Genre Classification of Paintings

Eva Cetinic [1], Sonja Grgic [2]
[1] Institute Ruder Boskovic, Bijenicka 54, 10000 Zagreb, Croatia
[2] University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

ecetinic@irb.hr

Abstract **- Extensive digitization efforts in the recent years have led to a large increase of digitized and online available fine-art collections. With digitization of artworks, we aim to preserve all those valuable evidences of various human creative expressions, as well as make them available to a broader audience. The digitalization process of artworks should not constrain only to fulfilling the purpose of preservation, but also serve as a starting point for exploring of this type of data in a novel way, which is made possible with the rise of new achievements in computer vision. In the domain of computer analysis of visual art there are various ongoing research challenges. In this paper, we explore different image feature extraction methods and their applicability in the task of classifying painting by genre. Our dataset includes paintings of various styles grouped in seven genre categories. We achieved an accuracy of 77.57% for the task of genre classification. We concluded that the best performance is achieved when using features derived from a pre-trained deep convolutional neural network.**

Keywords **- Painting Classification; Genre; Image Features; Visual Art**

I. INTRODUCTION

The availability of large collections of digitized artworks triggers the need to effectively archive, retrieve, browse and analyze this type of data. Most of the available online collections include specific metadata in the form of annotations done by art historians and curators. Those annotations usually contain information about the painting's author, style, genre, date, location, etc. Art experts can easily identify the author, style and genre of a painting using their experience and knowledge of specific features. However, one great current challenge is to automate this process using computer vision and machine learning techniques. Generating metadata by hand is time consuming and requires the expertise of art historians. Therefore, automated recognition of the painting's characteristics would enable not only a faster and cheaper way of generating already existing categories of metadata such as style and genre in new collections, but also open the possibility of creating new types of metadata which relate to the content of the paintings or its specific stylistic properties. To be able to generate new types of metadata, we first need to master the classification problem within existing and well known categories such as genre.

Most of the earlier studies which addressed the problem of painting classification did not use one common dataset, but different datasets that highly varied in size and content, therefore it was difficult to adequately compare classification results. More recently, one dataset has become more ubiquitous namely the dataset of WikiArt[1], which comprises a large number of paintings annotated with a broad set of labels (e.g. style, genre, artist, technique, date, etc.). Furthermore, because of the use of different datasets and the lack of one unified annotation terminology, a confusion occurred when using the terms "genre" and "style". Many studies refer to "genre classification", but they actually focus on classifying paintings by art movements which would correspond better to the term "style". The term "genre" refers to the traditional division of paintings based on the type of content they depict. Paintings are traditionally divided into five genres: history painting, religious painting, genre painting, landscape and portrait. However, the WikiArt dataset includes a broader set of genre annotations to which we refer in this paper. We focus on those genre categories which correspond to specific objects or types of scenes, therefore our dataset consist of paintings from the following categories: portrait, landscape, cityscape, still life, nude painting, animal painting and flower painting (Fig. 1).

The next section provides an overview of related work. The third section presents the extracted image features. The fourth section includes the database description and classification implementation details. Classification results are presented and analyzed in the fifth section, while the final conclusion is given in the sixth section of this paper.

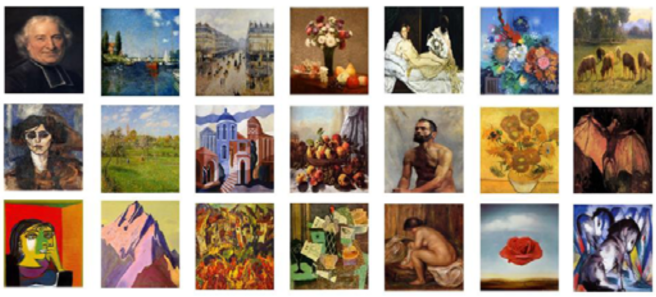

Figure 1. Examples of paintings from seven genres categories: portrait, landscape, cityscape, still life, nude painting, flower painting and animal painting

[1] Wikiart, www.wikiart.org/

II. Related work

The development of image processing techniques and machine learning algorithms led to significant achievements in classification and object recognition tasks when applied on photographic examples. Eventually this raised the question whether those techniques can also be used for painting related problems.

One of the first attempts in this direction was related to the challenge of classifying paintings according to their corresponding author [1]. The problem of classifying paintings by artist has later been addressed in several studies [2, 3], as well as the task of visualizing similarities and exploring influential links among painters [4, 5]. Identifying of the artist usually implies recognizing the artist's personal style - a set of specific characteristics of the painting perceived as formal elements of style such as color, light, line, texture, composition, etc. Most of the research concerning classification of paintings is based on extracting various low-level image features and then use those features as inputs for different machine learning classifiers. For example, Lombardi [6] presented a study of the performance of different low-level features, which correspond to the painting's light, line, texture and color elements, for the task of artist classification using several supervised and unsupervised learning techniques.

Similar methodology as in the task of classifying paintings by artist can be found in a considerable number of studies which address the problem of style classification [7, 8, 9]. Most of these studies provide an analysis of the impact of the different image features on the overall classification accuracy. Recent progress in computer vision achieved using deep neural networks, showed advantage of features which were "learned" from data on behalf of engineered features. Karayev et al. [10] describe an approach of recognizing the style of paintings using features extracted from a deep convolutional neural network. One of the most interesting outcomes of this study is that features derived from a deep convolutional neural network, trained for object detection on photographic images, show very good results when used in the task of painting style classification. Recently, Gatys et al. [11] proposed a method which uses the layered features of a deep convolutional neural network to separate the painting's style from content. Features derived from CNNs, combined with other well-known texture and color features, show a remarkable performance in the task of style and artist classification [12, 13].

III. Image features

The main challenge in automated recognition of painting genres is the transformation of painting characteristics into numerical descriptors. In solving this task, various approaches are possible and different types of image features can be extracted. We decided to use six different features, namely CNN-derived features, SIFT, GIST, HOG, GLCM and HSV color histograms.

A. *CNN-based features*

Deep convolutional neural networks have recently attained significant interest in the computer vision community due to the fact that they showed an impressive performance for the task of large-scale image classification [14]. Convolutional neural networks consist of multiple layers of small neuron collections which hierarchically process small portions of the input image. Each layer can be understood as a collection of image filters where each of them extracts a particular feature from the input image. The output of a given layer consists of differently filtered versions of the input image, corresponding to different levels of abstraction. First-layer features appear not to be specific to a particular dataset, but more general and referring to a lower level of abstraction in terms of image content representation. Onward the CNN, features transition from general to specific and therefore features derived from the last layer correspond to a higher level of abstraction such as specific object parts or the object's category. Following this observation and the conclusion that convolutional networks trained on one dataset of images can also be used as feature extractors from a different dataset [10], we used the MatConvNet [15] framework to extract features from a pre-trained VGG-F Network described in [16]. We used the outputs of the last hidden layer (conv7), which is forming a 4096 dimensional feature vector, and the last layer (conv8) which has output dimensionality of 1000, equal to the number of classes for which the network was trained.

B. *SIFT*

The Scale-Invariant Feature Transform (SIFT) [17] has shown very good performance for object recognition tasks, as well as good performance in classifying paintings by style and genre in [18]. Using the VLFEAT library [19], we extracted for each painting the 128-dimensional dense sift features and computed a Bag-Of-Word histogram of these descriptors, using a K-means vocabulary of 1000 words.

C. *GIST*

The GIST descriptor was initially proposed in [20]. The main of this description is to capture the spatial structure of an image and to create a low dimensional representation of the image scene, therefore the GIST descriptor is known to perform well for retrieving images that are visually similar at a low resolution scale. To extract GIST features, we use the GIST implementation available on the project website.

D. *HOG*

Histogram of Oriented Gradients (HOG) [21] is a widely used descriptor for the purpose of object detection in images. The HOG descriptor is constructed by counting occurrences of gradient orientation in localized portions of an image. The image is first decomposed into cells and for each cell a histogram of oriented gradients is extracted. Finally, the HOG descriptor is constructed by concatenating all histograms. To extract HOG features, we used the VLFEAT library.

E. GLCM

The gray-level co-occurrence matrix (GLCM), introduced by Heralick et al. [22], is a statistical approach that gives information about positions of pixels with similar gray level values. The total descriptor is a 256-dimensional vector, concatenated from 64 dimensional vectors obtained for each direction.

F. HSV histograms and statistical measures

Color is probably the most noticeable part of information we obtain when observing a painting. To obtain information about color, we use the HSV (hue, saturation, value) color model representation of the image. The extracted features include a 120-bin hue histogram and 100-bin saturation and value histograms, as well as the mean, variance, skewness and kurtosis values calculated for each histogram.

IV. IMPLEMENTATION

A. Dataset

Our dataset consists of paintings downloaded from the publicly available dataset of WikiArt, which is at the moment the largest online public collection of digitized artworks. When choosing the categories for our genre classification experiment, we focused on categories which include more than 1000 paintings. We downloaded all the paintings included in following the categories: portrait, still life, landscape, cityscape, nude, flower and animal painting. From those paintings we randomly chose a subset of 1000 images per category. A subset from one specific genre includes paintings from a variety of different styles and artists.

B. Classification

We experimented with various classifiers and performed hyperparameters optimization using grid search to select the best model. By comparing classification results for different models, we found out that the best overall accuracy was achieved with the use of support vector machines (SVM) [23] with a radial basis function (RBF) kernel and hyper parameters γ=1 and C=10. SVM maps features non-linearly into n dimensional feature space where the features vectors define the hyper plane. The margin represents the distance between hyper plane and support vectors. The margin should be positioned in such a way to maximize the distance between support vectors. The gamma parameter defines the distance which a single training example can reach, where lower values mean 'far' and higher values mean 'close'. The C parameter regulates the compromise between misclassifying training examples and the simplicity of the decision boundary. A model with lower C values ensures a smoother decision surface, while with higher value it aims at classifying all training examples correctly.

To measure the model's accuracy, we performed 5-fold cross validation. We sampled the images into 5 different training and testing sets so that the each training set contained 5600 images (800 images from every genre category) and testing set contained 1400 images (200 images per category). The model selection process was performed based on combining all the features by concatenating them into one high-dimensional feature vector. After selecting the best model, we tested the performance of all the individual feature subsets, as well as features combined in two different ways. Besides testing the classifier using as inputs the concatenated feature vectors, another way was to combine outputs from several SVM classifiers which were trained using different feature subsets and then apply a majority voting scheme on the predictions obtained from those classifiers. To implement the classification task we used the scikit-learn package [24].

V. RESULTS

Features obtained from the seventh layer of a pre-trained convolution neural network (conv7) achieved the best performance, as show in Fig. 2. Those features perform well because they are optimized to distinguish specific shapes, but are still more general than the features obtained from the last layer of the CNN (conv8), which are quite object and category specific. When combining features by simply concatenating them into one feature vector, we achieve worse results than when using only conv7 features. By using the second method, namely combining the outputs of classifiers trained on separated feature subsets, we achieve better result, but not a significant improvement in comparison with conv7 features.

Although combining different types of features outperforms the use of only CNN derived features in some classification tasks such as style or artist classification, our work shows that combining this particular sets of features doesn't significantly improve genre classification. Because the genre of paintings corresponds more to objects or types of scenes depicted in paintings, low-level color and texture features don't contribute to the genre classification task as much as they do in style or artist classification tasks.

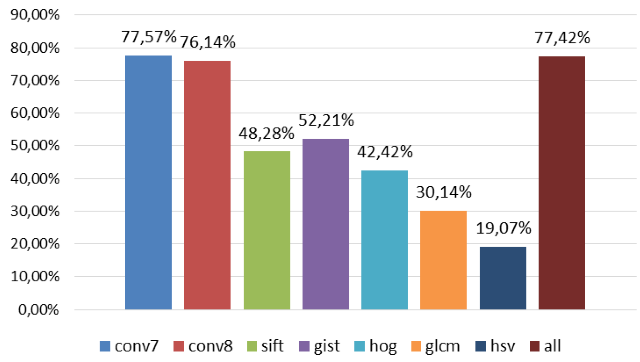

Figure 2. Classification accuracy of different features

Our results show better performance on the task of genre classification than those presented in [13], where Saleh et al. tested CNN features derived only from the last layer of a pre-trained CNN (1000 dimensional vectors). Agarwal et al. [18] achieved classification accuracy of 84.56% without using CNN derived features, but they performed their experiment on a smaller dataset divided into a smaller number of genre categories, which included the class of abstract paintings and the class of sculptures, which makes the overall distinction among classes considerably easier.

Interpretation of the misclassified paintings indicate general similarity between landscape and cityscape paintings, still life and flower painting, as well as and portrait and nude painting categories (Fig. 3.).

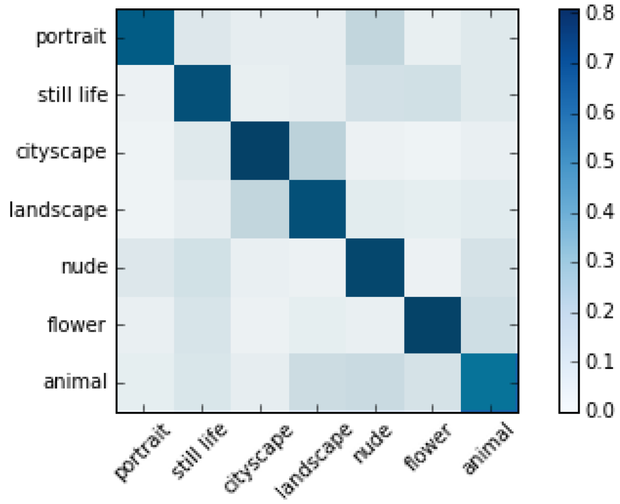

Figure 3. Confusion matrix for genre classification

These confusions are expected because landscape and cityscape both include outdoor scenes, whilst still life and flower paintings mostly include indoor scenes, as well as similar objects. Also, nude painting often depicts faces and are therefore often misclassified as portrait paintings

.

VI. Conclusion

In this work, we proposed an approach to image feature extraction for the purpose of classifying paintings by genre. We analyze the impact of different features and experiment with various classifiers. We conclude that using CNN-based features outperform all other image feature types for this particular classification task. This indicates that features derived from CNNs, trained for object recognition in photographs, can also very well distinguish scenes and objects in paintings, regardless of the various artistic techniques and styles. This leads to the conclusion that higher accuracy in the problem of genre classification could be achieved if the CNN is trained on paintings dataset. Training such a network requires a very large dataset of annotated paintings which is the moment not available. However, future work could still evolve in this direction by fine-tuning existing CNNs, as well as exploring more sophisticated feature fusion and classifier ensemble techniques.

References

[1] D. Keren, “Painter identification using local features and naive bayes,” in Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 2, , pp. 474–477, 2002.

[2] E. Cetinic, S. Grgic “Automated painter recognition based on image feature extraction”, Proceedings of the 55th International Symposium ELMAR, pp. 19-22. IEEE, 2013.

[3] N. van Noord, E. Hendriks, and E. Postma. “Toward Discovery of the Artist's Style: Learning to recognize artists by their artworks.” Signal Processing Magazine, IEEE 32, no. 4, pp. 46-54., 2015.

[4] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I.G. Goldberg, “Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art, ” ACM Trans. Appl. Percept. 7, 2, 2010.

[5] M. Bressan, C. Cifarelli, F. Perronnin, "An analysis of the relationship between painters based on their work," Image Processing,15th IEEE International Conference on , vol., no., pp.113-116, October 2008.

[6] T. Lombardi. “The classification of style in fine-art painting”, Ph.D. thesis, Pace University, 2005.

[7] J. Zujovic et al. “Classifying paintings by artistic genre: An analysis of features & classifiers,” in Multimedia Signal Processing, MMSP’09. IEEE International Workshop, pp. 1–5., 2009.

[8] R.S. Arora, A. Elgammal, "Towards automated classification of fine-art painting style: A comparative study," Pattern Recognition (ICPR), 21st International Conference on , pp.3541-3544, Nov. 2012.

[9] F. Corneliu et al. “Pandora: Description of a Painting Database for Art Movement Recognition with Baselines and Perspectives.” arXiv preprint arXiv:1602.08855, 2016.

[10] S.Karayev, et al. "Recognizing image style." arXiv preprint arXiv:1311.3715 , 2013.

[11] L. A Gatys., A.S. Ecker, and M. Bethge. “A Neural Algorithm of Artistic Style.” arXiv preprint arXiv:1508.06576, 2015.

[12] Y. Bar et al., “Classification of artistic styles using binarized features derived from a deep neural network”, In Computer Vision-ECCV Workshops, Springer International Publishing, pp. 71-84., 2014.

[13] B. Saleh and A. Elgammal, “Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature.” arXiv preprint arXiv:1505.00855, 2015.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, In: Advances in neural information processing systems, pp. 1097–1105, 2012.

[15] A. Vedaldi and L.Karel, “MatConvNet: Convolutional neural networks for matlab.”, In Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, ACM, pp. 689-692. 2015.

[16] K. Chatfield, K. Simonyan, A.Vedaldi, and A, Zisserman, “ Return of the devil in the details: Delving deep into convolutional nets”, arXiv preprint arXiv:1405.3531, 2014.

[17] D.G Lowe., “Distinctive image features from scale-invariant keypoints”, International journal of computer vision, 60(2), pp.91-110., 2004.

[18] S.Agarwal, et al. “Genre and Style Based Painting Classification.” Applications of Computer Vision (WACV), IEEE Winter Conference, pp. 588-594., 2015.

[19] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms, www. vlfeat.org, 2008.

[20] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision, 42(3), pp.145–175, 2001.

[21] N. Dalal and B. Triggs., “Histograms of oriented gradients for human detection” In International Conference on Computer Vision & Pattern Recognition, volume 2, pp. 886–893, June 2005.

[22] R. M. Haralick. "Statistical and structural approaches to texture," Proceedings of the IEEE , vol.67, no.5, pp.786-804, May 1979.

[23] J. Shawe-Taylor and N. Christianini, An Introduction to Support Vector Machines and Other Kernel Based Methods, Cambridge University Press, 2000.

[24] F. Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.