognition, texture recognition and action recognition tasks [22,28,31,50]. The Dominant orientation is not used for the computation of SIFT. The SIFT descriptor works on gray level images while ignoring the color information. The resulting descriptor is a 128 dimensional vector.

Color SIFT [47]**:** To incorporate color information with shape in an early fusion manner, we employ the popular Color SIFT descriptors. In our experiments, we use RGB-SIFT, OpponentSIFT and CSIFT which were shown to be superior by [47].

SSIM [43]**:** Unlike other descriptors mentioned above, the self similarity descriptor provides a measure of the layout of an image. The correlation map of a 5×5 size patch is computed with a correlation radius of 40 pixels. Consequently, a 30 dimensional feature vector is obtained with a quantization of 3 radial bins and 10 angular bins.

4.2.3 Visual vocabulary and histogram construction

The feature extraction stage is followed by vocabulary construction step within the bag-of-words framework. To construct a visual vocabulary, we employ the K-means Algorithm. The quality of a visual vocabulary depends on the size of the vocabulary. Generally, a larger visual vocabulary results in improved classification performance. We experimentally set the size of visual vocabulary for different visual features. The optimal performance for SIFT descrip-

tor is obtained when a visual vocabulary of 1,000 words is employed. The performance gets saturated over 1,000 visual words. Similarly, we use a visual vocabulary of 500 for complete LBP, Color names, SSIM, respectively. A visual vocabulary of 1,500 visual words is used for Color SIFT methods. Finally, a histogram over the visual words is constructed by counting the occurrence of each visual word in an image. We also constructed a late fusion representation of SIFT and Color names (CN-SIFT) by concatenating the two histogram representations. The concatenated image representation is then input to the SVM classifier.

4.3 Bio-inspired object recognition

In this work, we also investigate how a bio-inspired object recognition framework [33] performs for the task of paining categorization. To the best of our knowledge, this is the first time such a framework is investigated for painting categorization task. The method starts with an image of grayscale pixels and then applies Gabor filters at all positions and scales. The Gabor filters are computed using 4 orientations with multiple units at each position and scale. Afterward, invariance with respect to position and scale is achieved by alternating template matching and max pooling operations. Each image is represented by constructing a fixed length feature vector. Finally, the method employs an all-pairs linear SVM. We use the same parameter settings, as in the original method, throughout our experiments.

4.4 Evaluation protocol

We follow the standard criteria used in image classification to evaluate the performance of our system on the tasks of artist and style classification. We fix the training size for each category to 25, and the rest of the instances are used for testing. We randomly sampled the images into training and testing. To ensure a fair comparison and reusability, we intend to release the train/test splits. For classification, we use a one vs all support vector machines (SVM) with a χ^2 kernel [50]. Each test instance is assigned a category label of the classifier providing the highest performance. The final categorization result is obtained as a mean recognition rate per category.

4.5 Combining multiple visual features

Combining multiple visual cues is shown to provide excellent results for object and scene recognition [11,47]. Here, we also investigate to what extent the visual features are complementary to each other for painting categorization task. To this end, we follow the standard protocol by combining the kernel outputs of individual visual features. In the work of Gehler and Nowozin [11], it has been shown that

a simple addition or product of different kernel responses provide excellent results compared to more sophisticated Multiple Kernel Learning (MKL) methods. In our experiments, we found that the addition of different kernel responses provide superior results compared to the product operation.

4.6 Artist classification

We first present our results on the task of artist classification. The dataset consists of 91 artists with 2,275 training images and 1,991 test instances. The first row of Table 3 shows the results obtained using different features for this task. As a single visual cue, color alone provides inferior performance with a classification accuracy of 18.1 %. The color variants of LBP, GIST and PHOG always provide improved performance. The three ColorSIFT descriptors achieve inferior performance compared to intensity-based SIFT descriptor for artist classification. The best single feature is CN-SIFT which provides a classification performance of 44.1 %. The results show that late feature fusion provides superior performance compared to channel-based early fusion (ColorSIFT). Previously, late fusion has also shown to provide superior performance compared to channel-based fusion for object recognition, object detection and action recognition [21,25,26]. In the work of Khan et al. [25], it has been shown that late fusion possesses the feature compactness property achieved by constructing separate visual vocabularies for both color and shape. This property is especially desirable for categories where one of the visual cues varies significantly. The ColorSIFT descriptors lack this property since a joint visual vocabulary of color and shape is constructed.

The bio-inspired HMAX approach provides inferior performance with mean classification results of 22.1 %. Combining all feature kernels additively provides a significant gain of 9.0 % over single best feature alone. This shows that the different features are complementary to each other. For example, paintings of Claude Lorrain category are ideal-landscape paintings providing a view of the nature. Most of the paintings in this category contains pastoral figures with an inspiration from the countryside around Rome. Color plays a pivotal role for this category. For example, the performance obtained using LBP and ColorLBP on this category is 64 and 80 % respectively.

The result shows that computer nowadays can correctly classify 50 % of the paintings to its painter in a large dataset of over 90 painters. Figure 4 shows artist categories with highest and lowest recognition rate. The method provides inferior performance for categories such as Titian, Manet, Courbet, Ernst, Giorgione, Velazquez, etc. The best performance is achieved on Claude Lorrain, Roy Lichtenstein, Mark Rothko, Fernand Leger, Frans Hals and Franz Marc, etc.

Table 3 Recognition performance on our dataset

Task	LBP	Color-LBP	PHOG	Color-PHOG	GIST	Color-GIST	SIFT	CLBP	CN	SSIM	OPPSIFT	RGBSIFT	CSIFT	CN-SIFT	Combine	HMAX
Artist CLs	28.5	35.0	18.6	22.8	23.9	27.8	42.6	34.7	18.1	23.7	39.5	40.3	36.4	44.1	**53.1**	21.7
Style CLs	42.2	47.0	29.5	33.2	31.3	36.5	53.2	46.4	33.3	37.5	52.2	47.4	48.6	56.7	**62.2**	32.8

Top row: results on the artist classification task. Bottom row: results on style classification problem. Combining all the visual cues significantly increase the overall performance. Note that in both tasks HMAX provides inferior results to single best visual cue

Bold values indicate the best results (higher the better)

4.7 Style classification

Style classification associates many different painters with respect to a specific painting genre. School of arts or genres are higher level semantic classes. The problem is different to artist categorization in that a style is shared across many different painters. In our study, we have categorized 50 painters into 13 different painting styles. The painters in our dataset are divided into 13 diversified styles namely: abstract expressionism, baroque, constructivism, cubbism, impressionism, neoclassical, popart, postimpressionism, realism, renaissance, romanticism, surrealism and symbolism. Several painters are categorized into different school of arts. For example, Abstract Expressionism is represented by Jackson Pollock, Willem de Kooning and Mark Rothko; Constructivism by Kazimir Malevich, Wassily Kandinsky and El Lissitzky etc.

The second row in Table 3 shows the results obtained using different visual features for style categorization. Color names provide a classification accuracy of 33.3 %. SIFT provides a recognition performance of 53.2 %. The late fusion of Color names and SIFT improves the performance compared to SIFT alone. For style categorization again, the color variants of LBP, PHOG and GIST significantly improve the performance compared to their intensity-based counterparts. Combining all the visual features provides a significant improvement with a mean classification accuracy of 61.2 %. The results are very encouraging since the images are collected from different sources, across different illumination, compression methods and artifacts. All these factors and the inherent complexity of the problem makes this task extremely challenging.

Figure 5 shows the confusion matrix for the style categorization task. The confusions are visually understandable, and are often with periods which are adjacent in time. The largest confusions are between painting styles baroque, renaissance and neo-classical, between realism and romanticism, and between impressionist and postimpressionists.

## 5 Computational saliency estimation in paintings

We provide eye fixation data for 182 paintings (two per painter) from the dataset. Here, we look at one of the potential applications of such data. We turn the fixation data of the three observers per image into a single saliency map by averaging their fixation time maps, and use this data to evaluate existing saliency methods. The computed saliency maps are thresholded at different values and compared to the average fixation map to construct an ROC curve. From this, we compute the area under curve (AUC) score for each saliency method.

(a) best results

(b) lowest results

Fig. 4 Artist categories with best and lowest classification performance on our dataset. The best performance is achieved on painting categories such as Claude Lorrain, Roy Lichtenstein, Mark Rothko, Fernand Leger etc. Whereas, on categories such as Titian, Manet, Courbet, etc. inferior results are obtained

Fig. 5 The confusion matrix for style categories. The *Y* axis shows the style categories in our dataset. The confusion among categories seem to agree with the art historians. For example, there exist confusion between Baroque and Renaissance artwork since most paintings contain interior of buildings
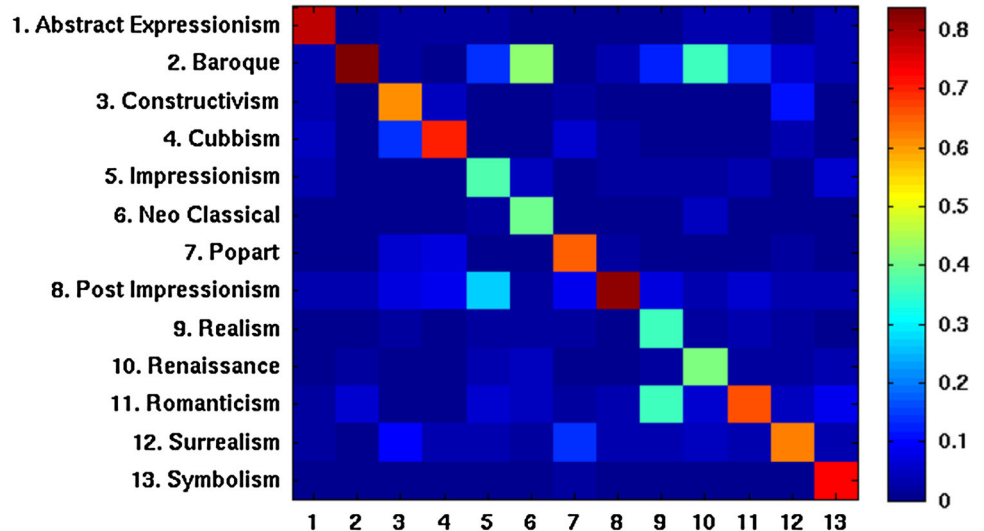
We evaluate several state-of-the-art saliency algorithms. These approaches have shown to obtain good results on several natural image datasets, but their generalization to art images has not yet been investigated. Here, we provide a brief introduction to the saliency methods used in our evaluation. **Center-surround method** [46]**:** This approach provides a saliency map by computing local saliencies over a rectangle region of interest in an image. The method is able to capture local contrast without requiring any training priors.
Salient object method [29]**:** This approach works by formulating the saliency estimation as image segmentation problem by separating the salient objects from the background. A variety of features such as multi-scale contrast, center-surround histogram and color spatial distribution are employed to describe a salient object at local, regional and global level. These features are then combined using Conditional Random Field to detect salient regions in an image.
Itti and Koch method [17]**:** This approach employs a bottom-up feed-forward feature extraction mechanisms. The method combines multi-scale image features into a single topographical saliency map. The scales are created using dyadic Gaussian pyramids.

Table 4 Comparison of different saliency approaches on the paintings

CH	CS	SO	Ik	GB	SB	SR
0.50	0.76	0.74	0.77	0.79	0.76	0.63

CH denotes to chance. Whereas CS, SO, IK, GB, SB, SR denote to center surround, salient object, Itti and Koch, graph-based, signature-based, and spectral residue-based approach, respectively. Note that most of the saliency algorithms provide improved performance compared to baseline

Graph-based method [14]**:** This approach is bottom-up and based on graph computations. The method employs Markov chains over various image maps by treating distribution over map locations as activation and saliency values. The activation maps are generated from raw features.

Signature-based method [16]**:** This approach proposes a binary, holistic image descriptor which is defined as a sign function of the discrete cosine transform (DCT) of an image. The method approximates the spatial location of a sparse foreground for saliency estimation.

Spectral residue based [15]**:** This approach works by analyzing the log-spectrum of an input image by extracting the spectral residual of an image in spectral domain. The proposed technique is fast and generates saliency maps in spatial domain.

To evaluate the performance of saliency algorithms, we use area under the ROC curve as in Judd et al. [20]. As a baseline, we use the chance model. Table 4 shows the performance of different saliency algorithms on the paintings. Other than the spectral residue approach [15], all the other techniques provide significantly improved results compared to the baseline (chance). The best results are obtained using the Graph-based method [14] (GB) with a score of 0.79.

To further validate the statistical significance of the performance obtained using different saliency methods, we perform a paired nonparametric Wilcoxon signed-rank test on the ROC results. Let A and B be variables representing area under the ROC curve obtained on all the images by two different saliency methods. The median values of the two random variables are represented as μ_A and μ_B. The Wilcoxon test is then used to test the null hypothesis $N_0 : \mu_A = \mu_B$ compared to the alternative hypothesis $N_1 : \mu_A \neq \mu_B$ at 95 % significance level. Table 5 shows a comparison of different approaches using the Wilcoxon signed-rank test. A positive number (1) at a location (r,c) indicates that the improvement obtained using the saliency method r is statistically significant compared to the results obtained using saliency method c. A negative number (−1) indicates that the improvement obtained using saliency method r is statistically worse compared to method c. A zero (0) indicates that the two saliency methods are considered statistically equivalent. Other than the Spectral residue method (SR), all saliency approaches

Table 5 Wilcoxon signed-rank test on the saliency methods presented in Table 4

	CH	SB	SO	SR	CS	IK	GB
CH	0	-1	-1	-1	-1	-1	-1
SB	1	0	1	1	0	0	-1
SO	1	-1	0	1	-1	-1	-1
SR	1	-1	-1	0	-1	-1	-1
CS	1	0	1	1	0	0	-1
IK	1	0	1	1	0	0	-1
GB	1	1	1	1	1	1	0

A positive number (1) at a location (r, c) shows that the results obtained using the saliency method r has been considered statistically better than that obtained using the method c (at 95 % confidence level). A negative number (−1) shows the opposite where as a zero (0) indicates no significant difference between the two saliency methods

are significantly better statistically compared to the baseline chance method (CH) (95 % confidence level).

6 Conclusion

In this work, we have proposed a new large dataset of 4,266 digital paintings from 91 different artists. The dataset contains annotations for artist and style categorization tasks. We also perform experiments to collect eye-tracking data and evaluate state-of-the-art saliency methods for saliency estimation in art images.

6.1 Computational painting categorization

We study the problem of both artist and style classification on a large dataset. A variety of local and global visual features are evaluated for the computational painting categorization tasks. Among the global features, the best results are obtained using Color-LBP with a accuracy of 35.0 and 47.0 % for artist and style classification, respectively. Within the bag-of-words-based framework, the best results are obtained using CN-SIFT with an accuracy of 44.1 and 56.7 % for artist and style classification, respectively. We further show that combining multiple visual cues further improves the results with a significant gain of 9.0 and 5.5 % for artist and style classification, respectively. Furthermore, we also evaluate the bio-inspired framework (HMAX) for computational painting categorization tasks. In summary, we show that the computer can correctly classify the painter for 50 % of the paintings, and in 60 % of the images can correctly assess its art style.

While significant amount of progress has been made in the field of object and scene recognition, the problem of computational painting categorization has received little attention. This paper aims at moving the research of computational painting categorization one step further in two ways:

The introduction of a new large dataset of art images and a comprehensive evaluation of visual features commonly used in the object recognition literature for computational painting categorization. The results obtained on both artist and style classification tasks clearly suggest the need of designing more powerful visual features specific to the painting categorization tasks.

6.2 Computational saliency estimation in paintings

We also investigate the problem of saliency estimation in digital paintings. We provide eye fixation data for a subset of our dataset. We further evaluate several state-of-the-art saliency algorithms for estimating saliency in art images. These methods have shown to provide excellent results for natural images. Our results show that these standard saliency algorithms provide similar performance for saliency estimation in art images.

In this work, we evaluate the state-of-the-art saliency methods on a subset of painting images. We hope that our work will motivate other researchers to further investigate the problem of saliency estimation in digital paintings. Future work involves a comprehensive evaluation of saliency methods on a large number of painting images together with more subjects for obtaining the ground-truth. Another potential research direction is to use the saliency estimation results for the task of computational painting categorization. The saliency estimation can be used either in a bottom-up or top-down manner to accentuate salient regions in an image with respect to a certain category label.

References

1. Arora, R., Elgammal, A.: Towards automated classification of fine-art painting style: a comparative study. In: Proceedings of the ICPR (2012)
2. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the CIVR (2007)
3. Carneiro, G., da Silva, N.P., Bue, A.D., Costeira, J.P.: Artistic image classification: an analysis on the printart database. In: Proceedings of the ECCV (2012)
4. Cinzia Di Dio, G.R., Macaluso, E.: The golden beauty: brain response to classical and renaissance sculptures. Plos One **2**(11), e1201 (2007)
5. Condorovici, R.G., Vranceanu, R., Vertan, C.: Saliency map retrieval for artistic paintings inspired from human understanding. In: Proceedings of the SPAMEC (2011)
6. Csurka, G., Bray, C., Dance, C., Fan, L.: Visual categorization with bags of keypoints. In: Proceedings of the Workshop on Statistical Learning in Computer Vision, ECCV (2004)
7. Danelljan, M., Khan, F.S., Felsberg, M., van de Weijer, J.: Adaptive color attributes for real-time visual tracking. In: Proceedings of the CVPR (2014)
8. Deng, J., Dong, W., Socher, R., Jia, L., Kai, L., Li, F.F.: Imagenet: a large-scale hierarchical image database. In: Proceedings of the CVPR (2009)
9. Elfiky, N., Khan, F.S., van de Weijer, J., Gonzalez, J.: Discriminative compact pyramids for object and scene recognition. Pattern Recognit. **45**(4), 1627–1636 (2012)
10. Everingham, M., Gool, L.J.V., Williams, C.K.I., Winn, J.M., Zisserman, A.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
11. Gehler, P.V., Nowozin, S.: On feature combination for multiclass object classification. In: Proceedings of the ICCV (2009)
12. Goguen, J.: Art and the Brain. Imprint Academic, UK (1999)
13. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. Trans. Image Process. **19**(6), 1657–1663 (2010)
14. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Proceedings of the NIPS (2006)
15. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: Proceedings of the CVPR (2007)
16. Hou, X., Harel, J., Koch, C.: Image signature: highlighting sparse salient regions. Pattern Anal. Mach. Intell. **34**(1), 194–201 (2012)
17. Itti, L., Cristof, K., Ernst, N.: A model of saliency-based visual attention for rapid scene analysis. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
18. Jiang, S., Huang, Q., Ye, Q., Gao, W.: An effective method to detect and categorize digitized traditional chinese paintings. Pattern Recognit. Lett. **27**(7), 734–746 (2006)
19. Johnson, R., Hendriks, E., Berezhnoy, I.J., Brevdo, E., Hughes, S.M., Daubechies, I., Li, J., Postma, E., Wang, J.Z.: Image processing for artist identification. IEEE Signal Process. Mag. **25**(4), 37–48 (2008)
20. Judd, T., Ehinger, K.A., Durand, F., Torralba, A.: Learning to predict where humans look. In: Proceedings of the ICCV (2009)
21. Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A.D., Vanrell, M., Lopez, A.M.: Color attributes for object detection. In: Proceedings of the CVPR (2012a)
22. Khan, F.S., van de Weijer, J., Ali, S., Felsberg, M.: Evaluating the impact of color on texture recognition. In: Proceedings of the CAIP (2013b)
23. Khan, F.S., van de Weijer, J., Bagdanov, A.D., Vanrell, M.: Portmanteau vocabularies for multi-cue image representations. In: Proceedings of the NIPS (2011)
24. Khan, F.S., van de Weijer, J., Vanrell, M.: Top-down color attention for object recognition. In: Proceedings of the ICCV (2009)
25. Khan, F.S., van de Weijer, J., Vanrell, M.: Modulating shape features by color attention for object recognition. Int. J. Comput. Vis. **98**(1), 49–64 (2012b)
26. Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A., Lopez, A., Felsberg, M.: Coloring action recognition in still images. Int. J. Comput. Vis. **105**(3), 205–221 (2013a)
27. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the CVPR (2006)
28. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. Pattern Anal. Mach. Intell. **27**(8), 1265–1278 (2005)
29. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. Pattern Anal. Mach. Intell. **32**(2), 353–367 (2011)
30. Lowe, D.G.: Distinctive image features from scale-invariant points. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
31. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
32. Murray, N., Marchesotti, L., Perronnin, F.: Ava: a large-scale database for aesthetic visual analysis. In: Proceedings of the CVPR (2012)

33. Mutch, J., Lowe, D.G.: Object class recognition and localization using sparse features with limited receptive fields. Int. J. Comput. Vis. **80**(1), 45–57 (2008)

34. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)

35. Oliva, A., Torralba, A.B.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001)

36. Quiroga, R.Q., Pedreira, C.: How do we see art: an eye-tracker study. Front. Hum. Neurosci. **5**, 98 (2011)

37. Ramachandran, V., Hirstein, W.: The science of art: a neurological theory of aesthetic experience. J. Conscious. Stud. **6**(7), 15–51 (1999)

38. Rojas, D., Khan, F.S., van de Weijer, J., Gevers, T.: The impact of color on bag-of-words based object recognition. In: Proceedings of the ICPR (2010)

39. Sablatnig, R., Kammerer, P., Zolda, E.: Hierarchical classification of paintings using face- and brush-stroke models. In: Proceedings of the ICPR (1998)

40. Shamir, L., Macura, T., Orlov, N., Eckley, D.M., Goldberg, I.G.: Impressionism, expressionism, surrealism: automated recognition of painters and schools of art. ACM Trans. Appl. Percept. **7**(2), 8–16 (2010)

41. Shamir, L.: Computer analysis reveals similarities between the artistic styles of Van Gogh and Pollock. Leonardo **45**(2), 149–154 (2012)

42. Shamir, L., Tarakhovsky, J.A.: Computer analysis of art. J. Comput. Cult. Herit. **5**(2), 1–7 (2012)

43. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Proceedings of the CVPR (2007)

44. Shen, J.: Stochastic modeling western paintings for effective classification. Pattern Recognit. **42**(2), 293–301 (2009)

45. Siddiquie, B., Vitaladevuni, S.N.P., Davis, L.S.: Combining multiple kernels for efficient image classification. In: Proceedings of the WACV (2009)

46. Vikram, T., Tscherepanow, M., Wrede, B.: A saliency map based on sampling an image into random rectangular regions of interest. Pattern Recognit. **45**(9), 3114–3124 (2012)

47. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010)

48. van de Weijer, J., Schmid, C., Verbeek, J.J., Larlus, D.: Learning color names for real-world applications. Trans. Image Process. **18**(7), 1512–1524 (2009)

49. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: large-scale scene recognition from abbey to zoo. In: Proceedings of the CVPR (2010)

50. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object catergories: a comprehensive study. Int. J. Comput. Vis. **73**(2), 213–218 (2007)

51. Zujovic, J., Gandy, L., Friedman, S., Pardo, B., Pappas, T.: Classifying paintings by artistic genre: an analysis of features and classifiers. In: Proceedings of the MMSP (2009)

Fahad Shahbaz Khan is a post doctoral research fellow in Computer Vision Laboratory, Linköping University, Sweden. He has a master in Intelligent Systems Design from Chalmers University of Sweden and a Ph.D. degree in Computer Vision from Autonomous University of Barcelona, Spain. His research interests are in color for computer vision, object recognition, action recognition and visual tracking. He has published articles in high-impact computer vision journals and conferences in these areas.

Shida Beigpour is an Associate Professor at the Faculty of Computer Science and Media Technology in Gjovik University College (Hgskolen i Gjvik) and Researcher at the Norwegian Colour and Visual Computing Laboratory. She received her M.Sc. degree on Artificial Intelligence and Computer Vision in 2009, and her Ph.D. of Informatics (Computer Vision) in 2013 from Universitat Autonoma de Barcelona, Spain. Her main research interests are Color Vision, Illumination and Reflectance analysis, Color Constancy and Color in Human Vision. She has published articles in high-impact journals and conferences and is a member of the IEEE.

Joost van de Weijer is a senior scientist at the Computer Vision Center. Joost van de Weijer has a master in applied physics at Delft University of Technology and a Ph.D. degree at the University of Amsterdam. He obtained a Marie Curie Intra-European scholarship in 2006, which was carried out in the LEAR team at INRIA Rhne-Alpes. From 2008 to 2012 he was a Ramon y Cajal fellow at the Universitat Autonma de Barcelona. His research interest are in color for computer vision, object recognition, and color imaging. He has published in total over 60 peer reviewed papers. He has given several postgraduate tutorials at major ventures such as ICIP 2009, DAGM 2010, and ICCV 2011.

Michael Felsberg received PhD degree (2002) in engineering from University of Kiel. Since 2008 full professor and head of CVL. Research: signal processing methods for image analysis, computer vision and machine learning. More than 80 reviewed conference papers, journal articles, and book contributions. Awards of the German Pattern Recognition Society (DAGM) 2000, 2004, and 2005 (Olympus award), of the Swedish Society for Automated Image Analysis (SSBA) 2007 and 2010, and at Fusion 2011 (honourable mention). Coordinator of EU projects COSPAL and DIPLECS. Associate editor for the Journal of Real-Time Image Processing, area chair ICPR and general co-chair DAGM 2011.