Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Inferring Painting Style with Multi-task Dictionary Learning

Gaowen Liu[1], **Yan Yan**[1,2,*], **Elisa Ricci**[3], **Yi Yang**[4], **Yahong Han**[5], **Stefan Winkler**[2], **Nicu Sebe**[1]
[1]DISI, University of Trento, Italy [2]ADSC, UIUC, Singapore [3]Fondazione Bruno Kessler, Italy
[4]QCIS, University of Technology Sydney, Australia [5]Tianjin University, China
gaowen.liu@unitn.it, yan@disi.unitn.it, eliricci@fbk.eu, yee.i.yang@gmail.com,
yahong@tju.edu.cn, Stefan.Winkler@adsc.com.sg, sebe@disi.unitn.it

Abstract

Recent advances in imaging and multimedia technologies have paved the way for automatic analysis of visual art. Despite notable attempts, extracting relevant patterns from paintings is still a challenging task. Different painters, born in different periods and places, have been influenced by different schools of arts. However, each individual artist also has a unique signature, which is hard to detect with algorithms and objective features. In this paper we propose a novel dictionary learning approach to automatically uncover the artistic style from paintings. Specifically, we present a multi-task learning algorithm to learn a style-specific dictionary representation. Intuitively, our approach, by automatically decoupling style-specific and artist-specific patterns, is expected to be more accurate for retrieval and recognition tasks than generic methods. To demonstrate the effectiveness of our approach, we introduce the DART dataset, containing more than 1.5K images of paintings representative of different styles. Our extensive experimental evaluation shows that our approach significantly outperforms state-of-the-art methods.

1 Introduction

With the continuously growing amount of digitized art available on the web, classifying paintings into different categories, according to style, artist or based on the semantic contents, has become essential to properly manage huge collections. In addition, the widespread diffusion of mobile devices has led to an increased interest in the tourism industry for developing applications that automatically recognize the genre, the art movement, the artist, and the identity of paintings and provide relevant information to the visitors of museums.

Imaging and multimedia technologies have progressed substantially during the past decades, encouraging research on automatic analysis of visual art. Nowadays, art historians have even started to analyse art based on statistical techniques, *e.g.* for distinguishing authentic drawings from imitations [Hughes *et al.*, 2010]. However, despite notable attempts [Li *et al.*, 2012; Carneiro, 2013; Wang and Takatsuka, 2013; Karayev *et al.*, 2014], the automatic analysis of paintings is still a complex unsolved task, as it is influenced by many aspects, *i.e.* *low-level* features, such as color, texture, shading and stroke patterns, *mid-level* features, such as line styles, geometry and perspective, and *high-level* features, such as objects presence or scene composition.

Figure 1: Given the images belonging to the *Baroque*, *Renaissance*, *Impressionism*, *Cubism*, *Postimpressionism*, *Modern* art movements, can you detect which ones correspond to the same style[1]?

In this paper we investigate how to automatically infer the artistic style, *i.e.* *Baroque*, *Renaissance*, *Impressionism*, *Cubism*, *Postimpressionism* and *Modernism*, from paintings. According to Wikipedia, an artistic style is a "tendency with a specific common philosophy or goal, followed by a group of artists during a restricted period of time or, at least, with the heyday of the style defined within a number of years". Referring to paintings, the notion of *style* is more difficult to define than to perceive. Looking at Fig. 1, where images representative of six art movements are shown, can you guess which

*Corresponding Author. This work was supported in part by the MIUR Cluster project Active Ageing at Home, the EC project xLiMe, in part by A*STAR Singapore under the Human-Centered Cyber-physical Systems (HCCS) grant, in part by the Australian Research Council Discovery Projects, and in part by the NSFC (under Grant 61202166 and 61472276)

ones belong to the same style? At the first glance, it may not be hard to group these images into different styles, *i.e.* (1) and (9), (4) and (8), even if you have never seen these paintings before. Indeed, human observers can easily match artworks from the same style and discriminate those originated from different art movements, even if no a-priori information is provided. That is because humans *recognize the style* by implicitly using both low-level cues such as lines or colors and more subtle compositional patterns.

Recently, statistical methods have shown potential for supporting traditional approaches in the analysis of visual art by providing new, objective and quantifiable measures that assess the artistic style [Carneiro, 2013; Wang and Takatsuka, 2013; Karayev *et al.*, 2014]. In this paper we propose a dictionary learning approach for recognizing styles. Dictionary learning, which has proved to be highly effective in different computer vision and pattern recognition problems [Yang *et al.*, 2009; Elad and Aharon, 2006], is a class of unsupervised methods for learning sets of over-complete bases to represent data efficiently. The aim of dictionary learning is to find a set of basis vectors such that an input vector can be represented as a linear combination of the basis vectors. In this paper we propose a novel framework unifying multi-task and dictionary learning in order to simultaneously infer artist-specific and style-specific representations from a collection of paintings. Our intuition is that if we can build a style-specific dictionary representation by exploiting common patterns between artists of the same style with multi-task learning, more accurate results can be obtained for painting retrieval or recognition. For example, by automatically learning a dictionary for *Cubism* which captures the features associated to straight lines, we expect to easily detect that the paintings (1) and (9) in Fig.1 belong to the same category. Our experiments, conducted on the new DART (Dictionary ART) dataset, confirm our intuition and demonstrate that the learned dictionaries can be successfully used to recognize the artistic styles.

To summarize, the main contributions of this paper are: (i) We are the first to introduce the idea of learning style-specific dictionaries for automatic analysis of paintings. (ii) A novel multi-task dictionary learning approach is proposed through embedding all tasks into an optimal learned subspace. Our multi-task learning strategy permits to effectively separate artist-specific and style-specific patterns, improving recognition performances. The proposed machine learning framework is a generic one and can be easily applied to other problems. (iii) We collected the DART dataset which contains paintings from different art movements and different artists.

2 Related Work

2.1 Automatic Analysis of Paintings

In literature, [Cutzu *et al.*, 2005] were the first to borrow ideas from classification systems for automatic analysis of visual art and studied the differences between paintings and photographs. Image features such as edges, spatial variation of colors, number of unique colors, and pixel saturation were used for classification. [Li *et al.*, 2012] compared van Gogh with his contemporaries by statistical analysis of a massive set of automatically extracted brushstrokes. [Carneiro, 2013] introduced the problem of artistic image annotation and retrieval and proposed several solutions using graph-based learning techniques. [Wang and Takatsuka, 2013] proposed a SOM-based model for studying and visualizing the relationships among painting collections of different painters. [Yanulevskaya *et al.*, 2012] presented an analysis of the affective cues extracted from abstract paintings by looking at low-level features and employing a bag-of-visual-words approach. Few works focused specifically on inferring style from paintings [Shamir *et al.*, 2010; Karayev *et al.*, 2014]. However, none of these works have studied the problem of decoupling artist-specific and style-specific patterns as we do with our multi-task dictionary learning framework.

2.2 Dictionary and Multi-task Learning

Dictionary learning has been shown to be able to find succinct representations of stimuli. Recently, it has been successfully applied to a variety of problems in computer vision, pattern recognition and image processing, *e.g.* image classification [Yang *et al.*, 2009], denoising [Elad and Aharon, 2006]. Different optimization algorithms [Aharon *et al.*, 2006; Lee *et al.*, 2006] have also been proposed to solve dictionary learning problems. However, as far as we know, there is no research work on learning dictionary representations for recognizing artistic styles.

Multi-task learning [Argyriou *et al.*, 2007; Yan *et al.*, 2013; 2014] methods aim to simultaneously learn classification and regression models for a set of related tasks. This is typically advantageous as compared to considering single tasks separately and not exploiting their relationships. The goal of multi-task learning is to improve the performance by learning models for multiple tasks jointly. This works particularly well if these tasks have some commonality while are all slightly under-sampled. However, there is hardly any work on combining multi-task and dictionary learning problems. [Ruvolo and Eaton, 2014] developed an efficient online algorithm for dictionary learning from multiple consecutive tasks based on the K-SVD algorithm. Another notable exception is [Maurer *et al.*, 2013] where theoretical bounds are provided to study the generalization error of multi-task dictionary learning algorithms. [Yang *et al.*, 2010; Chang *et al.*, 2014; 2015] proposed different convex formulations for feature selection problems. These works are very different from ours, since we focus on a specific applicative scenario and propose a novel multi-task dictionary learning algorithm.

[1] Answers: Cubism (1,9), Impressionism (2,7), Postimpressionism (3,10), Renaissance (4,8), Baroque (6,11), Modern (7,12).

Names and authors of paintings: 1) Bottle and Fishes, *Braque*; 2) Bouquet of Sunflowers, *Monet*; 3) Portrait of the Postman Joseph Roulin, *van Gogh*; 4) Christ Falling on the Way to Calvary, *Raphael*; 5) The Disintegration of the Persistence of Memory, *Dali*; 6) The Adoration of the Golden Calf, *Poussin*; 7) Portrait of Claude Renoir Painting, *Renoir*; 8) Death of Actaeon, *Titian*; 9) Bananas, *Gris*; 10) Vegetation Tropicale, Martinique, *Gauguin*; 11) The Night Watch, *Rembrandt*; 12) Living Still Life, *Dali*.

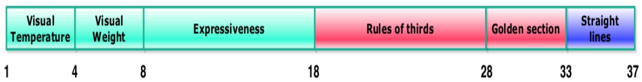

Figure 2: Extracted features: color (light blue), composition (red) and lines (blue).

3 Learning Style-specific Dictionaries

In this section we present our multi-task dictionary learning approach for inferring style-specific representations from paintings. In the following we first describe the chosen feature descriptors and then the proposed learning algorithm.

3.1 Feature Extraction from Paintings

Color, composition and brushstrokes are considered to be the three most important components in paintings. Therefore, to represent each painting, we construct a 37-dimensional feature vector as proposed in [Wang and Takatsuka, 2013], including color, composition and lines informations (Fig.2).

Color. Following [Wang and Takatsuka, 2013], the color features are computed as a function of luminance and hue. They are: (i) The visual temperature of color (the feel of warmth or coldness of color), as the wavelengths of the visible color light waves are considered to be related to the human perception of color temperatures. Different emotions can be expressed by using cold or warm color temperatures. (ii) The visual weight of color (the feel of heaviness of color). From the perspective of psychology, people usually feel that a darker color is heavier and a lighter color is lighter. (iii) The expressiveness of color (the degree of contrast including the contrast between luminance, saturation, hue, color temperature, and color weight). Global and local contrast features are both used to measure the differences between pixel and image regions.

Composition. The composition represents the spatial organization of visual elements in a painting. For each image we compute a saliency map. The saliency map is divided into three parts both horizontally and vertically and we consider the mean salience for each of the nine sections to compute the "rule of thirds". Additionally, properties of the most salient region such as size, elongation, rectangularity and the most salient point are used to represent properties of 'golden section' composition principles. In details, elongation measures the symmetricity along the principal axes, rectangularity measures how close it is to its minimum bounding rectangle, the most salient point is the global maximum of the saliency map.

Lines. Lines in paintings are generally perceived as edges. Different styles of paintings or different painters may favor a certain type of line. To interpret the concepts of lines, the Hough Transform is adopted to find straight lines that are above a certain threshold (longer than 10 pixels). The mean slope, mean length, and standard deviation of slopes of all the detected straight lines are calculated.

3.2 Multi-task Dictionary Learning

Intuitively, in this and in many other applications [Kong and Wang, 2012; Mairal *et al.*, 2008], it is reasonable to expect that more accurate recognition results are achieved if class specific dictionaries are adopted rather than generic ones. To this end, in this paper we demonstrate that better classification performance are obtained when we consider a style-specific dictionary for each artistic style. In details, we propose to jointly learn a set of artist-specific dictionaries and discover the underlying style-specific dictionary projecting data in a low dimensional subspace.

Algorithm 1: Learning artist-specific and style-specific dictionaries.

Input:
Samples $\mathbf{X}_1, ..., \mathbf{X}_k$ from K tasks
Subspace dimensionality s, dictionary size l, regularization parameters λ_1, λ_2.
Output:
Optimized $\mathbf{P} \in \mathbb{R}^{d\times s}, \mathbf{C}_k \in \mathbb{R}^{n_k \times l}, \mathbf{D}_k \in \mathbb{R}^{l\times d}, \mathbf{D} \in \mathbb{R}^{l\times s}$.
1: Initialize $\mathbf{P}$ using any orthonormal matrix
2: Initialize $\mathbf{C}_k$ with l_2 normalized columns
3: **repeat**
Compute $\mathbf{D}$ using Algorithm 2 in [Mairal *et al.*, 2009]
for $k = 1 : K$
Compute $\mathbf{D}_k$ using Algorithm 2 in [Mairal *et al.*, 2009]
Compute $\mathbf{C}_k$ using FISTA [Beck and Teboulle, 2009]
end for
Compute $\mathbf{P}$ by eigendecomposition of
$\mathbf{B} = \mathbf{X}'(\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}')\mathbf{X}$;
until *Convergence*;

More formally, for each painting style we consider K tasks and the k-th task corresponds to the k-th artist. Each task consists of data samples denoted by $\mathbf{X}_k = [\mathbf{x}_k^1, \mathbf{x}_k^2, ..., \mathbf{x}_k^{n_k}]$, $\mathbf{X}_k \in I\!R^{n_k \times d}, k = 1, ..., K$, where $\mathbf{x}_k^i \in I\!R^d$ is a d-dimensional feature vector and n_k is the number of samples in the k-th task. We propose to learn a shared subspace across all tasks, obtained by an orthonormal projection $\mathbf{P} \in I\!R^{d\times s}$, where s is the dimensionality of the subspace. In this learned subspace, the data distribution from all tasks should be similar to each other. Therefore, we can code all tasks together in the shared subspace and achieve better coding quality. The benefits of this strategy are: (i) We can improve each individual coding quality by transferring knowledge across all tasks. (ii) We can discover the relationship among different tasks (artists) via coding analysis. (iii) The common dictionary among tasks, *i.e.* the style-specific dictionary, can be learned by embedding all tasks into a good sharing subspace. These objectives can be realized solving the optimization problem:

$$\min_{\mathbf{D}_k,\mathbf{C}_k,\mathbf{P},\mathbf{D}} \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{C}_k\mathbf{D}_k\|_F^2 + \lambda_1 \sum_{k=1}^{K} \|\mathbf{C}_k\|_1 + \lambda_2 \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{P} - \mathbf{C}_k\mathbf{D}\|_F^2 \quad (1)$$

$$\text{s.t.} \quad \begin{cases} \mathbf{P}'\mathbf{P} = \mathbf{I} \\ (\mathbf{D}_k)_{j\cdot}(\mathbf{D}_k)'_{j\cdot} \le 1, \quad \forall j = 1, ..., l \\ \mathbf{D}_{j\cdot}\mathbf{D}'_{j\cdot} \le 1, \quad \forall j = 1, ..., l \end{cases}$$

where $\mathbf{D}_k \in I\!R^{l\times d}$ is an overcomplete (artist-specific) dictionary ($l > d$) with l prototypes of the k-th task, $(\mathbf{D}_k)_{j\cdot}$ in the constraints denotes the j-th row of $\mathbf{D}_k$, and $\mathbf{C}_k \in I\!R^{n_k \times l}$ corresponds to the sparse representation coefficients of $\mathbf{X}_k$. In the third term of Eq.1, $\mathbf{X}_k$ is projected by $\mathbf{P}$ into the subspace to explore the relationship among different tasks. $\mathbf{D} \in I\!R^{l \times s}$ is the (style-specific) dictionary learned in the tasks-shared subspace and $\mathbf{D}_{j\cdot}$ in the constraints denotes the j-th row of $\mathbf{D}$. Moreover, $\mathbf{I}$ is the identity matrix, $(\cdot)'$ denotes the transpose operator and λ_1 and λ_2 are regularization parameters. The first constraint guarantees the learned $\mathbf{P}$ to be orthonormal, and the second and third constraints prevent the learned dictionary to be arbitrarily large. In our objective function, we learn a dictionary $\mathbf{D}_k$ for each task k and one shared dictionary $\mathbf{D}$ among k tasks. When $\lambda_2 = 0$, Eq.1 reduces to the traditional dictionary learning on separated tasks. It is fundamental to underline the difference between $\mathbf{D}$ and $\mathbf{D}_k$: $\mathbf{D}$ is the learned style-specific dictionary and $\mathbf{D}_k$ is the dictionary associated the k-th artist in each style. In Eq.1, we share the same coefficient $\mathbf{C}_k$ in the global and in the task-specific reconstruction error terms. This is actually meant to enforce the coherence between artist-specific and style-specific dictionaries found in the low dimensional subspace.

Optimization

To solve the problem in Eq.1, we adopt an alternating optimization algorithm. The proposed algorithm is summarized in Algorithm 1. The source code for the optimization will be made available online. In details, we optimize with respect to $\mathbf{D}$, $\mathbf{D}_k$, $\mathbf{C}_k$ and $\mathbf{P}$ respectively in four steps as follows:

Step 1: Fixing $\mathbf{D}_k$, $\mathbf{C}_k$, $\mathbf{P}$, compute $\mathbf{D}$. Considering the matrices $\mathbf{X} = [\mathbf{X}_1', ..., \mathbf{X}_k']'$, $\mathbf{C} = [\mathbf{C}_1', ..., \mathbf{C}_k']'$, we obtain $\sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{P} - \mathbf{C}_k\mathbf{D}\|_F^2 = \|\mathbf{XP} - \mathbf{CD}\|_F^2$. Therefore Eq.1 is equivalent to:

$$\begin{aligned} \min_{\mathbf{D}} \quad & \|\mathbf{XP} - \mathbf{CD}\|_F^2 \\ \text{s.t.} \quad & \mathbf{D}_{j\cdot}\mathbf{D}_{j\cdot}' \leq 1, \quad \forall j = 1, ..., l \end{aligned}$$

This is equivalent to the dictionary update stage in traditional dictionary learning algorithms. We adopt the dictionary update strategy of Algorithm 2 in [Mairal *et al.*, 2009] to efficiently solve it.

Step 2: Fixing $\mathbf{D}$, $\mathbf{C}_k$, $\mathbf{P}$, compute $\mathbf{D}_k$. To compute $\mathbf{D}_k$ we solve:

$$\begin{aligned} \min_{\mathbf{D}_k} & \|\mathbf{X}_k - \mathbf{C}_k\mathbf{D}_k\|_F^2 \\ \text{s.t.} \quad & (\mathbf{D}_k)_{j\cdot}(\mathbf{D}_k)_{j\cdot}' \leq 1, \quad \forall j = 1, ..., l \end{aligned} \tag{2}$$

Similarly to Step 1, solving (2) corresponds to the update stage for dictionary learning in case of k tasks. Then, to compute $\mathbf{D}_k$ we also use the approach described in Algorithm 2 in [Mairal *et al.*, 2009].

Step 3: Fixing $\mathbf{D}_k$, $\mathbf{P}$, $\mathbf{D}$, compute $\mathbf{C}_k$. Eq.1 is equivalent to:

$$\begin{aligned} \min_{\mathbf{C}_k} \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{C}_k\mathbf{D}_k\|_F^2 + \lambda_1 \sum_{k=1}^{K} \|\mathbf{C}_k\|_1 \\ + \lambda_2 \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{P} - \mathbf{C}_k\mathbf{D}\|_F^2 \end{aligned}$$

This problem can be decoupled into $n' = n_1 + n_2 + ... + n_k$ distinct problems:

$$\min_{\mathbf{c}_k^i} \|\mathbf{x}_k^i - \mathbf{c}_k^i\mathbf{D}_k\|_2^2 + \lambda_1\|\mathbf{c}_k^i\|_1 + \lambda_2\|\mathbf{x}_k^i\mathbf{P} - \mathbf{c}_k^i\mathbf{D}\|_2^2 \tag{3}$$

We adopt the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [Beck and Teboulle, 2009] to solve the problems in Eq.3. FISTA solves the optimization problems in the form of $\min_{\boldsymbol{\mu}} f(\boldsymbol{\mu}) + r(\boldsymbol{\mu})$, where $f(\boldsymbol{\mu})$ is convex and smooth, and $r(\boldsymbol{\mu})$ is convex but non-smooth. We adopt FISTA since it is a popular tool for solving many convex smooth/non-smooth problems and its effectiveness has been verified in many applications. In our setting, we denote the smooth term part as $f(\mathbf{c}_k^i) = \|\mathbf{x}_k^i - \mathbf{c}_k^i\mathbf{D}_k\|_2^2 + \lambda_2\|\mathbf{x}_k^i\mathbf{P} - \mathbf{c}_k^i\mathbf{D}\|_2^2$ and the non-smooth term part as $g(\mathbf{c}_k^i) = \lambda_1\|\mathbf{c}_k^i\|_1$.

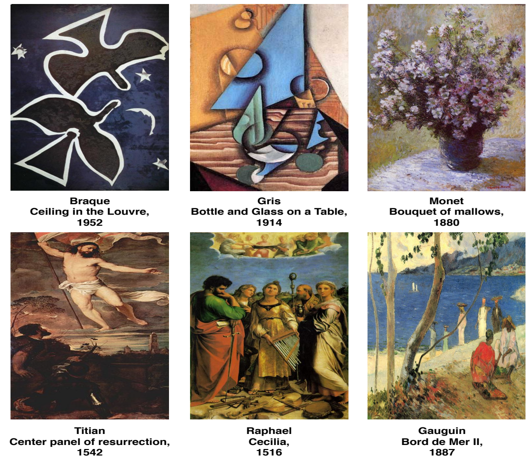

Figure 3: Examples of paintings from the DART dataset. Each image is associated with a detailed description containing year, artist and painting name.

Step 4: Fixing $\mathbf{D}_k$, $\mathbf{C}_k$, $\mathbf{D}$, compute $\mathbf{P}$. Considering $\mathbf{X} = [\mathbf{X}_1', ..., \mathbf{X}_k']'$, $\mathbf{C} = [\mathbf{C}_1', ..., \mathbf{C}_k']'$, we solve:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \|\mathbf{XP} - \mathbf{CD}\|_F^2 \\ \text{s.t.} \quad & \mathbf{P}'\mathbf{P} = \mathbf{I} \end{aligned} \tag{4}$$

Substituting $\mathbf{D} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{XP}$ back into the above function, we obtain:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \mathrm{tr}(\mathbf{P}'\mathbf{X}'(\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}')\mathbf{XP}) \\ \text{s.t.} \quad & \mathbf{P}'\mathbf{P} = \mathbf{I} \end{aligned}$$

The optimal $\mathbf{P}$ is composed of eigenvectors of the matrix $\mathbf{B} = \mathbf{X}'(\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}')\mathbf{X}$ corresponding to the s smallest eigenvalues.

After the optimized dictionaries are obtained for styles and artists, the final classification of a test image is based on computing its sparse coefficient and calculating the minimal reconstruction error, similarly to [Yang *et al.*, 2011; Mairal *et al.*, 2008].

4 Experimental Results

In this section we introduce the DART dataset and evaluate the effectiveness of our method.

4.1 Dataset

The DART dataset contains paintings collected from the web representing six different artistic styles, *i.e.*, *Baroque*, *Cubism*, *Impressionism*, *Postimpressionism*, *Renaissance* and *Modern*. Examples with a detailed description for artists,

Table 1: Structure of the DART dataset.

Artistic Style	Artists	# of paintings
Baroque	Rubens	60
	Rembrandt	104
	Poussin	117
Cubism	Braque	113
	Gris	119
	Picasso	62
Impressionism	Monet	108
	Renoir	109
	Manet	58
Post-impressionism	van Gogh	134
	Gauguin	136
	Odilon	69
Renaissance	Raphael	67
	Titian	92
	Bosch	46
	Caravaggio	59
Modern	Mondrian	60
	Frida	45
	Dali	58

Table 2: Comparison with baseline methods.

Methods	Average accuracy
SVM	0.564 ± 0.004
EN [Karayev *et al.*, 2014]	0.624 ± 0.007
AT-DL	0.595 ± 0.003
LLC [Wang *et al.*, 2010]	0.642 ± 0.003
GSMTL [Zhou *et al.*, 2012]	0.681 ± 0.010
DMMTL [Jalali *et al.*, 2010]	0.651 ± 0.005
RMTL [Chen *et al.*, 2011]	0.672 ± 0.006
Ours	$\mathbf{0.745 \pm 0.003}$

Table 3: Evaluation on different features combinations.

Features	Average accuracy
Raw Pixels	0.527 ± 0.004
Color	0.533 ± 0.002
Composition	0.571 ± 0.008
Lines	0.489 ± 0.003
Color + Composition	0.632 ± 0.005
Color + Lines	0.598 ± 0.004
Composition + Lines	0.675 ± 0.006
Color + Composition + Lines	$\mathbf{0.745 \pm 0.005}$

painting name and year as recorded in DART are shown in Fig.3. For each style the painting of at least three artists have been collected. As shown in Table 1, there are totally 1616 paintings in the DART dataset. There is a high variability in paintings as each artist typically developed different painting techniques and styles as time passed. Therefore, for each painter, we ensured that the selected artworks cover a wide range of techniques and subjects. We also ensured that the paintings are from different periods of the artist life. To the best of our knowledge, DART is the largest high quality art dataset available with paintings and associated descriptions so far.

4.2 Experimental Setup and Baselines

In our experiments we randomly split the dataset into two parts, half for training and half for testing. We repeated the experiments ten times. The average results and associated standard deviations are reported. We set the regularization parameters, the subspace dimensionality s and the dictionary size l with cross-validation.

We compare the proposed method with several state-of-the-art single-task dictionary learning and multi-task learning methods. Specifically we consider (1) *Support Vector Machine* (SVM); (2) *Elastic Net* (EN), as it is the classifier used for painting style analysis in [Karayev *et al.*, 2014]; (3) *Dictionary Learning by Aggregating Tasks* (AT-DL), *i.e.* performing single task dictionary learning by simply aggregating data from all tasks; (4) *Locality-constrained Linear Coding* (LLC) [Wang *et al.*, 2010], a method which uses the locality constraints to project each descriptor into its local-coordinate system and integrates the projected coordinates by max pooling to generate the final representation; (5) *Graph Structure Multi-Task Learning*[3] (GSMTL) [Zhou *et al.*, 2012], a state-of-the-art multi-task learning method imposing graph structure to exploit tasks relationship; (6) *Dirty Model Multi-Task Learning*[3] (DMMTL) [Jalali *et al.*, 2010], a multi-task learning algorithm based on ℓ_1/ℓ_q-norm regularization; (7) *Robust Multi-Task learning*[3] (RMTL) [Chen *et al.*, 2011], a multi-task learning approach which imposes a low rank structure capturing task-relatedness and detects outlier tasks.

4.3 Quantitative Evaluation

We conduct extensive experiments to evaluate the effectiveness of the proposed method in recognizing artistic styles. Table 2 compares our approach with different single-task dictionary learning and multi-task methods. From Table 2, the following observations can be made: (i) Our proposed style-specific dictionary learning method significantly outperforms generic single task methods such as SVM and EN. (ii) Multi-task learning approaches (GSMTL, DMMTL, RMTL) always perform better than single-task dictionary learning (AT-DL, LLC) since they consider the correlation among paintings of different artists with the same style. (iii) Our approach performs better than the other multi-task learning methods, due to its unique ability of combining multi-task and dictionary learning. By introducing style-specific dictionaries a more discriminative data representation is obtained.

Fig. 4(left) shows the confusion matrix obtained with the proposed method. *Cubism* achieves relative high recognition accuracies compared with other styles, which is reasonable since the paintings belonging to *Cubism* contain many "long lines" compared with other styles. This aspect is evident observing Fig. 1. Moreover, many *Impressionism* and *Postimpressionism* paintings are misclassified into the other class because these styles are more correlated. In the literature, *Postimpressionism* was influenced by *Impressionism*. Indeed, *Postimpressionism* was meant to extend *Impressionism*. The painters continued to use vivid colors and brushtrokes and focused on real-life subjects, but they were more interested to emphasize geometric forms, use unnatural colors and distort the original forms for more expressive effects.

We also evaluate our approach with respect to different parameters, namely the dictionary size l and the different subspace dimensionality s. Fig. 4(middle) shows that the proposed method achieves the best results when the dictionary size is 100 and the subspace dimensionality is 25. Too large

[3] http://www.public.asu.edu/~jye02/Software/MALSAR/

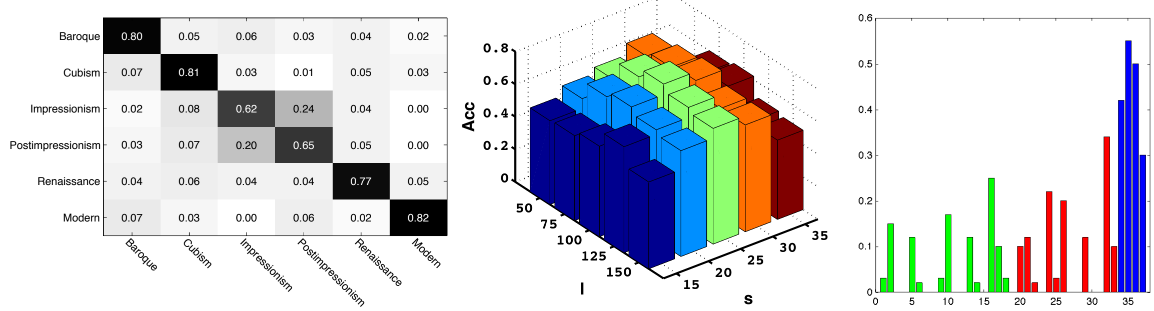

Figure 4: (Left) Confusion Matrix on DART dataset. (Middle) Performance at varying dictionary size l and subspace dimensionality s. (Right) Visualization of contributions of each component for the *Cubism* style. Different colors represent different components, *i.e.* color (green), composition (red) and lines (blue).

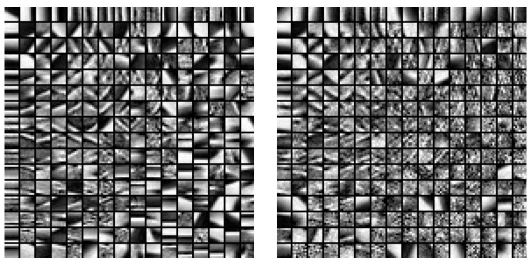

Figure 5: Visualization of learned dictionaries when using raw pixels features for (left) *Cubism* and (right) *Renaissance*.

or too small values for dictionary size and subspace dimensionality tend to decrease the performance. We also analyze the convergence of the proposed approach.

It is also interesting to investigate the contributions of each component (color, composition, and lines) for painting style classification. To evaluate this, we set the dictionary length equal to the dimensions of the feature vector and averaged the learned sparse codes for each style. Fig. 4(right) visualizes the contribution of each component for the *Cubism* style. We observe that the line features contribute the most to the recognition of the *Cubism* style. We also quantitatively evaluate the importance of different features on recognizing all styles as shown in Table 3. Raw pixels, color, composition, lines and their combinations are considered. Experimental results shows that using high-level features is advantageous with respect to simply using raw pixels. Moreover, combining all the heterogeneous features is greatly beneficial in terms of accuracy. While raw pixels are not appropriate for classification, to give a better idea of the output of our method, we use pixel values as features to learn the dictionary for each specific style. Fig. 5 visualizes the qualitative learned dictionaries for the *Cubism* and the *Renaissance* style, respectively. It is interesting to notice that the learned dictionaries share some similarity while many visual patterns are different. This clearly implies the necessity of learning style-specific dictionaries for paintings classification.

Finally, to further validate the proposed feature representation, we show a phylogenetic tree reflecting the similarities among artists (Fig. 6). The similarities are measured by euclidean distance among the average values of our feature

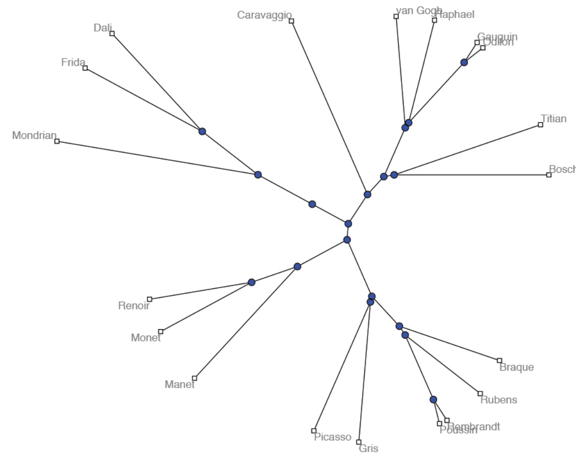

Figure 6: The phylogenetic tree reflecting the similarities among artists. (Figure is best viewed under zoom).

vectors. Then a hierarchical clustering algorithm is applied. We can clearly see that painting collections of the same artistic styles are much more similar to each other than painting collections of different art movements (*e.g.* Dali is clustered with Frida Kahlo and Mondrian rather than with Rubens or Picasso).

5 Conclusions

In this paper we investigated how to automatically infer painting styles from the perspective of dictionary learning and we proposed a novel multi-task dictionary learning approach to discover a low dimensional subspace where a style-specific dictionary representation can be computed. We conducted extensive experiments to evaluate our algorithm on the new DART dataset. Our results show that our style-specific approach performs significantly better than a generic one and that the proposed multi-task method achieves higher accuracy than state of the art dictionary learning algorithms.

References

[Aharon *et al.*, 2006] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TIP*, 54(11):4311–4322, 2006.

[Argyriou *et al.*, 2007] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Multi-task feature learning. In *NIPS*, 2007.

[Beck and Teboulle, 2009] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[Carneiro, 2013] Gustavo Carneiro. Artistic image analysis using graph-based learning approaches. *IEEE TIP*, 22(8):3168–3178, 2013.

[Chang *et al.*, 2014] Xiaojun Chang, Feiping Nie, Yi Yang, and Heng Huang. A convex formulation for semi-supervised multi-label feature selection. In *AAAI*, 2014.

[Chang *et al.*, 2015] Xiaojun Chang, Fieping Nie, Zhigang Ma, Yi Yang, and Xiaofang Zhou. A convex formulation for spectral shrunk clustering. In *AAAI*, 2015.

[Chen *et al.*, 2011] Jianhui Chen, Jiayu Zhou, and Jieping Ye. Integrating low-rank and group-sparse structures for robust multi-task learning. In *SIGKDD*, 2011.

[Cutzu *et al.*, 2005] Florin Cutzu, Riad Hammoud, and Alex Leykin. Distinguishing paintings from photographs. *CVIU*, 100:294–273, 2005.

[Elad and Aharon, 2006] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representation over learned dictionaries. *IEEE TIP*, 15(12):3736–3745, 2006.

[Hughes *et al.*, 2010] James M Hughes, Daniel J Graham, and Daniel N Rockmore. Quantification of artistic style through sparse coding analysis in the drawings of pieter bruegel the elder. *Proceedings of the National Academy of Sciences*, 107(4):1279–1283, 2010.

[Jalali *et al.*, 2010] Ali Jalali, Pradeep Ravikumar, Sujay Sanghavi, and Chao Ruan. A dirty model for multi-task learning. In *NIPS*, 2010.

[Karayev *et al.*, 2014] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style. In *BMVC*, 2014.

[Kong and Wang, 2012] Shu Kong and Donghui Wang. A dictionary learning approach for classification: separating the particularity and the commonality. In *ECCV*, 2012.

[Lee *et al.*, 2006] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. Efficient sparse coding algorithms. In *NIPS*, 2006.

[Li *et al.*, 2012] Jia Li, Lei Yao, Ella Hendriks, and James Z. Wang. Rhythmic brushstrokes distinguish van Gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE TPAMI*, 34(6):1159–1176, 2012.

[Mairal *et al.*, 2008] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Discriminative learned dictionaries for local image analysis. In *CVPR*, 2008.

[Mairal *et al.*, 2009] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *ICML*, 2009.

[Maurer *et al.*, 2013] Andreas Maurer, Massimiliano Pontil, and Bernardino Romera Paredes. Sparse coding for multitask and transfer learning. In *ICML*, 2013.

[Ruvolo and Eaton, 2014] Paul Ruvolo and Eric Eaton. Online multi-task learning via sparse dictionary optimization. In *AAAI*, 2014.

[Shamir *et al.*, 2010] Lior Shamir, Tomasz Macura, Nikita Orlov, D Mark Eckley, and Ilya G Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*, 7(2):8, 2010.

[Wang and Takatsuka, 2013] Ying Wang and Masahiro Takatsuka. SOM based artistic styles visualization. In *ICME*, 2013.

[Wang *et al.*, 2010] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2010.

[Yan *et al.*, 2013] Yan Yan, Elisa Ricci, Ramanathan Subramanian, Oswald Lanz, and Nicu Sebe. No matter where you are: Flexible graph-guided multi-task learning for multi-view head pose classification under target motion. In *ICCV*, 2013.

[Yan *et al.*, 2014] Yan Yan, Elisa Ricci, Ramanathan Subramanian, Gaowen Liu, and Nicu Sebe. Multi-task linear discriminant analysis for multi-view action recognition. *IEEE TIP*, 23(12):5599–5611, 2014.

[Yang *et al.*, 2009] Jianchao Yang, Kai Yu, Yihong Gong, and Thomas Huang. Linear spatial pyramid matching using sparse coding for image classification. In *CVPR*, 2009.

[Yang *et al.*, 2010] Yi Yang, Dong Xu, Feiping Nie, Shuicheng Yan, and Yueting Zhuang. Image clustering using local discriminant models and global integration. *IEEE TIP*, 19(10):2761–2773, 2010.

[Yang *et al.*, 2011] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, 2011.

[Yanulevskaya *et al.*, 2012] Victoria Yanulevskaya, Jasper Uijlings, Elia Bruni, Andreza Sartori, Elisa Zamboni, Francesca Bacci, David Melcher, and Nicu Sebe. In the eye of the beholder: Employing statistical analysis and eye tracking for analyzing abstract paintings. In *ACM MM*, 2012.

[Zhou *et al.*, 2012] Jiayu Zhou, Jianhui Chen, and Jieping Ye. Malsar: Multi-task learning via structural regularization. In *Technical Report of Arizona State University*, 2012.