

# Understanding Art with AI: Our Research Experience

Giovanna Castellano, Gennaro Vessio

*Department of Computer Science, University of Bari Aldo Moro, Italy*

## Abstract

Artificial Intelligence solutions are empowering many fields of knowledge, including art. Indeed, the growing availability of large collections of digitized artworks, coupled with recent advances in Pattern Recognition and Computer Vision, offer new opportunities for researchers in these fields to help the art community with automatic and intelligent support tools. In this discussion paper, we outline some research directions that we are exploring to contribute to the challenge of understanding art with AI. Specifically, our current research is primarily concerned with visual link retrieval, artwork clustering, integrating new features based on contextual information encoded in a knowledge graph, and implementing these methods on social robots to provide new engaging user interfaces. The application of Information Technology to fine arts has countless applications, the most important of which concerns the preservation and fruition of our cultural heritage, which has been severely penalized, along with other sectors, by the ongoing COVID pandemic. On the other hand, the artistic domain poses entirely new challenges to the traditional ones, which, if addressed, can push the limits of current methods to achieve better semantic scene understanding.

## Keywords

Digital Humanities, Visual arts, Artificial Intelligence, Deep Learning

## 1. Introduction

Artificial Intelligence is revolutionizing numerous fields of knowledge and has established itself as a key enabling technology. Among the various domains that have been powered by AI-based solutions there is also the artistic one. In fact, in recent years, a large-scale digitization effort has been made, which has led to the increasing availability of huge digitized artwork collections. And this availability, combined with the recent advances in Pattern Recognition and Computer Vision, has opened up new opportunities for researchers in these fields to assist domain experts, particularly art historians, in the study and analysis of visual arts. Among other benefits, a deeper understanding of visual arts can favor their use by an ever wider audience, thus promoting the spread of culture. Visual arts, and more generally our cultural heritage, play a role of primary importance for the economic and cultural growth of our society [1, 2].

The ability to recognize characteristics, similarities and, more generally, patterns within and between digitized artworks, in order to favor a deeper study, inherently falls within the domain of human aesthetic perception [3]. Since this perception is highly subjective, and influenced by various factors, not least the emotion the artwork evokes in the observer, it is extremely difficult to conceptualize. However, representation learning techniques, such as those on which

---


*AIXIA 2021 Discussion Papers*

✉ giovanna.castellano@uniba.it (G. Castellano); gennaro.vessio@uniba.it (G. Vessio)

🆔 0000-0002-6489-8628 (G. Castellano); 0000-0002-0883-2691 (G. Vessio)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

state-of-the-art neural network models are based [4], appear to hold promise for automatically extracting meaningful features from artworks based on their elementary pixel encoding. Such representations can be useful for carrying out numerous tasks of interest from an artistic point of view, such as the automatic categorization of a painting based on artist, style and genre (e.g., [5, 6, 7]), or the retrieval of works similar to a given artwork based on visual features, textual descriptions, etc. (e.g., [8, 9, 10]). Interest is also growing around the use of generative paradigms, in order to create some form of art without the intervention of human artists in the generation process (e.g., [11, 12, 13]).

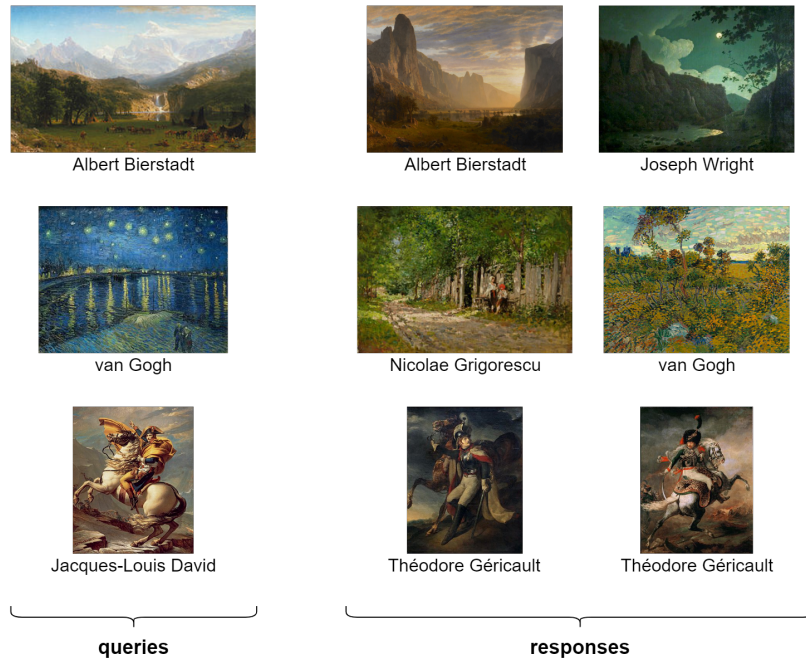
At the Computational Intelligence laboratory (CILab) of the Department of Computer Science of the University of Bari Aldo Moro we are contributing to the interdisciplinary research in this field, which is now very active and fertile, in order to propose new techniques, methodologies and tools for the automatic and “intelligent” analysis of visual arts. The remainder of this discussion paper will be devoted to briefly outlining the main research directions we are currently exploring at CILab.

## 2. Visual Link Retrieval

One of the building blocks of most analysis in the visual arts is finding similarity relationships, that is retrieving links, between paintings by different artists and painting schools. These relationships can help art historians to discover and better understand the influences and changes from one art movement to another. Indeed, art experts rarely analyze artworks as isolated creations, but typically study paintings within broad contexts, involving influences and connections between different schools.

Traditionally, this type of analysis is done manually by inspecting large collections of human annotated photos. However, the manual search over thousands of images, spread across different periods and painting schools, is a very time-consuming and costly process. Along this direction, we have recently proposed a method for *visual link retrieval*, which relies on the use of a deep convolutional neural network to perform feature extraction and a fully unsupervised nearest neighbor mechanism to retrieve links between the digitized paintings [14]. Searching for visual links is completely unsupervised, making the method especially useful in cases where metadata are scarce, unavailable or difficult to collect. It is worth noting that the proposed method not only provides those images that are most similarly related to the input query, but also allows the user to study historical patterns by analyzing the “influence graph” built on the retrieved links. In fact, by applying graph measures on the network built on the links obtained, the proposed method performs a form of *historical knowledge discovery* on the artists. To provide an illustrative example of the behavior of the system, in Fig. 1 we provide some image queries along with the corresponding top visually linked artworks retrieved by the system.

The proposed method can be beneficial not only for art historians. Enthusiasts, in fact, can benefit from the automatic link retrieval when visiting the digital collections of online museums and art galleries. This can favor a sort of interactive navigation capable of favoring the enjoyment of art.



**Figure 1:** Query examples and corresponding visually linked paintings.

### 3. Artwork Clustering

While the approach described in the previous section is suitable for finding visually linked artworks, it is not effective to cluster artworks into different groups, since the data appears to be uniformly distributed within a single homogeneous cluster in the feature space. Having a model that can cluster artworks without depending on hard-to-collect labels or subjective knowledge can be really useful for many domain applications. For example, the model can be used to discover different periods in the production of the same artist. Likewise, it can help experts classify contemporary art, which cannot be richly annotated.

To this end, we have proposed a method that uses a pre-trained deep convolution neural network to perform feature extraction, and uses a deep embedded clustering model [15], based on an auto-encoder neural network, to perform clustering [16]. The choice of this fully deep pipeline was motivated by the difficulty of applying traditional clustering algorithms and feature reduction techniques to both the highly dimensional input pixel space and the feature space resulting from CNN embedding, especially when input images are very complex artistic images. Quantitative and qualitative experimental results showed that the proposed method is able to find well-separated clusters both when considering an overall data set spanning different periods and when focusing on artworks produced by the same artist. In particular, from a qualitative point of view, it seems that the model look not only at stylistic features to group artworks, but also especially at the semantic attributes relating to the content of the scene represented. For example, Fig. 2 shows sample images of clusters found in Pablo Picasso’s artistic production and it can be seen that the model places semantically related works, such as portraits and still



**Figure 2:** Sample images from the clusters found among Picasso’s artworks.

lives, in the same clusters, despite their stylistic representation. This capacity seems to hold promise for addressing the well-known “cross-depiction” problem, which still poses a challenge for the research community [17]. Indeed, this capacity could be exploited to find similarities between artworks despite the way they are depicted.

#### 4. Computer Vision & Knowledge Graphs

Our research then moved on from another consideration: much of the work in the literature relies solely on the pixel information inherent in the digitized paintings and drawings. Unfortunately, this approach leads to ignoring a large amount of domain knowledge, as well as known relationships and connections between artworks and/or artists that could increase the quality of existing solutions. Artworks, in fact, cannot be studied on the basis of their visual appearance alone, but also considering various other historical, social and contextual factors that allow us to frame them in a more complex framework. Therefore, having a knowledge base where not only artworks, but also a rich plethora of metadata, contextual information, textual descriptions, etc., are unified within a structured framework can provide a valuable resource for more powerful information retrieval and knowledge discovery tools in the artistic domain. Such a framework would be beneficial not only for enthusiastic users, who can exploit the encoded information to navigate the knowledge base, but also especially for art experts, interested in

finding new relationships between artworks and/or artists for a better understanding of the past and modern art.

To fill this gap, we have developed ArtGraph: an artistic knowledge graph (KG) [18]. A KG [19] provides a more expressive and flexible representation to incorporate relationships of arbitrary complexity between entities related to art, which cannot be obtained by considering only the visual content [20]. The proposed KG integrates the information collected by WikiArt<sup>1</sup> and DBpedia<sup>2</sup> and exploits the potential of the Neo4j<sup>3</sup> database management system, which provides an expressive graph modeling and query language. In this way, the NoSQL database already helps provide a powerful knowledge discovery framework without explicitly training a learning system. The user, in fact, can query the graph for example to study the influences between artists, retrieve the works stored in a specific place, etc. Furthermore, the contextual knowledge encoded in ArtGraph can be integrated with visual features automatically learned by deep neural networks to develop more powerful learning models in the art domain. Several tasks, in fact, could be addressed, such as artwork attribute prediction, multi-modal retrieval and artwork captioning, which are attracting increasing interest in this domain.

## 5. Social Robotics

As applications of Computer Vision algorithms to artistic tasks become more mature, an interesting implementation of these techniques in real-world cases is to incorporate them into *social robots*. These represent an emerging field of research focused on developing a “social intelligence” that aims to maintain the illusion of dealing with a human being [21]. In this context, recent advances in Computer Vision allow researchers to equip robots with new and powerful capabilities. In our research we are using a social robot, Pepper, as a museum tour guide. In particular, we are developing a vision-based approach to support people during a museum visit [22]. Pepper is a semi-humanoid robot on wheels, equipped with several cameras and sensors. The vision module allows Pepper to perceive the presence of visitors and locate them in space, estimating their age and gender. Additionally, the same visual link retrieval module previously described gives Pepper the ability to take the image of the painting viewed by the visitor as a visual query to search for visually similar paintings in the museum database. The robot uses these data and other information acquired during the dialogue to provide the visitors with advice on similar artworks that they would like to see in the museum.

Designing the behaviors of a social robot that acts as a museum guide requires equipping it with different skills that provide visitors with an engaging and effective experience during the visit. These capabilities are intended to allow the robot to detect and locate people in the museum, recognize the artwork the visitor is looking at, profile the user during the visit in order to generate adequate recommendations and finally involve people in the interaction using adequate conversation skills. We have tested the proposed approach in our research laboratory and preliminary experiments have shown its feasibility.

---

<sup>1</sup><https://www.wikiart.org/>

<sup>2</sup><https://www.dbpedia.org/>

<sup>3</sup><https://neo4j.com/>

## 6. Conclusion

The growing availability of large collections of digitized artworks has given rise to an intriguing new area of research where Computer Vision and the visual arts meet. The new research area is framed as a constantly growing subfield of *Digital Humanities*, which aims to bring together digital technologies and humanities. The applications are innumerable and range from information retrieval in digital databases to the synthetic generation of new forms of art.

Encouraged by the growing literature that has emerged on the topic, we are working to make a contribution. We are confident that this exciting field of research will be strengthened in the future by leveraging the rapid advances in Deep Learning approaches. We believe that these approaches will continue to evolve rapidly, thus paving the way for the realization of amazing scenarios in which computer systems will be able to analyze and understand fine arts on their own. In fact, one of the final objectives of this research is the ability of the machine, once a picture has been taken, to autonomously derive an understanding of what the scene depicts, what it metaphorically represents, what are the possible historical implications, etc., without any human guidance.

However, the artistic domain is significantly different from the natural/photo-realistic domain Computer Vision researchers usually work on. First, there is inherent variability between the stylistic and figurative characteristics of the two domains, as well as between works by different artists belonging to the same period, if not between works by the same artist. Furthermore, the datasets with which we now pre-train Deep Learning models are affected by “recentism”, and are not representative of situations, ways of being/dressing, iconographic and mythological scenes, etc., from the past, which never existed or simply no longer exist. In other words, the cultural heritage, given its historical background over the centuries, poses entirely new and intriguing scientific challenges which, if addressed, can push beyond the semantic scene understanding achieved by current models.

## Acknowledgments

G.V. acknowledges the financial support of the Italian Ministry of University and Research through the PON AIM 1852414 project.

## References

- [1] G. Castellano, G. Vessio, Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview, *Neural Computing and Applications* 33 (2021) 12263–12282.
- [2] E. Cetinic, J. She, Understanding and creating art with AI: Review and outlook, *arXiv preprint arXiv:2102.09109* (2021).
- [3] E. Cetinic, T. Lipic, S. Grgic, A deep learning perspective on beauty, sentiment, and remembrance of art, *IEEE Access* 7 (2019) 73694–73710.
- [4] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M. S. Lew, Deep learning for visual understanding: A review, *Neurocomputing* 187 (2016) 27–48.

- [5] E. Cetinic, T. Lipic, S. Grgic, Fine-tuning convolutional neural networks for fine art classification, *Expert Systems with Applications* 114 (2018) 107–118.
- [6] S. Liu, J. Yang, S. S. Agaian, C. Yuan, Novel features for art movement classification of portrait paintings, *Image and Vision Computing* 108 (2021) 104121.
- [7] S.-h. Zhong, X. Huang, Z. Xiao, Fine-art painting classification via two-channel dual path networks, *International Journal of Machine Learning and Cybernetics* 11 (2020) 137–152.
- [8] N. Garcia, B. Renoust, Y. Nakashima, ContextNet: representation and exploration for painting classification and retrieval in context, *International Journal of Multimedia Information Retrieval* 9 (2020) 17–30.
- [9] B. Seguin, C. Striolo, F. Kaplan, et al., Visual link retrieval in a database of paintings, in: *European conference on computer vision*, Springer, 2016, pp. 753–767.
- [10] X. Shen, A. A. Efros, M. Aubry, Discovering visual patterns in art collections with spatially-consistent feature learning, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9278–9287.
- [11] A. Elgammal, B. Liu, M. Elhoseiny, M. Mazzone, CAN: Creative adversarial networks, generating art by learning about styles and deviating from style norms, *arXiv preprint arXiv:1706.07068* (2017).
- [12] X. Gao, Y. Tian, Z. Qi, RPD-GAN: Learning to draw realistic paintings with generative adversarial network, *IEEE Transactions on Image Processing* 29 (2020) 8706–8720.
- [13] Y. Liu, Improved generative adversarial network and its application in image oil painting style transfer, *Image and Vision Computing* 105 (2021) 104087.
- [14] G. Castellano, E. Lella, G. Vessio, Visual link retrieval and knowledge discovery in painting datasets, *Multimedia Tools and Applications* 80 (2021) 6599–6616.
- [15] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, in: *International conference on machine learning*, PMLR, 2016, pp. 478–487.
- [16] G. Castellano, G. Vessio, A deep learning approach to clustering visual arts, *arXiv preprint arXiv:2106.06234* (2021).
- [17] H. Cai, Q. Wu, T. Corradi, P. Hall, The cross-depiction problem: Computer vision algorithms for recognising objects in artwork and in photographs, *arXiv preprint arXiv:1505.00110* (2015).
- [18] G. Castellano, G. Sansaro, G. Vessio, Integrating contextual knowledge to visual features for fine art classification, in: *Workshop on Deep Learning for Knowledge Graphs (DL4KG 2021)*, 2021.
- [19] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, *ACM Computing Surveys (CSUR)* 54 (2021) 1–37.
- [20] C. Brahim El Vaigh, N. Garcia, B. Renoust, C. Chu, Y. Nakashima, H. Nagahara, GCNBoost: Artwork classification by label propagation through a knowledge graph, *arXiv e-prints* (2021) arXiv-2105.
- [21] G. Castellano, B. De Carolis, F. D’Errico, N. Macchiarulo, V. Rossano, PeppeRecycle: Improving children’s attitude toward recycling by playing with a social robot, *International Journal of Social Robotics* 13 (2021) 97–111.
- [22] G. Castellano, B. De Carolis, N. Macchiarulo, G. Vessio, Pepper4Museum: Towards a human-like museum guide., in: *AVI<sup>2</sup>CH@AVI*, 2020.