S.I.: VISUAL PATTERN RECOGNITION AND EXTRACTION FOR CULTURAL HERITAGE

Training a shallow NN to erase ink seepage in historical manuscripts based on a degradation model

Pasquale Savino[1] (ID) · Anna Tonazzini[1]

Abstract
In historical recto–verso manuscripts, very often the text written on the opposite page of the folio penetrates through the fiber of the paper, so that the texts on the two sides appear mixed. This is a very impairing damage that cannot be physically removed, and hinders both the work of philologists and palaeographers and the automatic analysis of linguistic contents. A procedure based on neural networks (NN) is proposed here to clean up the complex background of the manuscripts from this interference. We adopt a very simple shallow NN whose learning phase employs a training set generated from the data itself using a theoretical blending model that takes into account ink diffusion and saturation. By virtue of the parametric nature of the model, various levels of damage can be simulated in the training set, favoring a generalization capability of the NN. More explicitly, the network can be trained without the need for a large class of other similar manuscripts, but is still able, at least to some extent, to classify manuscripts with varying degrees of corruption. We compare the performance of this NN and other methods both qualitatively and quantitatively on a reference dataset and heavily damaged historical manuscripts.

Keywords Ancient manuscript virtual restoration · Degraded document binarization · Registration of recto–verso documents · Shallow multilayer neural networks

1 Introduction

Historical and archival manuscripts are almost always damaged by the natural degradation of the materials over time or by other accidental factors such as fires, floods and poor preservation. Typically these ancient manuscripts appear as the overlay of a number of different patterns, or layers of information. In addition to the main text and paper texture, they may contain other informational elements, such as annotations, thumbnails, stamps, or non-informative interferences due to damage, such as damp and mold stains, or ink infiltrations from the back.

An important goal of digital image processing techniques is to provide scholars with digital versions that can help them in their work of reading, transcription and interpretation. Therefore, virtual restoration algorithms are required that attempt to restore the manuscripts to their original appearance, eliminating only the degradation without destroying the other informative characteristics. In this sense, the plurality of the content of the manuscript should be analyzed and discriminated, in such a way as to be able to preserve and highlight the useful patterns, and remove the extra, useless patterns that can disturb or even make impossible the academic study [1].

Another goal of digital image processing of a manuscript is to prepare it for automatic word and character recognition. In this case, the common approach is to first perform binarization, to extract the foreground text of interest against all other features that are considered,

Pasquale Savino and Anna Tonazzini have contributed equally to this work.

✉ Pasquale Savino
pasquale.savino@isti.cnr.it

Anna Tonazzini
anna.tonazzini@isti.cnr.it

[1] Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Via G.Moruzzi, 1, 56124 Pisa, Italy

⊘ Springer

collectively, complex background or noise. There is widespread interest in binarizing degraded documents, and a variety of methods have been proposed so far [2, 3]. Among them, local and adaptive thresholding, or recurrent, convolutional, or deep neural networks can, to some extent, handle degradation such as non-uniform lighting, image contrast variation, changes in stroke width and connection, faded or seeping ink [4–8].

Virtual restoration and binarization are often complementary or preparatory to each other. Indeed, a manuscript from which the strongest degradation has been removed can be binarized more effectively, as in [8], where a NN learns the degradation and iteratively refines the output, which is then binarized using Otsu's global threshold. Conversely, binary maps, where the foreground text has been located and extracted, can form the basis for accurate restoration, as in [9], where the main text is mapped onto a clean inpainted background.

In this paper we tackle the virtual restoration of manuscripts affected by the bleed-through degradation, which can occur when both sides of the paper are written, and poor storage conditions or the natural aging of the materials have made the ink penetrate from one side to the other. We assume that both sides of the manuscript are available, which is a common condition in digitized archives and libraries.

We propose a procedure based on a multilayer shallow neural network with back-propagation training [10]. The NN, by exploiting the information from both sides of the manuscript, classifies the pixels into three classes: foreground text, bleed-through noise or background. Then, the noisy pixels are suppressed and replaced by background texture inpainting. The NN is implemented in such a way that it automatically adapts to the manuscript to be restored, i.e.,, without requiring prior learning from a large class of other similar manuscripts.

To generate the training set from the data images, an intuitive way might be to extract a set of degraded patches and then somehow estimate the corresponding ground truths. When an analytical data model exists that describes the degradation, the training set can be instead self-generated starting from ground truths drawn from the clean zones of the manuscript. Indeed, the corresponding degraded patches can be obtained by feeding these ground truths into the model. This second way of working may be somewhat simpler. In particular, we use the theoretical model of blending proposed in [11], which approximates the physical phenomenon of ink diffusion through the paper fiber and its filtration on the reverse of the sheet. We use the model in direct mode, i.e.,, to generate data consistent with the degraded manuscript we want to restore. The point of view is to learn the degradation affecting the manuscript in question, so that the trained NN is specialized in classifying, as clean or noisy, the pixels of this and other manuscripts affected by that specific degradation.

The paper is organized as follows. In Sect. 2 we briefly present the state of the art in the field of bleed-through removal, and highlight the improvement with respect to our previous work [12]. Section 3 describes the fundamentals of our classification technique, i.e.,, the data model to construct the training set, and the architecture of the shallow NN. Section 4 is devoted to an overall block-by-block procedure for joint registration and virtual restoration of misaligned recto–verso pairs. In Sect. 5, we present the experimental results, producing a quantitative analysis on a reference dataset, and a qualitative analysis on historical manuscripts. Finally, Sect. 6 concludes the paper, with prospects on future developments.

2 Related works

The removal of bleed-through degradation has attracted some interest individually and outside the context of binarization, as a major issue for the enhancement of ancient manuscripts and degraded documents in general. Indeed, heavy bleed-through cannot be completely removed by binarization alone, due to its significant overlap with the foreground text and the wide variation in its extent and intensity. Methods specifically designed for bleed-through reduction have therefore been proposed. The so-called blind methods exploit only the information of the front side, such as the blind source separation technique proposed in [13], the unsupervised recursive segmentation suggested in [14], and the conditional random field presented in [15]. More recently, [16] and [17] propose segmentation of multiple color representations via Gaussian mixture models and the EM algorithm.

In general, images of both sides of the manuscript are available in the modern documental repositories, and their joint use is recommended because the verso page brings additional data that is complementary to that of the primary page. This wealth of information allows for the design of algorithms capable of selectively removing only unwanted interference, leaving the rest of the manuscript unaltered, thus performing a very fine virtual restoration. On the other hand, the need for perfect alignment of the two images greatly complicates the problem, especially in the presence of document skews, different image resolutions or wrapped pages when scanning books. Thus, dedicated registration algorithms have been developed for recto–verso manuscripts (see [18–20]).

For carefully aligned recto–verso pairs, restoration was then proposed according to a variety of different approaches. For example, in [21] a, classification is performed by

segmenting the recto–verso joint histogram with the aid of available ground truths, in [22] a regularized energy uses a data term derived from small sets of user-labeled pixels and a smoothness term based on dual-layer Markov Random Fields, and in [23] correlated component analysis is used to separate the information layers. The work in [9] extends data decorrelation [24] for deriving a fast, practical procedure, which incorporates a binarization step as well.

Another delicate aspect concerns the filling-in of the pixels classified as degradation with values that simulate the grain of the paper, in order to guarantee the fidelity of the restored manuscript to the original one. Usually, a substitution with the average value of the surrounding pixels or a random fill-in [15] are suggested for the inpainting of the degraded pixels, but they result in visual imprints, especially when the size of the degraded area is large. In [16, 25], the surrounding context is accounted for through sparse image representation with dictionary learning, and Gaussian conditional simulation.

In the conference paper referenced in [12], we experimented with the approach, further developed here, of using a shallow neural network trained with back-propagation in order to classify recto–verso pixels into foreground text, bleed-through noise or background, and then of replacing noisy pixels with close-up background texture. The peculiarity of the approach is that the NN is trained on a dataset built from the same degraded images through the theoretical model in [11], which approximates the physical phenomenon of bleed-through.

In addition to outline and motivate this approach, in [12] we discussed from a qualitative point of view the experimental results obtained on real, heavily damaged manuscript. In this paper, the method is described in more detail and a quantitative evaluation is included as well. In particular, we conducted a systematic, quantitative analysis on the 25 recto–verso pairs of a public database containing ground truths [26, 27]. Furthermore, we describe the block-by-block procedure used for jointly performing registration and restoration of originally misaligned recto–verso pairs.

3 The classification phase through a shallow NN

Our approach to manuscript virtual restoration entails two processing steps: a classification step of the recto–verso pixels into safe and corrupted, and a proper restoration step, based on the inpainting of the corrupted pixels with the texture of the surrounding background. The classification step is performed through a shallow NN employing a model-based learning dataset.

3.1 The data model used to construct the training set

In most of the manuscripts that we have examined, the seeped ink has also diffused through the fiber of the paper. As a result, the bleed-through pattern usually appeared as a smudged, lighter version of the opposing text that generated it. This does not mean that, on the same side, the bleed-through pattern was always clearer than the foreground text. Indeed, in each side the intensity of the bleed-through is usually highly variable, i.e.,, highly non-stationary, and can sometimes be as dark as the foreground text.

We thus adopt a measure of optical density, defined for each pixel t of any observation channel as: $D(t) = -\log\left(\frac{s(t)}{b}\right)$, where $s(t)$ is the intensity, and b is the mean background value. To model bleed-through degradation in front-to-back manuscripts using this measure, given a pair of registered recto–verso images acquired in any observation channel, we propose the following non-stationary linear model:

$$
\begin{aligned}
D_r^{\text{obs}}(t) &= D_r(t) - q_v(t)\log\left(\frac{h_v(t) \otimes s_v(t)}{b_v}\right) \\
D_v^{\text{obs}}(t) &= D_v(t) - q_r(t)\log\left(\frac{h_r(t) \otimes s_r(t)}{b_r}\right)
\end{aligned}
\tag{1}
$$

In Eq. (1), D^{obs} and D are the observed and ideal optical densities, with the subscripts r and v indicating the recto side and the reflected verso side respectively. The symbol $\otimes$ indicates the convolution between the ideal intensity s and a Point Spread Functions (PSF), h, which describes the smearing of the ink that penetrates the paper. Finally, the space-variant quantities q_r and q_v, in the interval [0, 1], have the physical meaning of percentages of ink penetration from one side to the other.

In previous works [11, 20, 28], we proposed inverting the above model to virtually restore the recto–verso pair. Based on the observed densities of the two sides, we first inverted the model assuming an ideal density identically zero on the opposite side, thus obtaining estimates of the percentages of ink penetration at each pixel. After some simple adjustments of these percentages, the system can then be solved with respect to ideal density maps, resulting in virtually restored sides of the manuscript.

The model approximates the phenomenon of ink transparency quite well almost everywhere, apart from the occlusion areas where the inks of the two sides overlap. In those areas, it is not possible to estimate the ink penetration percentage as a ratio of the observed densities, since the ideal density is not truly zero. With the two observed densities almost equal, small fluctuations make the value of their ratios unpredictable. Consequently, during the

restoration phase, one of the two sides will present a sort of "hole" (values close to those of the background) in correspondence with the occlusion zones.

Here we aim to solve the direct problem of Eq. (1) to generate the data, rather than solving the inverse problem to estimate the unknown ideal densities. Thus, it is easier to extend the model to adequately describe areas of occlusion, e.g., assuming that the density of the foreground text does not increase due to ink leaks. In practice, since we know the nature of each pixel, when a pixel is foreground text in both sides, the density value is saturated to that of the original side.

3.2 The neural network: learning and recall

We adopted a simple feedforward network having the architecture of a multilayer shallow neural network with a single hidden layer and a backpropagation training [10]. This network is a pattern recognition NN that can be trained to classify inputs according to target classes.

Given a recto–verso pair of degraded manuscripts, the network is first trained from the data itself, and then used to classify each pixel of the reference side into three different classes: *background*, *foreground* and *bleed-through*.

Figure 1 illustrates the general architecture of such a network.

In the specific, our network has an input layer D consisting of 2 nodes, a hidden layer composed of 10 neurons, and a final output layers that returns a 3-value vector C. The two input values consist in the density of the pixel in the reference side and its aligned pixel in the opposite side. These values are mixed through the synaptic weights W^1 and the bias b^1 and fed to the 10 hidden neurons. The resulting values are passed through non-linear transfer functions f_1, of the kind of *sigmoid* functions, to account for possible non-linear relationships across the data, and then mixed again with the weights W^2 and bias b^2 to give the three output values. These are then normalized via a *softmax* function f_2, which furnishes the vector C of the probabilities that the reference pixel is classified as one of the three classes. As usually done, the pixel is assigned with the class of maximum probability.
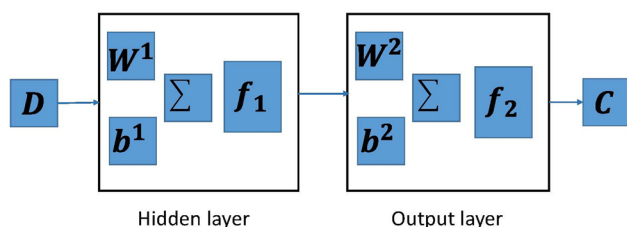
Fig. 1 Diagram illustrating the general architecture of a shallow NN with one hidden layer

To build the training set, we manually select in the manuscript a few pairs of patches containing clean text, binarize them to generate the target maps, and mix them symmetrically using the model described in the previous section to generate the input maps. The generation of the input and target maps for one pair of patches is sketched in Fig. 2.

The two patches are fed to the system in Eq. (1) in a forward manner, with different values of the percentage of ink infiltration, in order to synthetically generate recto–verso text samples with bleed-through (Fig. 2). For the generation of a single pair of input maps the model is taken as stationary, i.e.,, with fixed percentage of ink leakage. However, the construction of multiple pairs with different percentage values means that, taken together, samples of non-stationary degradation will be presented to the network. Figure 3 shows the training set obtained from one pair of clean patches through mixing with parameters in the range (0, 1).

After building the NN, in the recall phase the network is applied sequentially to each pixel of each side, according to a visiting scheme that, as described in next section, allows to associate the pixel at hand with the corresponding one in the other side. The output of the NN consists in the classification of the pixel as foreground text, bleed-through noise or background. When all pixels of a side are classified, it is immediate to obtain the binarized version of the manuscript, combining pixels classified as noise and background in the same class. If the goal is instead to obtain a virtually restored version of the manuscript, the foreground text pixels and the background pixels are given their original values, and the noisy pixels are replaced with samples estimated from the closest safe background region.

4 Virtual restoration of misaligned recto–verso manuscripts

As mentioned, pixel classification and inpainting are performed in the context of a block-by-block procedure that allows to treat misaligned recto–verso pairs without the need for a real global registration.

This overall block-by-block procedure works on the sequence of small adjacent blocks in the images. In brief, at each block, let us say in the recto, an initial phase of search for the best matching block in the reflected verso is performed. The matching opposite block is found based on the relative displacement between the reference block and the opposite block at the same position. NN-based pixel classification is then applied to the pair of aligned blocks, followed by inpainting of the degraded pixels.

The full virtual restoration procedure is illustrated in the diagrams of Figs. 4 and 5.
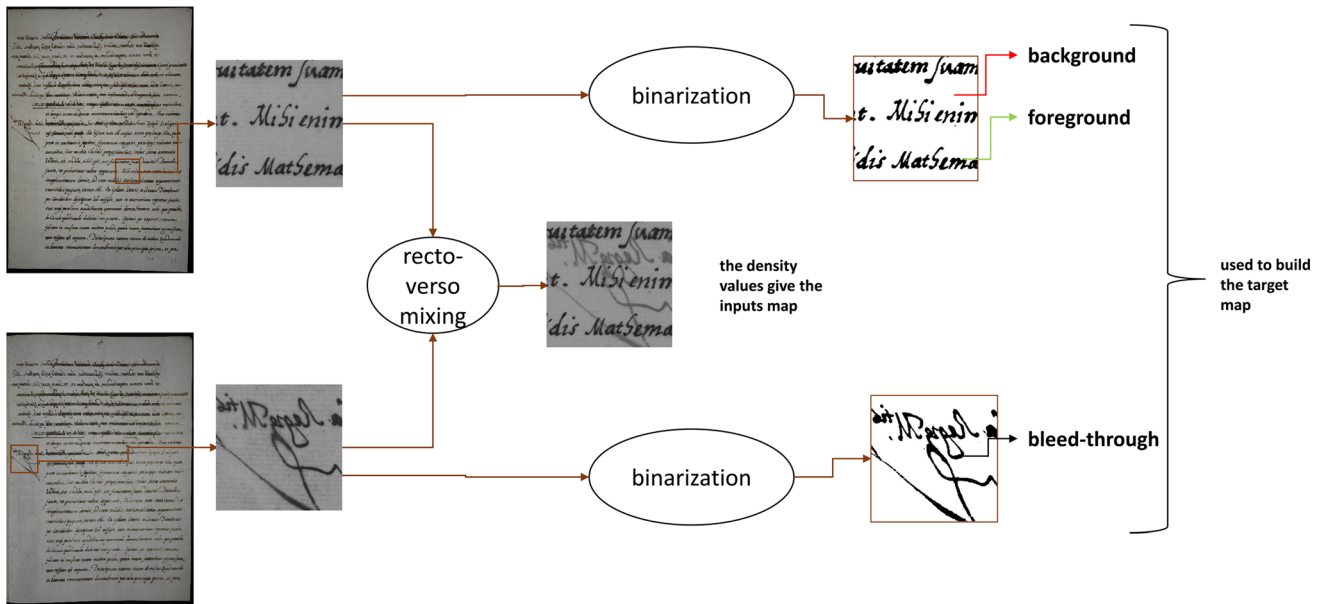
Fig. 2 Diagram illustrating the generation of one of the input maps and the related target maps, to be used in the learning phase

Fig. 3 Example of construction of the data set based on the selection of a single pair of clean patches and ten values of the ink seeping percentage


Figure 4 summarizes the block-by-block mechanism. Each block is restored independently from the others and, after restoration, it is placed in the proper position in the restored image.

The whole restoration procedure is applied twice. The first time we consider the recto as a reference image and detect pairs of blocks by simultaneously moving a square window on the two sides so as to cover the entire image domain. The only block of the recto is classified, based on
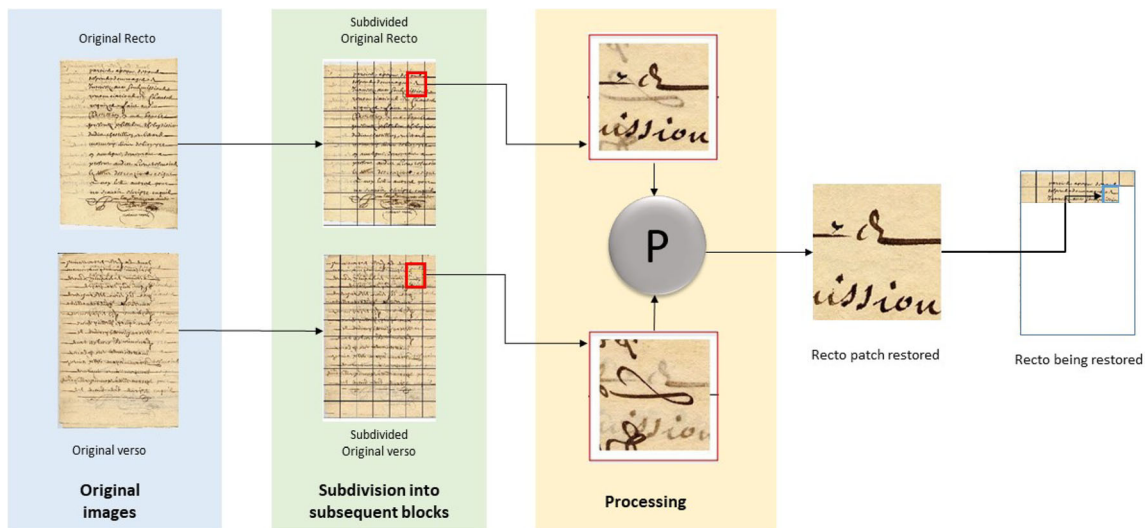
Fig. 4 The block-by-block processing mechanism

Fig. 5 The virtual restoration at the level of one single block
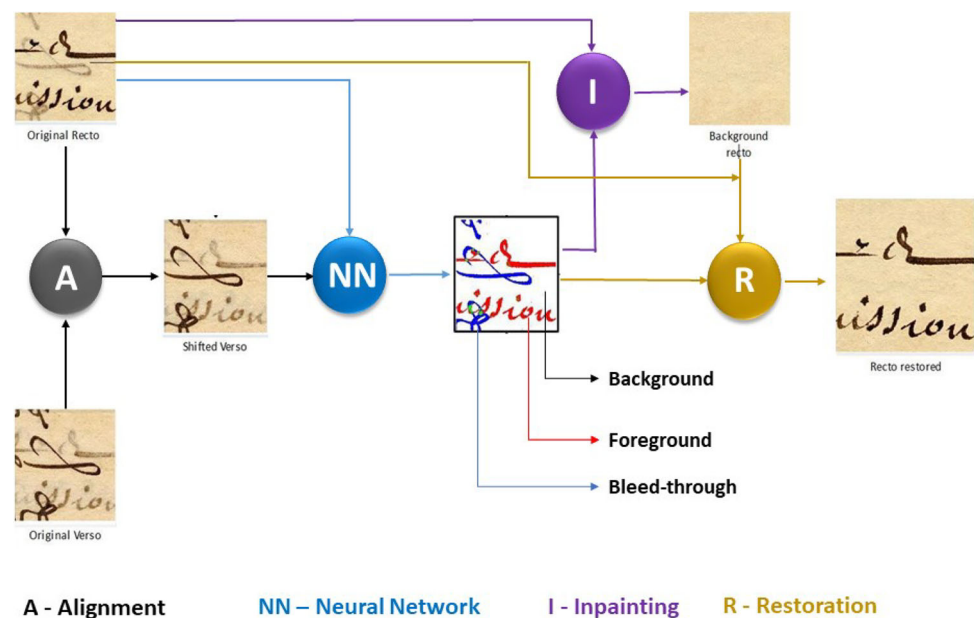
information from both, and then restored. The second time the procedure is repeated by inverting the role of recto and verso.

Although the classification of the verso block could be deduced as well from the classification of the recto block, the former refers to a displaced block, and therefore different from that detected by the moving square window. Direct use of this classification of the verso would involve solving a mosaicing problem to reconstruct the entire verso image, and this is not cost effective. Therefore, since one side is restored at a time, by virtue of the block-by-block mechanism the restored side remains geometrically unaltered, i.e.,, misaligned like the original. Since there is no

global registration between recto and verso, the resolution of the images is not reduced by any interpolation process.

For RGB manuscripts, we assume a perfect alignment of the three color planes of a side, so that we can compute the relative shifts of the blocks in a single pair of channels only.

Figure 5 illustrates the procedure at the level of each pair of detected recto–verso blocks. Virtual recovery of each block of the side under processing is achieved by first looking for the truly matching block in the opposite side. Then, pixel classification is performed using the trained NN. This approach to the virtual restoration of misaligned back-to-front manuscripts is similar to that proposed in [20], where block restoration is performed through the

inversion of the model described in Sect. 3.1 instead of using a NN. It assumes that, at the very local level, the reciprocal deformation between the two sides amounts to a translation only. For each small block in a side, its cross-correlation with the opposite block at the same location gives their relative displacement, which can be used to locate the best matching opposite block by a simple shift.

The NN is applied to the two aligned blocks, in order to discriminate the pixels of the reference block in background, foreground and bleed-through. Foreground and bleed-through pixels are removed from the original reference block and replaced with background pixels using the exemplar-based image inpainting technique described in [29]. Finally, the pixels recognized as belonging to the foreground class are taken from the original image and placed on the inpainted image.

5 Discussion of the experimental results

Our experiments were conducted on images of a benchmark dataset comprised of ground truths, and on images of historical manuscripts. In both cases we used the shallow NN implemented by the function `patternnet` of the Matlab Deep Learning Toolbox, with a single hidden layer consisting of 10 nodes. A schematic diagram of this network is shown in Fig. 1.

We have chosen the scaled conjugate gradient as minimization algorithm (`training function`), and the cross-entropy to measure the network performance (`performance function`) during training. The transfer function f_1, applied to the outputs of the hidden layer, is chosen as a *logsinc* function, equal for all the 10 neurons. The transfer function f_2, applied to the vector of the three output neurons, is the *softmax* function.

For each experiment we built the training set from a single pair of images, by selecting 8 pairs of clean patches (8 in the recto and 8—non corresponding—in the verso) of size 30×30 pixels. Each pair of patches was then mixed using the model in Eq. (1), with 10 different values of ink penetration percentage. The dataset was randomly split into training set (the 70% of patch pairs) and validation set (the remaining 30%).

Tests performed with a number of neurons higher than 10 did not provide valuable improvements in the quality of the results.

5.1 Quantitative analysis on a benchmark dataset with ground truths

We tested our procedure on a dataset of high-resolution images of ancient documents affected by bleed-through, which is a benchmark in the field [26, 27]. The dataset

includes 25 pairs of already registered recto–verso images, taken from larger manuscripts, with various levels of bleed-through. The dataset is very varied, both in terms of the severity of the degradation and the morphological characteristics of the written texts.

In addition to the degraded images, a binary ground truth mask of the foreground text is provided for each image. Although the foreground text is labeled manually in these ground truth images, they are commonly used for a quantitative analysis of the binarized version of the restored images and then, indirectly, of the bleed-through suppression quality. In our case, we can thus quantitatively evaluate the classification performance of our NN.

As quality indices we adopted the probability $\mathrm{Fg_{err}}$ that a pixel in the foreground text was classified as background or bleed-through, the probability $\mathrm{Bg_{err}}$ that a background or bleed-through pixel was classified as foreground, and T_{err}, i.e.,, the weighted mean of $\mathrm{Fg_{err}}$ and $\mathrm{Bg_{err}}$, with the weights being the numbers of the foreground pixels and the background pixels as they result from the corresponding ground truth images. The T_{err} error indicates the probability that any pixel in the image was misclassified. According to [26], these quality indices are defined as:

$$\mathrm{Fg_{err}} = \frac{1}{N_{\mathrm{Fg}}} \sum_{t \in \mathrm{GT(Fg)}} | \mathrm{GT}(t) - B(t) |$$

$$\mathrm{Bg_{err}} = \frac{1}{N_{\mathrm{Bg}}} \sum_{t \in \mathrm{GT(Bg)}} | \mathrm{GT}(t) - B(t) |$$

$$T_{\mathrm{err}} = \frac{N_{\mathrm{Fg}} \mathrm{Fg_{err}} + N_{\mathrm{Bg}} \mathrm{Bg_{err}}}{N}$$

where GT is the ground truth, B is the binarized classification map, GT(Fg) is the foreground region of the ground truth image constituted of N_{Fg} pixels, GT(Bg) is the complementary background region of the ground truth image constituted of N_{Bg} pixels, and N is the total number of pixels in the image.

We conducted two types of experiments. The first experiment was totally synthetic, in the sense that the recto–verso images to restore were numerically built based on the ground truths available in the benchmark dataset. In the second experiment, the NN was trained on some of the degraded recto–verso pairs of the dataset, i.e.,, a number of different NN's were built. Then, each NN was used to classify all the 25 pairs, and the results were comparatively discussed.

5.1.1 Synthetic data

In the synthetic experiments, we built a clean recto–verso pair by placing clean foreground texts on a textured background obtained by scanning a clean sheet of paper. The foreground texts were obtained from two images of the

dataset, by picking up the graylevels of the pixels that are "0" (black) in the corresponding ground truths.

Figure 6a, c shows the degraded pair, obtained by mixing the ideal one through the data model of Eq. (1), where the percentage of penetrating ink has been increased from 0.1 to 0.6 (left to right). Figure 6b, d shows instead the result of applying the NN (training and recall) plus inpainting on those images. As expected, the reconstructions are very good, with the following quality indices: $T_{\text{err}}^r = 0.0083$, $\text{Fg}_{\text{err}}^r = 0.0052$, $\text{Bg}_{\text{err}}^r = 0.023$, $T_{\text{err}}^v = 0.0058$, $\text{Fg}_{\text{err}}^v = 0.0027$, $\text{Bg}_{\text{err}}^v = 0.016$, where the superscripts r and v indicate the side.

5.1.2 Real data

In the second experiment, we performed eight tests, each time using a different pair of benchmark images for learning. The eight chosen training images are shown in Fig. 7. In the figure, we only show the recto side, but training requires both recto and verso. These images are representative of the entire dataset in terms of characters morphology and bleed-through intensity. The trained NN was then applied to the sample pair and the other 24 pairs. All 25 pairs were thus first classified and then virtually restored by inpainting.

Figure 8 shows the foreground error, the background error, and the total error when training the NN with the recto–verso pair corresponding to the first image shown in Fig. 7. These errors are compared with those of the method proposed in [9].

It can be observed that the ranges of errors are not very wide, demonstrating the NN's generalization capability. In general, the total error depends on the level of degradation of the images, and on the difficulties their restoration may present (e.g., when many overlapping pixels between recto and verso are present, or when a signifying ink smearing exists, or when bleed-through presents a high variability in the entire image). Thus, it may happen that the network built with a given sample pair performs better in the "simpler" images than in the sample pair used to train it. However, the comparison of the total errors obtained with the networks built using the eight test images shows that in general the best result is obtained with the self-trained NN, as we will see in Figs. 9 and 10.

We can also observe that the variability of the errors on the entire dataset is similar for the two compared methods (the shape of the two plots is similar), and when a different training pair is used. As already mentioned, this mainly depends on the intrinsic difficulty that restoration of some images may present. Overall, the NN-based method outperforms the [9] method for foreground estimation but not

Fig. 6 Synthetic experiment: **a** degraded recto; **b** recto restored with NN; **c** degraded verso; **e** verso restored with NN

(a)

(b)

(c)

(d)
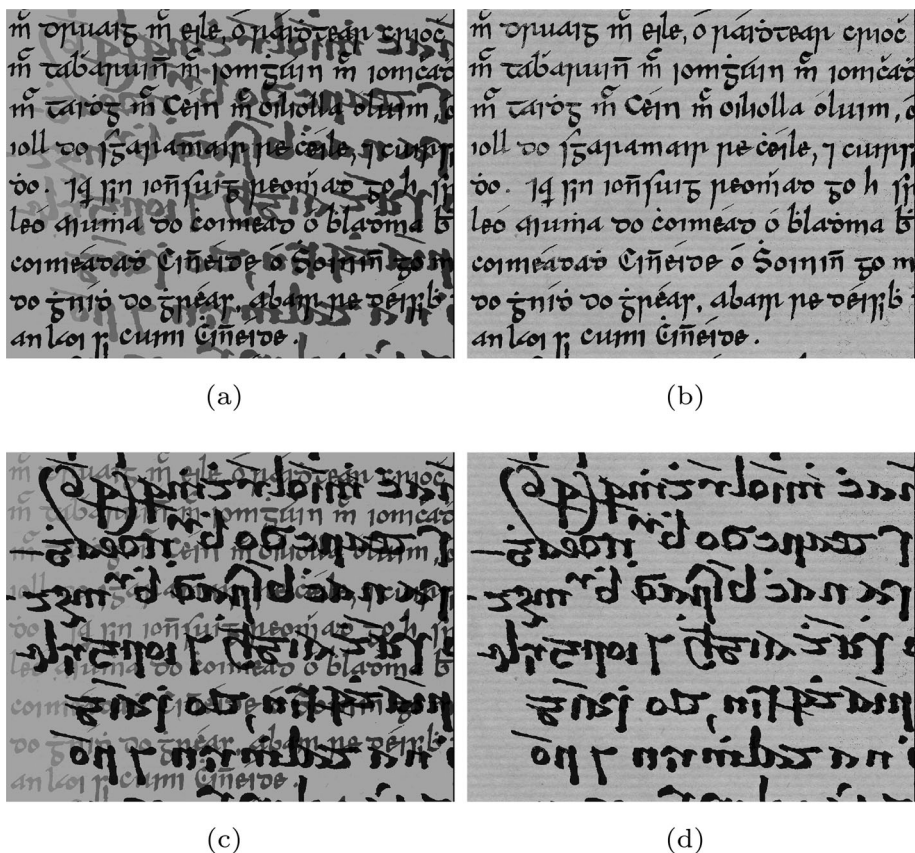
Fig. 7 Recto of the 8 pairs used for training 8 different networks to be applied to the restoration of the benchmark dataset

(a)　(b)　(c)　(d)

(e)　(f)　(g)　(h)

(a)

(b)

(c)
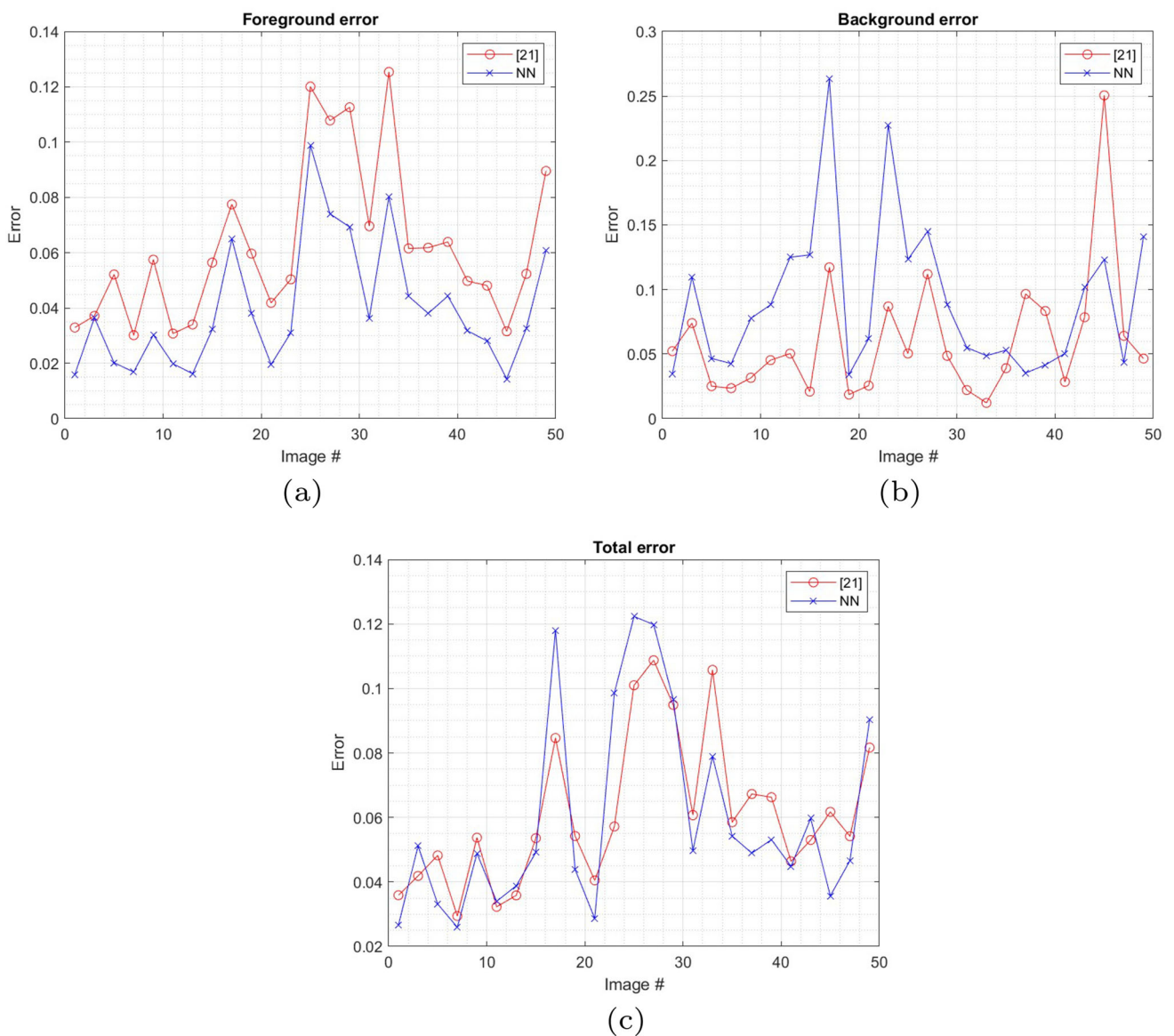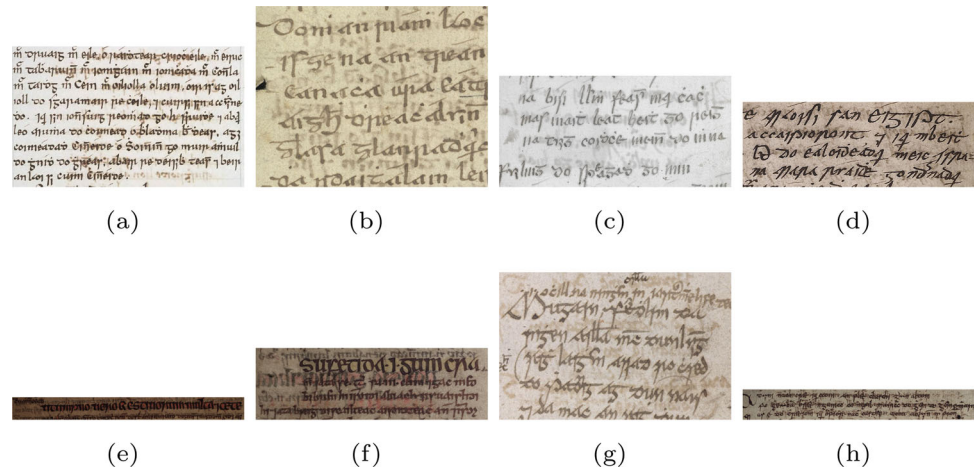
Fig. 8 Plots of the errors for all the 25 pairs when the NN has been trained with the sample pair of Fig. 7a: **a** Foreground error; **b** Background error; **c** Total error

Fig. 9 Restoration of the training pair in Fig. 7a: **a**, **b** degraded recto and verso; **c**, **d** NN trained with the training pair in Fig. 7a ($T_{err} = 0.027$); **e**, **f** NN trained with the training pair in Fig. 7e ($T_{err} = 0.059$)

(a)

(b)

(c)

(d)

(e)

(f)
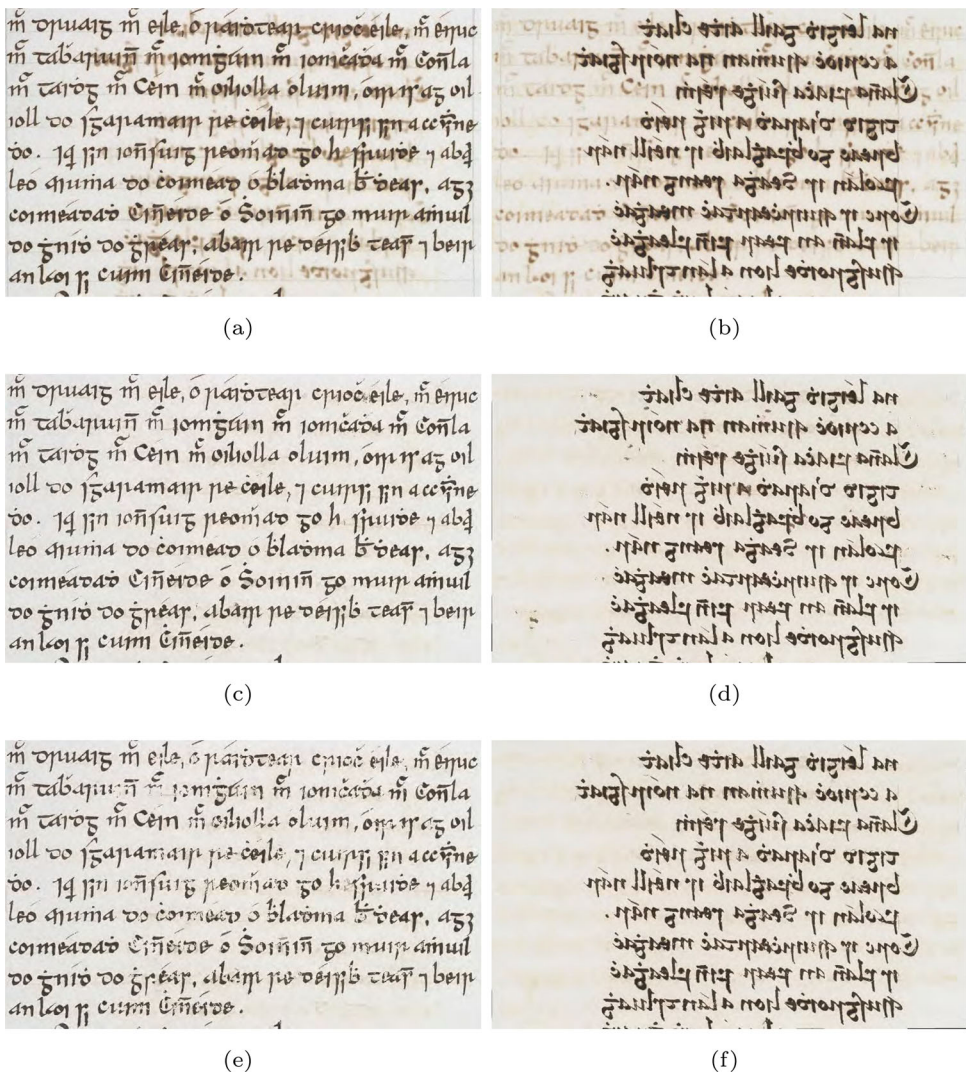
Fig. 10 Restoration of the training pair in Fig. 7e: **a**, **b** degraded recto and verso; **c**, **d** NN trained with the training pair in Fig. 7e ($T_{err} = 0.10$); **e**, **f** NN trained with the training pair in Fig. 7a ($T_{err} = 0.12$)

(a)

(b)

(c)

(d)

(e)

(f)
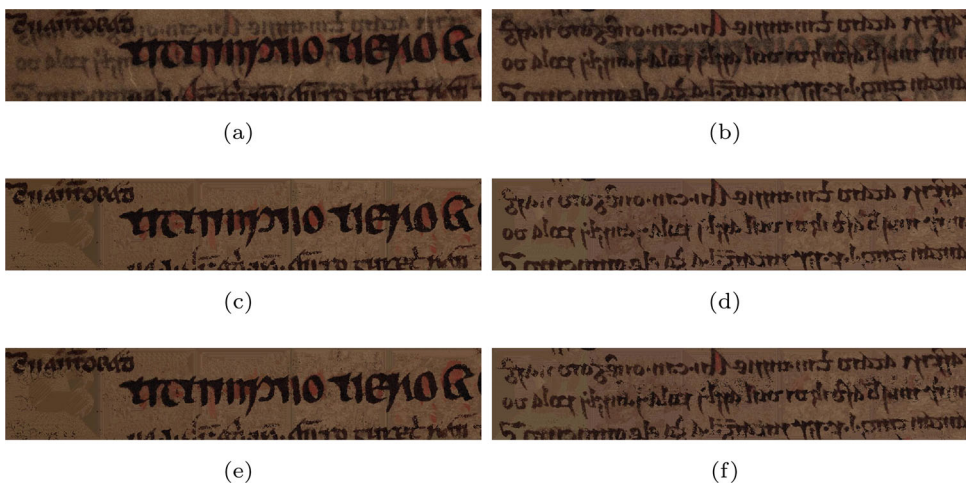
for background estimation, where that method slightly prevails. In practice, true foreground pixels are captured fine, but some false foreground pixels are estimated instead of bleed-through pixels. Regarding the total error, the two methods are comparable, with some images better restored by method [9], and others by the proposed NN approach.

However, for a given pair, the self-trained NN outperforms the method [9] (in Fig. 8c the error on image in Fig. 7a used to train the network is 0.027 compared to 0.034).

In Fig. 9, we show the results of the virtual restoration of the sample pair corresponding to Fig. 7a when using the self-trained network or the network trained with the sample pair corresponding to Fig. 7e (best and worst total error). We may observe that the self-trained network gives significantly better restorations.

In Fig. 10, we present a similar comparison for the virtual restoration of the sample pair in Fig. 7e when using the self-trained network or the network trained with the sample pair in Fig. 7a (best and worst total error). In this case, the image is heavily degraded, and the values of the total errors remain very high regardless of the sample pair used for training.

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:

$$\text{Precision} = \frac{Sum(\text{FT}_R \cap \text{FT}_{\text{GT}})}{Sum(\text{FT}_R)}$$

$$\text{Recall} = \frac{Sum(\text{FT}_R \cap \text{FT}_{\text{GT}})}{Sum(\text{FT}_{\text{GT}})}$$

$$F\text{-measure} = \frac{2 \times (\text{Precision})(\text{Recall})}{\text{Precision} + \text{Recall}}$$

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.

In our method, the binary map of the restored foreground text is a preliminary result. We computed the average values of the above metrics on the 25 pairs, for the most favorable setting, i.e.,, using the self-trained NN for each pair. From a computational point of view, this is not excessively heavy, due to the efficiency of the training phase. Of course, the performance of the method increases sensibly with this implementation, e.g., with respect to the implementation where the same NN is used for all the 25 pairs, as in the experiments that produced the plots of Fig. 8.

The obtained values are compared in Table 1 with those of the non-blind method [21], the blind method [15], and the non-blind method [9]. 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 that the resulting metric values can be affected by this choice. For the non-blind method [21] and the blind method[15], we

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

Method	Precision	Recall	*F*-measure
Non-blind 2013 [21]	0.92	0.87	0.89
Blind 2016 [15]	0.89	0.86	0.87
Non-blind 2021 [9]	0.93	0.82	0.90
Proposed	0.94	0.87	0.90

report here the values that were 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 correct. Also the recall value of our NN is much higher than that of the method in [9], which means that it is able to detect more foreground pixels, and comparable with those of [21] and [15]. It is to be noted that better values of precision and recall correspond to lower foreground error, as confirmed by the plot in Fig. 8a.

5.2 Qualitative analysis on historical misaligned recto–verso pairs without ground truths

The metrics considered in the previous section are highly suitable for an effective quantitative evaluation of a binarization/classification task. However, for a virtual restoration intervention, the quality of the background reconstruction, which replaces the bleed-through pattern, is also of great importance, and is related to the effectiveness of the inpainting algorithm used. This simulated texture, together with the fidelity to the original of the reconstructed foreground text, and the suppression of as much as possible bleed-through noise, contributes to the overall appearance of the restored manuscript. In general, this aspect is evaluated qualitatively, also given the complexity and variety of the characteristics to be considered. On the other hand, when dealing with real manuscripts, ground truths are not available.

The experiments illustrated in this section were conducted on rather demanding specimens, selected from the correspondence of Christoforus Clavius, conserved in the Historical Archives of the Pontifical Gregorian University in Rome. For the learning phase of the neural network used for the restoration, the manuscript was converted to grayscale, as the chromatic information in this case is not essential for classification purposes. Since the three RGB channels of a color manuscript share the same classes, the restored version of the color manuscript can be obtained directly.

Figure 11a, d show the original (unregistered) recto and verso of one of such letters, Fig. 11b, e show the results

Fig. 11 Application of the whole procedure: **a** original recto; **b** recto restored with NN; **c** recto restored with [9]; **d** original verso; **e** verso restored with NN; **f** verso restored with [9]. Original images **a**, **d**: reproduction by courtesy of The Historical Archive of the Pontificia Università Gregoriana, APUG 529/530, c. 131r/v (Fondo Clavius)

(a) (b) (c)

(d) (e) (f)

produced by our composite procedure using NN's, and Fig. 11c, f show the results obtained with the procedure described in [9]. As previously mentioned, the restored images are not registered, as both the present method and the algorithm in [9] use the block-by-block alignment and restoration mechanism described in Sect. 4.

With the NN the results might not be quantitatively perfect, however, from the qualitative point of view they are correct, as the two completely overlapping texts, almost indistinguishable in the originals, have been optimally separated. Furthermore, it is evident that the NN-based procedure outperforms the procedure in [9], which is still based on a recto–verso mixing model, but stationary linear in the intensity.

In Fig. 12 we compare the binarization results provided by our NN with those obtained by the segmentation method based on a Laplacian energy, which was the winner of the H-DIBCO-2018 competition [2, 30, 31]. In our results, the

bleed-through pattern was almost completely removed. Conversely, we can observe that the method in [30], while providing excellent results on documents with limited amount of transparency or with other types of degradation, gives unsatisfactory results with such a heavily damaged manuscript. However, we must specify that, to obtain the result on each side, the NN exploits twice of information compared to that used by the method in [30]. On the other hand, to binarize both sides, the amount of total information available and exploited by the two methods is the same. The crucial difference is that in our method the overall information is jointly exploited.

The generalization ability of our network was also tested. Figure 13 shows the results of applying the same NN constructed for the manuscript of Fig. 11a, d on a different recto–verso manuscript, this time presented in color.

However, our method still has some defects. For instance, in the binarization produced by the NN, the
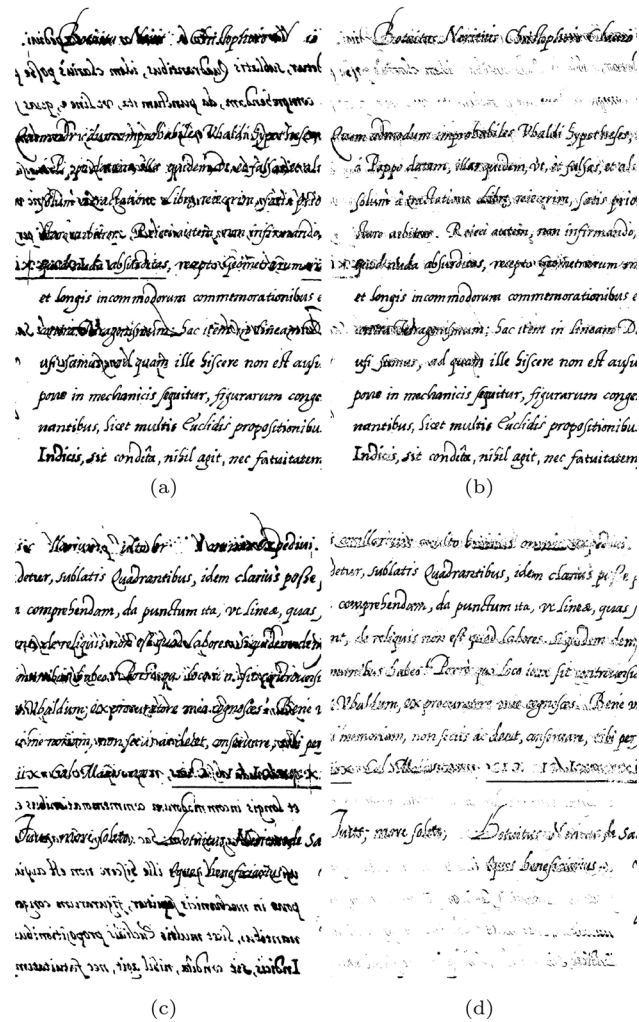
Fig. 12 Binarization of the original recto–verso manuscript of Fig. 11a, d: **a** recto binarized with [30]; **b** recto binarized with our NN; **c** verso binarized with [30]; **d** verso binarized with our NN

legibility of the foreground text extracted suffers from a sort of "corrosion" of the most compromised characters in correspondence with the occlusion areas. Probably, this is caused by a still unsatisfactory modeling of the overlap of recto and verso texts, but also, perhaps, by an insufficient presentation of samples where occlusions occur, or by the need to include a specific fourth class for occlusions.

More in general, our approach lends itself to variations and improvements, both at the data model level and at the network architecture level. For example, even the simple

network architecture currently considered could be modified to include more and different input features, which describe the texture of the text traits to be classified and which take into account the surrounding context [32]. Subsequently, deep and convolutive neural networks could be tested, and their implementations could be optimized [33, 34].

We intend to investigate in all these directions.

6 Conclusions

We showed that the common availability of scans of both sides of pages of ancient manuscripts affected by noise in the form of ink leaks allows for the use of a very simple shallow NN to classify pixels into text, background and noise. In fact, for recto–verso manuscripts it is possible to find analytical models that approximately describe the degradation with varying levels of intensity, in such a way to generate useful examples for training the network directly from the single document to be restored, without having available large sets of different manuscripts. The output of the NN, which is essentially a classifier, can be used to produce a binarization of the foreground text or a virtual restoration version of the manuscript that preserves the fullness of its information content and esthetics. For a performance evaluation of the method, we focused on the quality of binarization, which can be quantitatively judged on a public dataset of recto–verso manuscripts associated with ground truths. The results were also discussed in terms of the network's ability to generalize beyond the manuscript in question. Furthermore, we showed the binarization results of our NN on a heavily damaged historical manuscript in comparison with those provided by the H-DIBCO-2018 competition winning algorithm.

The method still has some shortcomings in pixel classification where the two texts overlap, which should be managed both at the data model level and at the network architecture level. This represents the main aspect of our future work on this topic. In this regard, we intend to study nonlinear and nonstationary diffusion models, modify the network architecture to include both more input features and more output classes, and experiment with deep and convolutive neural networks.

Fig. 13 Application of the same network to a different manuscript: **a** original recto; **b** recto restored; **c** original verso; **d** verso restored

(a)

(b)

(c)

(d)

Data availability The datasets analyzed during the current study are available in the *Irish Script on Screen* repository, https://www.isos. dias.ie, and at the Historical Archive of the Pontifical Gregorian University in Rome. The results generated during the current study by processing the above datasets/images are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Tonazzini A, Savino P, Salerno E, Hanif M, Debole F (2019) Virtual restoration and content analysis of ancient degraded manuscripts. Int J Inf Sci Technol (iJIST) 3:16–25

2. Pratikakis I, Zagori K, Kaddas P, Gatos B (2018) ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018). In: Proceedings of 16th international conference on frontiers in handwriting recognition (ICFHR), pp 489–493

3. Pratikakis I, Zagoris K, Karagiannis X, Tsochatzidis L, Mondal T, Marthot-Santaniello I (2019) ICDAR2019 competition on document image binarization (DIBCO 2019). In: Proceedings of the 15th IAPR international conference on document analysis and recognition (ICDAR 2019), pp 1547–1556

4. Pai Y, Chang Y, Ruan S (2010) Adaptive thresholding algorithm: efficient computation technique based on intelligent block detection for degraded document images. Pattern Recogn 43:3177–3187

5. Westphal F, Lavesson N, Grahn H (2018) Document image binarization using recurrent neural networks. In: 13th IAPR international workshop on document analysis systems (DAS2018), proceedings, pp 263–268

6. Tensmeyer R, Martinez T (2017) Document image binarization with fully convolutional neural networks. In: 14th IAPR international conference on document analysis and recognition (ICDAR 2017), proceedings, pp 99–104

7. Vo Q, Kim S, Yang H, Lee G (2018) Binarization of degraded document images based on hierarchical deep supervised network. Pattern Recogn 74:568–586

8. He S, Schomaker L (2019) DeepOtsu: document enhancement and binarization using iterative deep learning. Pattern Recogn 9:379–390

9. Savino P, Tonazzini A (2020) A procedure for the routinary correction of back-to-front degradations in archival manuscripts. In: Nguyen NT, Hoang BH, Huynh CP, Hwang D, Trawinski B, Vossen G (eds) Proceedings of IWCIM 2020, ICCCI 2020, computational collective intelligence, pp 838–849. Springer

10. Hagan MT, Demuth HB, Beale MH (1996) Neural network design. PWS Publishing, Boston

11. Tonazzini A, Savino P, Salerno E (2015) A non-stationary density model to separate overlapped texts in degraded documents. SIViP 9:155–164

12. Savino P, Tonazzini A (2022) A shallow neural net with model-based learning for the virtual restoration of recto–verso manuscripts. In: 1st International virtual conference on visual pattern extraction and recognition for cultural heritage understanding—VIPERC 2022. https://ceur-ws.org/Vol-3266/paper3.pdf

13. Tonazzini A, Bedini L, Salerno E (2004) Independent component analysis for document restoration. Int J Doc Anal Recogn 7:17–27

14. Fadoua D, Bourgeois FL, Emptoz H (2006) Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. Document Analysis Systems VII, Lecture Notes in Computer Science, vol 3872, pp 27–38. Springer

15. Sun B, Li S, Zhang XP, Sun J (2016) Blind bleed-through removal for scanned historical document image with conditional random fields. IEEE Trans Image Process 25:5702–5712

16. Hanif M, Tonazzini A, Hussain SF, Habib U, Salerno E, Savino P, Halim Z (2022) Blind bleed-through removal in color ancient manuscripts. Multimedia Tools and Applications, published online 27 September. https://doi.org/10.1007/s11042-022-13755-6

17. Hanif M, Tonazzini A, Hussain SF, Khalil A, Habib U, Restoration and content analysis of ancient manuscripts via color space based segmentation. PLoS One, published online 22 March 2023. https://doi.org/10.1371/journal.pone.0282142

18. Wang J, Tan CL (2011) Non-rigid registration and restoration of double-sided historical manuscripts. In: Proceedings international conference on document analysis and recognition (ICDAR), pp 1374–1378

19. Rowley-Brooke R, Pitié F, Kokaram A C (2013) Non-rigid recto–verso registration using page outline structure and content preserving warps. In: 2nd International workshop on historical document imaging and processing, proceedings, pp 8–13

20. Savino P, Tonazzini A (2016) Digital restoration of ancient color manuscripts from geometrically misaligned recto–verso pairs. J Cult Herit 19:511–521

21. Rowley-Brooke R, Pitié F, Kokaram A (2013) A non-parametric framework for document bleed-through removal. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2954–2960

22. Huang Y, Brown MS, Xu D (2010) User assisted ink-bleed reduction. IEEE Trans Image Process 19(10):2646–2658

23. Tonazzini A, Bedini L (2013) Restoration of recto-verso color documents using correlated component analysis. EURASIP J Adv Signal Process 2013:58

24. Tonazzini A, Salerno E, Bedini L (2007) Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. Int J Doc Anal Recognit 10:17–25

25. Hanif M, Tonazzini A, Savino P, Salerno E (2018) Non-local sparse image inpainting for document bleed-through removal. J Imaging 4:68

26. Rowley-Brooke R, Pitié F, Kokaram AC (2012) A ground truth bleed-through document image database. In: Zaphiris P, Buchanan G, Rasmussen E, Loizides F (eds) Theory and practice of digital libraries, lecture notes in computer science, vol 7489, pp 185–196

27. Irish Script On Screen Project (2012). www.isos.dias.ie

28. Savino P, Tonazzini A, Bedini L (2019) Bleed-through cancelation in non-rigidly misaligned recto-verso archival manuscripts based on local registration. Int J Doc Anal Recognit 22:163–176

29. Criminisi A, Pérez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13:1200–1212

30. Xiong W, Jia X, Xu J, Xiong Z, Liu M, Wang J (2018) Historical document image binarization using background estimation and energy minimization. In: Proceedings of 24th international conference on pattern recognition (ICPR 2018), Beijing, China, pp 3716–3721

31. Xiong W, Zhou L, Yue L, Li L, Wang S (2021) An enhanced binarization framework for degraded historical document images. EURASIP J Image Video Process 2021:13

32. Brodići D, Amelio A, Milivojević ZN (2015) Characterization and distinction between closely related south Slavic languages on the example of Serbian and Croatian. In: Lecture notes in computer science, vol 9256, pp 654–666. Springer

33. Amelio A, Bonifazi G, Cauteruccio F, Corradini E, Marchetti M, Ursino D, Virgili L (2023) Representation and compression of Residual Neural Networks through a multilayer network based approach. Expert Syst Appl 215:119391

34. Amelio A, Bonifazi G, Corradini E, Ursino D, Virgili L (2023) A multilayer network-based approach to represent, explore and handle convolutional neural networks. Cogn Comput 15(1):61–89

⧆ Springer