

Determining the Influence of Papyrus Characteristics and Data Augmentation on Fragments Retrieval with Deep Metric Learning

In practice, ancient papyri are ^{frequently} torn into several fragments and the task of papyrologists is to assemble and decipher these fragments. Once successfully reconstructed, ancient papyrus is an important historical document. It offers the opportunity to gather crucial information about ancient cultures. However, reassembling by hand is very time-consuming and challenging because fragments differ in color, structure, and shape. In other words, they do not fit together perfectly - like an artificially designed toy puzzle.

The ^{reassembling process} objective of this thesis is to make the work of papyrologists easier by partially automating the process of reassembling. To this end, an algorithm is designed ^{to} that infers a smaller sub-selection of fragments with a high likelihood of being a good fit given one particular fragment as input. In the following, this algorithm is called a puzzle-helper. Its purpose is to increase the effectiveness and efficiency of the reassembling task. Nevertheless, in a real-world scenario, a puzzle-helper must fulfill several constraints.

First, training such a helper ^{cannot} take advantage of a big data set of fragmented and reassembled (labeled) papyri, because only a tiny bit of such data exists. So far, Deep Metric Learning (DML) has been used to create puzzle helpers, achieving a top-1 accuracy (top-1) of 0.73 [1, 2]. The metric stands for the percentage of correct classified candidates among the single most likely candidate. Unfortunately, further improvement of DML models require an immense amount of data, otherwise the model will probably not converge. If fragments are reassembled manually to get an additional data, the purpose of such a puzzle-helper becomes ^{useless (obsolete=outdated, no longer useful)} obsolete because the total time required by a papyrologist increases. Therefore, the task of creating labeled data has to be automated. In literature and research there are many different possibilities and approaches to the problem [3]. ~~In the following, two possible techniques will be presented:~~ ^{In this thesis, these two}

- ~~1. Generative Adversarial Networks [3] (GANs) are a well-known possibility to generate more artificial data and thereby improve the overall performance in different domains [4].~~
2. Creating a extensive data set ^{by} while automatically and artificially ^{tearing} torn papyri into several pieces is plausible as well [5]. More fragments lead to a significant increase in overall model performance [2, 1]. ^{training}

~~Both possibilities will be used during this thesis~~ to alleviate problems which are arising within the low-data regime.

Another constraint is to ^{of course} keep the number of false-negatives low. If the puzzle-helper selects a subset of potential candidates, all potentially matching fragments ^{should} shall be included. If the perfect match is not enclosed in the subset, a papyrologist has no chance to identify the correct fragment unless he extends his search again to the entire data set. ~~Thus the puzzle-helper is obsolete again.~~ ^{useless} Another disadvantage is that the approach fails silently. The papyrologist might search long for a match ^{because if} the expert is convinced that the algorithm works fine. However, if the ground truth match is not among the candidates of this subset, the search is senseless. Metrics that measure the accuracy in a helpful

Good point

I disagree; if for some fragments the system finds the match, then the expert gains time. If the match isn't there, then the expert has to do the usual task. There would be no loss of time, in my opinion.

This part goes too much into details. Keep in mind that

manner are the mean average precision (mAP), top-1, precision at 10 (pr@10) and at 100 (pr@100). The proposed metrics show how accurate the model is in general and how precise the model is within a particular range of potential candidates. In the next paragraphs, the different ^{metrics} matrices are summarized as accuracy, for the sake of simplicity. The success of a model depends strongly on which part of the image is used for training the puzzle helper. More precisely, what is suggested to the model during training as equal and unequal determines the model's accuracy [2]. In this work, we will statistically measure how specific papyrus characteristics determine the accuracy of a DML model. For this purpose, various metrics of different models are compared by using only the foreground of a papyrus (text) or the background (fibers) as input for training.

Separating foreground and background is challenging because a fiber often has similar colors as the text. That circumstance makes thresholding more complicated. Several methods will be used and get evaluated on the overall accuracy. A separate evaluation is not straightforward and is misleading because we do not have pixel-wise ground truth [6]. Also, the focus of this thesis lies on the influence of specific characteristics and the power of more labeled data available instead of the separating process itself. For cutting out the text, a method based on a color threshold will be used (binarization), which separates the text and generates a text mask. The extracted areas can now be filled using an inpainting [7] approach. The inpainting procedure is also just evaluated based on overall accuracy.

Accuracy of what?

Finally, further experiments will examine whether specific characteristics (fibers) can be exploited in some way. Specifically, it will be tested if the fibers can be used to determine the exact position of the fragments to each other. Determining the position is the next logical step towards a fully automated helper. So, the final goal is to design a puzzle solver function by combining the DML model and a suitable method for position determination.

In summary, the thesis is divided into the following milestones:

- I. Separating text and papyrus fibers by using state-of-the-art binarization [6] and inpainting [8].
- II. Generating larger amounts of data through a GAN-based model and creating a base-line out of artificial ^{ly} torn fragments for evaluation. [3, 4].
- III. Evaluation by means of a DML model using the original data in the original state (text and fibers), in the processed state (text only or fibers only). Also, an evaluation of the artificial data together with the original data in the original state and processed state.
- IV. Review of the state of the art on the possibilities of using certain characteristics for position determination.

The following research questions emerge from the milestones:

- RQI:** Do the chosen metrics (mAP, top-1, pr@10, pr@100) differ significantly based on the binarization and inpainting methods used?
- RQII:** Do the selected metrics (mAP, top-1, pr@10, pr@100) differ significantly when only the text or only the fibers are used as input as opposed to the unprocessed data?
- RQIII:** Do the selected metrics (mAP, top-1, pr@10, pr@100) differ significantly when additional artificial data generated by a GAN are used for training?

The implementation will be done in Python.

Supervisors: Dr.-Ing. V. Christlein, Prof. Dr.-Ing. habil. A. Maier, Mathias Seuret M. Sc.
Student: Timo Bohnstedt
Start: November 8th, 2021
End: April 8th, 2022

References

- [1] C. Ostertag and M. Beurton-Aimar, “Using graph neural networks to reconstruct ancient documents,” in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds. Springer International Publishing, pp. 39–53.
- [2] A. Pirrone, M. Beurton-Aimar, and N. Journet, “Self-supervised deep metric learning for ancient papyrus fragments retrieval,” vol. 24, no. 3, pp. 219–234. [Online]. Available: <https://doi.org/10.1007/s10032-021-00369-1>
- [3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *arXiv preprint*, 1406.2661.
- [4] M. O. Tamrin and M. Cheriet, “Simultaneous detection of regular patterns in ancient manuscripts using GAN-based deep unsupervised segmentation,” in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds. Springer International Publishing, pp. 279–291.
- [5] M. Seuret, A. Nicolaou, D. Rodríguez-Salas, N. Weichselbaumer, D. Stutzmann, M. Mayr, A. Maier, and V. Christlein, “ICDAR 2021 competition on historical document classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, S. U. Josep Lladós, Daniel Lopresti, Ed., vol. 12824 LNCS. Springer Science and Business Media Deutschland GmbH, pp. 618–634.
- [6] C. Tensmeyer and T. Martinez, “Historical document image binarization: A review,” vol. 1, no. 3, p. 173. [Online]. Available: <https://doi.org/10.1007/s42979-020-00176-1>
- [7] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, journal Abbreviation: Proceedings of the ACM SIGGRAPH Conference on Computer Graphics Pages: 424 Publication Title: Proceedings of the ACM SIGGRAPH Conference on Computer Graphics.
- [8] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, “Image inpainting for irregular holes using partial convolutions,” vol. abs/1804.07723. [Online]. Available: <http://arxiv.org/abs/1804.07723>