

Identification of image circulation by AI in large collections of historical photographs

General information

- Keywords: History of image agencies, historical photographs, deep learning, computer vision, multimodal foundation model, text and image relations
- Institutes: Université Paris Cité, Laboratoire de recherche sur les cultures anglophones (LARCA UMR 8225) and Laboratoire d'Informatique Paris Descartes (LIPADE UR 2517)
- co-PIs: Daniel FOLIARD (LARCA), Florence Cloppet (LIPADE), Camille Kurtz (LIPADE)
- Duration: 6 months (standard stipend). To start between February and April 2024
- Location: 45 rue des Saints-Pères, 75006 Paris (LIPADE)
- Application: Please send a cover letter and a CV to Florence Cloppet and Camille Kurtz (first.lastname@u-paris.fr). The position is opened until filled.

Proposed topic

Motivation There is a blind spot for both the humanities and computer science when one considers the mass of heritage photographs now available in online repositories. Historians cannot analyze them on actual scale with their usual tools. Current computer vision and artificial intelligence (CVAI) systems face limitations when dealing with historical images and the visual cultures they reflect: they struggle to “see” the past. This historical bias stems from the datasets used to train vision and machine learning models (e.g. ImageNet), which mainly consist of recent pictures. As part of a collaboration between LARCA and LIPADE, we propose to combine our expertise to address this gap by pushing the boundaries of historical investigation and research in CVAI for large collections of historical photographs. Early agencies developed in the late 19th century and quickly became major actors of visual culture on a global scale by facilitating the circulation and reproduction of images among illustrated periodicals through standardized practices. Their archives are mostly composed of individual photographic prints and negatives with no adequate written documentation to trace their origins and distribution. The lack of suitable tools for their exploration favors internalist perspectives with a narrow focus on specific agencies, nation-states, or geographical areas. Until now, traditional approaches have been limited in capturing the true scale, cultural underpinnings, and economics of early mass visual culture. Our long term project aims to address these challenges by treating press photographs as both big data that can be digitally processed and unique documents made available for historians, archivists and the wider public. The switch from micro to macro perspectives using CVAI will open up new methodological and historical possibilities. To weave a quantum physics metaphor, images will be analysed not only as particles, but also as waves. Guided by an archival and historical treatment of the processed data, this approach will enable a truly interconnected history of visual cultures on a global scale. The complex nature of these visual and semantic data presents an exciting field of exploration for research in CVAI, with potential for industrial applications. GLAM institutions are actively seeking solutions to enhance their capacity to process the ever-expanding number of visual materials they curate. The project will provide transferable solutions for semi-automated approaches to visual heritage. Advances in image similarity estimation, layout analysis, and automated enrichment of archival metadata will contribute to the growing field of applied CVAI for archives and digital humanities.

Background and state of the art The accelerating digitization of major archival institutions' holdings worldwide has provided internet users with unprecedented access to tens of millions of heritage images. However, the instruments for exploring and analyzing these documents have not kept pace with this mass availability. AI tools are leading to a dramatic improvement of the automatic extraction of visual information from archival records. Making historical archival records machine readable is now the goal of large-scale research initiatives such as L. Tilton's and A. Taylor's research on "distant viewing" applied to heritage photography. The Visual Contagions project team has also shown how extracting images from printed material can aid in searching for visual patterns across vast numbers of pages and thus offer new insights into twentieth-century industrial visual cultures. Given exciting early results, the potential of CVAI for supporting historical inquiry is clear. AI has demonstrated its ability to automatically discover relationships across records using metadata. There is an extensive literature on image similarity, classification, and automated captioning, yet existing approaches rarely apply to historical data. Historically informed approaches can provide fresh insights into the early visual culture of news. While the development of twentieth-

century news agencies and the impact of photography have been extensively studied, the exploration of the early news image ecosystem on a large scale has remained uncharted territory. For example, Victor Forbin's significant role in France has been disregarded because the lack of textual records documenting his activities has made it impractical to study his large picture archive using traditional methods. A similar situation affects the other picture agency collections that will be analyzed by the High-Vision project; though historians of photography are aware of their importance, they have lacked adequate methods to appreciate the scope of their activities until now.

Proposed work and implementation

Proposed approach and contributions to the state-of-the-art Given a collection of images, our objective in this project is to develop a methodology allowing us to follow the trajectory of a photographic image following the vagaries of its use in newspapers, its sales by press agencies and of its potential re-use in other journalistic contexts. From a computer science point of view, the starting point will be to consider this problem as a content-based image retrieval task where a user can, from a given image, query a database to find similar images.

To go beyond the state of the art in computer vision (based for several years on the use of convolutional neural networks (CNN) for the learning of a discriminative image representation), we will consider pre-trained “foundation” models (based on Transformers), with a careful control of the level of supervision of the algorithms to deal with the limited amount of annotated data. The available data (images, textual descriptions extracted from the document layouts and annotations already present in the metadata of well-curated collections) will be employed, via weakly supervised and few-shot strategies, to fine-tune the multi-modal foundation models, making it possible to simultaneously learn joint visual and textual representations via a contrastive paradigm. Once the image representations have been optimized, they can be used to find similar images in the datasets with metadata (and vice versa from texts and annotations to images). The members of LIPADE involved in this project have already achieved promising results by implementing such strategies in the context of medical imaging. As a supplementary originality, we plan also in this internship project to automatically extract image metadata by employing object detection, visual saliency detection within the images to enrich the image captions with more semantic and contextual information, allowing ultimately the foundation model to learn richer and more discriminative representations for the given task and image circulation search.

By conducting a large-scale search for close, retouched, or similar images, we can facilitate a successful historical investigation of pioneering news photo agencies. The collection of the circulation contexts of the photographs and a better understanding of their textual environments (such as re-captioning and resignifications in different editorial contexts) will offer new insights to analyze these vast visual data, at the crossroad of digital sciences research and work in the humanities focused on heritage images.

Considered data

- Agence Rol photographic archive (20000 + digitized historical photos from the 1900s to the 1920s) - Bibliothèque Nationale de France
- Forbin collection (1000 + digitized photos from the 1890s to the 1920s) -EyCon Project
- Bain News Photo collection (Library of Congress, Washington)
- Chronicling America Newspapers database (<https://chroniclingamerica.loc.gov/>)
- EyCon database - <https://eycon.sempiternelia.com/s/fr/page/corpora>
- Gallica - Digitized illustrated newspapers from the 1890s to the 1920s and relevant datasets (<https://api.bnf.fr/fr/documents-de-presse-numerises-en-mode-article>)

References

Benjamin Charles Germain Lee et al., “The Newspaper Navigator Dataset: Extracting and Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America,” ArXiv:2005.01583 [Cs], May 4, 2020, <http://arxiv.org/abs/2005.01583>
K. Chumachenko, A. Männistö, A. Iosifidis, J. Raitoharju, Machine learning based analysis of Finnish world war II photographers, IEEE Access 8 (2020) 144184–144196.

Maud Ehrmann, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, and Antoine Doucet. 2023. Named Entity Recognition and Classification in Historical Documents: A Survey. ACM Comput. Surv. 56, 2, Article 27 (February 2024), 47 pages. <https://doi.org/10.1145/3604931>

Xu, Jiaolong, Liang Xiao, and Antonio M. López. "Self-supervised domain adaptation for computer vision tasks." IEEE Access 7 (2019): 156694-156706.

M. Lincoln et al. (2020). CAMPI: Computer-Aided Metadata Generation for Photo archives Initiative. Carnegie Mellon University. Preprint. <https://doi.org/10.1184/R1/12791807.v2>

Nanne van Noord (2022). A survey of computational methods for iconic image analysis. Digital Scholarship in the Humanities, vol. 37/4. <https://doi.org/10.1093/llc/fqac003>

N. Vincent and J.M. Ogier (2019). Shall deep learning be the mandatory future of document analysis problems?. Pattern Rec. 86. <https://doi.org/10.1016/j.patcog.2018.09.010>

M. T. Smits, M. Wevers (2023). A multimodal turn in Digital Humanities. Digital Scholarship in the Humanities, 38/3.

<https://doi.org/10.1093/llc/fqad008> J. Xu et al. (2019). Self-supervised Domain Adaptation for Computer Vision Tasks. *IEEE Access*, vol. 7. <https://doi.org/10.1109/2Faccess.2019.2949697>

Wevers, M. J. H. F., Vriend, N., & De Bruin, A. (2022). What to do with 2.000.000 Historical Press Photos? The Challenges and Opportunities of Applying a Scene Detection Algorithm to a Digitised Press Photo Collection. *TMG – Journal for Media History*, 25(1) <https://doi.org/10.18146/tmg.815>