Towards a Tool for Visual Link Retrieval and Knowledge Discovery in Painting Datasets

Giovanna Castellano and Gennaro Vessio(✉)

Dipartimento di Informatica, Università degli Studi di Bari, Bari, Italy
{giovanna.castellano,gennaro.vessio}@uniba.it

Abstract. This paper presents a preliminary investigation aimed at developing a tool for visual link retrieval and knowledge discovery in painting datasets. The proposed framework is based on a deep convolutional network to perform feature extraction and on a fully-unsupervised nearest neighbor approach to retrieve *visual links* among digitized paintings. Moreover, the proposed method makes it possible to study influences among artists by means of graph analysis. The tool is intended to help art historians better understand visual arts.

Keywords: Cultural heritage · Deep learning · Computer Vision · Visual link retrieval · Knowledge discovery · Paintings

1 Introduction

The cultural heritage, in particular visual arts, have invaluable importance for the cultural, historic and economic growth of our societies. One of the building blocks of most analysis in visual arts is to find similarity relationships, i.e. link retrieval, among paintings of different artists and genres. These relationships can help art historians discover and better understand influences and changes from an artistic movement to another. Traditionally, this kind of analysis is done manually by inspecting large collections of human annotated photos. Unfortunately, manually searching over thousands of pictures, spanned across different epochs and painting schools, is a very time consuming and expensive process.

In the last years, large-scale digitization efforts have been made, leading to a growing availability of digitized fine art collections (e.g., WikiArt[1] and the MET collection[2]). This has opened new opportunities for computer science researchers to assist the art community with automatic tools to further understand visual arts. Automated painting analysis, in fact, is becoming increasingly important for several tasks, ranging from object detection [5] to artistic style classification [15]. The purpose of our research is to develop an automatic tool to be used to retrieve *visual links* within large digitized collections of paintings by using simple

[1] https://www.wikiart.org.
[2] https://www.metmuseum.org/art/collection.

M. Ceci et al. (Eds.): IRCDL 2020, CCIS 1177, pp. 105–110, 2020.
https://doi.org/10.1007/978-3-030-39905-4_11

image queries. Relying only on visual patterns makes the approach desirable when difficult to collect textual metadata are either scarce or unavailable.

The ability to recognize artistic styles and similarities in fine art paintings inherently falls within the domain of human perception. Understanding semantic attributes of a painting, such as content and meaning, in fact, originates from the composition of the shape, colour and texture features visually perceived by the human expert. Recent breakthroughs in Computer Vision, particularly in Convolutional Neural Networks (CNNs), proved to be very effective to tackle the problem of learning meaningful representations from the low-level colour and texture features (e.g., [3,14]). For this reason, the proposed tool is mainly based on visual attributes automatically learned by a well-known CNN architecture, i.e. VGG. The resulting high dimensional representation is then embedded in a more compact feature space through the use of Principal Component Analysis (PCA). Finally, similarities among paintings, i.e. visual links, are obtained through a distance measure in a completely unsupervised nearest neighbor fashion. The proposed method not only provides the nearest neighbors for each query image, i.e. those images more similarly linked to the input query, but it also allows the user to study historical patterns by means of graph analysis.

2 Proposed Method

The method we propose is partly inspired by the research presented in [11] and [12]. In both works, a deep learning-based approach is followed to retrieve common visual patterns shared among paintings and to discover near duplicate patterns in large collection of artworks, respectively. In both cases, the authors used a supervised approach in which the labels to be predicted are manually provided by human experts. Conversely, our method works in a fully unsupervised fashion, making the laborious acquisition of annotations unnecessary.

An overall scheme of the proposed framework is depicted in Fig. 1. The method assumes to have a large collection of digitized paintings of different artists and genres, as those nowadays collected in several museum Websites. The goal is to transform the raw pixel images into a new, numerical feature space in which to search for similarities among paintings. In order to obtain meaningful representations of visual attributes of paintings, we used *transfer learning* from a pre-trained deep Convolutional Neural Network. This practice is common and it is now usually preferred over classic approaches based on hand-crafted features in several perceptual domains. For example, it has been recently applied to the problem of Parkinsonian handwriting classification [7].

The input to our system is thus represented by a 224×224 three-channels painting image, normalized within the range $[0, 1]$: this is the typical input expected by the CNN we used. In fact, each input image is propagated through a VGG16 architecture [13], pre-trained on the very large ImageNet dataset [6]. The main assumption is that if the original dataset is large and general enough, then the weights learned by the network on this set of data can be used to new, even completely different image datasets. VGG16 is a well-known CNN architecture

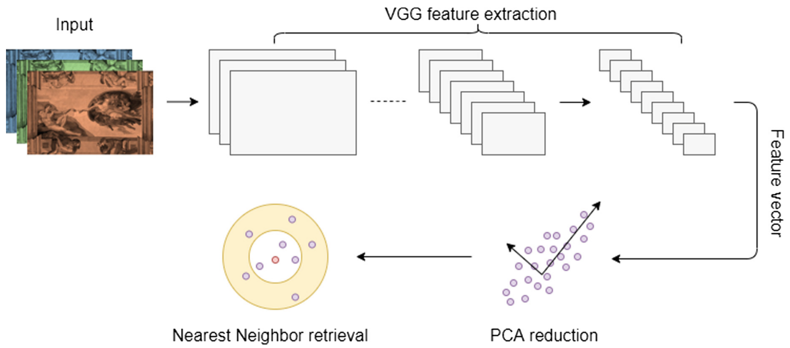

Fig. 1. Workflow of the proposed method.

which adopts 3×3 convolution and 2×2 max pooling throughout the network. All hidden layers are equipped with the ReLU activation function. To achieve transfer learning, we used the common practice to ignore the fully-connected layers stacked on top of the convolutional base and to extract the output features from the last convolutional layer. The network is able to construct a hierarchy of visual features, starting from simple edges and shapes at the earlier layers to higher-level concepts at the later layers. This approach is thus apt to obtain useful, semantic representations for the problem at hand.

Once the features extracted by the deep network are flattened, they have still too high dimensionality (i.e., 25, 088 dimensions) to make the use of distance measures feasible. For this reason, we need to transform this high dimensional feature space into a more compact low dimensional representation given by the application of PCA. Dimensionality reduction techniques are tailored to this goal (e.g., [2]); PCA, in particular, minimizes the mean ℓ_2 distance between data points and their linear projections [8]. For instance, the first two principal components span the plane that is *closest*, in terms of average distance, to the data points provided as input. In our case, to achieve a good compromise between representation power and dimensionality, the original 25, 088 dimensional feature space is projected onto a reduced space of 50 features.

The final search for visual links among paintings is performed in the reduced feature space in a fully unsupervised nearest neighbor fashion. In other words, for each query point q the methods returns the k data points *closest* to q. "Closeness" implies a metric that, as in PCA, corresponds to the usual ℓ_2 distance: $\sqrt{\sum_{i=1}^{N}(q_i - p_i)^2}$, being q_i and p_i the query point and each other data point, respectively, with N their dimensionality. In our case, we set k to 3, i.e. for each query, the three most similar paintings are provided by the system. Clearly, when searching for a particular artist's query, the other paintings from the same artist are excluded from the research, otherwise obvious, self links are likely to be retrieved. As previously stated, relying on a completely unsupervised approach makes the proposed method simple and practical, as it excludes the necessity to acquire labels of visual links or similarities of local parts of paintings, which are very difficult to collect.

3 Experimental Evaluation

We preliminarily investigated the effectiveness of the proposed method on a dataset collecting paintings of 50 very popular painters. More precisely, we used data provided by the Kaggle platform,[3] scraped from an art challenge Website.[4] Artists belong to very different epochs and painting schools, ranging from Giotto di Bondone and Renaissance painters such as Leonardo da Vinci and Michelangelo, to Modern Art exponents, including Pablo Picasso, Salvador Dalí, and so on. Painting images are non-uniformly distributed among painters for a total of 8,446 images of different sizes.

Experiments were run on an Intel Core i5 equipped with the NVIDIA GeForce MX110, with dedicated memory of 2 GB. As deep learning framework, we used TensorFlow 2.0 and the Keras API. It is worth to note that we did not perform an execution time analysis. In fact, one advantage of the proposed method is that its most expensive part, i.e. the VGG-based feature extraction, can be done completely offline, thus making the visual link retrieval, i.e. the search over the reduced feature space, only dependent on the collection size.

Some image queries together with their corresponding three output neighbors are provided in Fig. 2. In the first row, we asked for paintings visually linked with the Romanticist "Fort Vimieux" by William Turner, depicting a classic red sunset of the author. It can be seen that the system was able to retrieve paintings similar in their content. In the second row, we searched for paintings visually linked with the Impressionist "Confluence of the Seine and the Loing" by Alfred Sisley. It can be noticed that the given neighbors, i.e. two artworks by Camille Pissarro and a work by Claude Monet, share the same painting style. Finally, the third sample query was the more classic "Virgin and Child with Six Angels and the Baptist" by the Renaissance artist Sandro Botticelli. As it was expected, the visual features provided by the deep network were able to retrieve paintings similar in composition (holy family) and shape (tondo).

Finally, it is worth remarking that by collecting the nearest neighbors of all painters' artworks and by retaining only the most occurring linked painters, the proposed approach makes it possible to build a graph, possibly showing influences among artists. In other words, also historical knowledge discovery can be done. As depicted in Fig. 3, three connected components can be observed together with some *hubs*, represented by very influencing painters, such as Vincent Van Gogh and Edgar Degas. As expected, painters belonging to the same school appear to be close to each other.

[3] https://www.kaggle.com/ikarus777/best-artworks-of-all-time.
[4] http://artchallenge.ru.

Fig. 2. Sample queries (on the left) and corresponding output neighbors (on the right).

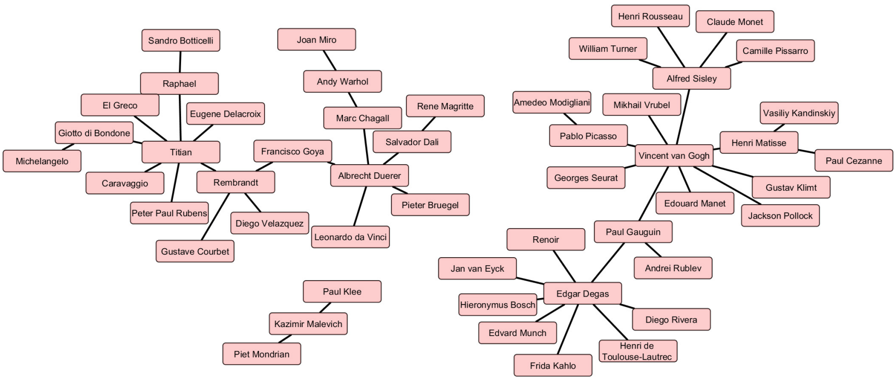

Fig. 3. Famous painters graph.

4 Conclusion and Future Work

In this paper, we have presented our preliminary research aimed at developing a tool for finding visually linked paintings among large digitized collections. The tool is based on a deep convolutional neural network as a feature extractor and on a fully-unsupervised nearest neighbor approach as image retrieval system. The proposed method provides the users with a very simple approach, as it requires only a sample image as a query; moreover, it does not need very difficult to collect human annotated labels. The proposed method may be helpful not only to historians for studying school paintings' evolution across centuries, but also to practitioners, for navigating through very large painting databases, as well as to forgery experts, for detecting suspicious plagiarism. Automatized tools can play a crucial role in managing large digitized collections and their use is receiving a lot of attention in the field of cultural heritage, for example also in the document layout analysis domain [4].

As future work, we plan to develop an easy-to-use user interface and to involve art historians to evaluate the effectiveness of the system. Finally, also a more refined graph analysis needs further work. Network analysis, in fact, is a powerful framework to study network systems and their interactions and several successful applications have been reported in the literature (e.g., [1,9,10]).

References

1. Carli, R., Dotoli, M., Cianci, E.: An optimization tool for energy efficiency of street lighting systems in smart cities. IFAC-PapersOnLine **50**(1), 14460–14464 (2017)
2. Casalino, G., Gillis, N.: Sequential dimensionality reduction for extracting localized features. Pattern Recogn. **63**, 15–29 (2017)
3. Cetinic, E., Lipic, T., Grgic, S.: Fine-tuning convolutional neural networks for fine art classification. Expert Syst. Appl. **114**, 107–118 (2018)
4. Corbelli, A., Baraldi, L., Balducci, F., Grana, C., Cucchiara, R.: Layout analysis and content classification in digitized books. In: Agosti, M., Bertini, M., Ferilli, S., Marinai, S., Orio, N. (eds.) IRCDL 2016. CCIS, vol. 701, pp. 153–165. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56300-8_14
5. Crowley, E.J., Zisserman, A.: In search of art. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8925, pp. 54–70. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16178-5_4
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
7. Diaz, M., Ferrer, M.A., Impedovo, D., Pirlo, G., Vessio, G.: Dynamically enhanced static handwriting representation for Parkinson's disease detection. Pattern Recogn. Lett. **128**, 204–210 (2019)
8. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, vol. 112. Springer, Heidelberg (2013)
9. Lella, E., Amoroso, N., Lombardi, A., Maggipinto, T., Tangaro, S., Bellotti, R.: Communicability disruption in Alzheimer's disease connectivity networks. J. Comp. Netw. **7**(1), 83–100 (2018)
10. Piccinni, G., Avitabile, G., Coviello, G.: An improved technique based on Zadoff-Chu sequences for distance measurements. In: 2016 IEEE Radio and Antenna Days of the Indian Ocean (RADIO), pp. 1–2. IEEE (2016)
11. Seguin, B., Striolo, C., diLenardo, I., Kaplan, F.: Visual link retrieval in a database of paintings. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9913, pp. 753–767. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46604-0_52
12. Shen, X., Efros, A.A., Mathieu, A.: Discovering visual patterns in art collections with spatially-consistent feature learning. arXiv preprint arXiv:1903.02678 (2019)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
14. Tan, W.R., Chan, C.S., Aguirre, H.E., Tanaka, K.: Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3703–3707. IEEE (2016)
15. Van Noord, N., Hendriks, E., Postma, E.: Toward discovery of the artist's style: learning to recognize artists by their artworks. IEEE Sign. Process. Mag. **32**(4), 46–54 (2015)