The Art Historian's Bicycle Becomes an E-Bike

Etienne Posthumus[1][0000−0002−0006−7542], Hans Brandhorst[2][0000−0001−8403−3552], and Harald Sack[3][0000−0001−7069−9804]

[1] FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Germany
etienne.posthumus@partners.fiz-karlsruhe.de
[2] Henri Van de Waal Foundation, The Netherlands
info@henrivandewaalfoundation.org
[3] FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Germany
harald.sack@fiz-karlsruhe.de

Abstract. This paper describes our experience in combining a large corpus of material that was classified manually over decades of Art Historical and Book History research using the ICONCLASS[4] subject classification system with heuristics based on a current state-of-the-art neural network (CLIP from OpenAI), leveraging visual similarity to provide suggestions for automated classification of cultural heritage content. The effectiveness of the approach is demonstrated through an evaluation of the underlying extreme multi-label classification problem.

Keywords: Art History · Book History · Subject Classification · Computer Vision

1 Introduction

"Computers are bicycles for the mind" - Steve Jobs

Since its creation in the 1940s, the ICONCLASS subject classification system has been considered as "the art historian's bicycle." It is not meant to replace the work done by cataloguers and art historians, but to make their work go "faster and farther". With the advances in computer vision tools that became practically accessible for working on image collections of reasonably large sizes, the art historian's bicycle has been given an upgrade to that enables us to explore even larger territories at a much greater speed. We were able to speed up the image description and information retrieval tasks used in Digital Art History research by using the CLIP[5] (Contrastive Language-Image Pre-Training) neural network from OpenAI trained on a variety of image/text pairs combined with a selection heuristic and a collection of c. 500K images classified using the ICONCLASS system.

ICONCLASS is a language independent subject classification system, originally devised in the Netherlands by Henri van de Waal in the 1940's. In 1968

[4] https://iconclass.org/
[5] https://openai.com/blog/clip/

the system was used to publish D.I.A.L the Decimal Index of Art of the Low Countries, and later as an independent series of 17 printed volumes between 1973 and 1985 [9]. The system has been applied to a large and diverse collection of images (and text) by independent scholars across the world.[6] Collections that are tagged with ICONCLASS concepts allow for sophisticated retrieval, smooth transition from narrow to broader queries, and multilingual word searches. But this flexibility comes at a price: for each image that is described, significant efforts involved at collection description time could be needed to determine which relevant notations to assign. There is no single "correct" notation, it all depends on how much time(money) is available to describe each image. In the paper version, this was a laborious process which has since been sped up by allowing fulltext searches in the web-based ICONCLASS versions. But the cataloguing activity is still a highly-specalized professional task that requires advanced iconographical expertise. It would benefit heritage institutions if we could dramatically accelerate the process with the help of new software tools, integrated in the online ICONCLASS browser. Devising a method that could speed up the classification process would enable a more extensive description of collections,and also assist the cataloguer with obscure or iconographically difficult to decipher items.

Efforts to use neural networks to study historical images have been made [10], but many of the pre-trained models were not suitable for historical material. Training large models were hobbled by the lack of good quality input data, and the prohibitively large number of trained classes needed to pose relevant art historical questions. We were able to speed up the image description and information retrieval tasks used in Digital Art History research by combining CLIP, which is a neural network trained on a variety of image/text pairs, with a selection heuristic and a collection of c. 500K images classified using the ICONCLASS system. [8, 5].

The digital revolution has only had a limited effect on the theory and practice of iconographic documentation in Art History. The introductory chapter to Erwin Panofsky's Studies in Iconology still serves as the main theoretical basis for the description of the subject matter of works of art. Seen from the perspective of information science, Panofsky's method has serious limitations. Originally published in 1932 and mostly widespread in its 1939 English translation, it did not possibly foresee the almost limitless expansion of our shared memory capacity caused by the massive digitization of visual resources.[1, 7] By extending the ICONCLASS browser with CLIP's pattern matching functionality, ICONCLASS concepts can be applied to images much faster and with far greater consistency. The corpus of already tagged images enables the user to switch from visual (pattern) matching to thematic comparisons. Visually similar images can be retrieved with the help of pictures, but as they have been tagged manually with ICONCLASS concepts the user can easily switch to thematic queries and retrieve visually dissimilar objects.The oscillation between pattern and thematic matching mimics the core process of iconographic research. Paradoxically, the 'pre-iconographic' perception of patterns, which Panofsky postulates as the first

[6] Systems using Iconclass for subject access, https://iconclass.org/help/aboutc

phase of iconographic analysis, is a theoretical rather than an actual option in human observation. However, it is a 'natural' starting point for pattern matching software. Our challenge is to optimize the software tools and create a laboratory for the study of image content that actually accommodates the working practice of a humanities researcher. By providing greater facilities for a quicker and a better description of images with a standardized, multilingual vocabulary at their disposal, it will be much easier to convince heritage institutions of the benefits of using ICONCLASS.

2 User Interface & Data

As a proof-of-concept user interface, the option to conduct visual searches using sample images has been added to the ICONCLASS website. A user can drag-and-drop an image from their local filesystem, or copy & paste an image from a website. The destination is the regular search box where one would normally type a textual search query, but this is enhanced to also accept visual input. See Fig. 1 for an example similarity search. Once an input image is chosen, a

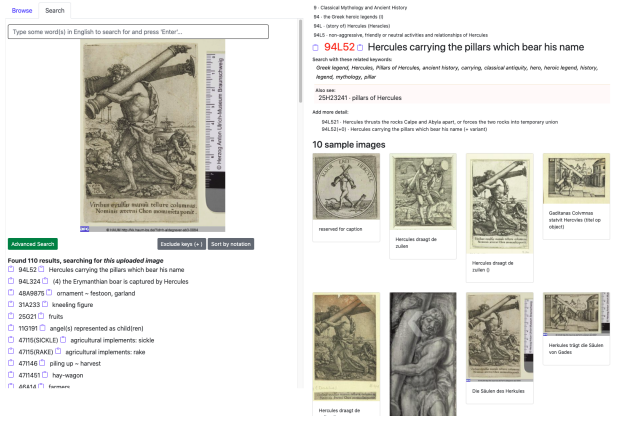

Fig. 1. Screenshot of the visual similarity search

search is performed over the collection of existent images after which the closest matches are shown. The notations are then matched using a simple frequency based heuristic and shown as classification suggestions.

3 Technical Details

The image collection used for this prototype contained: 531,172 images, with 2,526,145 assigned notations of which 90,347 were unique. A Faiss index of 2GB is needed to provide the KNN index [3].

Classification metric report An evaluation for the extreme multi-label classification problem [4] has been performed using a 10K item sample from the data which distributed over 4,341 different target classes. The ground truth of the 10k sample images contains overall 4,341 different ICONCLASS notations, each image manually annotated with 3 to 6 ICONCLASS codes. However, the class distribution, i.e. the number of annotated images per ICONCLASS notation, is highly imbalanced. Each notation from the 4,341 different ICONCLASS notations on the average is used for only 46.2 images (average class support) with a standard deviation of 1,218. The achieved results for this extreme multi-label classification problem are reported in table 1. The overall micro-F_1 score of 0.29 has been achieved with a simple similarity based search heuristics (cf. below) that takes into account the k nearest neighbors above a given threshold. Considering the high number of 4,341 imbalanced classes the proposed heuristics performs well in comparison with other deep learning based multi-label classification models for iconography classification that are typically applied on smaller samples or a smaller number of assigned icon codes per image [6, 2].

Search Heuristic When a user uploads a target image to the website, a search is first performed for the top n matching images from the KNN index, based on visual similarity. For each matched image, the assigned ICONCLASS notations are retrieved, and the notations are then sorted by the number of assigned images in reverse order. The list of possible notations are shown to the user as a list of possible suggested notations. Clicking on a matched image repeats the search process, using that selected image as the new seed of a similarity search.

4 Conclusion and Future work

This prototype has shown that it is feasible to satisfactorily do a speedy retrieval of visually matching items in a non-trivial iconographic dataset by using an off-the-shelf computervision model, without the costly training of a new neural network. Initial trials with practicing Art Historical researchers have shown a positive response with respondents reporting surprisingly salient suggestions in the complex domain of Iconographic research. From end user perspective, it is beneficial to improve the user interface to allow for multi-modal searches combining images and text. Moreover, the possibility of linking to the external sites from which the images are referenced as well as the possibility to upload new images with periodical updates to the search index and to assign new ICONCLASS notations can and will be investigated.

Table 1. Evaluation results for a 10K item data sample.

	precision	**recall**	F_1
micro avg	0.30	0.28	0.29
macro avg	0.19	0.20	0.18

References

1. Arnold, T.B., Scagliola, S., Tilton, L., van Gorp, J.: Introduction: Special issue on audiovisual data in DH. Digit. Humanit. Q. **15**(1) (2021), http://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html
2. Banar, N., Daelemans, W., Kestemont, M.: Multi-modal label retrieval for the visual arts: The case of iconclass. ICAART 2021 - Proceedings of the 13th International Conference on Agents and Artificial Intelligence **1**, 622–632 (2021)
3. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with gpus. CoRR **abs/1702.08734** (2017), http://arxiv.org/abs/1702.08734
4. Liu, W., Shen, X., Wang, H., Tsang, I.W.: The emerging trends of multi-label learning. CoRR **abs/2011.11197** (2020), https://arxiv.org/abs/2011.11197
5. Marinescu, M.C., Reshetnikov, A., López, J.M.: Improving object detection in paintings based on time contexts. In: 2020 International Conference on Data Mining Workshops (ICDMW). pp. 926–932 (2020). https://doi.org/10.1109/ICDMW51313.2020.00133
6. Milani, F., Fraternali, P.: A dataset and a convolutional model for iconography classification in paintings. J. Comput. Cult. Herit. **14**(4) (jul 2021). https://doi.org/10.1145/3458885, https://doi.org/10.1145/3458885
7. Panofsky, E.: Studies in Iconology: Humanistic Themes in the Art of the Renaissance. Harper & Row (1972)
8. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. CoRR **abs/2103.00020** (2021), https://arxiv.org/abs/2103.00020
9. Van de Waal, H.: Decimal index of the art of the Low Countries; D.I.A.L. Rijksbureau voor Kunsthistorische Documentatie, The Hague (1968)
10. Wevers, M., Smits, T.: The visual digital turn: Using neural networks to study historical images. Digit. Scholarsh. Humanit. **35**, 194–207 (2020)