

LayNet: End-to-End Layout Classification for Enhancing Image Preparation in Historical Documents

Sharva Gogawale Berat Kurar-Barakat Daria Vasyutinsky Shapira
Nachum Dershowitz

19.12.2023

Abstract

Within the domain of historical document image analysis, the process of identifying the spatial structure of a document image is an essential step in many document processing tasks, such as optical character recognition (OCR) and information extraction. Advancements in layout analysis promise to enhance efficiency and accuracy using specialized models tailored to distinct layouts.

We introduce LayNet, a new dataset for benchmarking layout classification algorithms in historical documents. It consists of 1352 images of pages of printed Hebrew books in a variety of styles, categorized into four different classes based on their layout (the number of text columns and regions). Ground truth was crafted manually by a paleographer at the page level.

Furthermore, we conduct an in-depth performance evaluation of various layout classification algorithms, which are based on deep-learning models that learn to extract spatial features from document images. We evaluate our algorithm on LayNet and achieve state-of-the-art results on the task of layout classification in historical documents. We explore dual classification methodologies: one by assigning a single label to each document and the other by assigning multi-labels to each document by representing each of the four classes as a vector of length five, where each dimension could be common among the four classes. This approach is advantageous as there are some overlapping attributes between our classes.

Our results demonstrate the potential of LayNet to be used as a benchmark for evaluating layout classification algorithms and training new classification models. LayNet is positioned to contribute to the advancement of research on layout classification in historical documents, facilitating the development of new and improved algorithms for challenging tasks like script counting, reading order prediction, and so on.

The images, sourced from the KTIV catalogue, were curated by Daniel Stoekl Ben Ezra. This research is conducted within the framework of the MiDRASH ERC Synergy project.

References

- [1] B. Kiessling, R. Tissot, P. Stokes and D. Stökl Ben Ezra, “eScriptorium: An open source platform for historical document analysis”. In *2nd International Workshop on Open Services and Tools for Document Analysis (OST@ICDAR)*, pp. 19–24, September 2019.
- [2] A. Droby, B. Kurar Barakat, D. Vasyutinsky Shapira, I. Rabaev, J. El-Sana, “VML-HP: Hebrew paleography dataset”. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, Lausanne, Switzerland, September 2021, pp. 205–220.

- [3] R. Shweka, Y. Choueka, L. Wolf, N. Dershowitz, “Automatic extraction of catalog data from digital images of historical manuscripts”, *Literary and Linguistic Computing*, volume 28, issue 2, June 2013, pages 315–330, <https://doi.org/10.1093/llc/fqt007>.

Biography

Sharva Gogawale is an M.Sc. student at Tel Aviv University. His research interests include historical image analysis, with a broader focus on leveraging AI for social good.

Berat Kurar-Barakat is a postdoctoral researcher at Tel Aviv University’s School of Computer Science, specializing in historical document image analysis.

Daria Vasyutinsky Shapira is a researcher at Tel Aviv University. Her field of expertise is digital Hebrew paleography.

Nachum Dershowitz is Professor Emeritus in the School of Computer Science at Tel Aviv University.