Impact of Image Enhancement Methods on HTR Trainings with eScriptorium

Pauline Jacsont, Elina Leblanc

Impact of Image Enhancement Methods on HTR Trainings with eScriptorium

Pauline Jacsont & Elina Leblanc

University of Geneva

Abstract:

This study stems from the project *Desenrollando el cordel* (*Untangling the cordel*), which focuses on 19th-century Spanish prints editing. It evaluates the impacts of image enhancement methods on the HTR transcriptions of low-quality documents, both in terms of printing and digitisation. We compare several methods (binarisation, deblur, defading) and present the results obtained during the training of models with the tool *eScriptorium*. We demonstrate that binarisation methods give better results than others, and that the combination of several techniques did not significantly improve the predictions of the HTR. This study shows the significance of using image enhancement methods with *eScriptorium* and paves the way for further experiments with larger and various corpora to help future projects design their HTR workflow.

Keywords: image enhancement methods, binarisation, printed documents, Spanish literature

1. Introduction

The Library of the University of Geneva holds a corpus of almost 1'000 Spanish chapbooks, printed during the 19th-century by several printers across Spain. Chapbooks, also known as *pliegos de cordel* consist of a few pages (4 to 8) and are in quarto. They relate real or fictitious events, share songs and poems, or display prayers and other religious writings (Figure 1). The Geneva collection is the object of the project *Untangling the cordel* [1], which aims at studying and promoting these documents through a digital library.

As one of the project's main objectives is to analyse the chapbooks' content, automatic transcription represents a significant step in our editorial workflow towards the publication of diplomatic digital editions with the standard XML-TEI. After testing several HTR tools, including *ABBYY FineReader* and *Transkribus* [2], we have chosen to use *Kraken* and its platform *eScriptorium*[1] [3].

During our first experiments with this tool, several challenges emerged. First, regarding the segmentation of the pages, we can mention the complex layout of the documents. Chapbooks' pages can display up to three columns. There can be variation in the layout of the pages even in the same chapbook (Figure 1).

[1] At the beginning of the project, we carried out our HTR experiments with *ABBYY FineReader* and *Transkribus*. These tools allowed us to build our Ground Truth (GT) and to transcribe a part of our collection. However, after one year, the University of Geneva developed an HTR platform called FoNDUE, based on the tools *Kraken*/*eScriptorium*. We became beta-testers of this new platform, which explains the use of different tools during our project. All the experiments we describe in this paper have been conducted exclusively with *Kraken/eScriptorium*.

Figure 1: Examples of ephemeral prints (From left to right: José María Moreno, Carmona, 1859; José María Moreno, Carmona, [s.d.]; Imp. *El Abanico*, Barcelone, [s.d.]; J. Jepús, Barcelone, 1884)

With respect to the transcription task, the multiple typography used represents a challenge for *eScriptorium*, especially with the titles which concentrate most of the errors produced by our model. Regarding the main text, the quality of both the paper and the ink makes characters recognition more difficult. Indeed, Spanish chapbooks – also known as cheap prints – were printed in mass on poor-quality papers. This culminates with bleed-through, which mixes the text on the recto with the one of the verso (Figure 2).

In our case, the quality of the digital facsimiles adds another layout of difficulty. Indeed, before the project's creation, several campaigns of digitisation were carried out with various scanners, by different operators from the Library and the Spanish Unit of the University. Therefore, the resulting digitised corpus is heterogeneous. If some chapbooks are in TIFF, most of the corpus is available in PDF, with a low resolution (72 dpi). The images can be highly pixelated, which causes additional noises for the recognition of characters by our model (Figure 2).

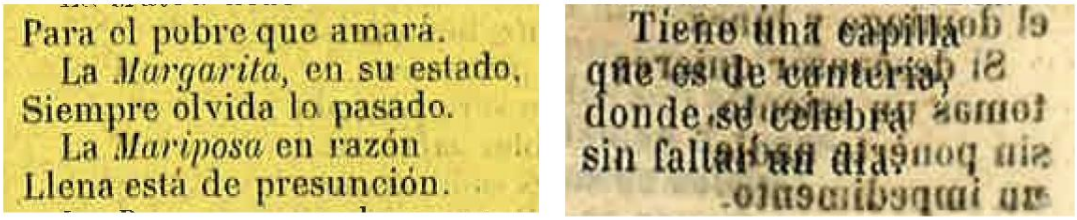

Figure 2: Examples of pixelated (left) and bleed-through (right) images

In this paper, we focus on one of the aforementioned issues: the low-quality of our data, both in terms of print and digitisation. We propose a comparison of several image enhancement methods to resolve it, namely binarisation and deblurring, and outline their impact on the accuracy of our HTR model.

After presenting related works in the next section, the third part describes our HTR workflow and the image enhancement methods we included in our study. The fourth section presents the effectiveness of each tested method to improve our HTR model, and the best method we have chosen for our data. Finally, we discuss the replicability of this experiment with other types of documents. We also address the opportunities offered by image enhancement methods to boost the results of HTR models trained with *eScriptorium*.

2. Related works

With the benefit of deep learning technologies, image enhancement is widely used to improve the efficiency of many computational tasks, such as image segmentation, face detection or optical character recognition. In a recent survey about image enhancement methods, [4] distinguish six different tasks to improve the quality of an image, depending on the type of damage: binarisation, deblur, denoise, defade, watermark removal and shadow removal. For instance, binarisation and denoising are favoured for document damages (wrinkles, stain, bleed-through); as for defading and deblur, they are mostly used to improve the quality of a digitisation and its exposure.

For printed or handwritten historical texts, binarisation appears as the main strategy which projects adopt to enhance their damaged images before the automatic transcription. Early binarisation approaches can be categorised into global and local thresholding methods. While global methods apply the same threshold to the entire image (Otsu [5]), local ones determine a threshold based on local statistics within a specific window (Niblack [6]). A comparison of these different solutions for the improvement of OCR predictions can be found in [7]. They noticed that the Otsu algorithm gives the best results for the transcription of English newspapers.

Lately, several studies have proposed to improve these methods, in particular for handwritten documents. [8] suggest modifying the Niblack method by dynamically defining the window used to determine the threshold instead of using a predefined value. They conclude that this method better harmonises the binarisation of images with inconsistent brightness repartition. In [9], a combination of global and local binarisation is used and proves its efficiency for detecting faint characters and removing bleed-through from highly damaged handwritten documents.

Regarding deblurring – which is an interesting task in our case to enhance our pixelated digitisations –, most studies are related to object or face detection. Usually, for this type of task, methods rely on blind deconvolution methods. However, several works showed that Convolutional Neural Network (CNN) is better at deblurring text images [10, 11]. Other approaches suggest Generative Antagonistic Network (GAN) methods to deal with a heterogeneous corpus composed of faces pictures and text images to reconstruct high-resolution images from low-resolution ones [12].

However, as pointed out by [4], while image enhancement methods prove their effectiveness in improving text images, few papers give details about their impact on the accuracy of OCR and HTR models. We can mention some experiments with *ABBYY FineReader* and *Tesseract* in [7, 10, 11, 13]. [8] and [9] propose an evaluation with HTR systems. Furthermore, these works focus their attention on only one method to enhance the images of their historical documents. Therefore, in this paper, we study and outline the benefit of several image enhancement techniques on the predictions of our model for Spanish printed documents. To achieve this, we use the HTR architecture of *Kraken/eScriptorium*, an endeavour that, to our knowledge, has never been carried out before with this specific tool.

3. Method

In order to evaluate the impact of image enhancement on the effectiveness of our HTR models, we performed a series of tests on the Ground Truth (GT): it consists of 198 pages

ocerized with *ABBYY FineReader*, then manually corrected. This corpus has been divided into three subsets[2]. Each set is used at a different stage:

- The first set, 80% of the GT, is used by the tool to train the model;
- The second set, 20% of the corpus, is used by the tool to evaluate each iteration during the training;
- The last set, 20% of the corpus, is used to evaluate the model's results on documents it has never encountered.

To compare the results, each training and each test were carried out with these same sets. The division was done manually to ensure that every set was equally difficult. Indeed, some pages have a more complex layout than others: first pages of each folder are more difficult to be dealt with HTR tools due to the variation of the fonts used.

Each time, the images were preprocessed on the whole corpus, thanks to a notebook[3]. Five different types of binarisation were tested: thresholding binarisation, Otsu, Niblack, Sauvola, and the binarisation proposed by *Kraken*[4]. Preprocessing affected image sharpness and gamma correction. To improve the sharpness of the images, we used a simple sharpening kernel with an OpenCV filter2D function. In addition, we test two types of Gamma correction: one to lighten the image (Gamma 1) and the other one to darken it (Gamma 2) to experiment with contrasts. These preprocessing techniques were chosen following the work of Zahra Anvari and Vassilis Athitsos [4] on historical documents which are, like those in our corpus, degraded and damaged.

Every HTR model[5] trained during this test phase was run following basic training[6]; the following command is used :

```
Ketos train -f alto -t train.txt -e eval.txt -d cuda data/*.xml
```

For each performed preprocessing, a training was followed by an evaluation test with the set prepared for this purpose. For every test, the same XML-ALTO files were used; only the images change according to the type of preprocessing.

Then, in a second phase, to evaluate the effectiveness of the preprocessing when combined, we conducted a series of tests by associating the preprocesses that had obtained the best results in the first phase.

---

[2] The details of the three sets are available on our Github repository: https://github.com/DesenrollandoElCordel/FoNDUE-Spanish-chapbooks-Dataset/tree/main/Grountruth/Split.

[3] The OpenCV, NumPy, SciPy and Scikit-image python libraries are used in this notebook. It can be consulted at this address: https://github.com/DesenrollandoElCordel/FoNDUE-Spanish-chapbooks-Dataset/blob/main/Varios-GroundTruth-TEST/ImagesTreatments.ipynb.

[4] The code for this binarization is available on Github: https://github.com/mittagessen/kraken/blob/master/kraken/binarization.py.

[5] The same tests were also carried out on the segmentation models, but this did not lead to improved results.

[6] The submission script is available at the following address: https://github.com/DesenrollandoElCordel/FoNDUE-Spanish-chapbooks-Dataset/blob/main/Grountruth/submission-script.sh.

4. Results

The results of the models trained during the first test phase (i.e., the models made with the images having undergone only one preprocessing) are presented in Table 1. The results are given for the accuracy per character of the training tests and the evaluation tests.

	Training test	*Evaluation test*
Without preprocessings	94.20 %	91.73 %
Thresholding binarisation	**95.99 %**	**94.42 %**
Otsu binarisation	96.04 %	94.13 %
Niblack binarisation	**97.80 %**	**94.36 %**
Sauvola binarisation	96. 29 %	94.33 %
Kraken binarisation	96.55 %	88.93 %
Deblur	95. 94 %	93. 47 %
Gamma 1	96. 37 %	92. 24 %
Gamma 2	96.54 %	93. 43 %

Table 1 : Models with single preprocessed images.

```
=== report  ===

27768   Characters
2296    Errors
91.73%  Accuracy

983     Insertions
75      Deletions
1238    Substitutions

Count   Missed  %Right
22240   1490    93.30%  Latin
5528    731     86.78%  Common

Errors  Correct-Generated
124     { , } - {  }
110     { . } - {  }
97      { . } - { , }
94      { SPACE } - {  }
76      { í } - { i }
63      { a } - {  }
49      { l } - {  }
48      { s } - {  }
44      { e } - {  }
39      { r } - {  }
```

```
=== report  ===

27768   Characters
1549    Errors
94.42%  Accuracy

658     Insertions
63      Deletions
828     Substitutions

Count   Missed  %Right
22240   1052    95.27%  Latin
5528    434     92.15%  Common

Errors  Correct-Generated
93      { , } - {  }
65      { SPACE } - {  }
60      { . } - {  }
46      { í } - { i }
45      { a } - {  }
43      { , } - { . }
32      { r } - {  }
31      { s } - {  }
28      { i } - {  }
27      { l } - {  }
```

```
=== report  ===

27768   Characters
1566    Errors
94.36%  Accuracy

630     Insertions
81      Deletions
855     Substitutions

Count   Missed  %Right
22240   1029    95.37%  Latin
5528    456     91.75%  Common

Errors  Correct-Generated
75      { SPACE } - {  }
66      { í } - { i }
64      { . } - {  }
62      { , } - {  }
46      { a } - {  }
36      { , } - { . }
34      {  } - { , }
34      { . } - { , }
32      { e } - {  }
31      { s } - {  }
```

Figure 3: Test reports of the model without preprocessing, on the left, with the thresholding binarisation in the middle, and with the Niblack model on the right.

The detailed results (Figure 3) show that the models trained with the binarised images succeed in detecting commas and full stop. However, some errors remain, such as confusion between 'i' and 'í', and problems in identifying certain letters such as 'a'.

A second series of tests was then carried out by combining a binarisation method with another image enhancement method. We did not launch some HTR model training when the image obtained was clearly unusable by the tool; this is the case, for example, of the association of the deblurring method (sharpening kernel) with the thresholding binarization (Figure 4).

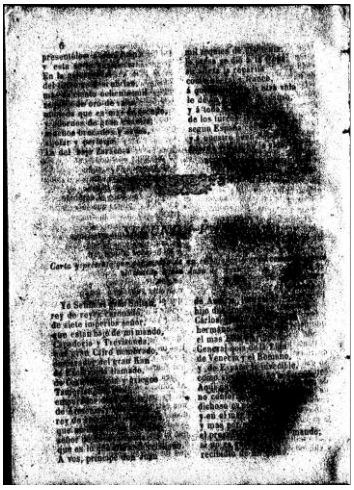

Figure 4: Example of unusable preprocessing (Combination of sharpening and thresholding binarisation)

Four other models were trained with double preprocessed images. The results obtained are given in Table 2, with the accuracy per character in percent.

	Training test	*Evaluation test*
Gamma 2 + Niblack	96.06 %	94.20 %
Gamma 2 + Sauvola	95.98 %	94.05 %
Gamma 2 + Otsu	95.53 %	93.44 %
Deblur + Niblack	96.02 %	94.05 %

Table 2 : Models with multiple preprocessed images.

Unlike our initial hypothesis, the multiple-image preprocessing method did not allow us to improve the results obtained in the first phase. However, the accuracy of the trained models remains higher than that of the model trained on images without any preprocessing.

These different tests show the importance and impact of image preprocessing methods on HTR predictions. In the end, the best method to improve our results, and the method that was applied to the rest of our corpus, is the thresholding binarisation. The model realized with the Niblack binarisation is also interesting as it has better results for the recognition of Latin alphabetic characters than the thresholding binarisation.

5. Conclusion and future works

In this paper, we have shown that using image enhancement methods can significantly improve the predictions of a model trained with *Kraken*/*eScriptorium*. This empirical approach has been conducted on a corpus of low-resolution printed documents, with a high level of noise caused by bleed-through and a high pixelation.

Image enhancement methods have been primarily thought for this type of damaged documents. To analyse their benefit for the transcription of non-damaged images, a similar experiment has been performed with a handwritten document, digitised with a high

resolution[7]. However, the experiments carried out with this manuscript revealed that image enhancement methods did not improve the prediction of the HTR models[8].

These results show that depending on the quality of the documents and its scanning, the same method will not be as effective. Furthermore, they pave the way for further experiments and research. Indeed, it would be useful to reproduce these experiments at a large scale to define recommendations about the best image enhancement methods to use with *Kraken*/*eScriptorium* and other user-friendly HTR tools, according to the type of texts and the quality of images.

Furthermore, for our binarisation experiments, we used traditional and well-known methods. However, new approaches are currently experimented with deep learning architectures (CNN [14] or GAN [15]). These new methods are still at their very beginning, but it would be interesting to include them in future works to analyse their impact on the accuracy of a model trained with *Kraken/eScriptorium* in comparison with thresholding methods.

Author's contributions

PJ has conceived and performed all the experiments with the different image enhancement methods. EL has elaborated the general workflow of the project and helped to design this study.

Acknowledgement

We sincerely thank the FoNDUE project team, Simon Gabay, Pierre Künzli, Christophe Charpilloz and Jean-Luc Falcone, for their constant support and invaluable responsiveness, which made all these experiments possible.

Bibliography

1. Desenrollando el cordel / Démêler le cordel / Untangling the cordel. 2022. Bibliothèque numérique. https://desenrollandoelcordel.unige.ch.
2. Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 2017. Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 04:19–24. https://doi.org/10.1109/ICDAR.2017.307.
3. Kiessling, Benjamin, Robin Tissot, Peter Stokes, and Daniel Stökl Ben Ezra. 2019. eScriptorium: An Open Source Platform for Historical Document Analysis. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2:19–19. https://doi.org/10.1109/ICDARW.2019.10032.
4. Anvari, Zahra, and Vassilis Athitsos. 2022. A Survey on Deep learning based Document Image Enhancement. arXiv. https://doi.org/10.48550/arXiv.2112.02719.
5. Otsu, Nobuyuki. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9: 62–66.

[7] This experiment has been done in the context of the digital edition project of the Cod. 174 of the Bern Burgerbibliothek: http://katalog.burgerbib.ch/detail.aspx?ID=340662. This Latin manuscript from the 16th century was digitised in TIFF format (400 dpi).

[8] The results of these tests are available in detail in Jacsont Pauline, *Mise en valeur du patrimoine textuel grâce aux éditions numériques « Le cas du codex 174 de la Bibliothèque de la Bourgeoisie de Berne »* (To come).

https://doi.org/10.1109/TSMC.1979.4310076.
6. Niblack, Wayne. 1986. *An Introduction to Digital Image Processing*. Prentice-Hall International.
7. Gupta, Maya R., Nathaniel P. Jacobson, and Eric K. Garcia. 2007. OCR binarization and image pre-processing for searching historical documents. *Pattern Recognition* 40: 389–397. https://doi.org/10.1016/j.patcog.2006.04.043.
8. Boiangiu, Costin-Anton, Alexandra Olteanu, Alexandru Stefanescu, Daniel Rosner, Nicolae Tapus, and Mugurel Andreica. 2011. Local Thresholding Algorithm Based on Variable Window Size Statistics. In *Proceedings of the 18th International Conference on Control Systems and Computer Science (CSCS)*, 2:647–652. Bucharest, Romania.
9. Ntirogiannis, K., B. Gatos, and I. Pratikakis. 2014. A combined approach for the binarization of handwritten document images. *Pattern Recognition Letters* 35. Frontiers in Handwriting Processing: 3–15. https://doi.org/10.1016/j.patrec.2012.09.026.
10. Hradiš, Michal, Jan Kotera, Pavel Zemcík, and Filip Sroubek. 2015. Convolutional Neural Networks for Direct Text Deblurring. In *Proceedings of BMVC*. Vol. 10. https://doi.org/10.5244/C.29.6.
11. Gangeh, Mehrdad J., Sunil R. Tiyyagura, Sridhar Dasaratha, Hamid Motahari, and Nigel P. Duffy. 2019. Document Enhancement System Using Auto-encoders. In *Proceedings of the 33rd Conference on Neural Information Processing Systems*. Vancouver, Canada.
12. Xu, Xiangyu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming-Hsuan Yang. 2017. Learning to Super-Resolve Blurry Face and Text Images. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 251–260. https://doi.org/10.1109/ICCV.2017.36.
13. Souibgui, Mohamed Ali, Yousri Kessentini, and Alicia Fornés. 2021. A Conditional GAN Based Approach for Distorted Camera Captured Documents Recovery. In *Pattern Recognition and Artificial Intelligence*, ed. Chawki Djeddi, Yousri Kessentini, Imran Siddiqi, and Mohamed Jmaiel, 215–228. Communications in Computer and Information Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-71804-6_16.
14. Akbari, Younes, Somaya Al-Maadeed, and Kalthoum Adam. 2020. Binarization of Degraded Document Images Using Convolutional Neural Networks and Wavelet-Based Multichannel Images. *IEEE Access* 8: 153517–153534. https://doi.org/10.1109/ACCESS.2020.3017783.
15. Khamekhem Jemni, Sana, Mohamed Ali Souibgui, Yousri Kessentini, and Alicia Fornés. 2022. Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement. *Pattern Recognition* 123: 108–141. https://doi.org/10.1016/j.patcog.2021.108370.