1st Symposium in Artificial Intelligence & Applied Mathematics for History & Archaeology
« IAMAHA » - Nice, 27-28 November 2023

Computer Vision and Historical Scientific Illustrations

AOUINTI Fouad[1], BALTACI Zeynep Sonat[2], AUBRY Mathieu[2],
GUILBAUD Alexandre[3], LAZARIS Stavros[4]
[1]ISCD, Sorbonne Université, Paris, France
[2]LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
[3]IMJ-PRG, UMR 7586, Sorbonne Université, CNRS, Université Paris-Cité
[4]UMR 8167, Collège de France, Sorbonne Université, CNRS, Paris, France
https://imagine.enpc.fr/~baltacis/illustrationExtraction

Abstract. The VHS project (computer **V**ision and **H**istorical analysis of **S**cientific illustration circulation) proposes a new approach to the historical study of the circulation of scientific knowledge based on new methods of illustration analysis. Our contributions in this paper are twofold. First, we present a semi-automatic interactive pipeline for scientific illustration extraction that allows and incorporates expert feedback from historians. Second, we introduce a new dataset of scientific illustrations from the Middle Ages to the modern era consisting of 8k illustrations validated by historians and a total number of 235k illustrations obtained from 405k corpora pages. We further discuss our current research for identifying a series of related illustrations from this data.

Keywords. historical document analysis; IIIF; illustration detection; deep learning

Introduction and context

Illustrations and their evolution in the scientific corpora of the Middle Ages and modern Western cultures have only been partially studied. More generally, the role of image in the construction and dissemination of scientific knowledge raises complex questions that remain historically delicate to grasp and for which adapted analysis tools are lacking. To fill this gap, we aim to develop automated methods for extracting illustrations and then analyzing their similarities, which will lead to the constitution of an iconographic series that can be interpreted by historians. One major difficulty lies in the heterogeneity characterizing these corpora and the impossibility of carrying out large-scale annotations for every specific task. We thus develop approaches to handle the heterogeneity of historical data with limited annotation effort.

Various deep learning methods have been developed for the visual analysis of historical documents. In this paper, we focus on our dataset creation and illustrations extraction. We introduce the digital platform developed to allow historians to leverage deep learning methods, improve them by collaborative annotation, and historically analyze the results.

Dataset and illustration extraction

We started by building a platform to enable historians to upload 54 manuscripts and 704 printed volumes of interest in varying formats (IIIF manifests, JPEG, PDF), and the associated metadata. This resulted in 405,594 document images. To ease the work of annotating illustrations, we developed a two-stage workflow (Fig. 1). First, we exploited YOLOv5s (Jocher et al., 2020) trained on SynDoc (Monnier et al., 2020) and predicted illustrations bounding-boxed on a subset of 4451 representative images. Our platform then enabled historians to easily verify and if necessary use SAS[1] to correct

[1]https://github.com/glenrobson/SimpleAnnotationServer

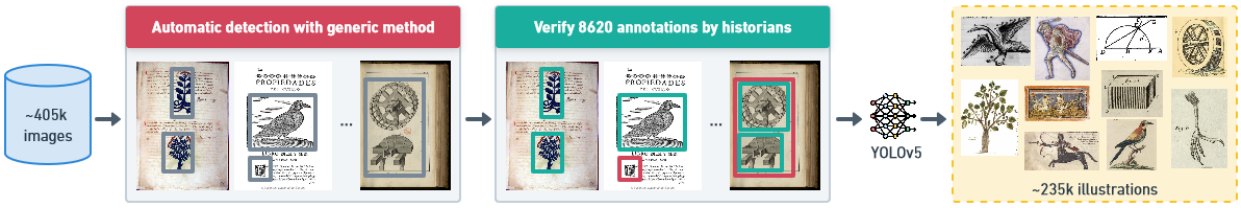

Fig. 1: Overview of our pipeline to refine illustrations detections in scientific historical documents.

Pre-training	Fine-tuning	Manuscripts				Prints			
		P	R	F-score	AP	P	R	F-score	AP
COCO	SynDoc	0.248	0.371	0.297	0.244	0.262	0.066	0.105	0.152
VHS	-	0.7	0.638	0.668	0.686	0.763	0.549	0.639	0.628
COCO	VHS	0.698	**0.682**	**0.69**	0.708	**0.831**	**0.635**	**0.72**	**0.719**
SynDoc	VHS	**0.75**	0.632	0.686	**0.715**	0.76	0.558	0.644	0.645

Tab. 1: Performance of YOLOv5s in precision (P), recall (R), F-score and average precision (AP).

and complete these detections. We split this curated 'VHS' data into train (60%), validation (20%), and test (20%) sets and use it to fine-tune YOLOv5s. We will release this data upon publication. We tested different pre-training and fine-tuning strategies and obtained a very significant boost over the network trained on synthetic data (see Table 1). We used the network pre-trained on SynDoc and fine-tuned on VHS to obtain 235,198 illustrations (including false positives) from our complete corpora.

Discussion

Based on this huge high-quality illustrations dataset, we work on developing similarity search algorithms and interfaces. We already obtained high-quality results using (Xi et al., 2021). Our next step is to equip the platform with tools that will enable historians to easily analyze and annotate the obtained image clusters, to assess their historical interest, and to refine the developed similarity search tools, in a spirit similar to the one we have presented for illustration detection.

Acknowledgements

This work was supported by the ANR (ANR project VHS ANR-21-CE38-0008). MA and SB were supported by ERC project DISCOVER funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101076028. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., ... Jain, M. (2022) *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (Version v7.0).*

Monnier, T. and Aubry, M. (2020) *docExtractor: An off-the-shelf historical document element extraction.* In 17^{th} International Conference on Frontiers in Handwriting Recognition (ICFHR).

Shen, X., Efros, A. A., Joulin, A., and Aubry, M. (2021) *Learning Co-segmentation by Segment Swapping for Retrieval and Discovery.* arXiv.