Automatic analysis of artistic heritage through Artificial Intelligence

Giovanna **Castellano**[1], Raffaele **Scaringi**[1,*] and Gennaro **Vessio**[1]

[1]*Department of Computer Science, University of Bari Aldo Moro, Italy*

Abstract
Technological improvements have resulted in a large-scale digitization effort in recent years, leading to the increasing availability of large digitized art collections. This provides an opportunity to develop AI systems capable of understanding art, thus supporting art historians and enjoying culture more generally. This paper briefly reviews our ongoing project on automatic art heritage analysis through AI. In particular, new graph representation learning approaches, combined with computer vision, are investigated to handle the complexity of visual arts.

Keywords
Digital humanities, Deep learning, Computer vision, Graph representation learning

1. Motivations and objectives

The mass digitization of cultural heritage, which is continuously increasing nowadays [1, 2], has offered the scientific community the opportunity to develop computational methods, from knowledge-based systems [3] to generative models [4], in order to address tasks in the art domain. However, solving these tasks is challenging.

To confirm artwork analysis complexity, consider the task of artwork attribute recognition. Traditionally, this task has been performed using hand-crafted features and classic machine learning algorithms (e.g., [5]). However, despite the good results achieved, extracting features proved to be very difficult, basically due to the subjective perspective of the individual human expert. This first limitation has been overcome with the advent of deep learning [6], which can automate the feature extraction stage thanks to its effective representation learning capability. One of the first attempts in this direction was made by Karayev et al. [7], who used a pre-trained Convolutional Neural Network to recognize the school of painting of a given artwork. Nevertheless, artists usually paint using different styles and can represent something unreal. Since standard pre-trained deep neural networks are biased towards natural domains, they could not capture certain aspects fundamental to analyzing cultural heritage. Indeed, as studied by Cetinic et al. [8], feature representation is vital when an artistic task must be solved.

The growing interest in this research field at the intersection of AI and art requires further efforts to achieve technology transfer. Developing solutions in cultural heritage could represent a non-trivial source for art experts and casual users. For example, extracting significant and art-oriented descriptions can be plugged into smart glasses to let blind people appreciate art. Moreover, throughout history, natural disasters have damaged many artworks: developing models capable of reconstructing these artworks could be a great tool for art experts. Our research at the Department of Computer Science, University of Bari, fits into this context. The rest of the paper outlines the approach we are exploring to address challenges in this domain. Finally, an overview of our current research and expected results concludes the paper.

2. Problem approach

As always, in the machine and deep learning community, the first step in creating effective models is to have a large and representative dataset available. A solid starting point is $\mathcal{A}rt\mathcal{G}raph$ [3], which consists of a large Knowledge Graph (KG) regarding cultural heritage, including information about artworks and their authors from different perspectives. However, it is extensible to include other data, such as textual descriptions, to enrich the encoded knowledge. $\mathcal{A}rt\mathcal{G}raph$ is saved in Neo4j, which already provides information retrieval and knowledge discovery capabilities even without training learning algorithms, using the Cypher query language. For example, Fig. 1 shows the subgraph related to "The Last Supper" by Leonardo da Vinci: all the metadata directly associated with the artwork include many different information, from the materials with which the artwork was made to the people depicted.

Once the data are available, some research problems can be addressed. As a first step toward a system that can understand art, we are interested in artwork attribute recognition and, more specifically, neuro-symbolic models. This model is suitable for the project's goal because it can jointly exploit different modes of information related to the images and metadata stored in the KG. In particular, Graph Neural Networks can be used to exploit graph features, a state-of-the-art approach to extract meaningful features from the graph and use them for downstream tasks [9]. Intimately related to attribute recognition is recognizing the emotion an image evokes in the observer. In [10], the authors have presented a full transformer-like architecture; in particular, they used the *ArtEmis* dataset [11], which provides utterances describing the motivation behind an elicited emotion.

Another relevant task, which could be helpful for historians and art experts, is the recognition of influences among artists. To this end, it is interesting to construct a method for obtaining such information by reconstructing the history of artistic influences. In particular, from a methodological point of view, also this task can be approached using the KG, in which all these relationships are stored, and solving a link prediction task.

Finally, generative algorithms can automate the generation of images and sequences. In this area, Diffusion Models [12] and Generative Adversarial Networks [13] represent the current state-of-the-art and can be considered to solve the task of artwork in-painting, the purpose of which is to reconstruct damaged artworks. Some generative methods are even multimodal. One example is DALL·E-2,[1] developed by OpenAI. This model can generate an image in many

---

[1]https://openai.com/dall-e-2/

Figure 1: Subgraph related to "The Last Supper". It contains information including style, genre, people depicted, and so on.

ways based on text prompts. On the other hand, image captioning can generate meaningful descriptions of artworks. For this purpose, methods based on natural language processing are crucial and will be explored. Unfortunately, image captioning systems that work well with natural images often fail when asked to generate output from an art image because they lack the richness and depth that a historical background would provide. This is confirmed by the study conducted by Cetinic et al. [14]. Developing a system that can mimic a human expert is a long-term goal of this community research line.

3. Current research and expected results

We are currently developing a tool for solving artwork attribute prediction, a preliminary version of which was presented in [15]. The main idea is to exploit "contextual" information provided by $\mathcal{ArtGraph}$, in combination with visual information of the given artwork, extracted respectively by a Graph Attention Network [16] and a Vision Transformer [17]. More specifically, we are trying to address the problem of predicting style, genre, and the evoked emotion of a painting. In this way, we are figuring out a practical approach to recognize those attributes, extending pure computer vision methods.

Regarding long-term objectives, we would like to develop a recommender system that could represent a digitized art gallery where general users can purchase a specific artwork based on what they like most. Alternatively, the same system can be used to provide customized tours in a museum. In fact, in this way, the tool can return and rank for the user a subset of the artworks, basing the decision on meta information. Along with artwork recommendation, there is another

essential task: artwork captioning. In fact, when returning a selected artwork, a significant description is needed. To this end, our goal is to develop a system capable of generating a caption that includes a description based on the visual content and adds some other information, such as the hidden message the artist wants to communicate to the observer.

Finally, we would like to develop an end-to-end method for in-painting to support art experts consistently. In particular, the tool should be able to reconstruct damaged images, varying the generation based on metadata or text prompts in which the experts specify some constraints for the reconstruction. For example, the user requires that the presence of a specific object or person in the reconstructed area is compulsory. Alternatively, the experts could be interested in a reconstruction based on a specific painting style. A promising starting point in this direction is the work of Cipolina-Kun et al. [18].

References

[1] N.-A. Ypsilantis, N. Garcia, G. Han, S. Ibrahimi, N. Van Noord, G. Tolias, The met dataset: Instance-level recognition for artworks, in: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), 2021.

[2] A. Reshetnikov, M.-C. Marinescu, J. M. Lopez, DEArt: Dataset of European Art, arXiv preprint arXiv:2211.01226 (2022).

[3] G. Castellano, V. Digeno, G. Sansaro, G. Vessio, Leveraging Knowledge Graphs and Deep Learning for automatic art analysis, Knowledge-Based Systems 248 (2022) 108859.

[4] R. Rombach, A. Blattmann, B. Ommer, Text-Guided Synthesis of Artistic Images with Retrieval-Augmented Diffusion Models, arXiv preprint arXiv:2207.13038 (2022).

[5] G. Carneiro, N. P. d. Silva, A. D. Bue, J. P. Costeira, Artistic image classification: An analysis on the printart database, in: European conference on computer vision, Springer, 2012, pp. 143–157.

[6] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.

[7] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, H. Winnemoeller, Recognizing Image Style, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2014.

[8] E. Cetinic, T. Lipic, S. Grgic, Fine-tuning Convolutional Neural Networks for fine art classification, Expert Systems with Applications 114 (2018) 107–118.

[9] W. L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications, arXiv preprint arXiv:1709.05584 (2017).

[10] D. Bose, K. Somandepalli, S. Kundu, R. Lahiri, J. Gratch, S. Narayanan, Understanding of Emotion Perception from Art, arXiv preprint arXiv:2110.06486 (2021).

[11] P. Achlioptas, M. Ovsjanikov, K. Haydarov, M. Elhoseiny, L. J. Guibas, ArtEmis: Affective Language for Visual Art, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 11569–11579.

[12] H. Cao, C. Tan, Z. Gao, G. Chen, P.-A. Heng, S. Z. Li, A survey on generative diffusion model, arXiv preprint arXiv:2209.02646 (2022).

[13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, Gen-

erative adversarial networks: An overview, IEEE signal processing magazine 35 (2018) 53–65.

[14] E. Cetinic, Towards Generating and Evaluating Iconographic Image Captions of Artworks, Journal of Imaging 7 (2021).

[15] S. Aslan, G. Castellano, V. Digeno, G. Migailo, R. Scaringi, G. Vessio, Recognizing the Emotions Evoked by Artworks Through Visual Features and Knowledge Graph-Embeddings, in: International Conference on Image Analysis and Processing, Springer, 2022, pp. 129–140.

[16] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, arXiv preprint arXiv:1710.10903 (2017).

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).

[18] L. Cipolina-Kun, S. Caenazzo, G. Mazzei, Comparison of CoModGans, LaMa and GLIDE for Art Inpainting Completing M.C Escher's Print Gallery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2022, pp. 716–724.