Published in final edited form as:

J Cult Herit. 2018 ; 31: 13–23. doi:10.1016/j.culher.2017.11.008.

Predicting and Grouping Digitized Paintings by Style using Unsupervised Feature Learning

Eren Gultepe,

Department of Electrical and Computer Engineering, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Thomas E. Conturo, and

Departments of Radiology, Physics, and Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

Masoud Makrehchi

Department of Electrical and Computer Engineering, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abstract

Objective—To create an system to aid in the analysis of art history by classifying and grouping digitized paintings based on stylistic features automatically learned without prior knowledge.

Material and Methods—6,776 digitized paintings from 8 different artistic styles (Art Nouveau, Baroque, Expressionism, Impressionism, Realism, Romanticism, Renaissance, and Post-Impressionism) were utilized to classify (predict) and cluster (group) paintings according to style. The method of unsupervised feature learning with K-means (UFLK), inspired by deep learning, was utilized to extract features from the paintings. These features were then used in: (1) a support vector machine algorithm to classify the style of new test paintings based on a training set of paintings having known style labels; and (2) a spectral clustering algorithm to group the paintings into distinct style groups (anonymously, without employing any known style labels). Classification performance was determined by accuracy and F-score. Clustering performance was determined by: 1) the ability to recover the original stylistic groupings (using a cost analysis of all possible combinations of 8 group label assignments); 2) F-score; and 3) a reliability analysis. The latter analysis used two novel ways to determine the distribution of the null-hypothesis: (a) a uniform distribution projected onto the principal components of the original data; and (b) a randomized, weighted adjacency matrix. The ability to gain insights into art was tested by a semantic analysis of the clustering results. For this purpose, we represented the featural characteristics of each painting by an N-dimensional feature vector, and plotted the distance between vector endpoints (i.e., similarity between paintings). Then, we color-coded the endpoints with the assigned lowest-

Correspondence to: Eren Gultepe.

Data sharing statement: All source code and analysis files will be available on Mendeley Data in the on-line journal.”

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cost style labels. The scatterplot was visually inspected for separation of the paintings, where the amount of separation between color clusters provides semantic information on the interrelatedness between styles.

Results—The UFLK-extracted features resembled the edges/lines/colors in the paintings. For feature-based classification of paintings, the macro-averaged F-score was 0.469. Classification accuracy and F-score were similar/higher compared to other classification methods using more complex feature learning models (e.g., convolutional neural networks, a supervised algorithm). The clustering via UFLK-extracted features yielded 8 unlabeled style groupings. In 6 of 8 clusters, the most common true painting style matched the cluster style assigned by cost analysis. The clustering had an F-score of 0.212. (There are no comparison methods for clustering paintings.) For the semantic analysis, the featural characteristics of Baroque and Art Nouveau were found to be similar, indicating a relationship between these styles.

Discussion/Conclusion—The UFLK method can extract features from digitized paintings. We were able to extract characteristics of art without any prior information about the nature of the features or the stylistic designation of the paintings. The methods herein may provide art researchers with the latest computational techniques for the documentation, interpretation, and forensics of art. The tools could assist the preservation of culturally sensitive works of art for future generations, and provide new insights into works of art and the artists who created them.

Keywords

painting styles; clustering; classification; unsupervised feature learning; k-means

1. INTRODUCTION

Artistic visual style is typically defined after the art has been created, and the style must be a significant break from other styles (1). However, the definitions of such styles are unclear since styles may overlap, or a painter may have multiple styles which causes difficulties in stylistic recognition (2). Experience in observing art objects and the ability to focus on technique despite the subject matter are skills necessary to differentiate styles of various artists (3). Furthermore, a concrete requirement for defining visual style does not exist, but some visual cues from paintings may be utilized, such as color palette, composition, scene, lighting, contours, and brush strokes (4).

Recently, computer vision has provided the opportunity to analyze paintings via advanced computational methods (5,6). These computational methods could augment the knowledge and abilities of artists, scholars, and curators (7) in the same way microscopes aid biologists (8). For this purpose, we present computational techniques based upon recent feature extraction algorithms for developing an automatic system of extracting the salient features of art. This is independent of any prior knowledge regarding the features and stylistic content of the art. Furthermore, we demonstrate methods to evaluate the grouping of paintings, which augment the analyses of art researchers. We implemented these methods on a large dataset of digitized paintings, which spanned various artistic movements. The methods presented in the study will allow researchers to use the latest computational techniques to

preserve, document, and provide new insights into the stylistic development of different artists and their stylistic movements.

There have been many studies focusing on the utility of computational techniques to analyze works of art. For a detailed review of the types of computational methods and analyses performed, such as on the brushstrokes, craquelure, or composition, please refer to (8,9). The number of studies which pertain to the classification of artistic style or art movements that involve large groups of paintings are minimal compared to the number of studies for specific works created by individual artists (1). To identify painting styles, previous studies extracted a variety of features to enable characterization of digitized paintings. Zujovic et al. (10) classified the five art movements of Abstract Expressionism, Cubism, Impressionism, Pop Art, and Realism. The features employed were edge maps for gray level and histograms for color, which were obtained from the hue, saturation, and value space. Gunsell et al. (11) completed similar studies except they proposed that luminance and the corresponding color features of images were more suitable. Spehr et al. (12) utilized 200 features to cluster and classify paintings from eight styles. The major features in their study consisted of color distributions based upon pixel histograms and contextual and semantic features derived from template matching and face detection. Lombardi et al. (13) deployed the k-nearest neighbor classifier with features, which captured light, line, and color to classify paintings.

Shamir et al. (14), developed and employed a feature extraction toolbox containing 4,027 image descriptors (15) to classify paintings from Impressionism, Expressionism, and Surrealism, but for only nine artists. Culjak et al. (16) employed features similar to Zujovic et al. (10) and Gunsell et al. (11), but for styles which were difficult to classify, such as Fauvism and Naïve art. Condorovici et al. (17) used features for six artistic styles of paintings which considered lightness perception, shape extraction, color distribution, texture, and edge analysis. Nonetheless, all of the aforementioned feature-based studies of style used handcrafted features, i.e., they were created via specific mathematical models and definitions, such as Gabor filters and statistical measures, which were decided by the experimenter empirically, prior to classification or clustering. To overcome the limitation in which prior knowledge is necessary to create features for paintings, deep neural networks and other similar techniques may be utilized.

Deep learning methods, which are based on traditional neural network techniques, contain many layers of hidden units (18) in contrast to traditional neural networks. These deep layers allow deep learning (also known as deep neural networks) of many complex features, which subsequently allow learning of intricate relationships within the data (19). Until recently, such deep learning models had limited implementations due to the lack of computational power and numerical techniques necessary for solving the numerous model parameters (20,21). However, the learning ability of deep neural network techniques has overcome the need to design (handcraft) classification features, as shown in different domains such as image classification and speech recognition (21).

Karayev et al. (22) were the first to utilize a deep learning algorithm to classify painting styles. They previously utilized a very deep version of Convolutional Neural Networks (CNNs) (23), which are feed-forward networks specialized for images (24). They trained

their CNN on paintings from the WikiArt gallery (available at www.wikiart.org). Bar et al. (1) also utilized CNNs and the same dataset to classify painting style, but combined them with “Picture Codes” features (25) which are descriptive image features extracted and compressed into binarized representations. Saleh and Elgammel (26) utilized the same CNN model and dataset as in (1,22), but their model was a combination of semantic-based features and CNN features, and further employing a feature selection framework. Although Karayev et al. (22), Bar et al. (1), and Saleh and Elgammel (26) achieved satisfactory classification accuracy, they did not analyze the features extracted from CNNs. Instead, their sole focus was improving style classification accuracy.

Since CNNs are supervised feature-learning algorithms (27,28), they require ground-truth labels (e.g., the known artistic movement/style of each painting) during the training phase of the algorithm. This requirement for prior knowledge leads what is termed the “grouping problem of paintings” (29), in which a naïve person is unable to cluster paintings with similar styles without expert knowledge of the art field. The grouping problem is due to the naïve person’s inability to provide the necessary prior style information during the training phase of deep neural networks such as CNNs (30). In our work, we overcame this limitation of CNNs in feature learning and artistic style clustering and classification by using unsupervised feature learning.

Unsupervised feature learning is an umbrella term describing recently developed feature extraction methods, which do not employ labeled data during model training. Examples of unsupervised feature learning are the deep learning methods of Restricted Boltzmann Machine (RBM) neural networks and Autoencoders (30). Specifically, an RBM neural network is composed of many stacked RBM units, with each unit composed of a “visible” input layer and a “hidden” feature-learning layer (18). An Autoencoder is composed of successively smaller stacked RBM units, which reduce the dimensionality by extracting the most important features for a specific dataset. Autoencoders and other neural network-like algorithms can be considered as a nonlinear generalization of principal components analysis (PCA) (18). The nonlinear aspect of neural networks refers to their ability to combine features in an intricate way, which cannot be performed by simple matrix multiplication.

A disadvantage of deep neural networks such as RBMs and Autoencoders is their complexity during the process of unsupervised feature learning. Recently, to reduce complexity, the traditional K-means clustering algorithm was utilized instead for the feature-learning part of the algorithm (31), which the authors applied to classifying photographs of objects. We implemented this procedure for feature extraction and style classification in paintings. In the study of (31), substitution of the K-means algorithm in place of RBMs and Autoencoders was found to greatly reduce the number of parameters needed to train an effective model, thus reducing training time. To perform this substitution, specialized transformations of the input data were found to be needed, such as selecting small random image patches from the training dataset and pre-whitening these image patches prior to the feature extraction step.

2. RESEARCH AIM

We seek to create an objective system to capture the overarching stylistic nature of paintings in different artistic movements by minimizing the expert knowledge required to characterize art. Thus, we use the deep learning-inspired method of unsupervised feature learning with K-means (UFLK) as the feature learner to extract features from digitized paintings, without any a-priori information regarding the nature of the extracted features. Furthermore, the features extracted by UFLK are learned independent of any labeled data, which for this study is the style category. If a new style were to develop, the algorithm would be able to learn the features without requiring any knowledge of its style. This is an important capability, since it can overcome any limitation arising from wrongly categorized paintings. Another key aspect of this study is that we also visualize the features learned by the UFLK algorithm to determine if the features resemble any basic painting techniques. Also, these features will be input into separate clustering and classification algorithms for the discrimination of painting styles. Specifically, with the UFLK features, we employed the linear classifier of support vector machines (SVM) to predict painting style and spectral clustering to solve the grouping problem of paintings.

3. MATERIALS AND METHODS

3.1 Overview of Processing Pipeline

The overall processing pipeline consists of three primary steps. (1) The first step is to preprocess all images of the paintings by downsampling to the desired pixel dimensions. (2) In the second step, the type of feature extraction that will be performed on the images is chosen. For comparison of the efficacy of different feature extraction algorithms in painting style classification and clustering, three types of features were used for analysis in this study. At this stage, one of three feature extraction processing steps can be chosen. The first case is when no feature extraction is performed, which means the raw pixels of the images are used as-is after the previous preprocessing step. The second case is that feature extraction is performed by the UFLK algorithm on the preprocessed images from step 1. In the third case, PCA feature extraction is performed on the preprocessed images from step 1. (3) Once the type of features to be used in classification and clustering have been decided, the third step is to employ the extracted features (if UFLK or PCA was performed) or the raw pixels (if no feature extraction is utilized) in the SVM algorithm for classification or the spectral clustering algorithm for clustering (Figure 1).

3.2 Dataset of Digitized Paintings and Image Preprocessing

For this study, we use the images from the digitized painting gallery database from abcgallery.com to extract paintings from the following eight styles: Art Nouveau, Baroque, Expressionism, Impressionism, Realism, Romanticism, Renaissance, and Post-Impressionism (Figure 2). For Realism, we refer to the artistic movement, which started after the French revolution and rejected the exotic themes of the Romantic style (32). These eight styles were chosen since they provide satisfactory coverage of the stylistic development of paintings in recent history. Each style category consisted of 847 randomly chosen paintings, providing 6,776 paintings in total. To classify the paintings for style, the

images are divided into the train and test datasets. Since the dataset is sufficiently large, an 80/20% split was performed by randomly selecting paintings into the train/test datasets (28), giving 5,421 and 1,355 paintings for training and testing cases, respectively. Clustering was performed using the whole dataset without any division of the data into train/test cases, since the style labels are not used to perform clustering.

Typically, for unsupervised feature learning and deep neural network learning, images are considered on the scale of 32×32 pixels (30,31,33,34), due to hardware and computation complexity. This scaling is applicable to grayscale or color images with RGB channels. Therefore, to facilitate the use of common unsupervised feature learning and deep learning algorithms, each painting was downsampled to 64×49 pixel dimensions using a bicubic interpolation. The average ratio between the width and height was ~ 1.3 among the paintings, which was determined from the mean of the histogram of the paintings dimensions. If a painting was taller than it was wide, it was rotated and then scaled accordingly.

3.3 Raw Pixel Data

The simplest representation of the digitized paintings is using the raw pixels obtained from the images. Each digitized painting is represented by a RGB (red, green, blue) color image having spatial dimensions of 64×49 pixels indicating the width $\times$ height of the painting, respectively. Each color channel of the images were reshaped into single vector with a dimension of $64 \times 49 = 3136$ pixels. Then, the vectors representing the three color channels were concatenated into one large vector with dimension $\mathbb{R}^{1 \times 9408}$, where $9408 = 3136 \times 3$. If no further feature extraction is to be performed, then this single vector would represent each image in the painting dataset and classification and clustering would be performed accordingly.

3.4 Nonlinear vs Linear Extraction of Features

PCA and similar implementations discover linear combinations of features that are extracted from the input data (35,36). This essentially means that input data can be represented by simply scaling and summing the newly extracted features, as demonstrated in subsection 5.5. This is also known as a linear mapping of the original feature space data into the newly extracted feature space (36). Features obtained by such algorithms can also be considered as being “linearly extracted” or obtained by a “linear feature extraction” method (36). Neural network algorithms find a nonlinear combination of features (18,36). In this case, the extracted features cannot simply be scaled and summed together to represent the data. This is also known as a nonlinear mapping of the original features of the data into the new feature space (36). We can consider these features being “nonlinearly extracted” or obtained from a “nonlinear feature extraction” method (36). As a result, neural networks can extract more complicated features because they are able to determine how complex features can be combined together to provide a meaningful explanation of the original input data.

3.5 Linear Extraction of Features Using PCA

PCA is a linear method of feature extraction and data compression (35). PCA is not a component of our nonlinear feature extraction approach, but we performed PCA for comparison. We performed PCA on the image dataset by using singular value

decomposition. For PCA, 100, 300, 500, and 1000 features were extracted to see the effect of varying the number of features. The PCA calculation of the painting images is very similar to the “eigenface” technique by Turk and Pentland (37).

3.6 Nonlinear Extraction of Features Using Unsupervised Feature Learning

Typically, unsupervised feature learning is used to learn a feature space representation from unlabeled data (38), rather than hand-craft or engineer features using expert or domain knowledge. The latter can be inherently difficult. However, some of the advanced learning algorithms, such as sparse restricted Boltzmann machines (RBMs) and sparse auto-encoders (SAEs), have many hyper-parameters and layers that need to be tuned, which can make training difficult and can severely affect performance.

For our study, to learn and nonlinearly extract features from the digitized paintings without using any information regarding the style of the art, we employed the UFLK algorithm, which is a fast and effective alternative to the aforementioned deep unsupervised learning algorithms (38). We alter the number of features learned and extracted in order to gauge its effect on classification and grouping of paintings. We extracted 500, 1000, 2000, and 3000 features with the UFLK algorithm.

It is worthwhile to compare the method herein to a current method commonly used for visual analysis. We choose to compare our method to the Principal component analysis (PCA) method because the PCA method is often used for dimensionality reduction, data compression, feature extraction, and visualization (39). The UFLK method is a very different algorithm compared to the PCA method, but this comparison may help put the two methods into context.

3.7 Classification

Using the training image features and the style labels, a linear classifier was trained using the linear L2-regularized SVM classifier with $C = 1$. For the test images, in the case of feature extraction, the feature mapping from the training images was simply re-applied. In the case of no feature extraction, the raw pixels of the images were employed as the features. Then, these test features with the style labels for the test images were input into an SVM classifier developed on the training data.

3.8 Clustering

After applying one of the three feature processing steps (raw pixels, PCA, or UFLK), painting styles were grouped using spectral clustering. Spectral clustering is preferred to other agglomerative or divisive clustering methods since it is able find non-spherical groupings whereas the other methods fail to do so (28). For this study, we use the normalized spectral clustering of Ng, Jordan and Weiss (40).

The style labeling of the final grouping of the paintings provided by the spectral clustering is blind to the actual painting style labels, hence the correspondence between painting style labels and clusters must first be established. To overcome the issue of cluster label assignment (41), the Hungarian method (42), a combinatorial optimization of the assignment

problem, was used to determine the lowest cost matching of the spectral clustering labels with the painting style labels.

3.9 Performance Evaluation

The outcome of interest is the positive (target) class, therefore it is preferable to use metrics such as precision (the ratio of clustered paintings that actually belong to the true style) and recall (also known as sensitivity; the ratio of paintings in the true style category that have been selected together). The F-score was used as the metric of interest as it is an unbiased combination (harmonic mean) of the precision and recall measures. The best-performing painting style or feature extraction method was defined by having the highest F-score value. Furthermore, to summarize the performance for a feature selection method, we calculated the average F-score of the eight styles to create the macro-averaged F-score (F-mac). Sometimes, to compare our results to other studies where the F-scores are not reported, we use the accuracy measure.

3.10 Clustering Null Distribution

To determine whether the spectral clustering of the paintings has grouped the paintings into meaningful stylistic clusters by virtue of the extracted features and not because of some random occurrence of the underlying properties of the data, the clustering performance was compared against a reference null distribution (43).

We implemented two different random data-generation schemes. The first method generated new data by using a uniform distribution projected onto the principal components of the original painting data, which has the benefit of preserving the shape and local structure of the paintings image data distribution (44). This was performed prior to inputting the painting pixel data into the spectral clustering algorithm. In the second method, we generated new data by randomizing the weighted adjacency matrix, which has the effect of preserving the magnitude of similarity among the pixels by assigning similarities to random image pairs. This generation method is implemented immediately after the first step of the spectral clustering algorithm.

The implementation of the above algorithms are detailed in the Supplementary Material linked to the on-line paper. To provide further details, transparency, and ability to replicate, the source code, input data, and output files used for this study are also deposited to Mendeley Data. The output files enable the users to confirm their implementation of the code. The comments in the source code give further details of the algorithm. A tutorial (README.txt file) is also included to aid the user in replicating the results. The README.txt file also describes the computer hardware/software environment and computation times for this study.

4. RESULTS

4.1 Feature Extraction

In Figure 3, we have visualized the nonlinear and linear extraction of features by the UFLK and PCA algorithms, respectively. We see that the UFLK features resemble lines, edges and

solid blocks of colors, whereas the PCA features are more similar to undefined blobs, with not much color differentiation. The UFLK algorithm has also extracted features that resemble Gabor filters, which have been used by Berezhnoy et al. (45) in the analysis of Van Gogh's paintings. Furthermore, when we visualize the representation of a specific painting, such as Gauguin's *Scene From Tahitian Life*, in UFLK feature space, we see that specific aspects of the paintings are activated. For instance, in some of the units in the feature space, it appears that the clothing of the girls is highlighted. However, in the PCA feature space, there is no discernable representation of artistic features. Our results in which we reconstructed the painting *Scene from Tahitian Life* from the PCA features indicates that many details of the original image are lost. Unfortunately, image reconstruction is not possible with UFLK features because they were extracted in a nonlinear way, as is the case with other nonlinear feature extraction methods (46).

4.2 Classification

For the classification of the eight different painting styles (Table 1), 3000 UFLK features provided the best overall style prediction performance with an F-mac of 0.469. Whereas the best PCA feature extraction (with 300 components) and the raw image pixels had an F-mac score of 0.278 and 0.288, respectively. All of these performances were better than a naive guess performance F-score of 0.167, especially the UFLK with 3000 features. This demonstrates that the UFLK with 3000 features performed 68.6% and 62.9% better than the PCA with 300 features and raw image features, respectively. In the confusion matrices for the UFLK 3000 features, the maximum painting counts were correctly attributed to all the painting categories (Figure 4a). The confusion matrices for the PCA with 300 features (Figure 4b) and the raw pixels (Figure 4c), also indicated all styles were correctly assigned the maximum painting counts. However, for classification with 3000 UFLK features, the maximum paintings counts are much higher than the PCA 300 and raw pixel features, as indicated by the higher recall rates in Table 1.

Furthermore, we compared the unsupervised learning feature classification to state-of-the-art convolutional neural networks. First, we used the popular AlexNet (23) and the LeNet (24) classification models which provided an accuracy of 0.326 and 0.401, respectively, which were both lower than the UFLK algorithm with 3000 features (accuracy = 0.475). Across the eight painting styles, the Baroque style had the highest F-score ($M = 0.512$, $SD = 0.128$) and Impressionism had the lowest ($M = 0.281$, $SD = 0.061$). A statistical analysis of the classification F-mac scores is in Figs. S1–S2 in the Supplementary Material.

4.3 Clustering

The effect of linear PCA, nonlinear UFLK, and absence of feature extraction on the clustering performance of paintings based on style is displayed in Table 2. Following feature extraction with UFLK and PCA, the best overall spectral clustering performance provided by feature extraction method was found using the UFLK method with 500 features (F-mac = 0.212). The best PCA based clustering performance (F-mac = 0.185) was demonstrated with only 100 principal components. Clustering only with the raw image painting pixels provided an F-mac of 0.197, which was better than using PCA for feature extraction. This indicates

that the best UFLK features performed 14.37% and 7.57% better than PCA features and raw pixels, respectively, for painting style clustering.

In Figure 5, the best performing clustering of the UFLK and PCA feature extraction method, along with the raw pixel features, are visualized as a scatterplot, where it is seen that the UFLK has more cohesive and disjointed groupings of paintings compared to the PCA based groups, which have overlapping and diffuse clusters. Also, the groups based on the UFLK features are organized in a way which suggests some semantic relationships among different art movements, whereas the raw pixel features have a linear organization, which is unlikely due to some stylistic overlap of art movements. For instance, we see in the bottom right of Figure 5a that Romanticism is a neighbor of Realism and Art Nouveau, which is not surprising given that the main Romanticism artistic movement occurred before the Realism and Art Nouveau movements (47). Also, the Realism stylistic movement developed in response to the Romantic ideals (32). At the top of Figure 5a, we see the Baroque style is surrounded by the Renaissance, Art Nouveau, and Realism styles. Generally, it is seen that the Baroque is highly ornate and eccentric in comparison to the Renaissance style (48), which preceded it. However, the Art Nouveau style is also known as a highly ornamental style, (49) similar to the Baroque style. These semantic organizations of the stylistic movements by using only the information extracted by UFLK from the content of the paintings is quite surprising, since it indicates that historical knowledge of the stylistic movement can be directly determined from the paintings without use of any other information.

The confusion matrix in Figure 4d demonstrates that the clustering of styles using 500 UFLK features produced the maximum number of paintings correctly in 6 out of 8 styles, as determined by the cost analysis. These styles include the Baroque, Impressionism, Art Nouveau, Romanticism, Expressionism, and Renaissance styles. The confusion matrix in Figure 4e shows that the clustering of paintings using 100 PCA features, also assigned the maximum number of paintings in style 6 out of 8 styles a well. These were the Baroque, Impressionism, Post-Impressionism, Realism, Art Nouveau, and Expressionism styles. However, using the 100 PCA features, the clustering algorithm predicted style groupings with poorer recall rates (Table 2). This means that only a few of the paintings that actually belong to a style group were selected. For instance, the precision of the Art-Nouveau cluster using 100 PCA features was 0.508, but the recall was 0.078, which is very poor. For the same Art Nouveau cluster using 500 UFLK features the precision, although lower, was 0.194. However, the recall was 0.28, which is much better than the recall rate of 0.078 obtained by the 100 PCA features. Clustering with raw pixels (Figure 4f) provided the maximum painting counts in the four painting styles of Baroque, Post-Impressionism, Realism, and Expressionism.

The significance tests for clustering, using the random PCA and similarity matrix distributions for simulated data, indicated that the Realism and Romanticism clusters formed by the 500 UFLK features and the raw pixels were both due to chance and not due to the features specific to those painting styles. Using 100 PCA features we again see that the Romanticism cluster was not statistically significant, as well as the Art Nouveau cluster.

Among the cluster of the eight different styles, across the three different types of feature extraction methods, the Baroque style had the overall highest F-score ($M = 0.248$, $SD = 0.169$), whereas Realism had the lowest F-score ($M = 0.088$, $SD = 0.064$). A statistical analysis of the clustering F-mac scores is in Figs. S3–S4 in the Supplementary Material.

5. DISCUSSION

We have demonstrated the feasibility of capturing stylistic information from a large dataset of digitized paintings without any prior knowledge of the painting styles or the types of features. We also demonstrated it may be possible to capture historical knowledge of the stylistic movements by utilizing only the UFLK features extracted from the content of painting images. Thus, we have introduced a method to capture the stylistic characteristics of art in an objective manner. This greatly simplifies the feature extraction process in art, as it has been indicated by Shamir (14,50,51) that a great variety of handcrafted features are necessary to successfully capture the characteristics of art. Our success is due to that fact that we utilized the recently proposed UFLK algorithm (31,38,52), which facilitates the use of a deep learning-like algorithm.

Shamir et al. (53) have shown that computers are able to exceed in accuracy the detection of abstract art created by children versus actual artists. In this study, Shamir et al. (53) used their specialized feature extraction algorithm (15,54), however, it may be possible to obtain similar results using the simple UFLK method demonstrated in our paper. Another similar use of the UFLK features could be to detect when an artist decides to change their painting style. For instance, it is known that Edouard Manet painted in both Realism and Impressionist styles.

Compared to other studies which classify painting styles by utilizing deep neural networks based on the supervised feature learning algorithm of CNNs (1,22,26), the UFLK algorithm with the SVM classifier provides a similar level of performance. Saleh and Elgammal (26) achieved the classification accuracy of 0.460 on 27 painting styles by combining features from CNNs (1000 dimensional vector) with semantic features from Classeme (55) (2659 dimensional vector) and Picodes (25) (2048 dimensional vector) by using a total of 5707 features, which is almost double our feature set. Bar et al. (1) achieved an F-macro of 0.414 by using 4096 CNN features with 2048 Picodes feature vector (total 6144 dimensional feature vector). Karayev et al. (22) achieved a mean precision of 0.473 using the AlexNet CNN (23).

Our top performing UFLK features had achieved an F-macro = 0.469, mean precision of 0.468, and accuracy of 0.475. Interestingly, for our dataset the AlexNet only achieved an accuracy of 0.326, which could be due the fact that aforementioned studies employed the WikiArt gallery, from which Karayev et al. (22), Bar et al. (1), and Saleh and Elgammal (26) used 78,449, 40,724, and 85,000 paintings, respectively. Typically, having an increased number of training data samples increases test accuracy in deep neural networks. The simpler LeNet CNN configuration achieved a better accuracy on our image dataset. Furthermore, by using handcrafted features and a classical SVM classifier, Spehr et al. (12) achieved an accuracy of 0.405 for eight painting styles.

Previous studies of clustering paintings by style (12,29) had not analyzed the accuracy and reliability of grouping paintings together by style, i.e., whether the style clusters were due to chance occurrence of the painting image data. Grouping paintings based on style is not a trivial problem (29), where naïve individuals can easily be influenced by the content of a painting, leading to incorrect style clusters (12). Spehr et al. (12) clustered 772 paintings, using kernel PCA feature transformation and K-means clustering, however, they employed handcrafted features, such as template matching, semantic content, shape segmentation, and color distribution features, which required an intermediate feature selection step to choose the relevant features. Furthermore, Spehr et al. (12) failed to provide a performance metric such as the accuracy or F-score, but rather used a derived measure, which only provided the relative discriminability among clusters, despite having the ground-truth style labels, most likely due to the issue of cluster label assignment (41).

Other clustering studies (13,50), which also utilized handcrafted features to group paintings by their creating artists, failed to provide quantitative metrics regarding the performance of the clusters. In Lombardi (13), clusters were only used as a visualization tool, demonstrating the divergence among artists by their paintings. In Shamir and Tarakhovsky (50), using only 994 paintings from a total of 34 painters, phylogenetic trees were created to demonstrate the relationships between artists, but the confidence level (56) of the tree branches was not provided. Without the confidence levels, it is not possible to objectively assess whether the relationships between artists existed due to a chance occurrence of the collected data, which could cause a false acceptance of the style clusters. For instance, in our clustering of the styles with our best performing algorithm using 500 ULFK features, the performance of the Realism and Romanticism clusters were better than the chance F-score of 12.5%, but were not significant due to the null distribution tests. As the confusion matrix in Figure 4d demonstrates, Realism was incorrectly clustered as Baroque, and Romanticism was incorrectly clustered as Impressionism, Post-Impressionism, and Art Nouveau. As a result, our null distribution methods provide a way to objectively assess whether groupings of paintings by a clustering algorithm are significant. These methods allow art researchers to determine the reliability of the groups of paintings independent of technical knowledge, in essence providing a second opinion regarding the clusters.

A limitation of this study is that the painting images, due to downsampling, are much smaller than the actual painting dimensions. In deep neural network type algorithms for painting style classification (1,22,26), this is a common limitation that can cause a loss of resolution and detail from the original paintings. Thus, reducing the detail of the features extracted using UFLK. Therefore, in future studies, we will employ images with much higher resolution than a typical deep neural network implementation.

6. CONCLUSIONS

We demonstrated that the deep neural network-inspired method of unsupervised feature learning with the K-means algorithm extracts features resembling edges, lines, and solid colors from digitized version of paintings. Also, we demonstrated that this nonlinear feature extraction algorithm may then be utilized to cluster and classify paintings based upon artistic style. When the clusters based upon the ULFK features were plotted as a scatterplots, and

the assigned styles were color-coded, we were able to ascertain semantic relationships between painting styles. We then demonstrated two new types of data distributions, which can assess the objective reliability of painting clusters. Therefore, the methods presented herein can provide art researchers with further insight, which may have been missed otherwise with traditional computational techniques. These procedures may also have use in other image data, such as sculptures, calligraphy or medical images. It may even be possible to extend this work into areas of art forensics such as forgery.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and was partially funded by National Institutes of Health (NIH) grant number R21 MH105823.

References

1. Bar Y, , Levy N, , Wolf L. Workshop at the European Conference on Computer Vision [Internet] Springer; 2014 Classification of artistic styles using binarized features derived from a deep neural network; 7184 [cited 2017 Jan 1]Available from: http://link.springer.com/chapter/10.1007/978-3-319-16178-5_5
2. Beckett W. The story of painting Dk Pub; 1994
3. Gardner H. Children's sensitivity to painting styles. Child Dev. 1970;813–21.
4. Mishory A. Art history: an introduction Open University of Israel; 2000
5. Stork DG, Coddington J. Computer image analysis in the study of art. 2008; 6810 publisher not identified.
6. Stork DG, Coddington J, Bentkowska-Kafel A. Computer vision and image analysis of art. 2010; 7531 publisher not identified.
7. Kirsh A, , Levenson RS. Seeing through paintings: physical examination in art historical studies Vol. 1. Yale University Press; 2002
8. Stork DG. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature Springer; 2009 924
9. Arora RS, Elgammal A. Towards automated classification of fine-art painting style: A comparative study. IEEE. 2012;3541–4.
10. Zujovic J, , Gandy L, , Friedman S, , Pardo B, , Pappas TN. Classifying paintings by artistic genre: An analysis of features & classifiers. Multimedia Signal Processing, 2009 MMSP'09 IEEE International Workshop on [Internet]; IEEE; 2009 15 [cited 2017 May 31]Available from: <http://ieeexplore.ieee.org/abstract/document/5293271/>
11. Günsel B, Sariel S, Içoglu O. Content-based access to art paintings. IEEE. 2005;II–558.
12. Spehr M, Wallraven C, Fleming RW. Image Statistics for Clustering Paintings According to their Visual Appearance. Comput Aesthet. 2009; 2009;1–8.
13. Lombardi TE. The classification of style in fine-art painting. 2005
14. Shamir L, Macura T, Orlov N, Eckley DM, Goldberg IG. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. ACM Trans Appl Percept. 2010 Feb; Jan; 7(2):1–17.
15. Shamir L, Orlov N, Eckley DM, Macura TJ, Goldberg IG. IICBU 2008: a proposed benchmark suite for biological image analysis. Med Biol Eng Comput. 2008; 46(9):943–7. [PubMed: 18668273]
16. uljak M, Mikuš B, Ješ K, Hadji S. Classification of art paintings by genre. IEEE. 2011;1634–9.

17. Condorovici RG, Florea C, Vrănceanu R, Vertan C. Perceptually-inspired artistic genre identification system in digitized painting collections Springer; 2013 68796
18. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *science*. 2006; 313(5786):504–7. [PubMed: 16873662]
19. Bengio Y. Learning deep architectures for AI. *Found Trends® Mach Learn*. 2009; 2(1):1–127.
20. Hinton G. A practical guide to training restricted Boltzmann machines. *Momentum*. 2010; 9(1): 926.
21. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 2015; 61:85–117. [PubMed: 25462637]
22. Karayev S, Trentacoste M, Han H, Agarwala A, Darrell T, Hertzmann A, et al. Recognizing image style. *ArXiv Prepr ArXiv13113715* [Internet] 2013 [cited 2017 Jan 10]; Available from: <http://arxiv.org/abs/1311.3715>
23. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks; *Advances in neural information processing systems* [Internet] 2012 10971105 [cited 2017 Jan 2] Available from: <http://papers.nips.cc/paper/4824-imagenet-classification-w>
24. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1989; 1(4):541–51.
25. Bergamo A, Torresani L, Fitzgibbon AW. Picodes: Learning a compact code for novel-category recognition; *Advances in Neural Information Processing Systems* [Internet] 2011 20882096 [cited 2017 Jan 1] Available from: <http://papers.nips.cc/paper/4319-picodes-learning-a-compact-code-for-novel-category-recognition>
26. Saleh B, Elgammal A. A unified framework for painting classification. *IEEE*. 2015:1254–61.
27. Unsupervised Feature Learning and Deep Learning Tutorial [Internet] [cited 2017 May 31]. Available from: <http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/>
28. Friedman J, Hastie T, Tibshirani R. *Springer series in statistics Vol. 1*. Springer; Berlin: 2001 The elements of statistical learning.
29. Wallraven C, Fleming R, Cunningham D, Rigau J, Feixas M, Sbert M. Categorizing art: Comparing humans and computers. *Comput Graph*. 2009; 33(4):484–495.
30. Le QV. Building high-level features using large scale unsupervised learning. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* [Internet]; IEEE; 2013 85958598 [cited 2017 May 18] Available from: <http://ieeexplore.ieee.org/abstract/document/6639343/>
31. Coates A, Ng A, Lee H. An analysis of single-layer networks in unsupervised feature learning. 2011:215–23.
32. Finocchio AR. Nineteenth-Century French Realism | Essay | Heilbrunn Timeline of Art History | The Metropolitan Museum of Art [Internet]. The Met's Heilbrunn Timeline of Art History [cited 2017 May 23]. Available from: http://www.metmuseum.org/toah/hd/rism/hd_rism.htm
33. Ciresan DC, Meier U, Gambardella LM, Schmidhuber J. Deep big simple neural nets excel on handwritten digit recognition. 2010:80. Cited On.
34. Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. 2009
35. Oja E. Principal components, minor components, and linear neural networks. *Neural Netw*. 1992; 5(6):927–35.
36. Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. *AICHE J*. 1991 Feb; 37(2):233–43.
37. Turk MA, Pentland AP. Face recognition using eigenfaces. *IEEE*. 1991:586–91.
38. Coates A, Ng AY. *Neural Networks: Tricks of the Trade* [Internet] Springer; 2012 Learning feature representations with k-means; 561580 [cited 2017 Jan 1] Available from: http://link.springer.com/chapter/10.1007/978-3-642-35289-8_30
39. Jolliffe I. *Principal component analysis* Wiley Online Library; 2002
40. Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. *Adv Neural Inf Process Syst*. 2002; 2:849–56.
41. Topchy AP, Jain AK, Punch WF. A Mixture Model for Clustering Ensembles. *SIAM*. 2004:379–90.

42. Kuhn HW. The Hungarian method for the assignment problem. *Nav Res Logist Q.* 1955; 2(1–2): 83–97.
43. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. *J Intell Inf Syst.* 2001; 17(2–3):107–45.
44. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Ser B Stat Methodol.* 2001; 63(2):411–23.
45. Berezhnoy I, Postma E, van den Herik J. Computer analysis of Van Gogh's complementary colours. *Pattern Recognit Lett.* 2007 Apr; 28(6):703–9.
46. Van Der Maaten L, Postma E, Van den Herik J. Dimensionality reduction: a comparative. *J Mach Learn Res.* 2009; 10:66–71.
47. Romanticism [Internet]. Encyclopedia Britannica [cited 2017 Jun 5]. Available from: <https://www.britannica.com/art/Romanticism>
48. Baroque art and architecture [Internet]. Encyclopedia Britannica [cited 2017 Jun 5]. Available from: <https://www.britannica.com/art/Baroque-art-and-architecture>
49. Art Nouveau | artistic style [Internet]. Encyclopedia Britannica [cited 2017 Jun 5]. Available from: <https://www.britannica.com/art/Art-Nouveau>
50. Shamir L, Tarakhovsky JA. Computer analysis of art. *J Comput Cult Herit.* 2012 Jul 1; 5(2):1–11.
51. Shamir L. What makes a Pollock Pollock: A machine vision approach. *Int J Arts Technol.* 2015; 8(1):1–10.
52. Coates A. Demystifying unsupervised feature learning [Internet] Stanford University; 2012 [cited 2017 Jun 5]. Available from: <https://pdfs.semanticscholar.org/59bd/d317abe8d87fb525eb4e3197a9311e2766e7.pdf>
53. Shamir L, Nissel J, Winner E. Distinguishing between Abstract Art by Artists vs. Children and Animals: Comparison between Human and Machine Perception. *ACM Trans Appl Percept.* 2016 May 28; 13(3):1–17.
54. Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recognit Lett.* 2008; 29(11):1684–93. [PubMed: 18958301]
55. Torresani L, Szummer M, Fitzgibbon A. Efficient object category recognition using classemes Springer; 2010 77689
56. Nei M, Kumar S. Molecular evolution and phylogenetics Oxford university press; 2000

Highlights

- Unsupervised feature learning with K-means (UFLK) can extract painting features.
- No prior knowledge of styles is used during the feature learning and extraction.
- Styles of test paintings are classified with better accuracy than other methods.
- Grouping by features shows semantic relations between styles that give insight.
- UFLK feature analysis may aid in preservation, interpretation and forensics of art.

Fig. 1. Processing pipeline of painting style grouping and prediction

First, the paintings, without using any stylistic or feature information, are input into one of three feature extraction processes: (1) nonlinear feature extraction with UFLK, (2) linear feature extraction with PCA, and (3) no feature extraction in which only the raw pixels are used. Then, the paintings, which are now represented either in feature space or in native pixels space, are input into a cluster or classification algorithm. The output from the cluster algorithm is grouping of paintings according to style. The output from the classification is a prediction of style for each individual painting.

Fig. 2. Types of painting styles
Representative paintings of each style type employed in this study are displayed.

Fig. 3. Visualization of UFLK and PCA features for *Scene from Tahitian Life*

(A) It is demonstrated at the top left that the extracted UFLK features resemble lines, edges and solid blocks of colors. The UFLK features are displayed as an image composed of a collection of 400 features organized into 20×20 grid. Each element in the grid contains a feature with dimension size of 6×6 pixels with RGB color channels. On the top right, Paul Gauguin's *Scene from Tahitian Life* is projected onto a few selected UFLK features, where it is demonstrated that the UFLK features highlight aspects such as the women's skirts or the background from the painting. (B) On the bottom left, the extracted PCA features are more similar to undefined blobs, with not much color differentiation. Here the features are displayed in 20×20 grid, each consisting of 64×49 pixels with RGB color channels. In the PCA feature space, the painting does not provide any selection of specific elements from the painting, as found in UFLK features. Furthermore, the reconstruction on the bottom left demonstrates that many details of the painting are lost in the PCA extraction of features.

Fig. 4. Confusion matrices of classification and clustering of painting styles

Each entry of the confusion matrices are scaled by the total paintings in a column, which in essence provides precision, i.e., the ratio of clustered paintings that actually belong to the true style.

Fig. 5. Scatterplot of painting styles clusters

The groupings of the painting styles obtained from spectral clustering are plotted as a scatterplot by projecting onto its first two multidimensional scaled components (MDS) (28) of the data. MDS is essentially a visualization of the similarity of observed data in a dataset. The size of each node, which represents a single painting in the dataset, is determined by its degree (i.e., the number of connections each painting has to other paintings). As a result, the larger a node, the more connections that painting has to other paintings. Also, the distance between nodes (paintings) demonstrates the similarity between paintings. Similar or related paintings and styles are in close proximity, while disparate paintings and styles are distant.

Table 1

Results of painting style classification

Painting Style	Raw Pixels	Feature Extraction Methods							
		PCA 100	PCA 300	PCA 500	PCA 1000	UFKL 500	UFKL 1000	UFKL 2000	UFKL 3000
Performance Metrics (F-score Precision Recall)	Baroque	0.412	0.424	0.409	0.402	0.629	0.64	0.656	0.644
		0.354	0.332	0.339	0.351	0.539	0.565	0.583	0.578
		0.492	0.585	0.514	0.47	0.754	0.738	0.749	0.727
Impressionism		0.23	0.233	0.251	0.204	0.317	0.353	0.347	0.338
		0.257	0.275	0.293	0.228	0.331	0.37	0.364	0.352
		0.208	0.202	0.219	0.185	0.303	0.337	0.331	0.326
Post-Impressionism		0.307	0.263	0.228	0.26	0.408	0.425	0.48	0.476
		0.322	0.3	0.234	0.277	0.428	0.444	0.5	0.479
		0.293	0.234	0.222	0.246	0.389	0.407	0.461	0.473
Realism		0.273	0.256	0.227	0.241	0.41	0.432	0.432	0.435
		0.259	0.255	0.23	0.227	0.394	0.424	0.417	0.422
		0.289	0.257	0.224	0.257	0.428	0.441	0.447	0.447
Art Nouveau		0.252	0.243	0.195	0.231	0.314	0.353	0.361	0.359
		0.292	0.278	0.221	0.242	0.391	0.423	0.425	0.431
		0.221	0.215	0.174	0.221	0.262	0.302	0.314	0.308
Romanticism		0.211	0.202	0.156	0.194	0.365	0.402	0.417	0.439
		0.222	0.224	0.177	0.221	0.388	0.423	0.436	0.446
		0.2	0.183	0.139	0.172	0.344	0.383	0.4	0.433
Expressionism		0.328	0.336	0.3	0.276	0.501	0.529	0.549	0.554
		0.305	0.307	0.271	0.257	0.475	0.5	0.538	0.559
		0.354	0.372	0.335	0.299	0.53	0.561	0.561	0.549
Renaissance		0.288	0.268	0.245	0.278	0.469	0.512	0.493	0.503
		0.294	0.286	0.245	0.28	0.466	0.491	0.466	0.48

	Painting Style	Feature Extraction Methods									
		Raw Pixels	PCA 100	PCA 300	PCA 500	PCA 1000	UFKL 500	UFKL 1000	UFKL 2000	UFKL 3000	
	F-mac [‡]	0.283	0.195	0.252	0.245	0.277	0.472	0.535	0.522	0.528	
		0.288	0.259	0.278	0.251	0.261	0.426	0.456	0.467	0.469	
		0.288	0.271	0.282	0.251	0.26	0.426	0.455	0.466	0.468	
		0.293	0.277	0.287	0.259	0.266	0.435	0.463	0.473	0.474	

[‡] F-mac is the average of the F-score; bold indicates the best F-mac score across the different feature extraction methods

Table 2

Results of painting style clustering

	Painting Style	Feature Extraction Methods								
		Raw Pixels	PCA 100	PCA 300	PCA 500	PCA 1000	UFLK 500	UFLK 1000	UFLK 2000	UFLK 3000
Performance Metrics (F-score Precision Recall)	Baroque	0.413	0.297	0.047	0.032	0.018	0.362	0.352	0.36	0.353
		0.389	0.237	0.22	0.195	0.211	0.306	0.304	0.304	0.308
		0.44	0.397	0.026	0.018	0.009	0.444	0.418	0.44	0.414
	Impressionism	0.165	0.217	0.005	0.038	0.106	0.199	0.198	0.197	0.188
		0.176	0.234	0.133	0.182	0.286	0.257	0.243	0.251	0.233
		0.155	0.203	0.002	0.021	0.065	0.162	0.168	0.162	0.158
	Post-Impressionism	0.21	0.184	0.226	0.247	0.281	0.199	0.106	0.104	0.108
		0.165	0.172	0.165	0.199	0.244	0.157	0.164	0.151	0.168
		0.288	0.197	0.36	0.323	0.332	0.273	0.078	0.079	0.079
	Realism	0.096	0.161	0.018	0.002	0	0.128	0.133	0.125	0.131
		0.202	0.218	0.17	0.077	0	0.144	0.141	0.137	0.141
		0.063	0.128	0.009	0.001	0	0.116	0.126	0.115	0.123
	Art Nouveau	0.24	0.135	0.206	0.15	0.16	0.229	0.22	0.223	0.219
		0.192	0.508	0.18	0.249	0.306	0.194	0.201	0.197	0.2
		0.32	0.078	0.24	0.107	0.109	0.28	0.243	0.259	0.242
Romanticism	0.107	0.031	0.036	0.245	0.009	0.157	0.175	0.163	0.168	
	0.177	0.226	0.121	0.157	0.174	0.325	0.323	0.333	0.321	
	0.077	0.017	0.021	0.557	0.005	0.104	0.12	0.107	0.113	
Expressionism	0.169	0.234	0.145	0.174	0.179	0.221	0.274	0.282	0.284	
	0.256	0.214	0.198	0.19	0.154	0.328	0.217	0.229	0.227	
	0.126	0.256	0.115	0.161	0.215	0.166	0.37	0.367	0.378	
Renaissance	0.175	0.222	0.246	0.2	0.256	0.197	0.203	0.21	0.208	
	0.153	0.159	0.158	0.175	0.156	0.186	0.181	0.186	0.178	

	Painting Style	Feature Extraction Methods									
		Raw Pixels	PCA 100	PCA 300	PCA 500	PCA 1000	UFLK 500	UFLK 1000	UFLK 2000	UFLK 3000	
	F-mac [‡]	0.204	0.367	0.558	0.231	0.714	0.208	0.231	0.242	0.25	
		0.197	0.185	0.116	0.136	0.126	0.212	0.208	0.208	0.207	
		0.214	0.246	0.168	0.178	0.191	0.237	0.222	0.223	0.222	
		0.209	0.205	0.166	0.178	0.181	0.219	0.219	0.221	0.22	

[‡] F-mac is the average of the F-score; bold indicates the best F-mac score across the different feature extraction methods