Artistic Movement Recognition by Boosted Fusion of Color Structure and Topographic Description

Corneliu Florea, Cosmin Ţoca
Image Processing and Analysis Laboratory,
University Politehnica of Bucharest, Romania
corneliu.florea@upb.ro; cosmin.toca@gmail.com

Fabian Gieseke
Department of Computer Science,
University of Copenhagen, Denmark
fabian.gieseke@di.ku.dk

Abstract

We[1] address the problem of automatically recognizing artistic movement in digitized paintings. We make the following contributions: Firstly, we introduce a large digitized painting database that contains refined annotations of artistic movement. Secondly, we propose a new system for the automatic categorization that resorts to image descriptions by color structure and novel topographical features as well as to an adapted boosted ensemble of support vector machines. The system manages to isolate initially misclassified images and to correct such errors in further stages of the boosting process. The resulting performance of the system compares favorably with classical solutions in terms of accuracy and even manages to match modern deep learning frameworks.

1. Introduction

The expansion of the digital data acquisition favors an eased access to works of art for the general public in parallel to the assembly of large collections over the web. While in the art domain it is often said that "precise formulations and rigorous definitions are of little help in capturing the meaning of art" [12, 33], in computer science there is a continuous effort to create autonomous systems that understand and replicate art concepts. For instance, there have been recent reports of algorithms that alter a digital image to replicate a painter style [16]. Alternatively and, arguably harder, is the task of automatic context recognition given a digitized painting. One of the broadest possible implementation of context recognition is the automatic art movement identification.

According to current online resources for art such as *Artyfactory*[2], the concept of *art movements* can be described as "collective titles that are given to artworks which share the same artistic ideals, style, technical approach or timeframe" [1]. While some pieces of work are clearly set into a single art movement, others are hard to classify—even for experts—as inceptive ideas sprung up randomly in different locales and they require contextual or background knowledge outside influence [12]. This work addresses the problem of an automatic categorization of digitized paintings into different types of art movements.[3] While other directions of image classification such as scene or object recognition benefit from large databases and agreed evaluation protocols, painting art movement recognition mainly lacks such aspects. Often, the performance assessment of a new method is carried out on a small database with only few paintings belonging to a certain art movement.

The contribution of this work is twofold: Firstly, we propose a new database consistent in terms of size and annotations. Secondly, we propose an adapted learning framework that is based on complementary feature extraction and boosted ensembles of support vector machines. The classification performance of our system is superior to those of other state-of-the-art models such as random forests and deep convolutional neural networks and reaches valuable accuracies for the task of art movement classification.

2. Related Work

Several solutions have been proposed for the automatic art movement recognition. Initially, systems were introduced along with an associated database. Later works resort to test images from publicly available visual art encyclopedias such as `wikiart`[4].

[1]The paper was published as C. Florea, C. Ţoca, F. Gieske "Artistic Movement Recognition by Boosted Fusion of Color Structure and Topographic Description", in IEEE Winter Application on Computer Vision, Santa Rosa, CA, USA, 2017. Please see the official version on IEEE Explore.

[2]`http://www.artyfactory.com`

[3]Depending on the source at hand, the concept of *art movement* considered in this work can also be named as "style", "genre" or "artistic current".

[4]`https://www.wikiart.org`

Method	Movements	Size	Test Ratio	CV	RR
Gunsel *et al*. [17]	3	107	53.5%	no	91.7%
Zujovic *et al*. [44]	5	353	10.0%	yes	68.3%
Siddiquie *et al*. [38]	6	498	20.0%	yes	82.4%
Shamir *et al*. [37]	3	517	29.8%	no	91.0%
Arora and Elgammal [3]	7	490	20.0%	yes	65.4%
Khan *et al*. [21]	13	2,338	46.5%	no	62.2%
Condorovici *et al*. [8]	8	4,119	10.0%	yes	72.2%
Agarwal *et al*. [2]	10	3,000	10.0%	yes	62.4%
Karayev *et al*. [20]	25	85,000	20.0%	yes	n/a
Bar *et al*. [5]	27	47,724	33.0%	yes	43.0%
This work	**18**	**18,040**	**25.0%**	**yes**	**50.1%**

Table 1. Overview of art movement recognition systems along with the sizes of the considered image databases. The sizes only refer to the database used for the art movement recognition system. The recognition rates (RR) are taken from the respective works (the test ratios depict the percentage of the databases being used for the systems' evaluations). Karayev *et al*. [20] report precision-recall values.

2.1. Databases

The most recent approaches and databases are listed in Table 1. In general, the sizes of the databases and the number of art movements considered increased over time, while the reported classification performances seem to have decreased. More recent works collected images from the web to create databases [5, 20, 21]. Khan *et al*. retrieved images from 91 painters for `Paintings-91` database [21]; here the movement annotation is available for painters associated with only one main art movement. In contrast, we allow the paintings of one author to be placed in different movements. For instance, Picasso authored more than 1,000 works, creating not only cubist, but also impressionist or surrealist works. Karayev *et al*. [20] collected an impressive number of images solely from `wikiart` and tested various combinations of descriptors and classifiers inspired from deep convolutional networks and metric learning while Bar *et al*. [5] retrieved a subset of images and performed a parallel experimentation. While `wikiart` was our main source, we also retrieved data from other websites and, more importantly, we have manually refined the collection and the labeling in two iterations (see below).

To conclude, many of the databases previously used tend to be quite small and are often based on non-standard evaluation protocols that might foster overfitting effects. Thus, a large-scale database with a fixed evaluation protocol should be beneficial for further development in this interesting and challenging field.

2.2. Art Movement Recognition

Again we refer to Table 1 for a systematic presentation of previously proposed solutions. Most systems addressed the problem via a classical approach: image description followed by potentially feature selection and classification. Typical texture-based description are often achieved via, e.g., Local Binary Patterns (LBP) [2, 5, 21], Histogram of Oriented Gradients (HOG) [2, 21], or Gabor Filters [8, 37]. For color description, either color variants of the gray levels texture descriptors (e.g., colorHOG or colorSIFT) or methods such as Color names [21] are used.

The great advance of machine learning in the last decade also impacted painting description. For instance, Arora *et al*. [3] rely on so-called Classemes descriptors, while Karayev *et al*. [20], Bar *et al*. [5] and Peng *et al*. [36] use convolutional filters from pretrained deep convolutional neural networks (ImageNet), as suggested by Donahue *et al*. [11]. Most other approaches proposed so far rely on the standard application of support vector machines (SVM) [2, 5, 21, 36].

3. Approach

The approach proposed in this work resorts to color structure and topographic features as descriptors. These descriptors are then processed by an ensemble of adapted boosted support vector machines.

3.1. Features

Given a structuring window of 8×8, the *color structure descriptor* (CSD) [30] counts the number of times a particular color is contained within the structuring element, as the structuring element scans the image. The CSD partially accounts for spatial coherence in the gross distribution of quantized colors. It has been used for color image description, yet in the later period other solutions have been preferred [39].

The concept of *complete topographical image description* is based on the following derivations: By interpreting a planar image as surface, i.e., as twice differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$, one can consider the Taylor series expansion to approximate $f(x, y)$ in a local region around a given point (x, y):

$$
\begin{aligned}
&f(x + \Delta_x, y + \Delta_y) \approx f(x, y) + \\
&\overrightarrow{\nabla f} \cdot [\Delta_x, \Delta_y] + \tfrac{1}{2}\,[\Delta_x, \Delta_y]\,\mathcal{H}(x, y) \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix},
\end{aligned}
\tag{1}
$$

where $\overrightarrow{\nabla f}$ is the gradient and $\mathcal{H}(x, y)$ the Hessian matrix of f. In a topographical interpretation, the vectorial gradient indicates inclination, while the Hessian provides cues about local curvature. Typically the gradient is presented in polar coordinates, $\overrightarrow{\nabla f(x, y)} = \left[|\overrightarrow{\nabla f}(x, y)| \cos(\Theta_f(x, y)), |\overrightarrow{\nabla f}(x, y)| \sin(\Theta_f(x, y)\right]^T$. From the 2×2 Hessian, one can retrieve the eigenvectors, $\overrightarrow{V^1_{\mathcal{H}}(x, y)}$, $\overrightarrow{V^2_{\mathcal{H}}(x, y)}$ and eigenvalues, $\lambda^1_{\mathcal{H}}(x, y)$, $\lambda^2_{\mathcal{H}}(x, y)$. Similarly to the gradient, one can express the Hessian eigenvectors in polar coordinates as magnitude and orientation (here, only the first eigenvector orientation matters as the second one is perpendicular). Thus a pixel (x, y), is described by the following components: $f(x, y)$, $|\overrightarrow{\nabla f}(x, y)|$, $\Theta_f(x, y)$, $|\overrightarrow{V^1_{\mathcal{H}}(x, y)}|$, $|\overrightarrow{V^2_{\mathcal{H}}(x, y)}|$, and $\Theta_{\mathcal{H}}(x, y)$.

In previous works, the local pixel value is the base, among others, for the *local invariant order pattern* (LIOP) descriptor [41]. Gradient orientation (and magnitude) is the basis of *histogram of oriented gradient* (HOG) [9]. The second derivative is used to locate key points in SIFT (Scale-invariant feature transform) or to describe shapes by means of principal curvature. Also Deng *et al.* [10] describe regions using principal curvature for object recognition, while Florea *et al.* [14] aggregate first and second derivative information to describe faces for pain intensity estimation. We propose to use all information for texture description and we stress that the use of curvature for this purpose is novel.

For a refined description, one uses the concept is *multi-scale topography*, which assumes the computation of the derivatives in the scale space [14, 27]. There, the image is replaced by the scale space of an image $F(x, y, \sigma)$:

$$
F(x, y, \sigma) = G(x, y, \sigma) * f(x, y), \tag{2}
$$

where $*$ stands for convolutions and $G(x, y, \sigma)$ is a Gaussian rotationally symmetric kernel with variance σ^2 (the scale parameter). The derivative of F is found by convolving the original image f with the derivative of G. A pyramidal version of the topographic descriptor requires merely consideration of multiple values for σ.

To provide global image descriptors it has been suggested to aggregate the topographic set into histograms [14]. This representation was named *histogram of topographic* (HoT) descriptor and has been successfully been applied for face analysis. We have implemented a similar version and our results show its usefulness in describing image texture.

3.2. Boosted Support Vector Machines

Our approach is based on support vector machines with *radial basis function kernels* (RBF). Yet to increase the overall performance, data from multiple features has to be brought together. As direct fusion into a single classifier failed, we consider a modified boosted fusion procedure inspired by the SAMME algorithm [43]: Given n training examples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\} \subset \mathbb{R}^d \times \{-1, +1\}$, a standard support vector machine aims at minimizing

$$
\begin{aligned}
&\Phi(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\textstyle\sum_{i=1}^N \xi_i, \;\; s.t. \\
&y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle + b) \geq 1 - \xi_i, \xi_i \geq 0, \;\; i \in \{1, \ldots, n\}
\end{aligned}
\tag{3}
$$

and can be extended with individual weights for the training patterns via [42]:

$$
\begin{aligned}
&\Phi(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\textstyle\sum_{i=1}^N W_i\xi_i, \;\; s.t. \\
&y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle + b) \geq 1 - \xi_i, \;\; \xi_i \geq 0, \;\; i \in \{1, \ldots, n\}.
\end{aligned}
\tag{4}
$$

Here, C is a cost parameter determining the trade-off between training loss and large margin and $W_1, \ldots, W_N$ are the weights associated with the training points. The feature mapping Φ stems from a kernel function; a prominent one is the RBF kernel defined as $k(\mathbf{x}, \mathbf{z}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{z})\rangle = \exp\left(\gamma^2\|\mathbf{x} - \mathbf{z}\|\right)$. Given the matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ consisting of the n training patterns, the associated weight vector $\mathbf{W} \in \mathbb{R}^n$, and the vector $\mathbf{Y} \in \{-1, +1\}^n$ consisting of the class labels, let $\mathcal{T}_{\gamma,C} = (\mathbf{X}, \mathbf{W}, \mathbf{Y}, \gamma, C)$ denote the resulting (trained) model. Accordingly, given two different sets of features with induced pattern matrices $\mathbf{X}_{(p)}$ and $\mathbf{X}_{(q)}$, the individual models can be denoted by $\mathcal{T}_{(p),\gamma,C}$ and $\mathcal{T}_{(q),\gamma,C}$, respectively. For simplicity of writing, $\mathcal{T}_{(p),\gamma,C} = \mathcal{T}_{(p)}$, with γ, C, being implicitly assumed.

The fusion procedure, for the general case with Q sets of features, is described in Algorithm 1. It was shown that *AdaBoost* with SVMs as component classifiers exhibits a better performance for binary classification compared to other approaches if the $\gamma = \frac{1}{\sqrt{\sigma}}$ parameter is iteratively increased [25]. We have found that in case bootstrapping is used instead of resorting to a single training set (as in [25]), this request may be avoided as a unique value γ suffices.

Algorithm 1 takes also inspiration from the principle of so-called *arcing classifiers* [6], with the major difference that instead of a full training set (i.e. all dimensions) we only use parts of it (feature oriented). Furthermore, various

1. Initialize the observation weights $W_i^{(1)} = 1,\ i \in \{1, \dots, n\}$;
2. Independently find the best parameters [15] for $\mathcal{T}_{(k),\gamma_k,C_k} = \mathcal{T}_{(k)}$ given $k \in \{1, \dots Q\}$;
3. for *m=1:M* **do**

a. Randomly select a classifier $\mathcal{T}_p^{(m)}$ with $p \in \{1 \dots Q\}$. Also select $\mathbf{X}_{(p)}$;
b. Select a random bootstrap sample of the data;
c. Fit the chosen classifier $\mathcal{T}_p^{(m)}$ to the training data using weights $\mathbf{W}^{(m)}$;
d. Compute the recognition error: ;

$$\varepsilon_m = \left(\sum_{i=1}^{n} W_i^{(m)} \left[c_i \neq \mathcal{T}_p^{(m)}(\mathbf{x}_i) \right] \right) / \sum_{i=1}^{n} W_i \tag{5}$$

d. Compute the update:

$$\alpha^{(m)} = \min \left(\log \frac{1 - \varepsilon_m}{\varepsilon_m} + \log(K-1), \alpha_{\max} \right) \tag{6}$$

e. Set

$$w_i \leftarrow w_i \cdot \beta^{\alpha^{(m)} \left[c_i \neq \mathcal{T}_p^{(m)}(\mathbf{x}_i) \right]} \tag{7}$$

end
Result: Boosted ensemble of partial SVMs given by:

$$C(\mathbf{X}) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \left[\mathcal{T}_p^{(m)}(\mathbf{X}_{(p)}) = k \right] \tag{8}$$

Algorithm 1: Fused SVMs: $[\mathbf{a}_i = \mathbf{b}_i]$ is the Iverson bracket notation for the number of occurrences; K=18 (number of classes), $\alpha_{\max} = 10$, $\beta = 1.2$ (so that $\beta^{\log(K-1)} \approx 2$).

solutions of SVM ensembles were previously introduced as it can be followed in the review of Wang *et al*. [19] and more recently in the work of Mayhua-Lopez *et al*. [32].

Among other details, we specifically differ by the supplementary regularization as additional randomness when we choose the next SVM for the overall ensemble. In fact this choice departs the proposed solution from the traditional approaches of boosting [6, 31], where improvement (i.e. next learner) is chosen as the steepest descent in the function space; here it is randomly chosen. Compensatory optimization is due to equation (5), where a large recognition error shows that a some randomly selected learner should not contribute much, thus will have less significance in the overall classifier, as shown by equation (8).

Algorithm 1 requires that the individual SVMs yield a reasonable performance; due to the RBF kernel, this implies that good parameter assignments for both γ and C have to be found. An alternative would be to consider linear SVMs (which require to optimize for C only), yet the potential decrease in performance is also transferred to the boosted ensemble. In contrast, considering the RBF case we model the parameters (γ, C) by a Gaussian process (i.e. collection of random variables that have a joint Gaussian distribution). In this case, we follow by Bayesian optimization as described in [15] so that to reach values close enough to the maximum in, at most, 10 iterations. In our experimental evaluation, this process led to the following approximate values for (γ, C) given two SVM models: $(90 \approx 2^{6.5}; 0.09 \approx 2^{-3.5})$ and $(2.8 \approx 2^{1.5}; 0.5 \approx 2^{-1})$, respectively.

4. Paintings Database: Pandora18k

One contribution of the work at hand is the collection of a new and extensive dataset of art images.[5] The database was formed in three steps: (1) collection, (2) image review, and (3) art movement review. The first step took all works from Table 1 into account (collection of images from the web along with an art movement label). `Wikiart` was used as a main source, but more than 25% was also collected from other sources. We specifically tried to balance the distribution among art movements.

The second step implied the manual review of all images. This was implemented by non-art experts and implied cropping to remove the painting framework and to eliminate images of sculptures or of 3D objects (often appearing in modern art); here were altered (image and/or label) about 15% of instances. In the third step, the entire database was reviewed by an art expert and all images that were considered to be "not artistic" were removed.

Following this review we note that: (i) There are works labeled with some style, while the author is known for his/her work for other styles; we have kept both. (ii) Multiple labels given to a work are eliminated and only the dominant one was kept. (iii) We try to replace parts of larger painting (parts which are abundant on Internet) with the full scale work. (iv) Modern art examples contain not only paintings, but also digitized graphics.

In contrast, the recently used `Wikiart` collection [5, 20] is more exhaustive, but also suffers from weak annotations. It contains images of sculptures, crops, images with non-original frameworks and, in many cases, the works have the style label of the movement to which their creator is associated, although they are, for instance, simply book illustrations. Due to these aspects, we consider it as being less suitable for rigorous artistic movement recognition.

The editing process resulted in a set of 18,040 images and 18 art movements in total, see Figure 1. The struc-

[5] The Pandora18k database with precomputed features data reported is available at `http://imag.pub.ro/pandora/pandora_download.html`.

Art movement	**Img.**	**Key dates**	**Main characteristics** [28]:
Byzantinism	847	500–1400	religious, aura
Early Renais.	752	1280–1450	ceremonial, divine, idealized
North. Renais.	821	1497–1550	detailed realism, tones, naturalism
High Renais.	832	1490–1527	rigor, antiquity, monumental, symmetry
Baroque	990	1590–1725	dramatic, allegory, emotion, strong colors, high contrast
Rococo	832	1650–1850	decorative, ludic, contemplative
Romanticism	895	1770–1880	rebellion, liberty emotion
Realism	1200	1880–1880	anti–bourgeois, real, social critique
Impressionism	1257	1860–1950	physical sensation, light effect, movement, intense colors, plein air
Post-Impress.	1276	1860–1925	meaningful forms, drawing, structure, strong edges
Expressionism	1027	1905–1925	strong colors, distortion, abstract, search
Symbolism	1057	1850–1900	emotion, anarchy, dream imagery
Fauvism	719	1905–1908	intense colors, simplified composition, flatness, unnatural
Cubism	1227	1907–1920	flat volumes, confusing perspective, angles, artificial
Surrealism	1072	1920–1940	irrational juxtaposition, subconscious, destruction
Abstract art*	1063	1910–now	geometric, simplified compositions
Naive art	1053	1890–1950	childlike simplicity, ethnographic, patterns, erroneous perspective
Pop art	1120	1950–1969	imagery from popular culture, irony

Table 2. Proposed database: The *Abstract art* class (*) encompasses Abstract Art (pure), Abstract expressionism, Constructivism, Neo-plasticism and Suprematism.

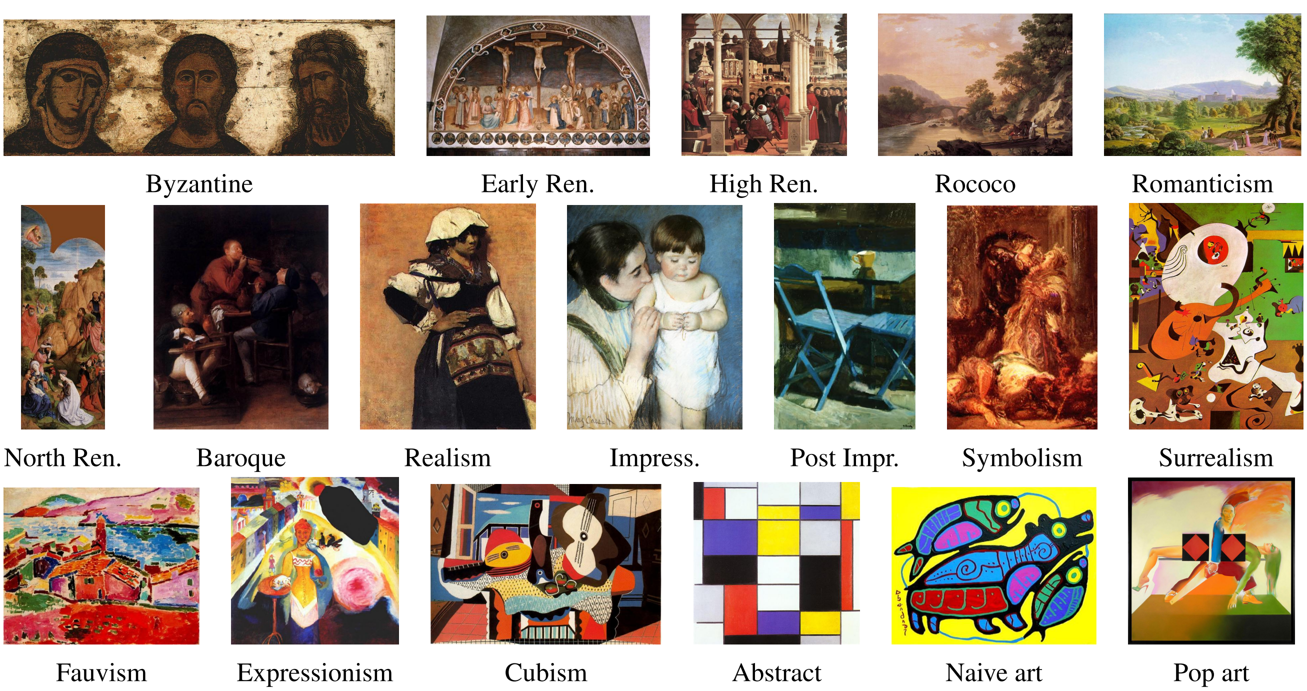

Figure 1. The 18 art movements illustrated in the proposed database.

ture overview may be followed in Table 2. The difficulties of automatic characterization are related to: (i) The quality of digitized images varies greatly, from high to low resolutions, further damaged by JPEG artifacts; (ii) The aspect ratio varies from 3:1 to 1:3 and some paintings have a circular frame; (iii) More importantly, following the short descriptions from Table 2, the main difference between various movements is subtle and more related to the content, that is not easy to measure it by formal characteristics.

5. Results

5.1. Wikiart

Although in the previous section we argued why `Wikiart` is less appropriate for art movement recognition, we have evaluated the proposed system on this database as well. The procedure is the same used by Karayev *et al.* [20]: Binary randomly balanced classes with 20% of the data as test set. The proposed algorithm (boosting SVM ensembles over pHoT and CSD) is detailed in Section 3, with the sole modification that for SVMs the convergence criterion is shrunk to 10^{-5} from 10^{-5} to cope with fewer data vs. higher dimensions. Overall, the proposed system obtained an average accuracy of 82.41% compared to 81.35% reported in [20] for the MC-bit variant as it was identified the top performer. A per class comparison is available in Figure 2.

5.2. Proposed Database

The remainder of the evaluation and discussions are related to the proposed database.

Training and Testing. To separate the database into training and testing parts, a 4-fold cross validation scheme was implemented. The division into 4 folds exists at the level of each art movement, thus each image is uniquely allocated to a fold. The same division was used for all further tests and it is part of the database annotations.

Features and Classifiers. As "there is no fixed rule that determines what constitutes an art movement" and "the artists associated with one movement may adhere to strict guiding principles, whereas those who belong to another may have little in common" [1], there cannot be a single set of descriptors able to separate any two art movements.

Prior works [3, 21] noted that multiple categories of feature descriptors should be used. For instance, to differentiate between impressionism and previous styles, one of the main difference is the brush stroke, thus *texture*; fauvism is defined by the *color palette*. Yet following Table 2, mainly the *composition* should be used.

To provide a baseline for further evaluation, we have tested various combinations of popular feature extractors and classification algorithms. The texture feature extractors used are: the previously mentioned HoT; HOG [9]; LPB [34]; LIOP [41]; Edge Histogram Descriptor (EHD) and Homogenous Texture Descriptor (HTD = Gabor filters; both are part of the MPEG-7) [30]; SIFT descriptor [29]. Initially, the features are computed on the equivalent grayscale image. The pyramidal versions implied four levels of a Gaussian pyramid. For HOG, LIOP, LBP and SIFT implementations, we relied on the VlFeat library [40]. MPEG-7 descriptors are computed with BilVideo-7 library [4].

While the pyramidal texture features should be able to describe the global composition, we also tested the GIST [35] for the same purpose. For color descriptions, we evaluated Discriminative Color Names (DCN) [22] and CSD.

For the initial evaluation, we have coupled each of those descriptors with two standard machine learning systems, which have been previously found [13] to be the best performers: support vector machines (using the `LibSVM` implementation [7]) and random forests (RF). For these tests, the SVM-RBF implied hyperparameter optimization, while the Random Forest contained 100 trees and $\sqrt{d}$ features were tested per internal node split.

The results do not contain any mid-level description; similar works on the topic showed that for the particular case of paintings, these do not help [5, 21]. We tested Fisher Vector over SIFT and the combination SIFT+FV+SVM lead to an overall decrease in performance with 1% compared to SIFT+SVM.

Systems. Noting the recent advances of deep networks, we have tested several alternatives and the results are presented in Table 4.[6] For LeNet and NiN, we used the MatConvNet library, while for AlexNet and ResNet, we resorted to the CNTK library. In all these cases, the mean was subtracted and image intensity was rescaled to [0,1]. We tested various image augmentations such as centering, cropping, and warping; warping to 224×224 gave the best performance, and we only report the corresponding results.

Given the results of the individual features, we have tested various alternatives to fuse the results; these are shown in Table 5. Following previous works on art movement recognition [5, 20, 36], convolutional filters from the Caffe version of the AlexNet trained on ImageNet were applied on the database and results are marked with DeCAF [11] and the layer subscript. Also, given the results from [36], we tried to use layers of the DeCAF filters in the boosting procedure.

Duration. The proposed system was implemented in Matlab with C code for feature extraction and the use of LibSVM. The feature independent SVM-RBF hyperparameter optimization was carried out in parallel, otherwise the

[6] The convolutional neural network (CNN) performance is taken after 40 epochs for LeNet and NiN and after 100 iterations for AlexNet and ResNet. ResNet reaches 49.1 accuracy in the course of the training process. More epochs (up to 500) did not improve the performance.

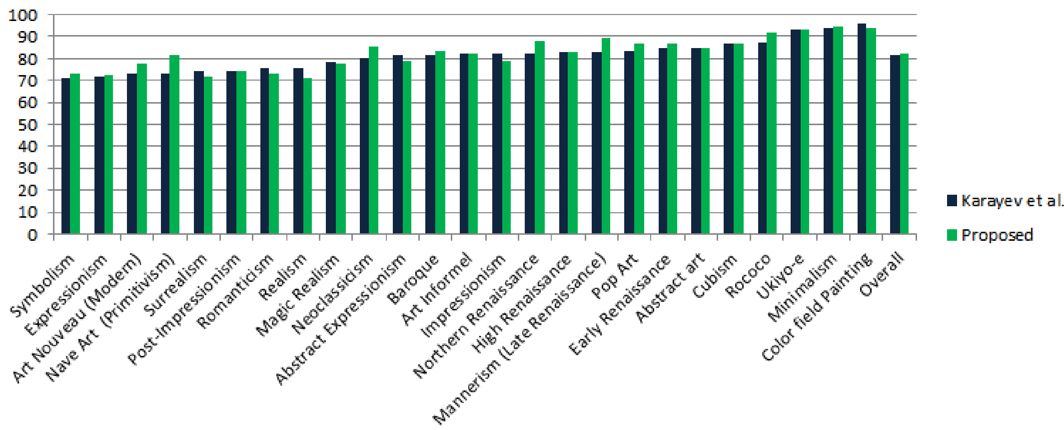

Figure 2. Results and comparison with [20] on the `Wikiart` database.

	HOG	pHoG	colHoG	HoT	pHoT	LBP	pLBP	SIFT	LIOP	HTD	EHD	GIST	DCN	CSD	pLBP+CSD	pHoT+CSD
RF	18.4	23.4	19.6	29.6	32.3	27.2	32.7	21.6	24.4	22.3	24.9	23.8	18.9	31.3	37.8	37.7
SVM	17.4	24.7	19.1	30.8	**42.5**	27.4	39.2	23.6	25.2	19.7	22.7	23.5	19.4	**33.8**	40.4	**47.1**

Table 3. Recognition rates (%) for various features and classifiers (Proposed database)

Deep CNN				
Type	Size	Layers	Time	RR
LeNet [24] - Rand	32	14	$<$ 1h	22.3
LeNet [24] - Rand	64	16	$<$ 1h	25.1
NiN [26] - Rand	64	17	$<$ 1h	26.5
AlexNet [23] - Rand	224	8	$<$ 1h	39.5
AlexNet [23] - FT-3	224	8	2h	39.5
AlexNet [23] - FT-4	224	8	60h	**56.5**
ResNet-34 [18] - Rand	224	34	2h	47.8

Table 4. Recognition rates (RR) for various CNN models (*size* refers to the width and height of the input images; *Layers* to the number of layers). *Rand* refers to the case when initialization was from scratch, while $FT-N$ refers to a pre-trained ImageNet instance with only the top N layers being retrained.

Features + Classifier			
Features	Classifier	Time	RR
DeCAF$_6$	SVM	1h	42.8
DeCAF$_5$	SVM	1h	41.7
All	RF	6h	44.5
All+PCA	RF	3h	38.5
All	SVM	1h	**50.0**
pHoT+CSD	SVM	2h	47.1
DeCAF$_6$	Boost	2h	49.4
DeCAF$_{All}$	Boost	2h	44.6
pHoT+CSD	Boost	*2.5h*	***50.1***
All	Boost	*6.5h*	48.5

Table 5. Recognition rates for a subset of the features/classifiers reported in Table 3. Here, *All* refers to all features listed in Table 3. For each approach, the training time is reported for a given fold. The proposed method is denoted by *Boost*.

code ran, unoptimized, on a single core Xeon E3-1280 in $\approx$ 100 minutes. The use of small in-bag sets and Bayes optimization instead of full search for (γ, C) allowed considerable acceleration. Reported time for the CNNs implied acceleration on Nvidia 980 Ti GPU. In all cases, testing requires up to 2 seconds per image.

6. Discussion

Features and Classifiers. Table 3 indicates that predominantly, SVM outperforms RF for the classification task. Typically SVMs perform well in high-dimensional feature spaces (where each feature is not "powerful", but the combination is). RFs perform well given single powerful features. Tested features are essentially histograms and the information carried by a single bin is not relevant, while the unbalance in the entire distribution is. Regarding the features, the texture category is dominated by HoT and LBP, with a slight edge for the first. In the color category, the CSD clearly reaches top accuracy.

With respect to the proposed boosted fusion method, one particular aspect that we found to be critical in achieving greater accuracy is the random choosing (Algorithm 1, 3.a)

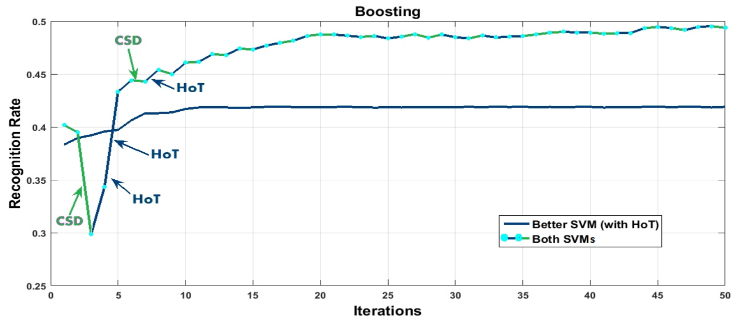

Figure 3. Accuracy while the system iterates. Using only the best classifier (HoT, blue solid line), the overall performance saturates early. Using both (segmented line), the increase is initially due to the strong classifier, while the weak one contributes positively only later.

of a classifier. The alternative is choosing the best classifier (according to the steepest descent w.r.t. a loss function [31]); yet this path, while initially yields steeper increase (see Figure 3), it also reaches the stationary point earlier. Also, choosing the best implies to train and test all classifiers at all steps to get the maximum, which adds considerably computational burden. Choosing the involved classifier randomly leads works if the base learners complements each other, which is achieved by using complementary features.

Comparisons. The numerical comparison with CNN shows that the proposed method exhibits a better accuracy in a comparable runtime. Yet we note that CNNs make extensive use of GPU acceleration, while the proposed method does not. Thus we may hope that further acceleration is reachable. Lower relative performance for CNN can be linked to the lack of repeatable objects into paintings; differences between various art movements are more subtle (as emphasized in Table 2).

Also another liability of the CNNs in the current evaluation is the limited size of the database. `Wikiart`, `ArtUK` and other Internet sources potentially contain significant supplementary data, but this needs to be further validated by art experts before being used with significant confidence. Alternatively pre-trained CNNs, that are fine-tuned on this database over many iterations outperform the proposed solution (as shown in table 4). Although, in such a case, the computational requirements are also significantly higher.

Using *all* the features with a simple SVM yields comparable results. Yet, this implies the computation of all features, which takes considerable more time than reported solely for training in Table 5. For this reason also, we opt to use only two types of features.

Art Movement Recognition. The resulting confusion matrix[7] is shown in Figure 4. The system recognizes very

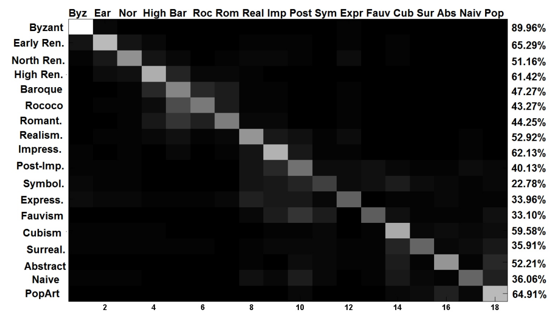

Figure 4. Confusion matrix (recognition rates)

accurately older movements like Byzantinism or High Renaissance and heavily confuses more modern ones. Worst recognition rates are for Expressionism (22.8%), Symbolism and Fauvism. Confusion is between older movements and, separately, between newer ones. There are no two classes being completely interfused.

Acknowledgment Corneliu Florea is supported by supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS UEFISCDI, number PN-II-RU-TE-2014-4-0733. Authors wish to express their gratitude to Iuliana Dumitru from University of Bucharest which patiently reviewed all paintings collected in the database and provided valuable feedback.

References

[1] What is an art movement ? `www.artyfactory.com/art_appreciation/art_movements/art_movements.htm`, Retrieved May 2016.

[2] S. Agarwal, H. Karnick, N. Pant, and U. Patel. Genre and style based painting classification. In *WACV*, pages 588–594, 2015.

[3] R. S. Arora and A. Elgammal. Towards automated classification of fine–art painting style: a comparative study. In *ICPR*, pages 3541–3544, 2012.

[7] Additional results are presented in the supplementary material.

[4] M. Baştan, H. Çam, U. Güdükbay, and O. Ulusoy. BilVideo-7: An MPEG-7-compatible video indexing and retrieval system. *IEEE MultiMedia*, 17(3):62–73, 2009.

[5] Y. Bar, N. Levy, and L. Wolf. Classification of artistic styles using binarized features derived from a deep neural network. In *ECCV Workshops - Visart*, pages 71–84, 2014.

[6] L. Breiman. Arcing classifiers. *The Annals of Statistics*, 26(3):801–824, 1998.

[7] C.-C. Chang and C.-J. Lin. LibSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3), May 2011.

[8] R. Condorovici, C. Florea, and C. Vertan. Automatically classifying paintings with perceptual inspired descriptors. *JVCI*, 26:222 – 230, 2015.

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[10] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro. Principal curvature-based region detector for object recognition. In *CVPR*, pages 2578–2585, 2007.

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *ICML*, 2014.

[12] D. Dutton. A naturalist definition of art. *Journal of Aesthetics and Art Criticism*, 64:367–377, 2006.

[13] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *JMLR*, 15(1):3133–3181, 2014.

[14] C. Florea, L. Florea, and C. Vertan. Learning pain from emotion: Transferred hot data representation for pain intensity estimation. In *ECCV workshops - ACVR*, pages 778–790, 2014.

[15] J. Gardner, M. Kusner, K. Weinberger, and J. Cunningham. Bayesian optimization with inequality constraints. In *ICML*, pages 937–945, 2014.

[16] L. Gatys, A. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016.

[17] B. Gunsel, S. Sariel, and O. Icoglu. Content-based access to art paintings. In *ICIP*, pages 558–561, 2005.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[19] S. jin Wang, A. Mathew, Y. Chen, L. feng Xi, L. Mab, and J. Lee. Empirical analysis of support vector machine ensemble classifiers. *Expert Systems with Applications*, 36:6466–6476, 2009.

[20] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. In *BMVC*, 2014.

[21] F. S. Khan, S. Beigpour, J. van de Weijer, and M. Felsberg. Painting-91: A large scale database for computational painting categorization. *Mach. Vis. App.*, 25(6):1385–1397, 2014.

[22] R. Khan, J. van de Weijer, F. Shahbaz Khan, D. Muselet, C. Ducottet, and C. Barat. Discriminative color descriptors. In *CVPR*, pages 2866–2873, 2013.

[23] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[24] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[25] X. Li, L. Wang, and E. Sung. AdaBoost with SVM-based component classifiers. *EAAI*, 21(5):785–795, 2008.

[26] M. Lin, Q. Chen, and S. Yan. Network in network. *CoRR*, abs/1312.4400, 2013.

[27] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):225–270, 1994.

[28] S. Little. *Isms: Understanding Art*. Turtleback, 2004.

[29] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 62(2):91–110, 2004.

[30] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Trans. Cir. and Sys. for Video Technol.*, 11(6):703–715, 2001.

[31] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean. Boosting algorithms as gradient descent. In *NIPS*, pages 512–518, 2000.

[32] E. Mayhua-Lopez, V. Gomez-Verdejo, and A. R. Figueiras-Vidal. A new boosting design of support vector machine classifiers. *Information Fusion*, 25:63–71, 2015.

[33] D. Novitz. Art by another name. *The British Journal of Aesthetics*, 38:19–32, 1998.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE T. PAMI*, 24(7):971–987, 2002.

[35] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[36] K. Peng and T. Chen. Cross-layer features in convolutional neural networks for generic classification tasks. In *ICIP*, pages 3057–3061, 2015.

[37] L. Shamir, T. Macura, N. Orlov, M. Eckley, and I. Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Trans. Appl. Percept.*, 7(2):1–17, 2010.

[38] B. Siddiquie, S. Vitaladevuni, and L. Davis. Combining multiple kernels for efficient image classification. In *WACV*, pages 1–8, 2009.

[39] K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE T. PAMI*, 32(9):1582–1596, 2010.

[40] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *ACM MM*, pages 1469–1472, 2010.

[41] Z. Wang, B. Fan, and F. Wu. Local intensity order pattern for feature description. In *ICCV*, pages 603–610, 2011.

[42] X. Yang, Q. Song, and Y. Wang. A weighted support vector machine for data classification. *IJPRAI*, 21(5):961–976, 2007.

[43] J. Zhu, H. Zou, S. Rosset, and T. Hastie. Multi-class AdaBoost. *Statistics and Its Interface*, 2:349–360, 2009.

[44] J. Zujovic, L. Gandy, S. Friedman, B. Pardo, and T. Pappas. Classifying paintings by artistic genre: An analysis of features & classifiers. In *IEEE MMSP*, pages 1–5, 2009.