iART: A Search Engine for Art-Historical Images to Support Research in the Humanities

Matthias Springstein
TIB – Leibniz Information Centre for Science and Technology
Hanover, Germany
matthias.springstein@tib.eu

Stefanie Schneider
Ludwig Maximilian University of Munich
Munich, Germany
stefanie.schneider@itg.uni-muenchen.de

Javad Rahnama
University Paderborn
Paderborn, Germany
javad.rahnama@uni-paderborn.de

Eyke Hüllermeier
Ludwig Maximilian University of Munich
Munich, Germany
eyke@ifi.lmu.de

Hubertus Kohle
Ludwig Maximilian University of Munich
Munich, Germany
hubertus.kohle@lmu.de

Ralph Ewerth
TIB–Leibniz Information Centre for Science and Technology;
L3S Research Center, Leibniz University Hannover
Hanover, Germany
ralph.ewerth@tib.eu

ABSTRACT

In this paper, we introduce iART: an open Web platform for art-historical research that facilitates the process of comparative vision. The system integrates various machine learning techniques for keyword- and content-based image retrieval as well as category formation via clustering. An intuitive GUI supports users to define queries and explore results. By using a state-of-the-art cross-modal deep learning approach, it is possible to search for concepts that were not previously detected by trained classification models. Art-historical objects from large, openly licensed collections such as Amsterdam Rijksmuseum and Wikidata are made available to users.

CCS CONCEPTS

• **Information systems** → *Multimedia and multimodal retrieval*; **Web applications**; • **Human-centered computing** → Information visualization.

KEYWORDS

Web application, Cross-modal retrieval, Deep learning, Art retrieval

1 INTRODUCTION

Basic art-historical techniques of analysis are essentially built on comparative processes. Heinrich Wölfflin, e.g., practised comparative vision in determining the stylistic history of the Renaissance and the Baroque, which he interpreted antagonistically [19]. However, open Web platforms that promote such image-oriented research processes by identifying objects that are similar to each

other are currently not available. Previous approaches lack either fine-tuning to the art-historical domain [16], flexible search query structures that adapt to users' needs [8], or the ability for users to upload their own datasets [12].

In this paper, we present iART[1]: an e-research-tool that analyses structures (or similarities) of a group of images by processing large, heterogeneous, and digitally available databases of art-historical objects through machine learning. Ordering criteria that were already common in early modern *Wunderkammern*, such as colour, material, or function, can be applied as well as more iconographically based classification principles that, e.g., examine objects for biblical motifs or Christian themes. Decisive is that these principles can be liquefied and reconfigured. The iART platform has been designed in such a way that it can be easily extended with the help of plug-ins to meet the various requirements of art historians. The retrieval of objects is not only performed with automatically generated keywords but also by utilizing state-of-the-art multimodal embeddings that enable search based on accurate, detailed scene descriptions given by the user.

2 SYSTEM ARCHITECTURE

The software was designed to be as modular as possible to facilitate adaptation to different research interests: the individual indexing steps are outsourced to plug-ins and user administration is separated from the search infrastructure. All models are accelerated with a RedisAI inference server to manage the resources needed for computation optimally. This step makes it easier to run different deep learning models on a single GPU and enables back end systems such as PyTorch [13] or TensorFlow [1]. The Web front end integrates the Vue framework and communicates with the indexer via a Django Web service. Figure 1a displays the architecture of iART.

3 IMAGE RETRIEVAL

Due to the diffuseness of the concept of similarity, iART supports different types of search queries that can be precisely targeted by the user. Hence, the underlying system was developed with the idea that each processing step should be extensible through a plug-in structure. This applies in particular to feature extraction, image classification, ranking of results, and various post-processing steps

---

[1]https://labs.tib.eu/iart, accessed: 2021-06-15.
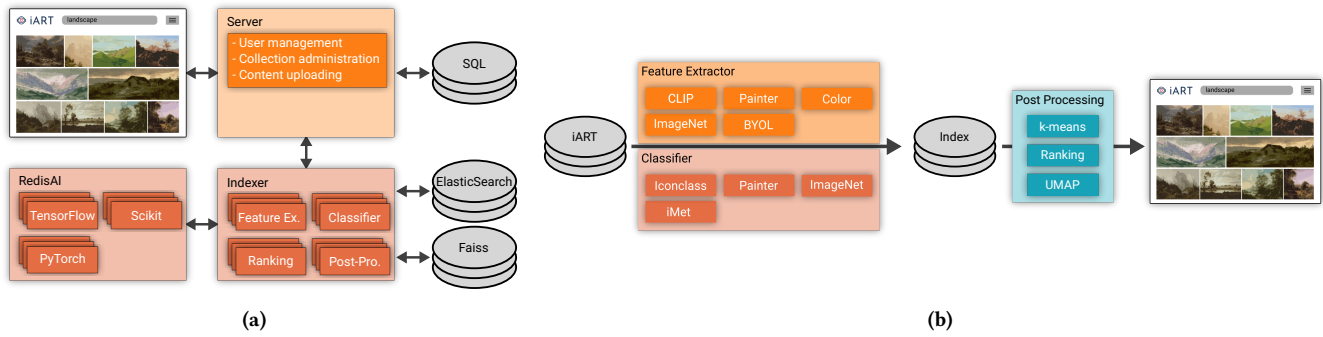
Figure 1: (a) Architecture with associated database structure and RedisAI inference server; (b) indexing and post-processing steps with relevant plug-ins. The generated keywords and features are stored with the help of Elasticsearch and Faiss [7].

that serve visualization and clustering. The complete pipeline is shown in Figure 1b.

Common feature extractors are integrated into iART, e.g., for ImageNet embeddings through a pre-trained ResNet [5]. These are complemented by models adjusted to the art-historical domain: (1) the self-supervised model BYOL (Bootstrap Your Own Latent) is trained on an adequate subset of Wikimedia images [4]; (2) while the Painter model utilizes the Painter by Numbers dataset [10] to extract features for style and genre; (3) and the transformer-based neural network CLIP (Contrastive Language-Image Pre-Training) learns visual concepts from natural language supervision [14]. Moreover, different classification models are trained to automatically predict art-historically relevant phenomena collected from Iconclass [6], iMet [11], and Painter by Numbers [10].

The extracted features primarily enable the user to retrieve similar images based on a query image. As the system extracts different embeddings for each image, the user can change the weighting of plug-ins and thus adjust the order of results according to his or her needs. Through its two decoder structures, CLIP creates a unified feature space for image and text, allowing the user to also enter textual descriptions. By weighting reference images and text queries, it is possible to create more complex search requests. For example, a reference image of Saint Sebastian can be combined with the text query "crucifixion" (Figure 2a). Using a faceted search, the list of results can be further narrowed down based on classified attributes and the metadata given by the respective collection. This helps users to filter their uploaded inventories and about one million openly licensed images, including examples from Amsterdam Rijksmuseum [15], Wikidata [18], Kenom [3], and ARTigo [17][2].

4 RESULT VISUALIZATION

To simplify the exploration of the results, different object views are implemented. By default, an image grid sorted by relevance is displayed, via which further details are provided on demand, such as metadata from the respective object. Results can be clustered with k-means and visualized, e.g., as image carousels vertically separated by groups. For more advanced use cases, it is possible to arrange the images on a two-dimensional canvas using the dimensionality reduction technique UMAP (Uniform Manifold Approximation and Projection) [9]. Zoom and filter operations, such as an interactive

(a)

(b)
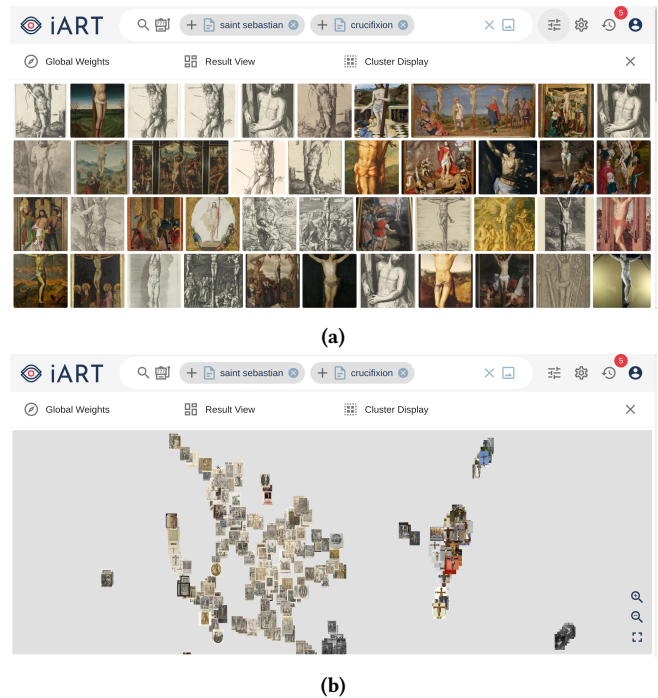
Figure 2: Search results for a reference image of Saint Sebastian, combined with the text query "crucifixion." (a) Default image grid; (b) two-dimensional canvas view.

drag-select to juxtapose multiple objects, are supported with the aid of VisJs (Figure 2b).

5 CONCLUSION

With iART, we introduced an open Web platform for art-historical research that facilitates the process of comparative vision. As the system is extensible and supports various classification plug-ins and feature extractors, users can adapt it to their needs. In the future, the system will be enriched with additional openly licensed datasets. Further plug-ins will also be integrated, e.g., feature extractors for human body pose or image composition.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR* abs/1603.04467 (2016). arXiv:1603.04467 http://arxiv.org/abs/1603.04467

[2] Matthias Becker, Martin Bogner, Fabian Bross, François Bry, Caterina Campanella, Laura Commare, Silvia Cramerotti, Katharina Jakob, Martin Josko, Fabian Kneißl, Hubertus Kohle, Thomas Krefeld, Elena Levushkina, Stephan Lücke, Alessandra Puglisi, Anke Regner, Christian Riepl, Clemens Schefels, Corina Schemainda, Eva Schmidt, Stefanie Schneider, Gerhard Schön, Klaus Schulz, Franz Siglmüller, Bartholomäus Steinmayr, Florian Störkle, Iris Teske, and Christoph Wieser. 2018. ARTigo – Social Image Tagging [Dataset and Images]. https://doi.org/10.5282/ubm/data.136..

[3] Reiner Diedrichs. 2021. *Kenom Digitaler Münzkatalog*. Retrieved June 15, 2021 from https://www.kenom.de

[4] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Ávila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. 2020. Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). https://proceedings.neurips.cc/paper/2020/hash/f3ada80d5c4ee70142b17b8192b2958e-Abstract.html

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[6] Iconclass. 2021. *Iconclass*. Retrieved June 15, 2021 from http://www.iconclass.org

[7] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2017. Billion-Scale Similarity Search with GPUs. *CoRR* (2017). http://arxiv.org/abs/1702.08734

[8] Sabine Lang and Björn Ommer. 2018. Attesting Similarity: Supporting the Organization and Study of Art Image Collections with Computer Vision. *Digital Scholarship in the Humanities* 33, 4 (2018), 845–856. https://doi.org/10.1093/llc/fqy006

[9] Leland McInnes and John Healy. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *CoRR* (2018). http://arxiv.org/abs/1802.03426

[10] Kiri Nichol. 2016. *Painter by Numbers*. Retrieved June 15, 2021 from https://www.kaggle.com/c/painter-by-numbers

[11] The Metropolitan Museum of Art. 2020. iMet Collection 2020 - FGVC7. Retrieved June 15, 2021 from https://www.kaggle.com/c/imet-2020-fgvc7

[12] Fabian Offert, Peter Bell, and Oleg Harlamov. 2020. imgs.ai. https://imgs.ai/. Accessed: 2021-06-15.

[13] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.). 8024–8035. https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

[14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *CoRR* (2021). https://arxiv.org/abs/2103.00020

[15] Rijksmuseum. 2021. *Rijksmuseum Amsterdam, Home of the Dutch masters*. Retrieved June 15, 2021 from https://www.rijksmuseum.nl/en

[16] Luca Rossetto, Ivan Giangreco, Claudiu Tanase, and Heiko Schuldt. 2016. vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *24th ACM International Conference on Multimedia*. 1183–1186. https://dl.acm.org/doi/10.1145/2964284.2973797

[17] Christoph Wieser, François Bry, Alexandre Bérard, and Richard Lagrange. 2013. ARTigo: Building an Artwork Search Engine with Games and Higher-Order Latent Semantic Analysis. In *Disco 2013, Workshop on Human Computation and Machine Learning in Games at the International Conference on Human Computation (HComp)*. https://www.en.pms.ifi.lmu.de/publications/PMS-FB/PMS-FB-2013-3/PMS-FB-2013-3-paper.pdf

[18] Wikimedia. 2019. *Wikidata*. Retrieved June 15, 2021 from https://www.wikidata.org/wiki/Wikidata:Main_Page

[19] Heinrich Wölfflin. 1915. *Kunstgeschichtliche Grundbegriffe*. Bruckmann, Munich.