

Identification of image circulation in large collections of historical photographs

Master internship

General information

- Keywords: History of image agencies, historical photographs, computer vision, multimodal foundation model, representation learning, self-supervised approaches, layout segmentation
- Duration: 6 months (standard stipend). To start between February and April 2025.
- Institutes: Université Paris Cité, Laboratoire d'Informatique Paris Descartes (LIPADE, team [Systèmes Intelligents de Perception](#) and Laboratoire de recherche sur les cultures anglophones (LARCA UMR 8225)
- Location: 45 rue des Saints-Pères, 75006, Paris
- Supervision: Camille Kurtz, Florence Cloppet and Daniel Foliard
- Applications: send an email with CV and cover letter to (camille.kurtz@u-paris.fr)

Note: Depending on the progress of the internship and the results obtained, **this internship may be continued by a Phd thesis in Computer Science** on a scientific subject in line with the internship subject and the objectives of the ANR HighVision project. Funding for the thesis is already guaranteed.

Proposed topic

Context

This internship is part of the High-Vision project, funded by the ANR, aiming to advance historical investigations and research in computer vision and artificial intelligence (CVAI) for large collections of historical photographs. It seeks to harness state-of-the-art models for AI training data with a view to shape little-curated photographic archives into digital assets so that they can be investigated by historians, archivists, and the public on a large scale and in innovative ways. The project brings together Humanities scholars, computer vision experts, archivists and other stakeholders to produce innovative research at the juncture of computer science, archival studies and history. Drawing on research already conducted by the partners, the hypothesis underlying this project is that cross-seeding approaches between historical epistemologies and CVAI expertise can unlock new perspectives on the historical and computational analysis of large and poorly curated photographic archives to provide new insights on the visibility of early mass visual culture of the news.

Work to be done

The work to be conducted during the proposed M2 internship will contribute to the ambition of the HighVision ANR project, in collaboration between LIPADE and LARCA laboratories. Given a collection of digitised newspaper, 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 ([Figure 1](#)). From a computer science point of view, we consider this problem as a content-based image retrieval (CBIR) task where a user can, from a given image, query a database to find similar images.

▷ **Contribution 1 – Document layout segmentation:** Prior to that, it is required to extract from the pages the targeted visual content. The first step relies then on a layout segmentation task, in order to extract both newspaper illustrations and their captions. Different approaches will be tested and assessed under the objective to extract and to classify detected page regions according to a fine-grained typology (text blocks, tables, images, titles, etc.) leading to multiclass segmentation maps. In particular, we will investigate the fine-tuning of existing models like LayoutParser (<https://layout-parser.github.io/>).

▷ **Contribution 2 – Self-supervised multimodal representation learning for CBIR:** Once the images/captions are extracted they can be used in the CBIR task mentioned above. 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

Photographies historiques



Journaux / documents historiques

Figure 1: Diffusion and circulation of a photograph in different newspaper pages over time and current events.

the level of supervision of the algorithms to deal with the limited amount of annotated data. The available data will be employed, via self-supervised approaches, to fine-tune the multi-modal foundation models. We will first investigate a purely visual approach with models such as SimCLR¹ by adapting pre-text tasks to the nature of the data and domain knowledge. In a second step, we will investigate models such as CLIP² involving a contrastive optimization strategy by aligning the embeddings of a visual and textual encoder, making it possible to simultaneously learn joint visual and textual representations. 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.

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>)

Desired background

We are looking for a Master 2 student or final year of MSc, or engineering school in computer science. The ideal candidate should have knowledge in image processing, computer vision, natural language processing, Python programming and an interest in handling large amount of data, in particular images.

¹Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey E. Hinton: A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020: 1597-1607

²Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. ICML 2021: 8748-8763