Digital painting database for art movement recognition

April 21, 2016

Abstract

In this paper we introduce a large paintings digitized collection that is annotated with art movement labels. The database consists of more than 18000 images from 18 art movements. Each genre is illustrated by a number of images varying from 700 to more nearly 1200. We investigate how standard local and global features and classification systems are able to discriminate between various art movements. Among the various descriptor used, we showed that the best performance is achievable by a pyramidal version of modified Histogram of Topographical features coupled with Color Structure descriptor.

1 Introduction

The remarkable expansion of the digital data during the last period favored a much easier access to works of art for the general public. While in the art domain it is often said that art is for human only to understand as "precise formulations and rigorous definitions are of little help in capturing the meaning of art" [20], [10], in computer science there is a continuous effort to create autonomous system that understand and replicate art concepts. For instance, recently there have been reported algorithms that alter a digital image to replicate a painter style [13]. Alternatively, more appropriate to the ultimate goal of computers is task the context recognition given a digitized painting. One of the broadest possible implementation of context recognition is the automatic art movement identification.

According to Artyfactory [1], *art movements* are "collective titles that are given to artworks which share the same artistic ideals, style, technical approach or timeframe". While some works 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 requires contextual or background knowledge outside influence. Also while the actual characteristics place a work in some art movement, its author, for personal reasons, refused to be categorized in such a way, giving birth to disputes.

This paper addresses the problem of computational categorization of digitized paintings into artistic genres (or art movements)[1]. While other directions of image classification, such as scene or object recognition, benefits from large databases and agreed evaluation protocols, paint art movement recognition mainly lack such aspects. Often, the performance assessment of a new method is carried on a small database with few paintings belonging to few movement. Yet larger databases and more rigid protocols are needed for actual advance.

2 Related work

Paintings databases. In the last period multiple solutions approached automatic art movement recognition and while testing used there introduced databases. The more recent are listed in table 1. As a general rule, the size of the databases and the number of art movements investigated increased with time, while the reported performance decreased. Out of the listed works, we will discuss in detail [16] and [5]. Khan et al [16] collected images from Internet to form the so-called Paitings-91 database. The database contain images from 91 painters and those authored by painters that are associated with 1 main art movement received the label of that movement. In contrast, we allow the paintings of one author to be placed in different movement. For instance, Picasso who authored more than 1000 works, creating not only cubist but also impressionist or surrealist works.

Bar et al. [5] collected an impressive number of images solely from wikiart by retrieving all available images at the moment of experimentation. While we also used as a main source wikiart, we also retrieved

[1]Depending on the source the "art movement" is also named "genre", "style" or "artistic current".

Table 1: Art movement recognition solutions with the size of used databases. The database size refers only to the database used for art movement recognition, as in some cases larger databases have been implied for other purposes. The value for recognition rate (RR) is the one reported by the respective work while the "test ratio" is the percentage used for testing from the overall database. For more details the reader is kindly asked to follow the respective work.

Method	**Move-ments**	**Db. size**	**RR.**	**Test ratio**	**Cross-Validation.**
Gunsel et al. [14]	3	107	91.66%	53.5%	No
Zujovic et al. [28]	5	353	68.3%	10%	yes
Siddiquie et al. [24]	6	498	82.4%	20%	yes
Shamir et al. [23]	3	517	91%	29.8%	no
Arora&ElGammal[3]	7	490	65.4%	20%	yes
Khan et al. [16]	13	2338	62.2%	46.53%	yes
Condorovici et al.[8]	8	4119	72.24%	10%	yes
Agarwal et al. [2]	10	3000	62.37%	10%	yes
Bar et al. [5]	27	47724	43%	33%	yes
Proposed	**18**	**18040**	**42.3%**	**25%**	**yes**

data from other websites and more important, as we will further discuss, we have manually refined the collection and labelling, in two iterations.

Concluding, many of the databases previously used, are small and contain non-standard evaluation protocols allowing overfitting. Thus, a larger scale database with fixed evaluation protocol should be beneficial for further development on the topic.

3 Paintings database

The first contribution is the collection of a new and extensive dataset of art images[2]. The database was formed in three steps: (1) collection; (2) image review; (3) art movement review. The first step exists in all reported works from table 1: we have collected images from Internet together with art movement label. While Wikiart was used as a main source, yet more than 25% of the paintings are collected from other locations. We specifically tried to balance the distribution among art movements, while to have all the important ones represented.

The second step implied the manual review of all images. This was implemented by non-art experts and the follow some guidelines:

- The digital image should focus on the image content in the sense that as much as possible the frame to be removed as it is not representative for the art movement. Yet, especially for older, religious, images (e.g. Byzantine or Early Renaissance), the frame is part of the paintings, or is highly curved. If the picture frame is part of the artistic composition, than it was kept. A consequence is that a polyptych, if its content is not part of the same scene, is divided into multiple images.

- Sculpture and modern art which contains 3D objects have been removed as shadows may play an important role. For older art, which commonly is mural, if the curvature of the wall it too great, we have removed the image.

- We have removed pencil or charcoal sketches. Also images with highly degraded/washed colors have been removed. In parallel, we noticed that paintings have been photographed with multiple white balance choices. We have removed those images that were obviously wrong.

Thirdly, the entire database was reviewed by an art expert. Consequently images that were found to be "not artistic" even under low acceptance were removed. Following this review some notes should be taken into account:

- There are works labelled with some style while the author is known for its work in other style. For instance, Kazimir Malevich is known as being the originator of the Suprematism movement, while he has realist works. We have kept both.

[2]The database with pre-computed features data reported will be available on the project page

Table 2: The structure of the Pandora database. * Under abstract art label we have grouped five directions: Abstract Art (pure), Abstract expressionism, Constructivism, Neo-plasticism and Suprematism. Cubo-futurist paintings are included in the cubist data.

Art movement	**Images**	**Key dates**	**Main characteristics** [18]:
Byzantine Iconography	847	500 - 1400	religious, aura,
Early Renaissance	752	1280 - 1450	ceremonial, divine, idealized
Northern Renaissance	821	1497 - 1550	detailed realism, tones, naturalism
High Renaissance	832	1490 - 1527	rigor, antiquity, monumental, symmetry,
Baroque	990	1590 - 1725	dramatic, allegory, emotion, strong colors, high contrast
Rococo	832	1650 - 1850	decorative, ludic, ornamental, 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-Impressionism	1276	1860 - 1925	meaningful forms, drawing, structure
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 abstraction, simplified compositions
Naive art	1053	1890 - 1950	childlike simplicity, ethnographic, patterns, erroneous perspective
Pop art	1120	1950 - 1969	imagery from popular culture, irony

- There are works which may get multiple labels. We have kept only the dominant one.
- On Internet there exist part/details of some larger painting that are presented as stand-alone works. In all the cases that we were able to recognize, only the original, full-work was kept.
- Many works from the more recent period also contain digitized parts. As long as they have artistic value they were kept.

Following this editing process, it resulted a set of 18040 images divided into 18 art movement. The structure overview maybe followed in table 2.

The difficulties of automatic characterization may come from the following aspects:

- The quality of digitized images varies greatly: from high resolution to low-resolution damaged further by JPEG artifacts
- The aspect ratio vary greatly from 3:1 to 1:3 as illustrated in figure 1. Also some painting have a circular frame, the minimum bounding box was kept. Yet while in some cases background information was digitally removed, in some it exists.
- Following the short description from table 2, the main difference between various movements is more related to actual content; often differences are subtle. Thus is rather hard for standard image descriptors to accurately encode relevant information.

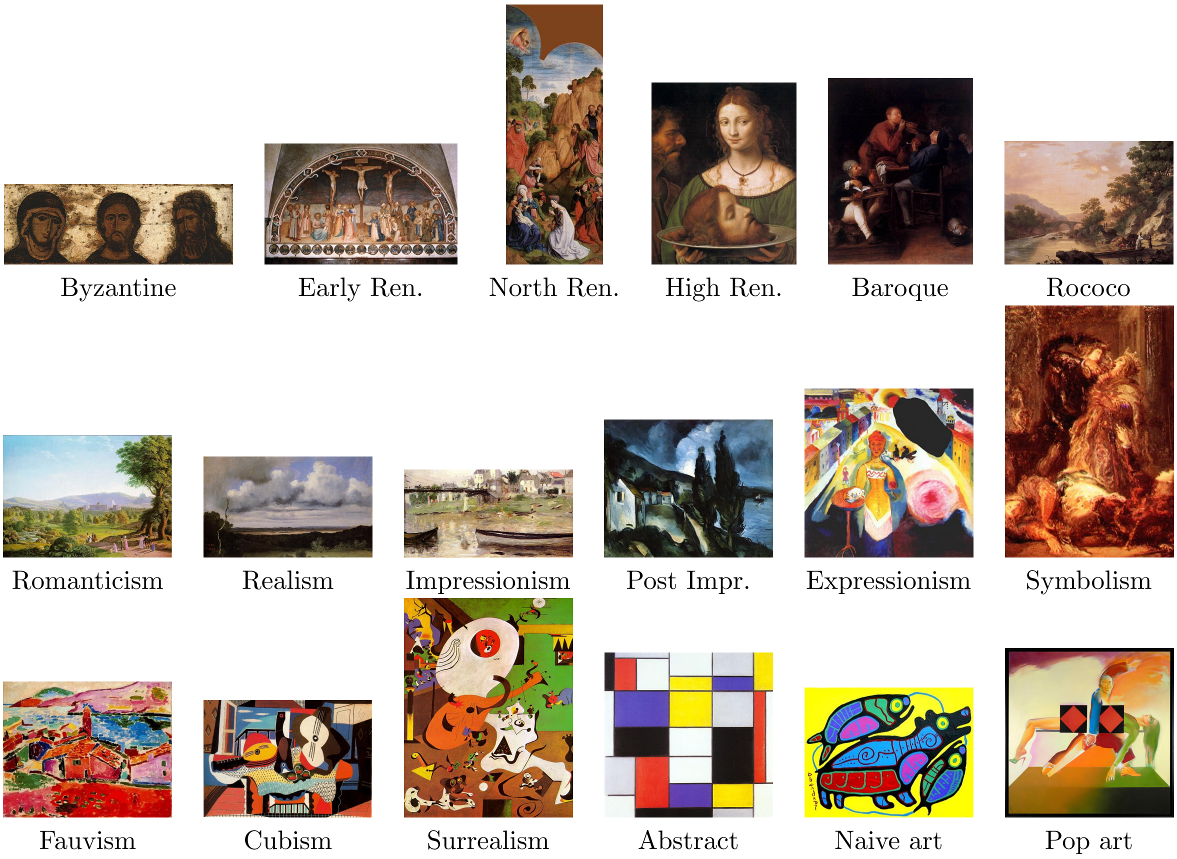

Figure 1: The 18 art movements illustrated in the proposed database.

4 Art movement recognition performance

4.1 Training and testing

To separate the database 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 being uniquely allocated into a fold. The same division was used for all further tests and it is part of the database.

4.2 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 that are able to separate any two art movements.

Prior works [3], [16] 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 mainly the thematic of 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 :

- **Histogram of oriented gradients** (HOG) [9] which computes the oriented gradient in each pixel and accumulates the weight of each orientation into a histogram. It has been previously used in painting analysis [16], [2].
- **Pyramidal HOG** (pHOG) the above mentioned HOG is implemented on 4 levels of a Gaussian pyramid.

- **Color HOG** - the above mentioned HOG descriptor applied on each color plane of the RGB color space.
- **Histogram of topographical features** (HoT) [12] which, supplementary to the gradient compute also 4 histograms derived from the local Hesian, thus quantifying local curvature.
- **Pyramidal HoT** (pHOG) the above mentioned HoT over 4 levels of a Gaussian pyramid.
- **Local Binary Pattern** (LPB) [21] is a histogram of quantized binary patterns pooled in a local image neighborhood of 3×3 and restrained to a total of 58 quantized non-uniform patterns. The LPB was used in painting description [16], [2].
- **Pyramidal LBP** (pLBP) - the above mentioned descriptor computed over 4 levels of a Gaussian pyramid.
- **Local Invariant Order Pattern** [27] - assume the order after sorting in the increasing intensity local samples.

  For HOG, LBP and LIOP we have relied on the implementation from the VLFeat library [25].
- **Edge Histogram Descriptor** (EHD) is part of the MPEG-7 standard. It accounts for the distribution of four basic gradient orientations within regular image parts. The implementation is based on BilVideo-7 library [4].
- **The spatial envelope, GIST** [22] describes the spatial character or shape of the painting and was previously used for painting categorization [2].

The color descriptors tested are:

- **Discriminative Color Names** (DCN) [17] - represents the dominant color retrieved through an information oriented approach. Here, we have used author provided code. The baseline form (Color Name) was successfully used to determine the style and the painter [16].
- **Color Structure Descriptor** (CSD) [19], which is based on color structure histogram, a generalization of the color histogram. The CSD accounts for some spatial coherence in the gross distribution of quantized colors within the image and it has been shown that is able to differentiate between various art movements [15]. We computed a 64 long CSD vector using the BilVideo-7 library [4].

Machine learning classification systems tested are:

- **Support Vector Machine**. We have relied for its implementation on the Lib-SVM [7]. We used on the radial basis function c-SVM with empirical found parameters.
- **Random forest** [6]. We have used 100 trees and unlimited depth. At each node we randomly look for a split in $N_1 = \sqrt{N}$ dimensions where N is the input feature dimension.

  Let us note that before the development of the deep networks the random forests and support vector machines have been found to be the most robust families of classifiers [11]. Also, for small and diverse databases SVM and RF out-compete deep networks.

Furthermore we have tested several systems that were previously used for art movement recognition. Inspired from previous work [3], we have run the Bag of Words (BoW) over SIFT keypoint detector with a vocabulary of 500. We have also tested a combination of color description, texture analysis based on Gabor filters and scene composition based on Gestalt frameworks [8].

Additionally, while the database is small for such a purpose and thus not really suited for deep learning, to have an indication of baseline performance, we have trained and evaluated a version of Deep Convolutional Neural Network (CNN). Our implementation is based on the MatConvNet [26]. Several alternatives were tested:

- Architecture:
  - LeNet (with 32×32 input);
  - LeNet with additional layers to permit increasing the size to 64×64 input;
  - Network in Network with 32×32 and 64×64 input patches.

Table 3: Recognition rates, [%] when various combinations of features and classifiers are used on the Pandora database. We marked with bold the best performance.

Feat. / Class.	Random Forest	SVM
HOG	18.4	17.4
pHOG	23.4	24.7
colorHOG	19.6	19.1
HoT	29.6	30.8
pHoT	32.3	37.2
LBP	27.2	27.4
pLBP	32.7	39.2
LIOP	24.4	25.2
EHD	24.9	22.7
GIST	23.8	23.5
DCN	18.9	19.4
CSD	29.4	31.6
pLBP + CSD	37.8	40.4
pHoT + CSD	37.7	**42.3**

Table 4: Recognition rates when various systems are used.

System	**Performance**
pHoT + CSD +SVM	**42.3**
BoW	25.2
Condorovici et al. [8]	22.78
Deep CNN	25.4

- Data formation. As the aspect ratio and original size varies greatly across the database, and deep CNN require standard images (square) we have experimented with several alternatives :
  - Resize, independently on width and height to get square image,
  - Resize, while keeping the aspect ratio. Thus the maximum size of the image has the required size
  - Resize while keeping the aspect ratio and extract patches of required size. This should have also the advantage to increase the database.

Yet overall, all these attempts had very little influence over the final recognition rate.

4.3 Results

We report first the results achieved when various combinations of features and classifiers are used (to be followed in table 3).

Secondly we report comparatively the best performance of aggregated systems in table 4. We note that for this particular database, the best performance is achieved by a standard combination of features (pyramidal HoT + Color Structure Descriptor) with a Support Vector Machine.

While one may find disappointing the performance of various established systems, this could be explainable. For the Bag of Words there is too much variability between keypoints to find a common ground; instead of the baseline version tested here, one should opt for much larger vocabularies with accurate compression to keep memory requirements low. Regarding the performance of the DeepCNN, the reported value hopefully is a lower boundary, as the database is too small for directly training nets with tens of thousands of variables, since no obvious data augmentation is reachable.

References

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

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

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

[4] M. Baştan, H. Çam, U. Güdükbay, and Özgür 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. *Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I*, chapter Classification of Artistic Styles Using Binarized Features Derived from a Deep Neural Network, pages 71–84. 2015.

[6] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[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. G. Condorovici, C. Florea, and C. Vertan. Automatically classifying paintings with perceptual inspired descriptors. *J. Vis. Commun. Image. Represent.*, 26:222 – 230, 2015.

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

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

[11] 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.

[12] C. Florea, L. Florea, and C. Vertan. *Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*, chapter Learning Pain from Emotion: Transferred HoT Data Representation for Pain Intensity Estimation, pages 778–790. 2015.

[13] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.

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

[15] Y.-F. Huang and C.-T. Wang. Classification of painting genres based on feature selection. In *Proc. of Multimedia and Ubiquitous Engineering, LNEE*, volume 308, pages 159–164, 2014.

[16] 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.

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

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

[19] 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.

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

[21] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, July 2002.

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

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

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

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

[26] A. Vedaldi and K. Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proc. of ACM MM*, pages 689–692, 2015.

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

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