imgs.ai. A Deep Visual Search Engine for Digital Art History

Offert, Fabian
offert@ucsb.edu
University of California, Santa Barbara, United States of America

Bell, Peter
peter.bell@uni-marburg.de
Philipps-Universität Marburg, Germany

Art history, as a discipline, is concerned with *multiple* images. A *singular* image can only become a historical entity in relation to its 'neighbors', to similar and dissimilar, related and unrelated original works. The significance of this comparative element has been emphasized already by Wölfflin (1917, see also Bruhn and Scholtz 2017) and has been epitomized by Aby Warburg's (2010, see also Gombrich 1970, Didi-Huberman 2017) idiosyncratic method of tracing iconographic elements across history. Any singular image, in other words, leads to an *image corpus*. Institutional art historical image corpora commonly consist of images and related metadata but search operations often target metadata exclusively. This is a result of the historical importance of periodization, attribution, and localization in art history. Many art historical questions, however, specifically those related to both semantic (e.g. iconography) *and* syntactic (e.g. style) aspects of an image, are irreducible to metadata. Such questions, then, can only be operationalized either as visual queries, that is, in the form of sample images that possess a unique combination of visual properties representing the query, or as fuzzy textual queries, that is, free-form descriptions of visual properties.

Imgs.ai, which has been in public beta since the fall of 2020, addresses this combined CV-HCI challenge of distant viewing (Arnold and Tilton, 2019, see also Wevers and Smits 2020) by means of deep visual search by providing a Web-based interface (fig. 1), and machine-learning based backend that allow the end user to both descriptively and visually search arbitrary image datasets, with multiple significant institutional datasets already indexed. It was the first publicly available digital art history application to implement a multimodal approach to deep visual search in early 2021. We first review the state of the art in three classes of deep visual search tools (toolkits, macro interfaces, and search interfaces) and then proceed to describe both the front- and backend aspects of imgs.ai.

Imgs.ai is set up to index datasets based on four kinds of feature extractors. VGG19 (Simonyan et al. 2015) is a convolutional neural network with a focus on stylistic similarity. The pose feature extractor is built on top of a Keypoint R-CNN model with a ResNet-50-FPN backbone (He et al. 2017), which is especially useful for datasets of figurative works, as shown for instance by Impett and Moretti (2017). The "raw" feature extractor simply treats a (resized) image's color data as its embedding, and thus allows searching for images that use similar pallets. Finally, the CLIP extractor uses the pre-trained model of the same name (Radford et al. 2021) and enables multimodal search.

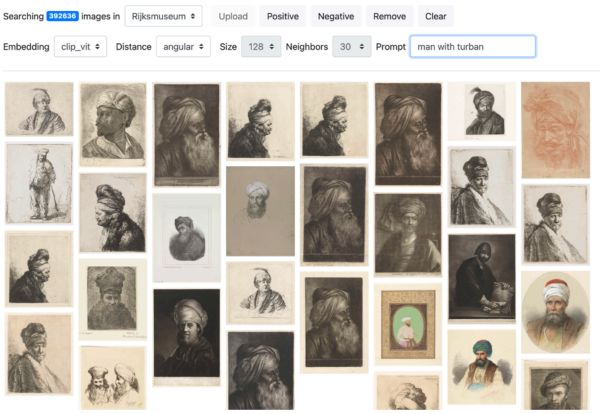

Figure 1: CLIP-based imgs.ai search in the Rijksmuseum collection. The toolbar allows to switch datasets or embeddings, add an image as a positive or negative example, or begin a new search.

One surprising exemplary result facilitated by CLIP that we describe in this paper is the list of search results for the prompt "Las Meninas", which references the famous 1656 painting by Diego Velásquez. The painting is famous, in particular, for its play on representation. Such metapictorial aspects, however, are rarely included in the metadata of a work. If we run a search for "Las Meninas" in the collection of the Museum of Modern Art, New York, the results show the conceptual depth that CLIP facilitates. Among them are two photographic works, Joel Meyerowitz' *Untitled* from *The French Portfolio* (1980) and Robert Doisneau's *La Dame Indignée* (1948). Both are explicit plays on representation and clearly pick up on the same themes as *Las Meninas*, especially the question of the gaze relation between people in, and people before the image, to use George Didi-Huberman's term. Another result is Richard Hamilton's *Picasso's Meninas* from *Homage to Picasso* (1973) which takes up the structure of the Velásquez original but fills it with figures from Picasso paintings. Here, imgs.ai, via CLIP, picks up on the compositional similarity of both works.

We finally discuss how imgs.ai is just one of potentially many solutions to the productive use of feature extraction (see Zhang et al. 2018) in digital art history. Given the increasing footprint of machine learning models it seems counterproductive to extract features more than once. We thus argue that the standardization of extracted features is thus the next big challenge the digital art history community has to solve.

Bibliography

Arnold, Taylor / **Tilton, Lauren** (2019): "Distant Viewing: Analyzing Large Visual Corpora", in: Digital Scholarship in the Humanities.

Bruhn, Matthias / **Scholtz, Gerhard** (2017): *Der vergleichende Blick*. Berlin: Dietrich Reimer Verlag.

Didi-Huberman, Georges (2017): *The Surviving Image: Phantoms of Time and Time of Phantoms: Aby Warburg's History of Art*. Pennsylvania State University Press.

Gombrich, Ernst H. (1970): *Aby Warburg*. Warburg Institute, University of London.

He, Kaiming / **Gkioxari, Georgia** / **Dollár, Piotr** / **Girshick, Ross B.** (2017): "Mask R-CNN." ArXiv Preprint 1703.06870.

Impett, Leonardo / **Moretti, Franco** (2017): "Totentanz. Operationalizing Aby Warburg's Pathosformeln", in: New Left Review 107.

Radford, Alec / **Kim, Jong Wook** / **Hallacy, Chris** / **Ramesh, Aditya** / **Goh, Gabriel** / **Agarwal, Sandhini** / **Sastry, Girish et al.** (2021): "Learning Transferable Visual Models from Natural Language Supervision", in: International Conference on Machine Learning (ICML): 8748–63.

Simonyan, Karen / **Zisserman, Andrew** (2015): "Very Deep Convolutional Networks for Large-Scale Image Recognition." ArXiv Preprint 1409.1556.

Warburg, Aby (2010): "Mnemosyne Einleitung", in: *Werke*, ed. by Martin Treml / Sigrid Weigel / Perdita Ladwig. Frankfurt am Main: Suhrkamp.

Wevers, Melvin / **Smits, Thomas** (2020): "The Visual Digital Turn: Using Neural Networks to Study Historical Images", in: Digital Scholarship in the Humanities 35, 1: 194–207.

Wölfflin, Heinrich (1917): *Kunstgeschichtliche Grundbegriffe. Das Problem der Stilentwicklung in der neueren Kunst*. München: Verlag Hugo Bruckmann.

Zhang, Richard / **Isola, Phillip** / **Efros, Alexei A.** / **Shechtman, Eli** / **Wang, Oliver** (2018): "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: 586–95.