Article

Mathematical Data Models and Context-Based Features for Enhancing Historical Degraded Manuscripts Using Neural Network Classification

Pasquale Savino * and Anna Tonazzini

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Via G. Moruzzi, 1, 56124 Pisa, Italy; anna.tonazzini@isti.cnr.it
* Correspondence: pasquale.savino@isti.cnr.it

Abstract: A common cause of deterioration in historic manuscripts is ink transparency or bleeding from the opposite page. Philologists and paleographers can significantly benefit from minimizing these interferences when attempting to decipher the original text. Additionally, computer-aided text analysis can also gain from such text enhancement. In previous work, we proposed the use of neural networks (NNs) in combination with a data model that characterizes the damage when both sides of a page have been digitized. This approach offers the distinct advantage of allowing the creation of an artificial training set that teaches the NN to differentiate between clean and damaged pixels. We tested this concept using a shallow NN, which proved effective in categorizing texts with varying levels of deterioration. In this study, we adapt the NN design to tackling remaining classification uncertainties caused by areas of text overlap, inhomogeneity, and peaks of degradation. Specifically, we introduce a new output class for pixels within overlapping text areas and incorporate additional features related to the pixel context information to promote the same classification for pixels adjacent to each other. Our experiments demonstrate that these enhancements significantly improve the classification accuracy. This improvement is evident in the quality of both binarization, which aids in text analysis, and virtual restoration, aimed at recovering the manuscript's original appearance. Tests conducted on a public dataset, using standard quality indices, reveal that the proposed method outperforms both our previous proposals and other notable methods found in the literature.

Keywords: ancient manuscript virtual restoration; degraded document binarization; shallow multilayer neural networks

MSC: 68U10

1. Introduction

Historical and archival manuscripts often suffer damage from a variety of factors, primarily due to the natural degradation of materials over time and the conditions in which they are conserved. The fundamental requirement for accessing these manuscripts is the removal of degradation to ensure the main text is fully understandable.

Binarization of degraded documents can effectively separate the primary text from complex background patterns that need to be removed entirely [1–5]. However, this process may not be sufficient for a complete appreciation of the manuscript. First, binarization produces a two-class, black-and-white image, inevitably losing important details that the manuscript may contain, such as annotations, miniatures, watermarks, and drawings, which should be preserved for their historical and informational value. Second, binarization struggles with severe degradation, particularly when the interference pattern is nearly as intense as the primary text. Thus, it is crucial to strike a proper balance between removing unnecessary and harmful elements while preserving the primary text and enhancing

features that, although they are not directly related to the primary text, are significant to the manuscript's history.

Ink bleed-through is one of the most common and damaging forms of degradation found in ancient manuscripts. This occurs when both sides of the paper are written on, resulting in the two texts appearing intertwined on both sides, albeit with varying intensities. Such degradation can be attributed to poor conservation conditions, water infiltration, or the natural composition of the materials. Methods designed to reduce bleed-through are categorized into blind methods, which utilize information from one side only [6,7], and non-blind methods, which take advantage of the two sides of the manuscript page often being available [8–10].

Non-blind methods can achieve highly effective virtual restoration, but they require precise alignment of the two images as a trade-off [11–13].

In a previous study (see [14]), we proposed a simple multilayer shallow neural network with backpropagation training [15] for addressing non-blind cases. We designed the neural network to auto-adapt to the manuscript being enhanced, meaning it does not require prior training on a large set of similar manuscripts that have already been classified. Among various potential solutions, we utilized a data model to generate simulated degradation samples for the training phase. This data model provided an approximate description of the degradation when recto–verso digitizations of the manuscripts are available [10]. A training set was created using ground truths from the clean areas of the manuscript and subsequently mixed according to the model. The experimental results presented in [14] on heavily damaged manuscripts are encouraging regarding degradation removal.

We accounted for variable degradation, including very severe cases. This enables our neural network, developed from a single exemplar manuscript, to potentially perform effectively on other manuscripts within the same corpus that exhibit different levels of degradation or on different pages of the same book.

However, despite these encouraging results, we encountered significant challenges with high and highly variable levels of bleed-through.

When bleed-through is particularly strong, its gray-level values can closely resemble those of the text, making it difficult, if not impossible, to distinguish it from the primary text without additional information. Specifically, although we know the true classification of each pixel during the training phase, a neural network trained on examples with similar features but different targets will produce random responses. This can lead to the unintended cancellation of true text where it overlaps with the opposite text (occlusion) or, conversely, the preservation of noisy bleed-through pixels, which are mistakenly recognized as text.

In addition to being intense, the level of ink penetration can exhibit significant spatial variability within the same manuscript due to localized factors, such as humidity. This may result in spurious spikes of extreme degradation within otherwise uniform areas of moderate degradation. The consequences again include the potential cancellation of true text or the erroneous preservation of bleed-through.

In conference paper [16], we devised some modifications to the network architecture and learning to cope with the above issues better and with occlusion in particular. We proposed adding an extra output class for overlapping text pixels in order to distinguish them from the ordinary foreground text pixels and explicitly included the conditions for occlusion to occur in the data model so that the construction of the training dataset was coherent with the inclusion of the new class. We discussed the improvement produced by these modifications from a qualitative point of view on both synthetic and real degraded manuscripts and compared the classification obtained by our NN with that obtained by state-of-the-art binarization algorithms.

In this paper, we further extend the work presented in [16]. First of all, we systematically assess the results of the method through a quantitative analysis on a popular dataset. Then, we try to improve the NN's classification further by also taking into account the context of the pixel to be classified. Our point of view is that it may be reasonable to favor, a priori, the same classification for pixels that are close to each other spatially considering

text images are locally homogeneous. More specifically, the classification of a pixel should also take into account its position within the text: in the center of a stroke, at its edge, or completely outside it. This information can partially be derived from the average characteristics of adjacent pixels, to be used as a further attribute of the current pixel. In the simplest way, each pixel can be thought as connected to its 3×3 neighborhood, which can predict or give insights into a pixel's nature.

So far, our NN has worked pointwise and has considered the value of optical density as the primary attribute of the pixel so that we have two features corresponding to each pixel, namely the optical densities of the front and back observations of that pixel. In view of the above considerations, here, we try to account for the local information and extend the method by adding extra features, namely the two average values, one at the front and the other at the back, of the optical densities of 8-neighboring pixels. Our aim is to show that accounting for different contexts can explain how pixels with similar pointwise densities may come with different classifications, thus helping us to resolve ambiguity. The experiments confirm that the final NN architecture with four classes and four features is superior to the NN architectures we proposed previously.

This paper is organized as follows. In Section 2, "Materials and Methods", we first describe the method adopted for the construction of the adaptive training set. This method exploits the idea of using a mathematical model to generate artificial examples by extending a data model we previously proposed for the peculiar degradation treated in this research. In this section, we also provide the operative details of the shallow NN architecture and the learning and recall phases. Section 3 analyzes and discusses the experimental results for both synthetic and real cases, showing the improvement that can be obtained using the proposed extended NN architecture. Finally, Section 4 concludes this paper.

2. Materials and Methods

The first step of the overall process of enhancing historical recto–verso manuscripts is to classify the pixels of each side into four different classes, which we call *foreground*, *background*, *bleed-through*, and *occlusion*, respectively. These classes represent the main text; the clean paper texture with, eventually, other marks; the seeping ink; and the areas in which the two sides have both been written on and the two texts overlap. In the previous work [14], we only considered three classes by merging occlusion pixels with text pixels. This reflects the appearance of each side individually, where the occlusions are actually text for that side and without knowledge of the opposite side, cannot be identified with certainty. As we will see in the experimental results, using only three classes resulted in an overestimation of the bleed-through class.

As a classifier, we propose a neural network (NN) that requires a training set with ground truths in order to learn to distinguish between pixels. Each pixel is described by four features: the two densities on the two sides of the manuscript and the two average densities of its 8-adjacent pixels, always on the two sides. In previous works on this subject, we only considered the two essential density values of a pixel. This left some ambiguities in the classification of pixels that, despite belonging to a homogeneous area, had anomalous density values, caused by large variability in the degradation or the inhomogeneity of the materials.

The overall workflow of the process of recto–verso manuscript enhancement is illustrated in Figure 1. In this diagram, the focus is on the recto side only.
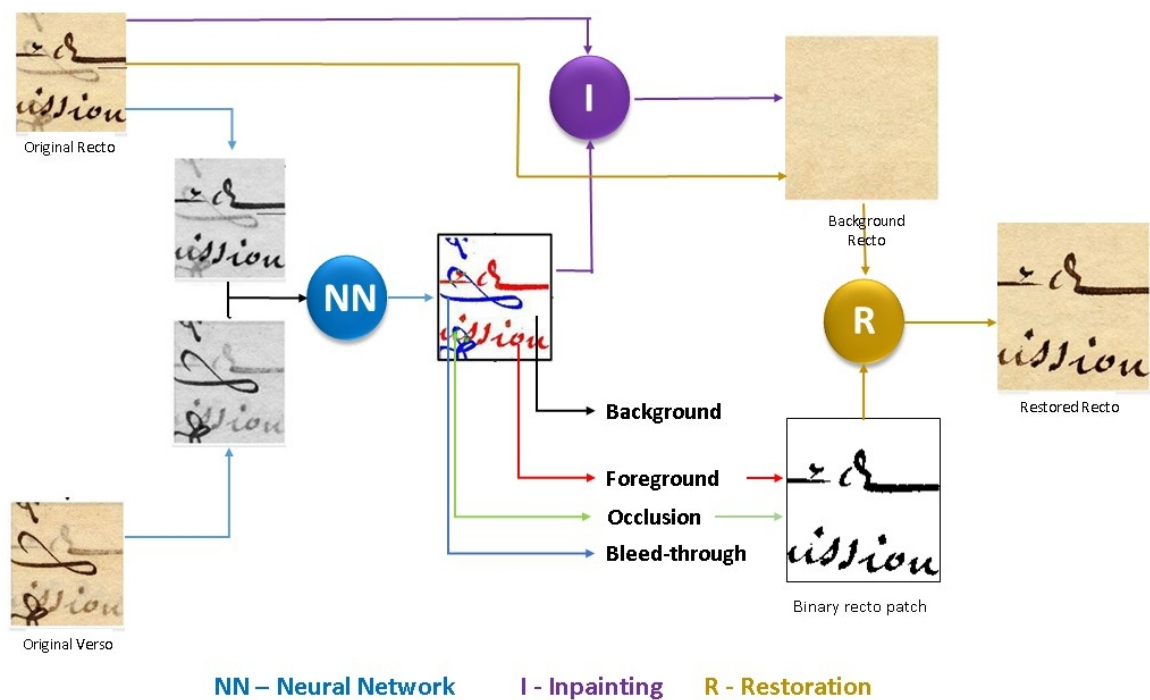
Figure 1. The workflow of the recto–verso manuscript enhancement process, focused on the recto side.

2.1. Construction of the Training Set

As mentioned, for the training phase, we do not use an external dataset based on similar manuscripts that have already been classified, but our neural network is trained using the manuscript we want to classify.

Thus, to build the training set, in the manuscript, we select N pairs of patches containing clean text and then symmetrically mix them using a data model for seeping ink that describes the observed optical density of each side as the weighted sum of the ideal densities of the two sides.

We define the optical density of pixel t as the minus log of the normalized intensity, i.e., $D_s(t) = -log\left(\frac{s(t)}{p}\right)$, with $s(t)$ being the intensity and p the mean intensity value of the paper support. This normalization allows the density to be independent of the color of the paper on the two sides. Based on this definition, the model is expressed in the following way:

$$D_x^{obs}(t) = \begin{cases} D_x(t), & \text{if } t \text{ is text on both sides} \\ D_x(t) + q_y(t)D_{h_y \otimes s_y}(t), & \text{elsewhere} \end{cases} \qquad (1)$$

where x and y indicate the two sides, which must be perfectly aligned after reflection of one of the two. Using Equation (1) for the opposite side, the roles of x and y are exchanged. In Equation (1), D^{obs} and D are the observed and ideal optical density, respectively, and $\otimes$ indicates the convolution between the ideal intensity s and a Point Spread Function (PSF), h, describing the smearing of ink penetrating the paper. Finally, the space-variant quantities q_x and q_y, whose allowed range is $[0, 1]$, represent the percentages of ink penetration from one side to the other. The first condition in the model Equation (1) means that we assume that the density of the foreground text does not increase due to ink seepage, as applies in the majority of the cases.

In previous works (see, e.g., [10]), we neglected the ink saturation effect and proposed inverting the equation in the second condition of the model to virtually restore the recto–verso pair. To make the inversion possible, we assumed that the hyperparameters q and h were known in advance. Based on the observed densities of the two sides, we first inverted the model by assuming an identically zero ideal density on the opposite side, thus obtaining

estimates of the ink penetration percentages for each pixel. The system could then be solved with respect to the ideal density maps, from which the virtually restored manuscript sides were obtained. To manage areas of text superposition (whose ideal density was not zero), the obtained images were corrected using some technicalities.

Here, we propose solving the direct problem of Equation (1) to generate the data necessary for the training set rather than solving the inverse problem to estimate the ideal densities, which are known in this case. Operatively, each patch out of the N selected pairs containing clean text is first binarized by the Sauvola algorithm in order to extract a map of the clean text and a map of the background. Comparing the binary map of both members of the pair allows us to locate the four classes on each side, including the occlusion. Then, the original, non-binary pairs of patches are fed into the system in Equation (1) in a forward manner, potentially with different ink seepage percentages, so that we numerically generate synthetic samples of recto–verso text with bleed-through. The first condition in Equation (1) permits us to simulate the saturation of the ink; that is, when a pixel is foreground text on both sides, the value of the density is set to that of the recto pixel (or the verso pixel, respectively). For the generation of a single pair of patches, the model is taken as stationary, i.e., with a fixed percentage of ink seeping. However, the construction of several pairs with different percentage values means that, as a whole, samples of non-stationary degradation will be presented to the network.

2.2. The Neural Network: Architecture, Learning, and Recall

We adopted a simple feedforward network with the architecture of a multilayer shallow neural network with one hidden layer and ten neurons and backpropagation training [15] (see Figure 2). To be specific, we used the function `patternnet`, available since the r2010b version of the Matlab Deep learning Toolbox. We run it on the 2023a version of Matlab. This network is a pattern recognition NN that can be trained to classify inputs according to target classes.
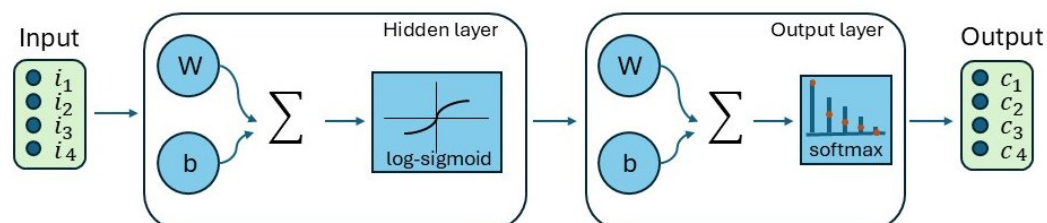
Figure 2. The basic network architecture.

The network processes the two sides of the manuscript simultaneously, on a pixel-by-pixel basis. For each pixel, we consider the two density values on the two sides as features, plus the two average recto and verso values of the densities of the 8-surrounding pixels. As target classes, we consider the four different classes of background, foreground, bleed-through, and occlusion.

Through construction, for the pairs of patches used to build the training set, we know the classification of each pixel on each side exactly. Thus, the target classes of the generated samples are directly available. The dataset is then randomly subdivided into a training set (70% of pairs) and a validation set (the remaining 30%). As said, the Matlab `patternnet` net is used with a single hidden layer constituting 10 nodes. As the minimization algorithm (`training function`), we chose scaled conjugate gradient and cross-entropy to measure the net performance (`performance function`) during training. Tests performed with a higher number of neurons did not provide a significant improvement in the quality of the results.

In the experiments, the number of patches N used to construct the dataset varied between 2 and 10; the size of the patches was chosen between 50×50 and 400×400; and the number of different values in $[0, 1]$ for the ink seepage percentage ranged from 10 to 20. The architectural simplicity of the network ensures very short learning times. The typical learning times were in the order of a few seconds when using the parameters given.

From the output of the NN, which consists of the classification of each pixel as one of the four classes, the binarized version of the manuscript can be obtained immediately by merging the pixels classified as text and occlusion into the same class and bleed-through noise and background into another single class. When the goal is instead to obtain a virtually restored version of the manuscript in which its original appearance and informative features are preserved as much as possible, the foreground text pixels, the occlusion pixels, and the background pixels are given their original values, whereas the noisy pixels are replaced with samples drawn from the closest safe background region. For this latter task, we tested various state-of-the art still image inpainting techniques and selected the exemplar-based image inpainting technique described in [17] as the best and simplest for our purposes.

3. Discussion of the Experimental Results

The results of manuscript enhancement using our self-trained NN approach were evaluated both qualitatively and quantitatively on degraded manuscripts contained in a popular dataset created within the Irish Script On Screen Project [18,19]. This database can be downloaded from the website at [20] and contains 25 pairs of aligned recto–verso manuscript portions affected by bleed-through. For each pair, the corresponding ground truths are available, which are manually constructed binary texts cleaned of degradation. The ground truths serve as a comparison for evaluating the performance of classification/binarization algorithms and, indirectly, that of virtual restoration algorithms.
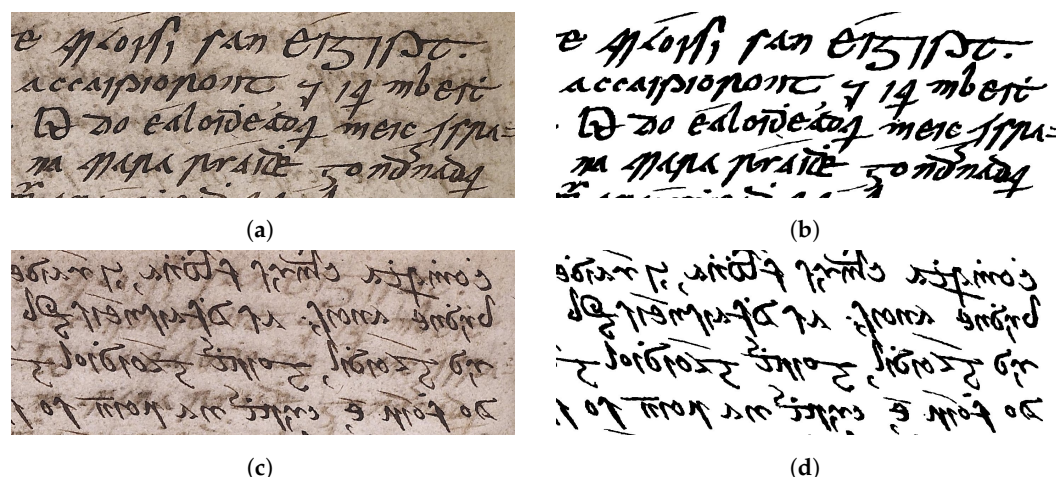
For the qualitative evaluation, we examined the accuracy of the virtual restoration from a perceptual point of view. For the quantitative evaluation, we compared the NN classification results with the ground truths. The standard error measures defined in [14] were used and displayed as plots versus the number of images. Additionally, the means of *Precision*, *Recall*, and *F-measure* were measured and compared with those of state-of-the-art methods.

At this stage, we did not exploit color information, neither in the training nor in the classification phase. The manuscripts were converted into grayscale because we assumed that the information from the verso, even from a single channel, would be much richer than the information provided by extra observations at different wavelengths on the recto side alone. This does not mean that including the spectral diversity of the individual sides would not improve the performance of the method further. In fact, we plan to add color information in our future work. In any case, even if the classification is performed on grayscale images, the virtually restored versions of the color manuscripts can be recovered directly since the three RGB channels share the same classes.

We processed the degraded pairs provided in the dataset, but we also processed synthetic pairs generated from the ground truths with different levels of degradation, with the purpose of testing the robustness of the method under conditions with wide variability and extreme intensity of degradation.

For the qualitative evaluation, we show images of the results obtained on the ninth pair in the dataset under different network architectures. We chose this manuscript because its rendering was agreeable and illustrated our problem simply and clearly. For the quantitative evaluation, we show comparative plots of the reconstruction errors for all the pairs in the dataset, still for different network configurations.

Figure 3a,c show the recto and the reflected verso of the ninth pair with the real degradation that affected the two images. Figure 3b,d show the manually built binary ground truths provided in the dataset. These ground truths represent the correct foreground texts on the two manuscript sides according to the perception of the human operator that constructed them.

(a)

(b)

(c)

(d)

Figure 3. Manuscript used for the qualitative evaluation: (**a**,**b**) color recto and corresponding ground truth; (**c**,**d**) color verso and corresponding ground truth (horizontally reflected).

3.1. From Two to Three Classes and from Two to Four Features: The Advantages of Distinguishing Occlusion Areas and Taking Context into Account

We processed all 25 RGB pairs in the dataset with NNs trained on clean patches extracted from the first pair in the dataset (not shown here). We used three different network architectures: (i) three classes (without the occlusion class) and two features (pointwise information); (ii) four classes (adding the occlusion class) and two features (pointwise information); and (iii) four classes (with the occlusion class) and four features (local information, adding the average density values for the eight surrounding pixels on the two sides).

The training set was constructed by mixing the selected pairs of clean patches according to the data model in Equation (1), using an ink penetration rate q ranging from 0 to 0.5. Figure 4 shows a subset of the training set, consisting of the six examples generated by mixing a single pair among the eight used.

As highlighted several times, we could have used the widest range of possible interference levels, say $[0, 0, 9]$, to build maximally general NNs. However, it is obvious that a network focused on effective degradation is more efficient. We therefore roughly estimated the maximum value for q for the entire dataset. In practice, we manually sampled a number of bleed-through pixels in the image for which we thought the interference was strong, then inverted Equation (1) for q based on the corresponding density values on the two sides, and finally averaged the obtained values.

An ink penetration rate between 0 and 0.5 is a good representation of the average amount of degradation across the entire dataset. However, since the estimation was manual, some images exhibiting peaks of stronger degradation that were not represented in the training set occurred.
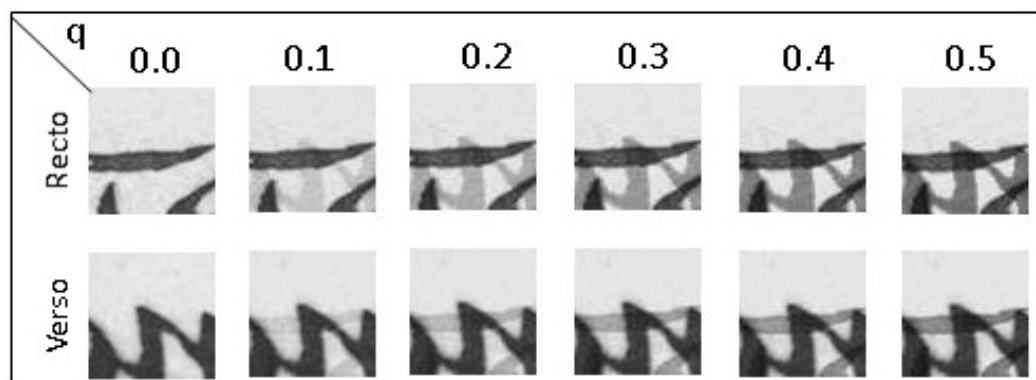
Figure 4. Part of the artificially constructed training set.

One solution would be to use a different NN for each image, basing the learning on the image itself. From a computational point of view, the learning phase is fast and has a minimal impact on the overall process, but the bottleneck is having to manually choose the patch pairs to use to generate the examples every time and carefully estimate the maximum amount of degradation.

In the following, we show the qualitative results of the binarization and the virtual restoration obtained on the ninth pair using the three different NNs (see Figure 5a,c,e). These results demonstrate that even with only three classes and two features, the NN is able to satisfactorily recognize most of the bleed-through pixels so that they can be removed (see Figure 5a,b). However, since the occlusion class is not considered, pixels that are text on both sides are normally classified as foreground on one side and bleed-through on the other, even with only negligible differences in their density. This ambiguity has the effect of producing small holes in the text characters. The introduction of the specific occlusion class is intended to eliminate this ambiguity by allowing pixels that have very similar and high densities to be classified as foreground on both sides. Indeed, already with four classes and two features, the number of holes in the characters decreases, as can be appreciated in Figure 5c,d. Also, by increasing the number of features from two to four, with the addition of the smoothness constraint, the resulting reconstruction should be smoother and flatter. Looking again at the characters in Figure 5e,f, this effect is very much evident and manages to correct the errors that remained in the correct classification of the occlusions. In summary, when using the final NN architecture, the reduction in the internal erosion of characters is particularly evident, and the reconstructed text characters are almost perfectly complete and filled.

The binary versions (see Figure 5b,d,f) highlight the behaviors already observed for the virtually restored images.

Besides our visual, qualitative results, we show also a quantitative evaluation of the performance of the different NN architectures on the entire dataset of 25 recto–verso pairs. Figure 6 shows the total errors in the binary reconstructions for the three different network architectures. Total error is measured according to the equations reported in [14], using the ground truths available in the dataset. The progressive improvement obtained by augmenting the number of classes and features used is apparent.

We finally consider the more conventional *Precision*, *Recall*, and *F-measure* metrics, where *Precision* indicates the percentage of how many of the detected foreground pixels are correct, and *Recall* indicates the percentage of how many of the correct foreground pixels are detected. These metrics are defined as follows:

$$
\begin{aligned}
Precision &= \frac{Sum(FT_R \cap FT_{GT})}{Sum(FT_R)} \\
Recall &= \frac{Sum(FT_R \cap FT_{GT})}{Sum(FT_{GT})} \\
F\text{-}measure &= \frac{2 \times (Precision)(Recall)}{Precision + Recall}
\end{aligned}
\tag{2}
$$

where FT_R is the binary map of the foreground text in the restored image, and FT_{GT} is the foreground text in the related binary ground truth mask.

We used the above measures to compare the final result of the $(4 + 4)$ NN architecture with state-of-the-art methods.

The values obtained are compared in Table 1 with those of the non-blind method [9] and the blind method [6] found in the respective papers. The proposed method exhibits a higher precision, which means that more of the pixels recognized as belonging to the foreground are correctly identified.
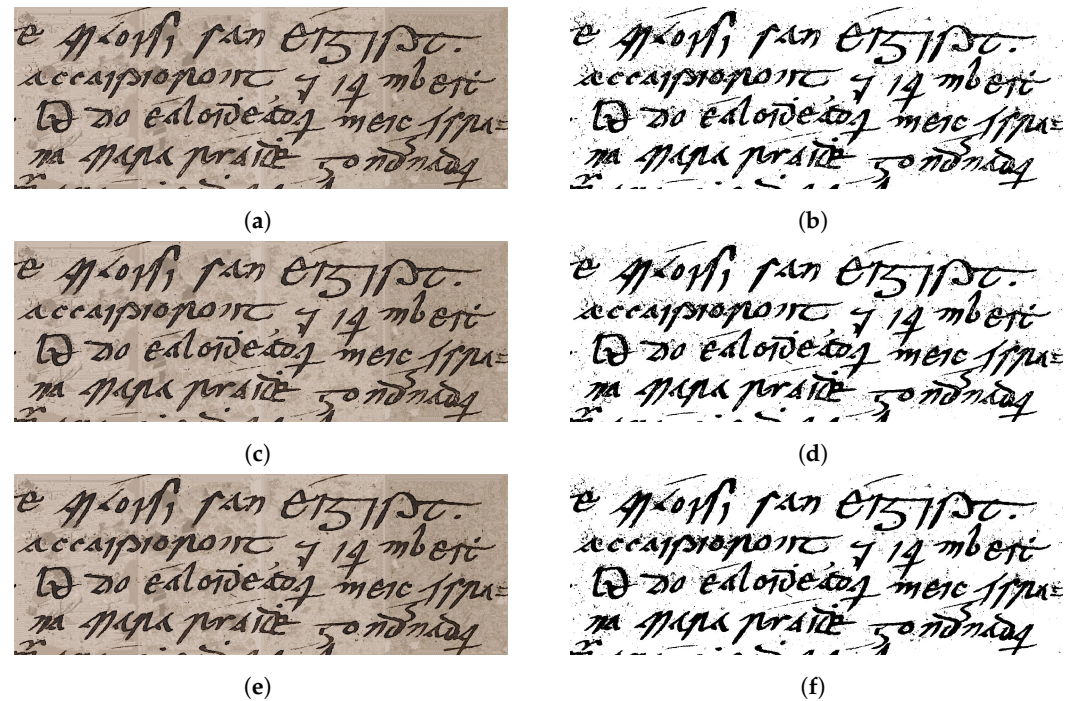
Figure 5. Binarization and virtual restoration of the real pair shown in Figure 3a,c: (**a**) recto restored with 3 classes and 2 features; (**b**) corresponding binarization; (**c**) recto restored with 4 classes and 2 features; (**d**) corresponding binarization; (**e**) recto restored with 4 classes and 4 features; (**f**) corresponding binarization.
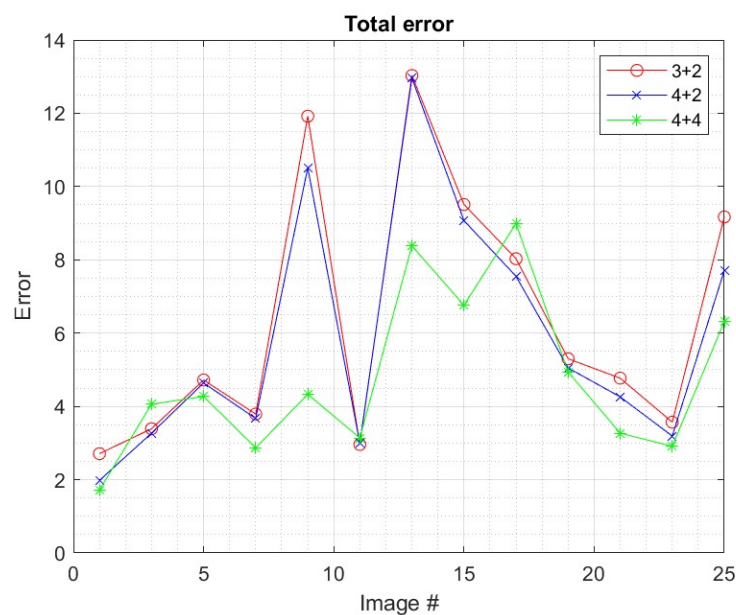
Figure 6. Plots of the total errors for all 25 pairs using the three different network architectures. The errors on the recto and verso sides of each pair have been averaged.

Table 1. Quantitative evaluation—average values of Precision, Recall, and F-measure on the whole dataset.

Method	Precision	Recall	F-Measure
non-blind, 2013 [9]	0.92	0.87	0.89
blind, 2016 [6]	0.89	0.86	0.87
proposed	0.94	0.92	0.93

It is to be noted that while the ground truth mask is fixed, in general, different binarization algorithms can be used to extract the binary mask of the foreground text from the restored images, so the resulting metric values may be affected by this choice. For instance, in our method and in [6], the binary map of the restored foreground text is a preliminary result, whereas in [9], the restored image is binarized using the Gatos algorithm [21].

3.2. Robustness of the Method for Space-Varying and Strong Degradation

In a final experiment, we tested the robustness of the method with respect to the strength of the degradation and its high space variability. This experiment also allowed us to qualitatively verify the effectiveness of the degradation model adopted.

Still based on the ninth pair of images, we built an artificial clean recto–verso pair by placing the clean foreground text onto a textured background obtained by inpainting. The foreground text was obtained by selecting the RGB values for the real degraded images in Figure 3a,c at the positions of the black pixels in the corresponding binary ground truth maps (Figure 3b,d). Figure 7a,b show this clean, ideal manuscript pair. An artificially degraded pair was then obtained by mixing the ideal pair using the data model in Equation (1), where the percentage of penetrating ink was increased from 0.1 to 0.9 from left to right (Figure 7c,d).


(**a**)

(**b**)

(**c**)

(**d**)
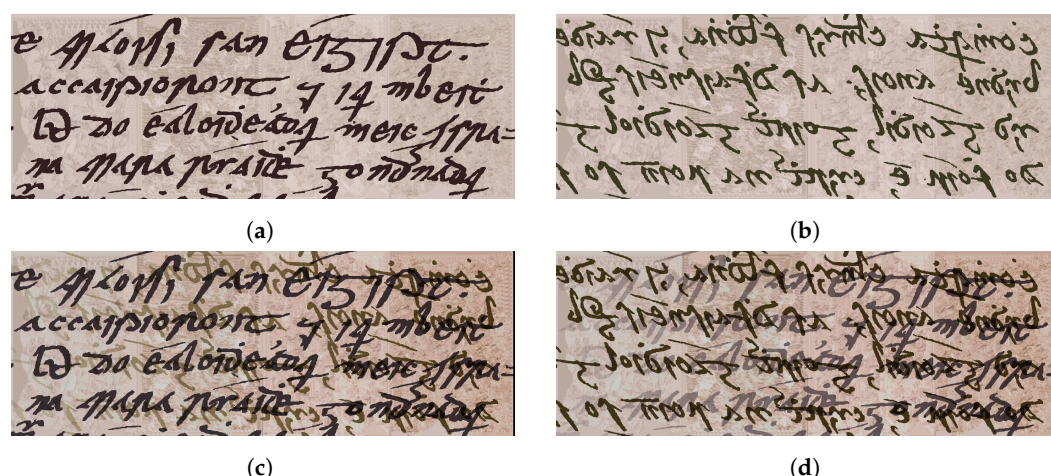
Figure 7. Generation of a synthetic manuscript pair: (**a**,**b**) clean, ideal recto and verso created from the images in Figure 3; (**c**,**d**) degraded recto and verso numerically constructed by feeding images (**a**,**b**) into the data model in Equation (1) with a spatially variable strength of degradation.

Note the plausibility of the visual aspect of the generated degraded images, which demonstrates the effectiveness of the degradation model adopted.

Figure 8 shows the results of applying the NNs with the three different architectures to the synthetically degraded images from Figure 7c,d. The training set was constructed by selecting pairs of clean patches from the original ninth pair and mixing them, this time with percentages of ink penetration spanning from 0.1 to 0.9, so as to cover the entire range of different amounts of degradation that was present in the generated degraded data.

The virtually restored recto for three classes and two features; four classes and two features; and four classes and four features is shown in Figure 8a,c,e, respectively. Note how when there are only three classes, the reconstructed text sometimes appears corroded and fragmentary, with missing strokes. Indeed, because the degradation is very strong here, reaching up to $q = 0.9$, the lack of a specific class for the occlusion causes text pixels on both sides to be attributed to the bleed-through class and then deleted. When the number of classes in the NN is extended to four, the text is reconstructed much better, and the characters are more complete and fuller. Note also the radical reduction in the remaining spikes in bleed-through. When the number of features is increased to four, the filling of the characters is even more apparent.
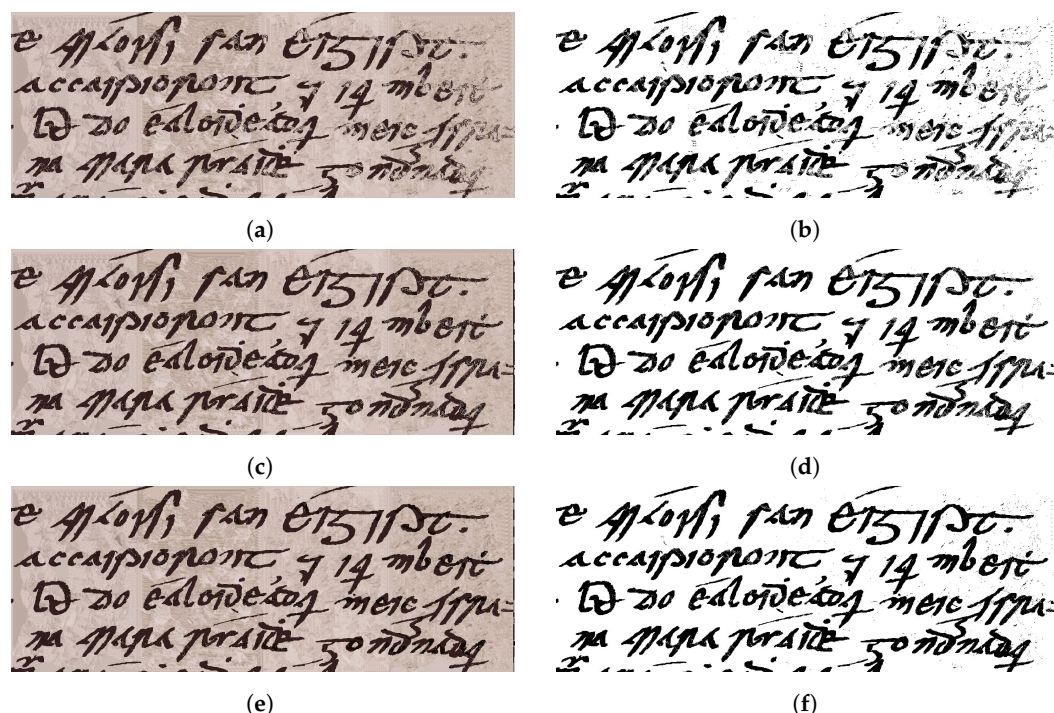
Figure 8. Virtual restoration of the synthetic pair shown in Figure 7c,d: (**a**) recto restored with 3 classes and 2 features; (**b**) corresponding binarization; (**c**) recto restored with 4 classes and 2 features; (**d**) corresponding binarization; (**e**) recto restored with 4 classes and 4 features; (**f**) corresponding binarization.

The corresponding binarizations are illustrated in Figure 8b,d,f and confirm the above considerations.

Looking at both real and synthetic experiments, it seems that the inclusion of the smoothness constraint, expressed by the third and fourth features, works much better in completing text characters than in reducing residual bleed-through noise, which is still present in the form of scattered and isolated spikes. We believe this depends on the extent of the neighborhood used, which is currently limited to immediately adjacent pixels only.

Definitely, our results show that exploiting the information provided by the reverse side of the page can allow for satisfactory binarization even of extremely degraded manuscripts. Currently, to the best of our knowledge, no algorithm is able to handle such a dramatic degradation using information from a single side alone. We already discussed this issue clearly in [16], where the high-performance algorithm proposed in [22,23] was used for comparison.

4. Conclusions

We demonstrated that joint processing of recto and verso pages of ancient manuscripts affected by ink penetration allows for significant improvements in terms of their binarization and virtual restoration thanks to the amount of information that the opposite side brings to the problem. This occurs without significant additional computational costs since scans of both sides of the sheet are normally available.

To this undoubted evidence, we have added the fact that the availability of an analytical data model describing degradation facilitates the use of NNs. Indeed, a data model allows us to artificially generate pairs of degraded recto–verso examples, whose targets are known by construction. In this way, it is possible to build a general training dataset without requiring real-world examples to be available.

We then trained a very simple shallow NN to correctly classify pixels as primary text, paper background, bleed-through noise, or overlaid text, using the data images themselves. After classification, the NN output can be used to binarize the foreground text or visually

restore the manuscript so as to maintain both the fullness of the information content and the aesthetics of the original.

The method described here improves our previous proposals in two respects: (i) the correct classification of overlaps between the two texts (occlusion) by adding a specific class and (ii) the disambiguation of pixels of different natures but with similar densities, taking into account the characteristics of the adjacent pixels.

The superiority of the four-class, four-feature network resulting from this proposal is evident, both in terms of binarization and in terms of virtual restoration. This confirms that categorization methods that exploit contiguity similarity constraints are clearly superior to point-based methods.

This method could be profitably applied in libraries and archives to processing entire manuscript books or a large corpus of manuscripts of the same typology. Indeed, in these cases, digitization of the verso side is readily available, so no new acquisitions need to be performed. Furthermore, the pages will presumably be homogeneous in the terms of the character font, ink composition, and damage entities, which makes the use of a unique training set, i.e., a unique NN, built on the basis of a single page, potentially chosen in a random way, effective.

The open problems that we intend to study in the immediate future concern the following: (i) the automatic extraction of the patches to use to construct the training examples; (ii) the introduction of color information and the extension of the method to multispectral observations; (iii) investigation of the generalization capabilities of the NN (e.g., with respect to overestimation of the degradation during the training phase); (iv) the introduction of further features to express other constraints or ameliorate the constraints already used; and (v) the use of other more sophisticated neural networks since the data generation model adopted is independent from the neural network paradigm.

Author Contributions: Conceptualization, P.S. and A.T.; methodology, P.S.; software, P.S.; validation, P.S. and A.T.; data curation, P.S.; writing—original draft preparation, A.T.; writing—review and editing, A.T. and P.S.; visualization, P.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in the study are included in the article/further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pai, Y.T.; Chang, Y.F.; Ruan, S.J. Adaptive thresholding algorithm: Efficient computation technique based on intelligent block detection for degraded document images. *Pattern Recognit.* **2010**, *43*, 3177–3187. [CrossRef]
2. Westphal, F.; Lavesson, N.; Grahn, H. Document image binarization using recurrent neural networks. In Proceedings of the 13th IAPR International Workshop on Document Analysis Systems (DAS2018), IAPR, Vienna, Austria, 24–27 April 2018; pp. 263–268.
3. Tensmeyer, R.; Martinez, T. Document image binarization with fully convolutional neural networks. In Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR 2017). IAPR, Kyoto, Japan, 9–15 November 2017; pp. 99–104.
4. Vo, Q.N.; Kim, S.H.; Yang, H.J.; Lee, G. Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognit.* **2018**, *74*, 568–586. [CrossRef]
5. He, S.; Schomaker, L. DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning. *Pattern Recognit.* **2019**, *9*, 379–3902. [CrossRef]
6. Sun, B.; Li, S.; Zhang, X.P.; Sun, J. Blind Bleed-Through Removal for Scanned Historical Document Image with Conditional Random Fields. *IEEE Trans. Image Process.* **2016**, *25*, 5702–5712. [CrossRef] [PubMed]
7. Hanif, M.; Tonazzini, A.; Hussain, S.; Khalil, A.; Habib, U. Restoration and Content Analysis of Ancient Manuscripts via Color Space based Segmentation. *PLoS ONE* **2023**, *18*, e0282142. [CrossRef] [PubMed]
8. Huang, Y.; Brown, M.S.; Xu, D. User Assisted Ink-Bleed Reduction. *IEEE Trans. Image Process.* **2010**, *19*, 2646–2658. [CrossRef] [PubMed]
9. Rowley-Brooke, R.; Pitié, F.; Kokaram, A.C. A Non-parametric Framework for Document Bleed-through Removal. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2954–2960.

10. Hanif, M.; Tonazzini, A.; Savino, P.; Salerno, E. Sparse representation based inpainting for the restoration of document images affected by bleed-through. *Proceedings* **2018**, *2*, 93. [CrossRef]

11. Wang, J.; Tan, C.L. Non-rigid registration and restoration of double-sided historical manuscripts. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), Beijing, China, 18–21 September 2011; pp. 1374–1378.

12. Rowley-Brooke, R.; Pitié, F.; Kokaram, A.C. Nonrigid Recto-Verso Registration Using Page Outline Structure and Content Preserving Warps. In Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing, Washington, DC, USA, 24 August 2013; ACM: New York, NY, USA, 2013; pp. 8–13.

13. Savino, P.; Tonazzini, A.; Bedini, L. Bleed-through cancellation in non-rigidly misaligned recto-verso archival manuscripts based on local registration. *Int J. Doc. Anal. Recognit.* **2019**, *22*, 163–176. [CrossRef]

14. Savino, P.; Tonazzini, A. Training a shallow NN to erase ink seepage in historical manuscripts based on a degradation model. In *Neural Computing and Applications, Topical Collection on Visual Pattern Recognition and Extraction for Cultural Heritage*; Springer: Berlin/Heidelberg, Germany, 2024.

15. Hagan, M.; Demuth, H.; Beale, M. *Neural Network Design*; PWS Publishing: Boston, MA, USA, 1996.

16. Savino, P.; Tonazzini, A. Mathematical models and neural networks for the description and the correction of typical distortions of historical manuscripts. In *Computational Science and Its Applications–ICCSA 2023 Workshops*; Gervasi, O., Ed.; Springer: Cham, Switzerland, 2023; Volume 14108, pp. 545–557.

17. Criminisi, A.; Pérez, P.; Toyama, K. Region filling and object removal by exemplar-based image inpainting. *EURASIP J. Adv. Signal Process.* **2004**, *13*, 1200–1212. [CrossRef] [PubMed]

18. Irish Script On Screen Project. 2012. Available online: http://www.isos.dias.ie (accessed on 30 August 2024).

19. Rowley-Brooke, R.; Pitié, F.; Kokaram, A. A ground truth bleed-through document image database. In *Theory and Practice of Digital Libraries*; Adn, G., Buchanan, P.Z., Rasmussen, E., Loizides, F., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7489, pp. 185–196.

20. Available online: https://www.isos.dias.ie/Sigmedia/Bleed_Through_Database.html (accessed on 30 August 2024).

21. Gatos, B.; Pratikakis, I.; Perantonis, S.J. Adaptive degraded document image binarization. *Pattern Recogn.* **2006**, *39*, 317–327. [CrossRef]

22. Xiong, W.; Jia, X.; Xu, J.; Xiong, Z.; Liu, M.; Wang, J. Historical document image binarization using background estimation and energy minimization. In Proceedings of the 24th International Conference on Pattern Recognition (ICPR 2018), Beijing, China, 20–24 August 2018; Volume 2083, pp. 3716–3721.

23. Xiong, W.; Zhou, L.; Yue, L.; Li, L.; Wang, S. An enhanced binarization framework for degraded historical document images. *EURASIP J. Image Video Process.* **2021**, *2021*, 13. [CrossRef]