Identification of image circulation in large collections of historical photographs

Two master internship positions (February or March 2025)

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 March 2025.
- Institutes:
  - Université Paris Cité, Laboratoire d'Informatique Paris Descartes (LIPADE)
  - Sorbonne Université, Laboratoire d'informatique de Sorbonne Université (LIP6 UMR 7606)
  - Université Paris Cité, Laboratoire de recherche sur les cultures anglophones (LARCA UMR 8225)
- Location: Paris
- Applications should be sent to camille.kurtz@u-paris.fr, florence.cloppet@u-paris.fr, isabelle.bloch@lip6.fr

Note: Within the ANR project High-Vision, coordinated by Daniel Foliard (Université Paris Cité, LARCA), two internship positions are open, with potential extension to PhD theses, one at LIP6 (Sorbonne Université) and one at LIPADE (Université Paris Cité).

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 visuality 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 LIP6, 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).

The objective of the two internships is to go beyond state-of-the-art developments in computer vision, focusing on layout segmentation as pre-processing steps, object recognition, image similarity and retrieval followed by metadata enrichment to improve the multimodal indexing steps of images from different data corpora. These works will establish a robust pipeline that can effectively trace the circulation of news photographs, document their changing editorial contexts and accurately shape a mass of digitized historical photos without pre-existing metadata into a usable collection. The main originality lies in both multi-modal and semantic analysis, as well as in taking into account historical specificities (type of documents, type of objects depending on the era, etc.). Models and architectures will be evaluated and made available in open-access to the research community. The methods will be developed and tested on data already gathered by the High-Vision consortium. The results of these internships are intended to

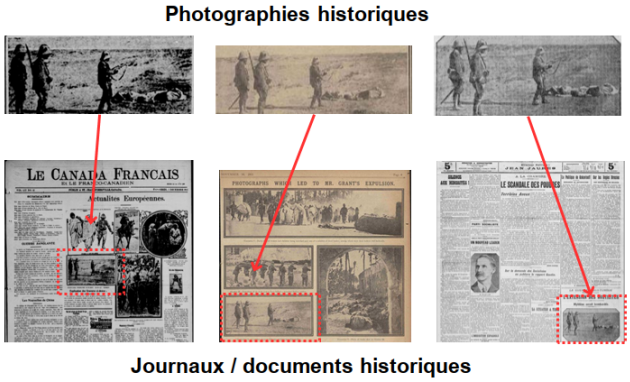

Figure 1: Several instances of the same photographs published in different newspapers of the Belle Époque.

support the historical analysis. Close interactions with the other partners of the project (historians and archivists) will allow integrating their feedback to adjust and improve the methods.

▷ **Contribution 1 – Document layout segmentation:** Prior to the identification of image circulation, 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 as a starting point the fine-tuning of existing Transformer 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 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. As a starting point, we will investigate a purely visual approach with models such as SimCLR[1] by adapting pre-text tasks to the nature of the data and domain knowledge. In a second step, we will investigate models such as CLIP[2] 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).

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 Master 2 students or final year of MSc, or engineering school in computer science. The ideal candidates 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 for digital humanities applications.

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

[2] Alec 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