Inferring Painting Style with Multi-task Dictionary Learning

Gaowen Liu\textsuperscript{1}, Yan Yan\textsuperscript{1,2,*}, Elisa Ricci\textsuperscript{3}, Yi Yang\textsuperscript{4}, Yahong Han\textsuperscript{5}, Stefan Winkler\textsuperscript{2}, Nicu Sebe\textsuperscript{1}

\textsuperscript{1}DISI, University of Trento, Italy \textsuperscript{2}ADSC, UIUC, Singapore \textsuperscript{3}Fondazione Bruno Kessler, Italy \textsuperscript{4}QCIS, University of Technology Sydney, Australia \textsuperscript{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, \textit{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.

In this paper we investigate how to automatically infer the artistic style, \textit{i.e.} Baroque, Renaissance, Impressionism, Cubism, Postimpressionism, Modern art movements, can you detect which ones correspond to the same style\textsuperscript{1}?

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{paintings.png}
\caption{Given the images belonging to the Baroque, Renaissance, Impressionism, Cubism, Postimpressionism, Modern art movements, can you detect which ones correspond to the same style?}
\end{figure}

\textsuperscript{*}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. [Yanulevskaia 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.

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

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 symmetry along the principal axes, rectangularity measures how close it is to its minimum bounding rectangle, the most salient point is used to measure the differences between pixel and image regions.

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.

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 \( X_k = [x_{1k}, x_{2k}, ..., x_{nk}] \), \( X_k \in \mathbb{R}^{n_k \times d} \), \( k = 1, ..., K \), where \( x_{ik} \in \mathbb{R}^d \) 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 \( P \in \mathbb{R}^{d \times s} \), 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:

\[
\min_{D_k, C_k, P} \sum_{k=1}^{K} \| X_k - C_k D_k \|_2^2 + \lambda_1 \sum_{k=1}^{K} \| C_k \|_1 + \lambda_2 \sum_{k=1}^{K} \| X_k P - C_k D_k \|_F
\]

s.t. \( \{ P'P = I \} \quad \text{for } j \leq l \), \( \forall j = 1, ..., l \)

\( \{ D_k'j \geq 1 \}, \forall j = 1, ..., l \)

where \( D_k \in \mathbb{R}^{l \times d} \) is an overcomplete (artist-specific) dictionary (\( l > d \)) with \( l \) prototypes of the k-th task, \( (D_k)_{j} \), in the constraints denotes the j-th row of \( D_k \), and \( C_k \in \mathbb{R}^{n_k \times l} \) corresponds to the sparse representation coefficients of \( X_k \). In
the third term of Eq.1, $X_k$ is projected by $P$ into the subspace to explore the relationship among different tasks. $D \in \mathbb{R}^{l \times s}$ is the (style-specific) dictionary learned in the tasks-shared subspace and $D_j$, in the constraints denotes the $j$-th row of $D$. Moreover, $I$ is the identity matrix, $()'$ denotes the transpose operator and $\lambda_1$ and $\lambda_2$ are regularization parameters. The first constraint guarantees the learned $P$ 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 $D_k$ for each task $k$ and one shared dictionary $D$ 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 $D$ and $D_k$: $D$ is the learned style-specific dictionary and $D_k$ is the dictionary associated the $k$-th artist in each style. In Eq.1, we share the same coefficient $C_k$ 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 $D$, $D_k$, $C_k$ and $P$ respectively in four steps as follows:

**Step 1:** Fixing $D_k$, $C_k$, $P$, compute $D$. Considering the matrices $X = [X_1', ..., X'_K']$, $C = [C_1', ..., C_K']$, we obtain

$$\sum_{k=1}^K \|X_kP - C_kD\|^2_F = \|XP - CD\|^2_F.$$ 

Therefore Eq.1 is equivalent to:

$$\begin{align*}
\min_D & \quad \|XP - CD\|^2_F, \\
\text{s.t.} & \quad D_jD_j' \leq 1, \quad \forall j = 1, ..., l
\end{align*}$$

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 $D$, $C_k$, $P$, compute $D_k$. To compute $D_k$ we solve:

$$\begin{align*}
\min_{D_k} & \quad \|X_k - C_kD_k\|^2_F, \\
\text{s.t.} & \quad (D_k)_j(D_k)_j' \leq 1, \quad \forall j = 1, ..., l
\end{align*}$$

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

**Step 3:** Fixing $D_k$, $P$, $D$, compute $C_k$. Eq.1 is equivalent to:

$$\begin{align*}
\min_{C_k} & \quad \sum_{k=1}^K \|X_k - C_kD_k\|^2_F + \lambda_1 \sum_{k=1}^K \|C_k\|_1 \\
& \quad + \lambda_2 \sum_{k=1}^K \|X_kP - C_kD\|^2_F.
\end{align*}$$

This problem can be decoupled into $n'$ distinct problems:

$$\begin{align*}
\min_{c_k} & \quad \|x_k - c_kD_k\|^2_F + \lambda_1 \|c_k\|_1 + \lambda_2 \|x_kP - c_kD\|^2_F
\end{align*}$$

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 $\min_{x_k} f(x) + r(x)$, where $f(x)$ is convex and smooth, and $r(x)$ 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 $f(c_k) = \|x_k - c_kD_k\|^2_F + \lambda_2 \|x_kP - c_kD\|^2_F$ and the non-smooth term part as $g(c_k) = \lambda_1 \|c_k\|_1$. 

**Step 4:** Fixing $D_k$, $C_k$, $D$, compute $P$. Considering $X = [X_1', ..., X'_K']$, $C = [C_1', ..., C_K']$, we solve:

$$\begin{align*}
\min_P & \quad \|XP - CD\|^2_F, \\
\text{s.t.} & \quad PP' = I
\end{align*}$$

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

$$\begin{align*}
\min_P & \quad \text{tr}(P'X'(I - C'(C'P')^{-1}C')XP) \\
\text{s.t.} & \quad PP' = I
\end{align*}$$

The optimal $P$ is composed of eigenvectors of the matrix $B = X'(I - C'(C'P')^{-1}C')X$ 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 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,
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 \( \ell_1/\ell_q \)-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 outliers 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 brushstrokes 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 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

---

\(^3\) http://www.public.asu.edu/~jye02/Software/MALSAR/
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 show 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 vectors. Then a hierarchical clustering algorithm is applied. We can clearly see that 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


