Zahra Ziran

Enhancing historical objects: Image Enhancement Approaches [1]

Image processing techniques play a crucial role in enhancing image quality and addressing various challenges in different datasets. The techniques of *edge detection* and *superpixel segmentation* are normally used to magnify the digital images of ancient written documents when these are poorly preserved and therefore present problems of legibility and interpretation. Such problems are encountered not only when the scholar examines these documents in person, but also (and especially) when dealing with their digital reproductions made available by the entities possessing them or created directly from the originals by scholars thanks to the technologies of reproduction that are now widespread and commonly used. It is to these reproductions that we shall hereafter refer with the expression "original image". In the case of the research project NOTAE these are images of written documents belonging to Late Antiquity (in Greek or Latin) or the early Middle Ages (Latin) that have been transmitted on various material supports (primarily papyrus or parchment).[2] Within this category of sources, the project dedicates especial attention to individual signs (whether simple or complex in terms of morphology) or groups of signs that were inserted in the graphic fabric of texts to serve as symbols, even if they are sometimes alphabetic signs.

Histogram equalization and adaptive histogram equalization improve contrast and handle illumination variations, resulting in more visible and interpretable symbols. This work explores two distinct image processing approaches and a deep learning method to enhance the quality of graphic symbols. These techniques aim to simplify readability, enhance contrast, mitigate illumination variations. By comparing the effectiveness of different methods, we can identify the most suitable approaches for our specific graphic symbols, enabling us to extract valuable insights and foster a deeper understanding of these ancient symbols. This comprehensive analysis emphasizes the importance of selecting appropriate image processing and deep learning techniques to ensure an accurate representation within the context of the NOTAE project.

1 Image Processing: The First Approach

The first proposed image enhancement approach based on image processing includes the following phases:

1.1 Edge Detection and Watershed Segmentation

Edge detection aims to identify boundaries and edges within an image. The Sobel operator is a widely used edge detection method that computes the gradient magnitudes in both the horizontal and vertical directions of an image. Mathematically, given an image *I (x, y)* where *(x, y)* are the

[1] Last revision 05.02.2024. © Zahra Ziran, ORCID ID: 0000-0002-3529-7380.This research is part of the project *NOTAE: NOT A writtEn word but graphic symbols*, which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Advanced Grant 2017, GA n. 786572, PI Antonella Ghignoli).

[2] The project also includes documentary texts transmitted on wooden tablets (the *Tablettes Albertini*) and pieces of slate (the Visigothic *pizarras*), but unfortunately no digital reproductions of these exist to date.

spatial coordinates, the Sobel operator calculates the gradient in the *x*-direction (*Gx)* and the gradient in the *y*-direction *(Gy)* as follows [1]:

$$Gx\ (x, y)\ =\ I\ (x + 1, y) - I\ (x - 1, y)$$

$$Gy\ (x, y) = I\ (x, y + 1) - I\ (x, y - 1)$$

The gradient magnitude *(G)* is then computed as:

$$G(x, y) = \sqrt{Gx(x, y)^2 + Gy(x, y)^2}$$

The result is an edge map highlighting regions with significant changes in intensity, representing edges in the image. Watershed segmentation leverages the concept of flooding basins to separate the foreground and background regions in an image. It treats the gradient magnitude image as a topographic surface, where each pixel value represents a height. Markers are placed to indicate the starting points of flooding, with the background and foreground markers assigned to specific regions. The flooding process starts from these markers and gradually fills basins, separating the image into distinct regions.

1.2 SLIC Superpixel Segmentation

SLIC (Simple Linear Iterative Clustering) [1] is a superpixel segmentation algorithm that efficiently groups pixels into compact and uniform clusters known as superpixels. The algorithm combines spatial distance and color similarity to form clusters. Mathematically, given an image *I (x, y)* and the corresponding pixel position *(x, y)* and color *(R, G, B)*, SLIC aims to find *(K)* clusters with a user-defined compactness parameter.

First, *(K)* cluster centers (representative colors) are initialized at regular grid intervals in the image. The cluster centers are updated iteratively by minimizing the distance between pixels and cluster centers. The distance metric considers both color similarity and spatial distance:

$$D = \sqrt{(R - R_K)^2 + (G - G_K)^2 + (B - B_K)^2} + m \times \sqrt{(x - x_K)^2 + (y - y_K)^2}$$

where $(R_K,\ G_K,\ B_K)$ are the color values of the K_{th}cluster center, $(x_K,\ y_K)$ are its spatial coordinates, *(R, G, B)* are the color values of the current pixel, (x, y) are its spatial coordinates, and *(m)* is the compactness parameter that balances color and spatial distances. The process continues until convergence, resulting in compact and meaningful superpixels.

1.3 Joining Segmentations

In this work, joining segmentations combines the results of edge-based watershed segmentation and SLIC superpixel segmentation to obtain more comprehensive segmentation results. The goal is to merge the global information from edge-based watershed with the local details from SLIC superpixels.

Mathematically, given two segmentation maps S_1 and S_2, the *join_segmentations* function combines them to create a new segmentation map S_j. The process involves assigning each pixel to the label with the highest frequency in its neighborhood. This merging process results in a more complete and accurate representation of the image, effectively combining the strengths of both segmentation techniques.

1.4 Histogram Equalization

In our approach, we employed histogram equalization [3], a contrast enhancement technique that enhances the dynamic range of pixel intensities in an image. This method is utilized to achieve a more visually appealing representation by redistributing the intensity values of the image and achieving a uniform histogram.

Mathematically, let *p(i)* be the probability of occurrence of pixel intensity *i* in the image. The cumulative distribution function (CDF), *F(i)*, is computed as the sum of probabilities up to intensity *i*:

$$F(i) = \sum_{k=0}^{i} p(k)$$

The new intensity value of each pixel is then calculated using the CDF:

$$\mathrm{new}_{\mathrm{intensity}} = (\mathrm{L} - 1) \times F(i)$$

where L is the total number of intensity levels (e.g., 256 for 8-bit images). The result is an image with enhanced contrast and improved visibility.

1.5 Adaptive Histogram Equalization

Because of the type of images in the NOTAE dataset, we applied Adaptive histogram equalization addresses the limitations of standard histogram equalization, which may amplify noise and ~~graphical~~ graphic symbols with varying illumination and contrast across different regions. In adaptive histogram equalization, the image is divided into small regions or tiles, and histogram equalization is applied locally to each tile.

Mathematically, for each tile, a local histogram is computed, and the cumulative distribution function (CDF) is obtained. The new intensity value of each pixel in the tile is then calculated using the local CDF. This adaptive approach prevents over-amplification of contrast in uniform regions and provides better enhancement results tailored to the local characteristics of the image. By combining these techniques, we effectively improved image quality, as can be shown in Figure1.

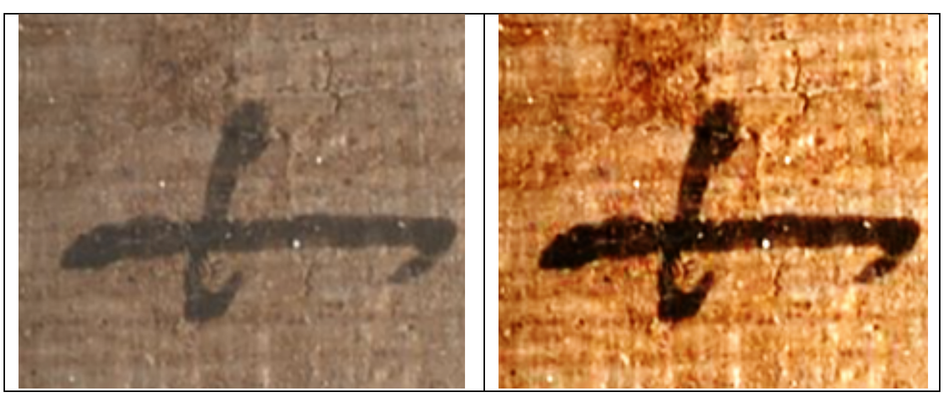

Figure 1.The original image (Left) and the enhanced version (Right) utilizing image processing (the first approach)

2 Image Processing: The Second Approach

In this approach, two different image processing techniques are applied to enhance the quality of images. The first technique is called *apply_filter*, which involves segmentation using edge-detection and watershed algorithms, combined with SLIC superpixels. The second approach, *remove_noise*, is used to remove noise from the background of the enhanced images. This process involves morphological opening, which consists of the following steps:

- Thresholding: The image is converted to a binary image by applying a threshold to distinguish between the background and foreground.
- Morphological Opening: Morphological opening is performed to remove small noise artifacts in the binary image. The opening operation involves erosion (shrinking) of the white regions followed by dilation (expanding). This effectively removes small objects and noise while preserving the overall structure of the objects.
- Combine Images: The cleaned binary image obtained from the morphological opening is combined with the original enhanced image, resulting in a final image with reduced background noise.

The choice of these techniques is motivated by their ability to enhance image quality, improve contrast, and mitigate noise, which are essential for better visualization and analysis of the graphic symbols in question. The *apply filter* technique effectively segments the symbols from the background, allowing for a clearer representation of the graphic content. Subsequently, the *remove_noise* technique further enhances the images by removing unwanted noise, which can significantly affect the interpretation and preservation of graphic symbols in papyri.

By applying and comparing these two techniques, we can assess their effectiveness in enhancing the quality of the graphic symbols in the NOTAE dataset. This comparative analysis enables us to select the most suitable image processing techniques for accurately representing and preserving ancient graphic symbols, thereby fostering a deeper understanding of cultural heritage from the historical period the project deals with. The combination of edge-based segmentation, superpixel clustering, and morphological noise removal empowers researchers and historians to explore and analyze these valuable historical objects with improved visual clarity as shown in Figure 2.

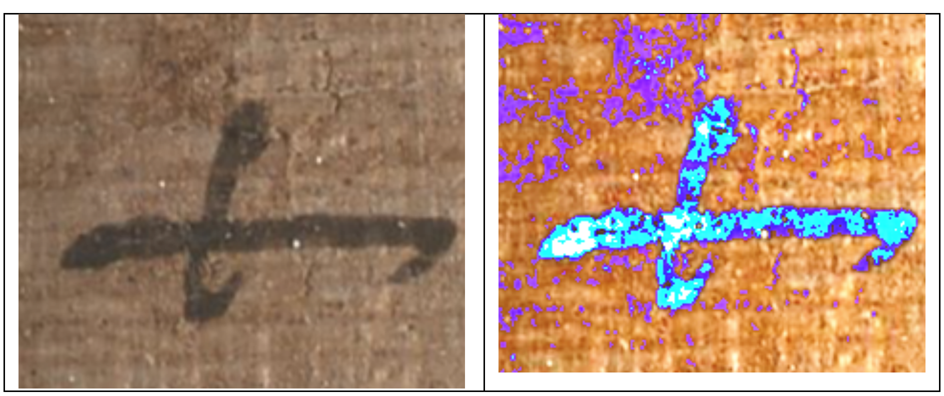

Figure 2.The original image (Left) and enhanced version (Right) utilizing image processing (the second approach)

3 Deep-learning-based approach

The task of accurately detecting objects in ancient documents poses significant challenges due to faded text, non-standardized writing, and complex layouts. To overcome these obstacles and enhance these historical resources, we employ deep learning techniques, including synthetic data generation, transformer-based models, and tokenization [4], [5], [6]. Our approach involves extracting visual features through tokenization and using an image classifier with a fast search method for gathering object detection annotations. We then train a Faster R-CNN model and generate a matching synthetic dataset alongside the original data. This enables us to approximate continuous functions that preserve the unique characteristics of the ancient documents[7], [8], [9], [10] and [11].

In our multi-step algorithm [12] for graphic symbol detection and classification, we start by preprocessing the input image using image processing techniques, as discussed earlier [13]. The image is then segmented into stroke segments and clustered using the OPTICS algorithm. To adaptively handle variable stroke thickness, we construct circles with varying radii for each cluster. These centroids and radii are utilized to create trapezoids that represent graphic symbols, with consistent orientation for easy reference. This comprehensive approach significantly improves the accuracy of object detection and classification in ancient documents [14], [15], and [16]. Through this method, the algorithm learns to identify and reconstruct ambiguous or missing symbols, as illustrated in Figure3 in the left corner of the enhanced images.

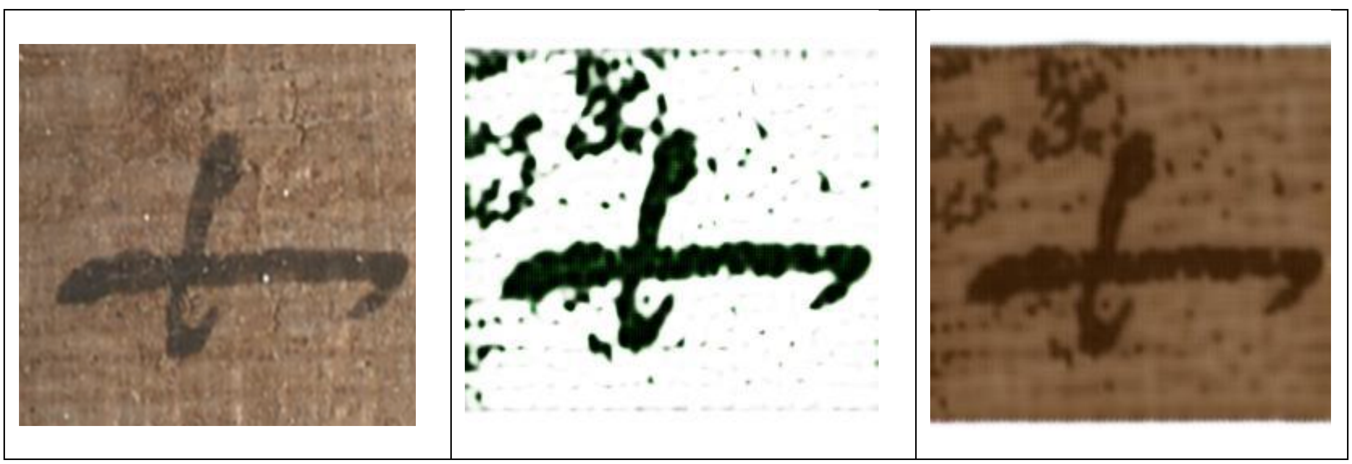

Figure 3.The original image (Left), the binarized version (middle), and the enhanced version based on deep learning (Right)

4 Conclusion

This study introduces two distinct image-processing techniques and a deep learning-based model. Our findings demonstrate that the first two approaches, utilizing image processing techniques, outperformed the deep learning model in the NOTAE project. Their success resides primarily in the fact that it is not a problem for the project's palaeographic researchers to identify graphic symbols within a documentary text written in Greek or Roman cursive on papyrus or parchment. The challenge is not such as to require advanced technological assistance such as that offered by deep learning, something that would have necessitated an immense amount of time spent in collaboration with IT specialists in machine learning and would therefore have subtracted time from research dedicated to the study and interpretation of graphic symbols and their contexts. What instead represents a significant contribution is the possibility of providing clearer images for the purposes of evaluating the composition and shape of a graphic symbol. These make it possible to discern more clearly the traces of ink that have flaked off or detect other elements useful for making a palaeographic conjecture. These are things that a precarious state of preservaton does not allow researchers to do with the originals or ordinary digital reproductions. The first two approaches are undoubtedly better suited to this goal.

5 References

[1] N. Salman, "Image segmentation based on watershed and edge detection techniques.," *Int Arab J Inf Technol*, vol. 3, no. 2, pp. 104–110, 2006.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012, doi: 10.1109/TPAMI.2012.120.

[3] H. D. Cheng and X. J. Shi, "A simple and effective histogram equalization approach to image enhancement," *Digit. Signal Process.*, vol. 14, no. 2, pp. 158–170, 2004, doi: https://doi.org/10.1016/j.dsp.2003.07.002.

[4] M. Boccuzzi *et al.*, "Identifying, Classifying and Searching Graphic Symbols in the NOTAE System," in *Digital Libraries: The Era of Big Data and Data Science*, M. Ceci, S. Ferilli, and A. Poggi, Eds., Cham: Springer International Publishing, 2020, pp. 111–122.

[5] Z. Ziran, E. Bernasconi, A. Ghignoli, F. Leotta, and M. Mecella, "Accurate Graphic Symbol Detection in Ancient Document Digital Reproductions," in *Document Analysis and Recognition – ICDAR 2021 Workshops*, E. H. Barney Smith and U. Pal, Eds., Cham: Springer International Publishing, 2021, pp. 147–162.

[6] H. Edwards and A. Storkey, "Towards a Neural Statistician." 2017.

[7] R. Girshick, "Fast R-CNN." 2015.

[8] J. Huang *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," *CoRR*, vol. abs/1611.10012, 2016, [Online]. Available: http://arxiv.org/abs/1611.10012

[9] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh, "Set Transformer," *CoRR*, vol. abs/1810.00825, 2018, [Online]. Available: http://arxiv.org/abs/1810.00825

[10] M. Phuong and M. Hutter, "Formal Algorithms for Transformers." 2022.

[11] A. R. Kosiorek, S. Sabour, Y. W. Teh, and G. E. Hinton, "Stacked Capsule Autoencoders." 2019.

[12] Z. Ziran, F. Leotta, and M. Mecella, “Enhancing Object Detection in Ancient Documents with Synthetic Data Generation and Transformer-Based Models.” 2023.
[13] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *CoRR*, vol. abs/1201.0490, 2012, [Online]. Available: http://arxiv.org/abs/1201.0490
[14] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “OPTICS: Ordering points to identify the clustering structure,” *ACM Sigmod Rec.*, vol. 28, no. 2, pp. 49–60, 1999.
[15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
[16] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, “Human-level concept learning through probabilistic program induction,” *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.