International Journal on Document Analysis and Recognition (IJDAR) (2021) 24:181–195
https://doi.org/10.1007/s10032-021-00379-z

SPECIAL ISSUE PAPER

Revealing a history: palimpsest text separation with generative networks

Anna Starynska[1] · David Messinger[1] · Yu Kong[2]

Received: 18 November 2020 / Revised: 28 May 2021 / Accepted: 8 June 2021 / Published online: 8 July 2021

Abstract

A palimpsest is a historical manuscript in which the original text (termed under-text) was erased and overwritten with another script in order to recycle the parchment. One of the main challenges in studying palimpsests is to reveal the under-text. Due to the development of multi-spectral imaging, the original text can sometimes be recovered through material differences of inks and parchment (Easton et al., in: 2011 19th European signal processing conference, IEEE, 2011). However, generally, the revealed text can be observed only partially due to the overlap with newer text and degradation of the material. In this work, we propose revealing the under-text completely using deep generative networks, by leveraging prior spatial information of the under-text script. To optimize the under-text, we mimic the process of palimpsest creation. This is done by generating the under-text from a separately trained generative network to match it to the palimpsest image after mixing it with foreground text. The mixing process is represented by a separate neural network, that is optimized with the under-text image to match the original palimpsest. We also add an additional background generative network to compensate for the unevenness of the background. We propose a novel way of training the background generative network, that does not require isolated background samples and can use any region with layers of text. This paper illustrates the first known attempt to solve palimpsest text layer separation with deep generative networks. We evaluate our method performance on artificial and real palimpsest manuscripts by measuring character recognition and pixel-wise accuracy of the reconstructed under-text.

Keywords Palimpsest · Blind image separation · Inverse problem · Deep neural network

1 Introduction

Palimpsested manuscripts are a type of handwritten document where the original text was erased and overwritten with another script. Historically, manuscripts were hand-written on animal skin-based parchment, which is a time-consuming and costly process. Also parchment is very sturdy material, that can easily sustain erasing. Both of these factors played a key role in the practice of manuscript recycling. The importance of palimpsested manuscripts is sometimes underestimated in the document processing community as palimpsests are believed to be rare findings. In reality palimpsested documents are not that rare. For instance 550 palimpsested documents were recently identified in the Vatican Library [42] and 160 were identified in St.Catherine's Monastery Library [34]. There likely exist thousands of palimpsests that potentially can lead to new discoveries about historic culture, everyday life, diplomacy, trade, and conflict history. Currently, the main issue with studying palimpsested manuscripts is the restoration and transcription of the original text (see Fig. 1).

The common practice for revealing palimpsest text is either by using Multispectral Imaging Systems, such as the system that was used for imaging the St. Catherine's Sinai Monastery Collection [34], or by making digital photos of the palimpsest under visible and ultraviolet (UV) light, as in the palimpsest Vatican library collection [42]. Iron-gall ink, which was the most commonly used ink in medieval documents, has a tendency to seep into the inner layers of parchment. So even after erasing, the ink residue can sometimes be seen under non-visible-spectrum light, especially ultraviolet. However, a majority of palimpsests need additional processing to make the under-text legible for tran-

✉ Anna Starynska
as3297@rit.edu

[1] Chester F. Carlson Center of Imaging Science, Rochester Institute of Technology, Rochester, NY, USA

[2] Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY, USA

scription. Previous methods proposed to separate the texts by using statistical methods that exploit the difference in the spectral signatures between older and newer inks [8,17,33]. The example of one of these methods, principal component analyses (PCA), can be seen on the right side of the fragment in Fig. 2. The problem of the aforementioned methods is that they could not separate the non-linear mixture of signals, which is long suspected about the palimpsest manuscripts. Instead, they are used to enhance the contrast of the areas that have the largest spectral difference (see Fig. 2).

In our work, we propose a Deep Learning method for under-text separation. Due to the specifics of the palimpsest text separation problem our method is unsupervised, which means that it does not require labeled data for training. Also, it is well-suited for one channel and multi-channel data such as multispectral images. Our main idea is to restore the original text by utilizing prior spatial information about text scripts (style of handwriting) modeled with a Deep Neural Network. In our case, we are trying to solve an inverse problem of finding the unknown under-text signal from the observed palimpsest image. We assume that the palimpsest

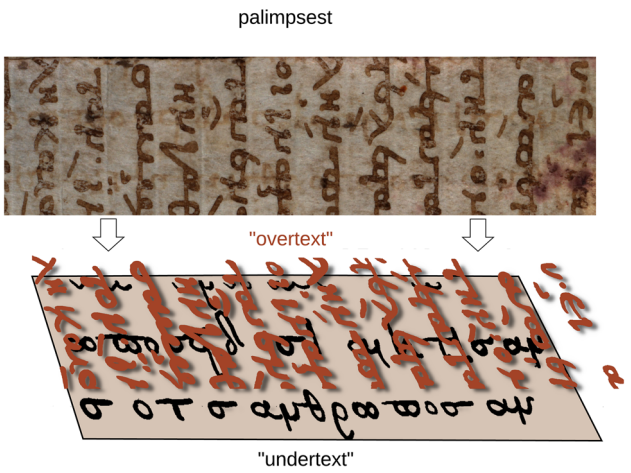


Fig. 1 Fragment of a historical palimpsest. The palimpsest is a document where the original text ("under-text") was erased and overwritten with another text ("over-text"). The top part of figure shows the scan of a palimpsest's fragment. In this work, we are aiming at separating layers of older and newer text, as is shown on the bottom. In particular, we are interested in reconstructing the older "under-text"

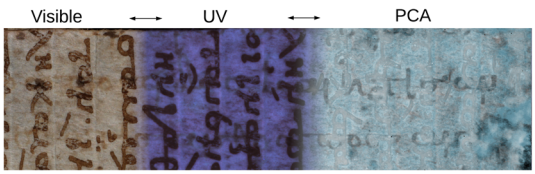


Fig. 2 Fragment of a historical palimpsest. The left side of the fragment depicts the palimpsest images in visible spectrum light, where the "over-text" is clearly seen in dark brown. The middle section is the same fragment but under UV illumination. The right side illustrates PCA processing. The "under-text" can be seen in dark gray there

image is a combination of layers of over-text, under-text and background transformed by some non-linear mixing function. Since the mixing function is unknown we adapt to our problem the work [2], where the transformation function is approximated by the neural network constrained by the original signal prior presented by a pre-trained Deep Generative Network.

In our problem, we cannot train our spatial prior model directly on the under-text, since we do not have access to the under-text layer from the palimpsest image. Therefore, we suggest to pre-train the under-text generative network on a separate "clean" manuscript. We assume that we can identify the text language and handwriting style from the raw palimpsest images. This allows us to choose the external non-palimpsested document with a similar script. Even though the text is hand-written, during medieval times, before the invention of printing, scribes usually practiced a calligraphic style of writing for copying books and writing official documents. This process caused a certain consistency in the writing styles of particular languages.

On a high level, our method can be described like this: At first we train the under-text generative model, then we use this model to optimize the under-text image combined with over-text through the mixing function until it matches the palimpsest image. In order to optimize the under-text, we optimize the input vectors of trained generative models. To match the combined layers of binarized text images to the image of the real palimpsest we also need to optimize the mixing model. Our palimpsest image reconstruction pipeline includes two major steps: (1) modeling the script with the generative network; (2) optimizing the under-text image, modeled by the generative network, and the mixing model to match the original palimpsest. The framework is shown in Fig. 3.

We also added the background generator to compensate for the unevenness of the parchment. In contrast to the under-text model, which we trained on the binarized samples of characters from the "clean" manuscript, we decided to train the background generator directly from the manuscript images with under and over-text. We choose this approach because creating an additional dataset of clean background samples from the palimpsest would require additional work and may not always be possible. To achieve this, we develop a novel method for training a Generative Adversarial Network (GAN) [13] on the image of mixed multiple signals, for one signal generation. Previously, the work of [6] demonstrated how a signal corruption model could be incorporated for training a GAN on a corrupted signal. We extended this model to the case of palimpsest signal mixing, where we do not know the exact form of the mixing model. To achieve this, we propose to train the mixing function and background generator in a complementary manner to the under-text and over-text.

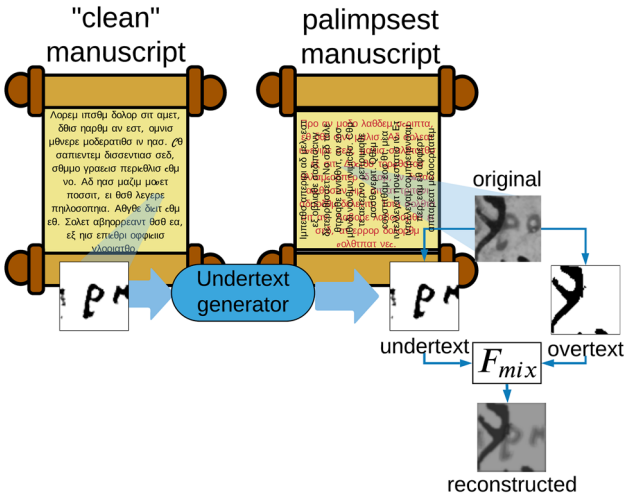


Fig. 3 To separate the palimpsest text, we cast the problem as palimpsest document reconstruction. Each layer of text is separately fed into the mixing function (F_{mix}) to imitate the original palimpsest. The under-text is modeled by the generative model that was pre-trained on samples of a "clean" document. The "clean" document is a un-palimpsest manuscript that is written in the same language and writing style as the under-text

The quality of the reconstruction was estimated on synthetic and real palimpsest images. Since we could not provide pixel-level ground truth for the real palimpsest data, we estimate the reconstruction quality indirectly by measuring the character classification accuracy.

The main contributions of our research are:

- We present a first attempt to use an unsupervised Deep Learning model for palimpsest text separation. Our framework can assist scholars in palimpsest text transcription as well as automatic text recognition algorithms;
- We propose a new framework for pre-training and fine-tuning generative models of "clean", unmixed signals from the mixed observations.

2 Related work

2.1 Palimpsest text separation

The main difficulty in digital restoration of palimpsested text is that due to the efforts of recycling old parchment, the ink of the original text has vanished. Sometimes, because of iron corrosion in the residual ink particles that were absorbed into the parchment, we can see the hints of the older text under the foreground text, as on the left side of the palimpsest fragment on Fig. 2. Guided by this information, researchers developed the idea of using imaging techniques that can enhance the older ink absorbed into the parchment. The most popular among them are multispectral, hyperspectral imaging, and X-ray fluorescence [9,12,29]. These techniques are usually based on the interaction of Electro-Magnetic (EM) radiation, such as UV light or X-ray radiation, with residual elements of the original ink. However, raw imaging results still usually are not suitable for human interpretation. Additional efforts are needed to separate the under-text.

Previously, to increase the visibility of "under-text," multispectral statistical methods such a PCA, Independent Component Analyses (ICA) or other linear unmixing approaches were used. These methods exploit the difference in the reflective (and sometimes fluorescence) properties of materials under controlled illumination. They assume that each pixel value corresponds to the reflectance of the superposition of different materials as a function of wavelength. In this method, the spatial information is usually discarded, and the image is analyzed as a set of vectors corresponding to each pixel, where the vector components represent a mixture of different sources. Despite the numerous reports of successful palimpsest text separation with eigenvalue decomposition, ICA, or other linear unmixing models [8,17,22,26,33,41], the text generally is still not fully visible (see Fig. 2). The main reason is that the mixing process of inks in parchments is usually non-linear and could not be fully reconstructed by using linear models.

The work [38] suggests the use of nonlinear, non-stationary physically inspired model for palimpsest modeling. The idea of this model is that it can be analytically inverted to extract the under-text writting. This model presents a palimpsest as mixture of overwritting and underwritting text layers. However, it assumes even and easily separable background. In practice historical palimpsest have very uneven background, and one of the biggest challenge lay in separating of under-text from background.

Some of the methods (see [36,44]) tried to perform automatic and semi-automatic text transcription directly on the palimpsest text processed by PCA. However, this method requires labeling of some portion of the same document, since the appearance of the under-text and mixing of components is usually unique to a particular manuscript. The labeling overall for palimpsest text is a challenging task since this text is very old and uses historic, unfamiliar script for writing.

2.2 Bleed-through removal models

It is also important to mention an adjacent topic in historical document processing called bleed-through removal. In this problem the goal is to remove the text that "bleeds-through" the page from the other side or is seen from the other side of the page due to the paper transparency. This is a well-studied problem with a long line of work. To solve this problem various methods were suggested that include thresholding, partial differential equations (PDE) applica-

tion [18,27], sparse representation [15], deep neural networks [16] and optical density models [38], Markov Random Fields (MRF) [37], etc. Sometimes a byproduct of this process is a restored text image from the opposite side that can be considered from the perspective of a palimpsest as an under-text.

For instance, in work [27], that uses an anisotropic diffusion method for text restoration, the opposite side is presented with a backward diffusion component. The idea is that starting from the current page, observation of the text is gradually diffused back to its appropriate side until some stable state is reached. However, in this algorithm the image of the opposite side, where the bleed-through text is a dominant signal as opposed to the face side, is required during the restoration. In the case of a palimpsest, we do not generally have access to the image where the under-text serves as a main source of signal from any side of the page or under any imaging modality. In the work [16] the bleed-trough image is also returned by a neural network to form the image, but is mixed with the main inverted text and requires ground-truth data for training. As the closest work to ours, we consider the work of [37] that uses an iterative algorithm with a mean field approximation to estimate both sources and the mixing matrix. However, it uses an MRF for the shape prior which is less expressive then generative networks, and it constrained the mixing process to a linear model for the signal mixture.

2.3 Inverse problems

In inverse problems we need to find the original signal that after applying a measurement operator (mixing function) would produce an image closest to the noisy observation of the corrupted image. Since there could be multiple original images that meet this requirement the problem is usually constrained by including a regularization term in the form of a signal prior into the cost function.

Current deep learning unsupervised approaches could be split into two main groups. In the first group are methods that learn the prior separately with a generative adversarial network [25,30]. The other type of methods are based on constraining reconstruction by generating original images from generative networks trained on uncorrected source images [5,20,46]. In this case the data distribution is parameterized by the generative network and optimization happens not in the original image space but in the latent vector space that is constrained to some particular data range. In all these algorithms, we know the explicit form of the measurement operator. However, in our case, the mixing process model is unknown. Works of [2,4] proposed a solution that does not require knowledge of the mixing process [2] suggested to approximate it with a shallow network that is alternatively optimized with source images. These methods require the source generators to be pre-trained on clean images separately. However in our work we do not have access to datasets of high quality clean images. Recently, papers were presented devoted to learning the original image distribution from its corrupted observations, such as [6,24,28,35]. Yet all these works assume a known mixing operation. In our work we propose an extension to these methods to account for the unknown mixing process during the training.

As an alternative to using a GAN network as a prior or direct source mapping for the unmixing problem, the works of [11,40] were considered. These works use untrained convolutional neural networks to imitate the mixture layers. This approach works well for a natural image separation with even background, but it does not deal well with a signal of similar structure, like the layers of text. It also assumes a linear supper-position of the image layers.

3 Proposed approach

3.1 Problem definition

The major difficulty of under-text reconstruction is the inability to access a ground-truth map of the under-text. The under-text labeling of a historical palimpsest is a tedious process that requires special qualification for reading historical scripts. The under-text corruptions make it extremely difficult to create the pixel-wise labeling even for a qualified scholar. In addition, variation in ink compositions, in manuscript storing conditions and the parchment material make any direct under-text segmentation model, such as [32], impractical, as they would work only on the document on which they were trained.

We propose to address these challenges by using an unsupervised method that accounts for differences in palimpsest appearances in the absence of labeled data. Formulating the task of under-text layer separation as an inverse problem allows us to develop such a method.

Generally, we can describe an inverse problem as a task of recovering the original signal X from its noisy transformed version $Y = \mathcal{F}(X) + n$, where $\mathcal{F}$ is the transformation function, X is a sample of the original signal that needs to be reconstructed, Y is its noisy observation, and n is a noise variable. One common way to estimate the original signal is by finding the maximum a posteriori (MAP) estimate of X. In [5] it was proposed to use pre-trained generative models to generate the MAP estimate by minimizing the reconstruction error with respect to the generator G input vector z: $\text{argmin}_z \|Y - \mathcal{F}(G(z))\|_2$. Later this approach was adapted to deblurring [4], inpainting [46], and blind source separation [2,20].

Our idea is to extend this framework to the assumption that the original signal X could be divided into three independent components: over-text X^{ot}, under-text X^{ut}, and background X^b, that were mixed together to create a palimpsest. The

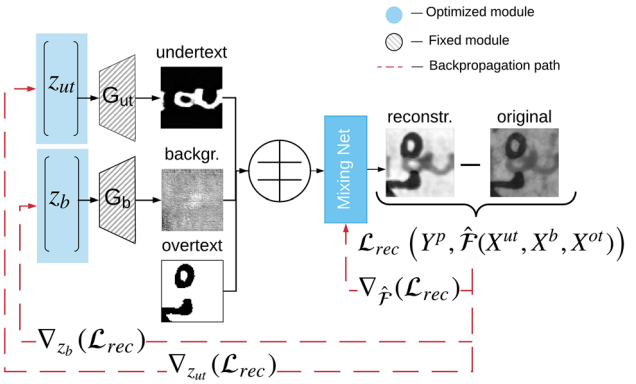
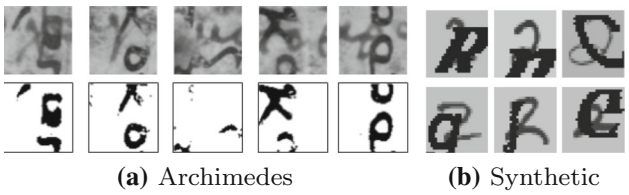


Fig. 4 Palimpsest text separation pipeline. At first, the under-text and background latent vectors z_{ut}, z_b, and Mixing Net weights are randomly initialized. The under-text and background image X^{ut}, X^b is generated by pre-trained generative networks on under-text G_{ut} and background G_b. Then X^{ut}, X^b are mixed with the foreground text mask X^{ot} through the Mixing Net to generate an estimation of palimpsest image $\hat{Y}^p$. The network is optimized by minimizing the reconstruction loss $\mathcal{L}_{rec}$ between generated $\hat{Y}^p$ and real palimpsest Y^p images. The red lines show a back-propagation path. The optimization is performed in alternating way for vectors z_{ut}, z_b and Mixing Net (trained parts represented by blue color belong to MixingNet, z^b, z^{ut}; fixed by gray belong to G_{ut}, G_b)



(a) Archimedes **(b)** Synthetic

Fig. 5 Example of synthetic and real palimpsest dataset: **a** Real palimpsest with example of foreground text image mask thresholding; **b** synthetic palimpsest,where bottom layer with MNIST characters is randomly overlayed by randomly generated characters from standard font libraries

intricate mixture of these components created by adding and erasing layers, that we present with a mixing function F, is unknown. In our approach, we suggest that in the case when the under-text can be at least partially distinguishable, we can reconstruct it in full view by finding original layers and their mixing function, which formed a palimpsest image. We can formulate the palimpsest formation process as, $Y^p = \mathcal{F}(X^{ut}, X^{ot}, X^b)+n$, where $\mathcal{F}$ is the unknown mixing function, and Y^p are samples of the observed palimpsest. The under-text and background X^{ut}, X^b images would be generated by pre-trained generative networks as in [6]. The pre-training process is described in Sect. 3.3. To find the original signal X^{ut}, X^b, we would minimize the reconstruction error with respect to the input vectors z_{ut} and z_b of generative networks. The over-text X^{ot} would be provided separately (see details as described in Sect. 3.2).

3.2 Palimpsest under-text reconstruction

The overall framework structure for palimpsest text-layer separation is illustrated in Fig. 4. The palimpsest image is reconstructed by finding the under-text, background, and mixture function that produce the closest image as possible to the observed palimpsest patch. The under-text X^{ut} and background X^b images are generated by pre-trained generators G_{ut} and G_b from latent vectors z_{ut} and z_b. The original signal solution space is constrained by modeling it with pre-trained generative models $X^{ut} \sim G_{ut}(z_{ut})$ and $X^b \sim G_b(z_b)$, so that during optimization, only the latent vectors (z_{ut} and z_b) are changed, as it is suggested in [5].

We present the mixing function as a shallow a convolutional neural network. The Mixing Net approximates the unknown mixing function $\mathcal{F}$ and is optimized during the reconstruction along with latent vectors z_{ut}, and z_b. This solution was inspired by [2], where the mixing function was represented with a shallow fully-connected network. In our case we only consider a location invariant architecture for a mixing function, such as a convolution neural network, as the mixing process depends only on the combination of source signals.

In our case, the second source of the text (the over-text X^{ot}) is not optimized during the training. We made this decision mainly because the over-text has a high contrast with the parchment and little variation so it can be easily found by a simple thresholding operation from an image taken under visible spectral light. An example is shown in Fig. 5a.

As a result, the reconstructed signal could be formulated as $\hat{Y}^p = \hat{\mathcal{F}}(G_b(z_b), G_{ut}(z_{ut}), X^{ot}) + n$, where the $\hat{\mathcal{F}}$ is an approximation of the mixing function $\mathcal{F}$. Then, the objective function for the under-text, background and mixing function reconstruction can be formulated as

$$\hat{\mathcal{F}}^*, z^*_{ut}, z^*_b = \underset{\hat{\mathcal{F}}, z_{ut}, z_b}{\operatorname{argmin}} \left\| Y^p - \hat{\mathcal{F}}(G_b(z_b), G_{ut}(z_{ut}), X^{ot}) \right\|. \quad (1)$$

3.3 Training image generators

To solve the inverse problem using a generative model, one would need to use pre-trained generators to provide a mapping from the latent vectors to unmixed "clean" images [5,20,46]. These generators have to be trained separately in advance of the unmixing. For the under-text and background generative model we chose a Generative Adversarial Network (GAN) [14], although other models such as Variational Auto-Encoder (VAE) [19] maybe considered. The next section will describe the procedures of training the under-text and background generators.

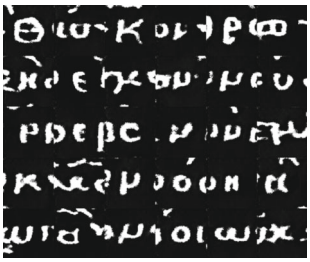


Fig. 6 Image samples of under-text characters generated by pre-trained generator

under-text Generative network In our approach we use an external historical document written in a similar style but without the topcoat of newer text (i.e., not palimpsested) to learn the under-text distribution. The generator is trained in a standard GAN framework on binary patches of characters extracted from the "clean" manuscript. You can see the examples of generated characters from trained GAN on Fig. 6.

Background generative network As was already mentioned, we did not prepare a separate dataset of background samples for background generator training. Instead we decided to develop a method that allows us to train the background generator directly from the palimpsest text image samples. However, our dataset of palimpsest samples is relatively small, just 1000 images. Thus, we decided at first to train the background generator on images of a "clean" manuscript due to the much larger size of the dataset. Then, we fine-tuned the background generator on the palimpsest images. The "clean" manuscript is also an old document written on a parchment, so it also has a distinctly uneven background. For training we used the pre-thresolded dataset of "clean" manuscript images, discussed below (see Fig. 7).

The gain from this approach is two-fold: first, we can learn the parchment background distribution from the mixed image samples that contain both background and text; second, we can further re-purpose the same framework for fine-tuning the under-text and background generator on the palimpsest document. The results from training the background generator on a "clean" manuscript and fine-tuning it on the palimpsest are shown on Fig. 8.

In [6] it was already shown that it is possible to train a GAN to learn the original image distribution from its corrupted observations. However, the authors assumed that the mixing process is known. In contrast, in our approach, we approximate the mixing process with a shallow network (Mixing Net, see Fig. 9) as in the reconstruction framework. The Mixing Net in this case is trained along with the generator.

To prevent the background generator from learning the text distribution, we add an additional "anti-shortcut" loss, as is suggested in [3]. The "anti-shortcut" loss is an inverse GAN generator loss, which is applied to the generated image without text. It should prevent the background generator from generating text. As a result, the objective function for training both the background GAN generator and the mixing network can be formulated as:

$$\mathcal{L}_{G_b,\mathcal{F}} = \mathcal{L}_{G_{b,\text{fake}}} + \mathcal{L}_{\text{antishortcut}} \tag{2}$$

where $\mathcal{L}_{G_{b,\text{fake}}} = \log(1 - D_m(\mathcal{F}_m(G_b(z_b), X^{ut}))$ is the background GAN generator loss, $\mathcal{L}_{\text{antishortcut}} = -\log(1 - D_m(\mathcal{F}_m(G_b(z_b), X^{ut} = X^{\text{emp}}))$ is the inverse background generator loss, D_m - discriminator of manuscript images, $\mathcal{F}_m$ - mixing function of "clean" manuscript, and X^{emp} - empty image.

The discriminator is trained in the usual way. At the stage of fine-tuning the generators on the palimpsest data, the images of thresholded over-text are simply added to feed into the mixing network along with the generated images of background and under-text. The training is performed with fixating weights of one of the generators. We also add a loss function between text edges before and after mixing to prevent the Mixing Net from ignoring one of the text inputs. The mixing function in this case is different because we have more input channels, so it is trained again.

3.4 Algorithm

We divide the palimpsest reconstruction process into two separate phases. In the first phase we train under-text and background generators; during the second phase we perform reconstruction. During the first phase, we train the background and under-text generator networks on character

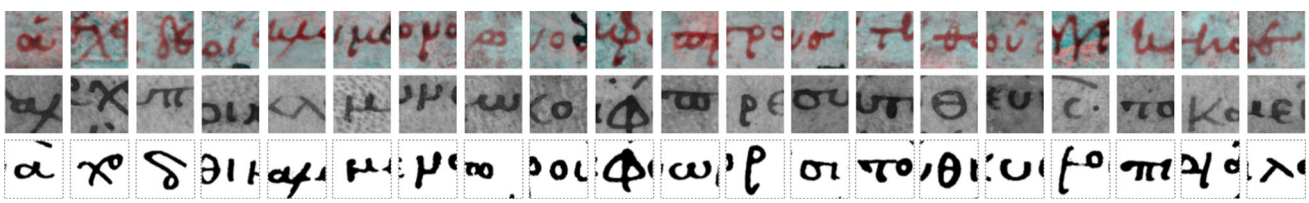


Fig. 7 Greek minuscule script characters from two different manuscripts. Top row shows characters from the "clean" section of Archimedes palimpsest. Middle shows corresponding characters from Sinai collection manuscripts. The bottom row is an example of binarized characters from Sinai manuscripts

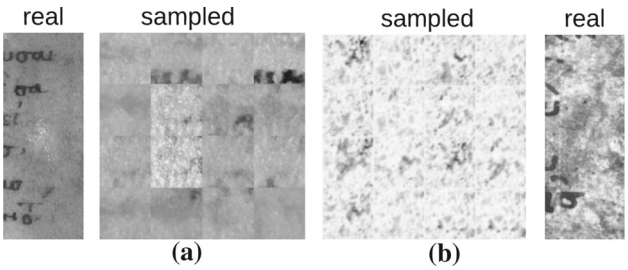


Fig. 8 Image samples of generated normalized background: **a** background generator trained on clean manuscript **b** background generator fine-tuned on the palimpsest manuscript. On the left and right showed fragments of the text margins from the "clean" manuscript and palimpsested accordingly

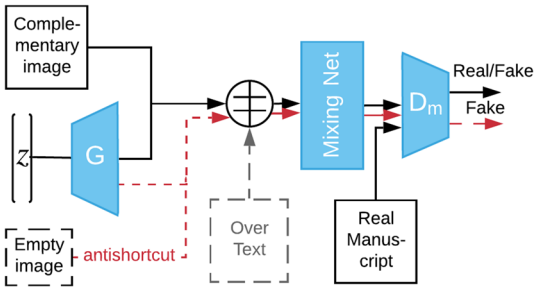


Fig. 9 Background generator training pipeline. The anti-shortcut path and complementary image of under-text are used for preventing generator from generating under-text. In the case of fine-tuning on palimpsest document, the image of over-text is additionally fed into the Mixing Net during training

Algorithm 1 Training the under-text GAN

Require: samples of binarized characters from "clean" manuscript X^{ut}

return Parameters of the manuscript discriminator D_{ut} the background generators G_{ut}
initialize parameters of D_{ut},G_{ut} to random values;
for number of iterations **do**
  Sample the mini-batch of n noise samples z_{ut} , sample X^{ut}
  Update the D_{ut} by ascending:
  $\frac{1}{n}\sum_{i=0}^{n}\log(D_{ut}(X_i^{ut})) + \log(1 - D_{ut}(G_{ut}(z_{ut,i})))$
  Sample the mini-batch of n noise samples z_{ut}
  Update the G_{ut} by descending:
  $\frac{1}{n}\sum_{i=0}^{n}\log(1 - D_{ut}(G_{ut}(z_{ut,i})))$
end for

Algorithm 2 Training the background GAN

Require: Samples of characters from "clean" manuscript Y^m, samples of binarized characters from "clean" manuscript X^{ut}
return Parameters of the manuscript discriminator D_m the background generators G_b
initialize parameters of Mixing Net $\hat{\mathcal{F}}$, D_m,G_b to random values;
for number of iterations **do**
  Sample the mini-batch of n noise samples z_b , sample Y^m and X^{ut}
  Update the D_m by ascending:
  $\frac{1}{n}\sum_{i=0}^{n}\log(Y_i^m) + \log(1 - D_m(\hat{\mathcal{F}}(X_i^{ut}, G_b(z_{b,i}))))$
  Sample the mini-batch of n noise samples z_b , X^{ut},X^{emp}
  Update the G_b and $\hat{\mathcal{F}}$ by descending:
  $\frac{1}{n}\sum_{i=0}^{n}\log(1 - D_m(\hat{\mathcal{F}}(X_i^{ut}, G_b(z_{b,i})))) - \log(1 - D_m(\hat{F}(X_i^{emp}, G_b(z_{b,i}))))$
end for

Algorithm 3 under-text reconstruction

Require: K samples of characters Y^p from palimpsest, K samples of thresholded over-text X^{ot} from the palimpsest, and parameters of pre-trained generators G_b, G_u, (if fine-tuned: parameters $\hat{F}$ of pre-trained Mixing Net)
return Parameters of the Mixing Net $\hat{F}$ model, estimated under-text character $\hat{X}^{ut}$ and background $\hat{X}^b$
Initialize set of K vectors z_b, z_{ut} to random values from uniform distribution $\mathcal{N}(0, 1)$
Initialize parameters of $\hat{F}$ to random values(if finedtuned: set to pre-trained parameters)
for number of iterations T **do**
  for number of iterations T_1 **do**
    Update $\hat{F}$ by descending:
    $\mathcal{L} = \sum_{i=1}^{K}\left\|Y_i^p - \hat{\mathcal{F}}(G_b(z_{b_i}), G_{ut}(z_{ut_i}), \mathcal{X}_i^{ot})\right\|$
  end for
  for number of iterations T_1 **do**
    for $i \leftarrow 0$ **to** K **do**
      Update z_{ut_i} by descending:
      $\mathcal{L}_i = \left\|Y_i^p - \hat{\mathcal{F}}(G_b(z_{b_i}), G_{ut}(z_{ut_i}), \mathcal{X}_i^{ot})\right\|$
      Clipping: $z_{ut_i} \leftarrow \mathcal{P}(z_{ut_i})$
    end for
  end for
  for number of iterations T_1 **do**
    for $i \leftarrow 0$ **to** K **do**
      Update z_{bi} by descending:
      $\mathcal{L}_i = \left\|Y_i^p - \hat{\mathcal{F}}(G_b(z_{b_i}), G_{ut}(z_{ut_i}), \mathcal{X}_i^{ot})\right\|$
      Clipping: $z_{b_i} \leftarrow \mathcal{P}(z_{b_i})$
    end for
  end for
end for

samples from the "clean" manuscript. The under-text generator is trained in a GAN framework on binarized text images from the "clean" palimpsest. Algorithm 1 shows the training procedure for under-text generator.

The background generator is trained in a more elaborate framework, as described in Sect. 3.3. We propose to train the background generator directly from mixed images in a complementary manner to the under- and over-text. Algorithm 2 shows the first stage of background generator training using the grayscale images of the "clean" manuscript. After that, the background generator is fine-tuned on the images of the palimpsest using the same algorithm but with the addition of an over-text layer, where the over-text layer is also treated as a complimentary source.

During the second phase, we perform a palimpsest under-text reconstruction (see Algorithm 3). To do that, we transfer the trained generator's network weights from the first phase into the image reconstruction framework. During the optimization, the generator's weights stay fixed. We use three stages of alternating gradient descent to optimize the latent variables of the under-text, background, and mixing network,

as is suggested in [2]. At the beginning of the image generation phase, we initialize the latent variables of the under-text and background with values sampled from a Gaussian distribution $\mathcal{N}(\mu = 0, \sigma^2 = 1)$, since this distribution was used for the under-text and background generator training. Then, we stack the generated images with the previously extracted foreground text mask and feed it into the Mixing Net. After that, we calculate the loss function using the loss 1 between the output of the Mixing Network and the original palimpsest image (see Eq. 1). Based on the loss gradient, we update the Mixing Net and latent variable vectors alternately for 50 iterations each. We repeat the alternating loop for from 100 to 400 epochs. The samples in the batch are not shuffled during the training since each latent variable vector corresponds to a specific image. The generated images are saved at the end of the training. As is suggested in [2,46], we also apply latent space parameter clipping, which prevents latent vector values from falling out of the range that is used for the pre-trained generative model.

Although, finding the over-text through thresholding of a visible band gives a good estimate, in many cases it still contains some imperfections. To compensate for the noisy estimates of the over-text layer, we added a pixel-wise weighted mask to the loss function $\mathcal{L}_w = W \odot \mathcal{L}$, such that areas that are estimated as over-text would have less influence over the reconstruction error. Our experiments show that setting the mask scale level to 0.5 is sufficient. In Fig. 10 we demonstrate how the quality of the reconstruction deteriorates if we increase an error in the over-text layer without adding a mask loss.

Fig. 10 Demonstration of influence of noise in over-text estimation on reconstruction performance. In the orange box, the palimpsest image created from MNIST's handwritten digit is shown. The top character represents under-text overlaid with over-text. The bottom-left image is the ground-truth image of the over-text, while the bottom-right image represents the ground-truth under-text digit. Outside the box, in the bottom row is the over-text layer with an increasing level of inaccuracies imitated by Gaussian noise (0.0;0.1;0.2;0.3;0.4) and reconstruction results with (middle row) and without a mask (bottom row). We can see that adding a mask level of 0.5 to the loss for the over-text layer makes reconstruction more robust to errors in the over-text layer estimation

4 Experiment

4.1 Datasets

To estimate the performance of the proposed method we used a synthetic palimpsest dataset and a dataset created from the scan of a real 12th-century palimpsest. The use of the synthetic dataset is motivated by the impossibility to estimate pixel-wise reconstruction accuracy and control palimpsest image parameters on real data.

Synthetic dataset. For the synthetic palimpsest we used the MNIST dataset of handwritten characters [21], as under-text text overlayed with randomly distorted characters from standard English font libraries (Arial, Times New Roman, Georgia, see Fig. 5b). The MNIST dataset was selected for the palimpsest under-text because it presents significant variability inherent to handwritten texts and, it was collected from subjects with various writing styles. We chose a "standard" font for the over-text since class variability does not matter in this case and the over-text is considered only as a random source of occlusion. Similar to [38], we assume that there is a band where the under-text becomes almost transparent and the over-text can be extracted through binarization using a modern historical document binarization algorithm (for example [16]). We used separate parts of the MNIST dataset for training the under-text generative model and for creating synthetic palimpsests. For training the generative network we used the MNIST training set of the 50,000 samples. The MNIST test set was used for the synthetic palimpsest creation.

Archimedes dataset The Archimedes dataset is the only known palimpsest character dataset with ground truth, and was created as a part of previous work [36]. It was created from the scans of the Archimedes palimpsest that has two layers of text. The under-text, a copy of Archimedes treatises, is written in 10^{th} century minuscule Greek script. The over-text is 13^{th} century text, also written in miniscule Greek, and is a religious text. This dataset consists of training, validation, and test sets of relatively "clean" characters and a separate test set with palimpsested characters. For text separation, we used a test set of palimpsested characters containing 200 samples.

As a "clean" manuscript for the under-text generator training, in this case we selected a manuscript also written in minuscule Greek. The one that seems suitable for this purpose was the scan of Sinai Greek 960 manuscript found in the Sinai Palimpsest collection[1]. This manuscript consists of a collection of clean and palimpsested pages. The clean pages with minuscule Greek script were used for the creation of a dataset

[1] Sinai Greek 960, a publication of St. Catherine's Monastery of the Sinai in collaboration with EMEL and UCLA, https://sinai.library.ucla.edu.

of unlabeled characters. Minuscule Greek script is not very consistent between the documents; because of that, some of the character's shapes differ between the two datasets. An example of corresponding characters is shown in Fig. 7. We created binarized and non-binarized versions of the "clean" manuscript characters dataset. The first one was used for the under-text generator training, while the second one was used for background generator training. The manuscript images were segmented and binarized into characters automatically by using a combination of morphological operations and connected component labeling [43,45]. The training set of binary threshold characters is 606 samples in the validation set with 7496 samples in the training set. The non-binarized section of the dataset contains 6082 train and 1238 validation images.

4.2 Metrics

For the synthetic palimpsest, we estimate the reconstruction accuracy using character level (character recognition accuracy) and pixel-level metrics. On a pixel level, the accuracy was estimated using Peak Signal to Noise Ratio (PSNR), F1-score, where the higher the score, the better the visibility of the under-text. Also, as a pixel-level metric we use Mutual Information (MI) in cases where the under-text became more visible but has a different shade, such that PSNR and F1-score could not be reliable accuracy estimators. For calculating the Mutual Information we discretize the image and then take the average result over all samples. The number of discrete levels varies for different experiments. As an alternative metric of reconstruction quality we suggest using character recognition accuracy, also called Goal-Directed Evaluation [39]. The idea is that higher reconstruction quality would positively influence the accuracy of character classification since the foreground text can be considered as an under-text image obstruction. For the real palimpsest data, we used only character recognition since we do not have access to pixel level ground-truth. For the synthetic palimpsest we used the MNIST network [7], trained on the handwritten digits of the MNIST training set, for character recognition.

For the Archimedes palimpsest we used a character recognition algorithm trained on a "clean" part of the Archimedes dataset (Fig. 7). For the most accurate performance, we manually binarized the images. Due to the small size of the Archimedes dataset, we used a transfer learning technique. At first, we pre-trained the convolution neural network (reused from the MNIST network) on the MNIST dataset to extract the features from the character images. Then we trained a support vector machine algorithm on the extracted features from the Archimedes characters dataset. That allowed us to achieve on the "clean", binarized, part of the Archimedes dataset 100% accuracy on training and validation sets and 93.5% on the test set.

4.3 Results on the synthetic palimpsest

Comparison with baselines. To compare our framework with a previous statistical approach, such as PCA and ICA for text separation, we created two synthetic palimpsest datasets. For this experiments, we used 200 samples from MNIST test set. The datasets have multiple bands, since the PCA and ICA algorithms perform decomposition along the spectral dimension.

The first dataset tests our framework for the case where the mixing function has an apparent nonlinear nature. We use this model because as was discussed before, it is suspected that the palimpsest text erasing and subsequent mixing with newer text could not be described by a linear function. For this case, the mixing function was the min operator between pixel values. The band images were generated by using pixel-wise minimum operator between three sources of signal: X^{ut}—original under-text (MNIST digit), X^{ot}—over-text, X^{b}—background image (plain gray value); $\min(\beta_{1,j}X^{ut}; \beta_{2,j}X^{ot}; \beta_{3,j}X^{b})$, where β_1, β_2, β_3 are randomly generated source coefficients ($\beta_1 = [0.512, 0.521]$, $\beta_2 = [0.237, 0.465]$, $\beta_3 = [0.576, 0.617]$.)

The second dataset was created using the non-stationary optical density model that was presented in work [38] for palimpsest document modeling (described above). The band images were generated by using $D^{obs} = D^{ut}X^{ut} + D^{ot}X^{ot}$, where D^{obs}, D^{ut}, D^{ot} - optical densities of band image, layer of under-text, and layer of over-text; the X^{ut}, X^{ot} - are binary maps of under-text and over-text. According to [38] the optical density is calculated as $D^{obs} = -\log \frac{s^{obs}}{R}$, where s^{obs} is the observed reflectance value and R is the mean reflectance value of the background. The images are generated by first randomly selecting reflectance values for over-, under-text and background, then converting them into the densities to calculate the observed optical density image. Lastly, the den-

Table 1 under-text separation methods comparison for synthetic palimpsest

Method	Acc. (%)	PSNR	Fscore	MI
Min-model				
PCA	65.71	13.445	0.943	0.194
ICA	66.67	13.534	0.939	0.194
Ours, 1–2nd band	89.05	18.045	0.973	0.333
Ours, 1st band	48.10	13.462	0.941	0.166
Optical density model				
PCA	65.71	9.240	0.898	0.207
ICA	90.00	14.220	0.959	0.365
Our, 1–5th band	91.90	17.350	0.970	0.381
Our, 1st band	90.00	17.781	0.974	0.366

Accuracy estimated by character recognition

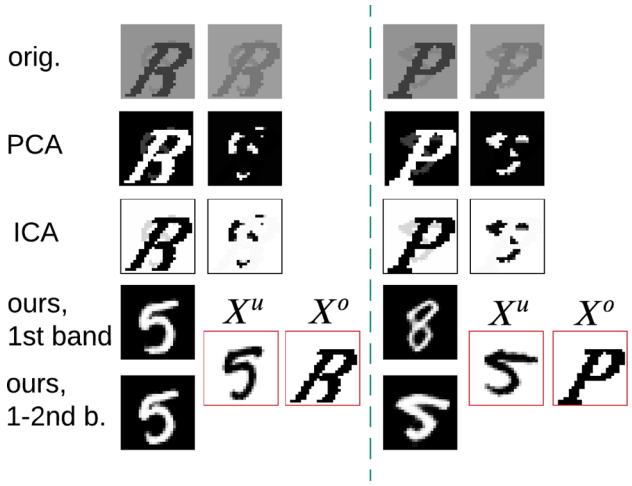
sities were converted to reflectance and stored as separate images.

For our methods, we used the Wasserstein GAN with gradient penalty[2] generator pre-trained on the MNIST data and the Mixing net, with two types of architecture for reconstructing multi-spectral and 1-channel (1-band) images. In our experiments with the MNIST dataset, we did not use the background generator since the background is uniform. The results are shown in Table 1, we define our method in the Table as "ours, 1-nth" for multispectral data, and 1st band for one channel images. We illustrated reconstruction results for both models on Fig. 11.

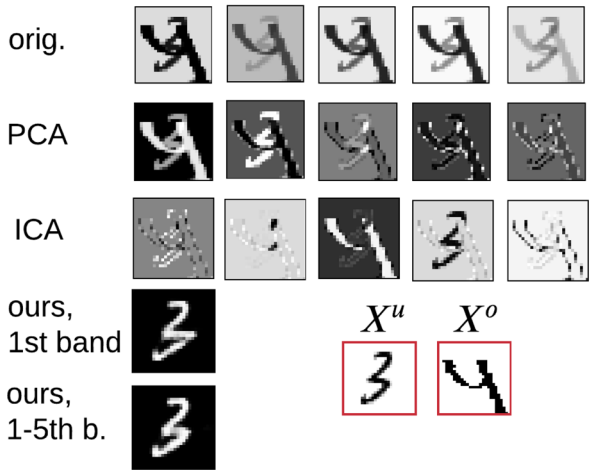
We compared our method to two traditional methods for palimpsest processing. The image transformation was performed along the channel dimension. We show here the best accuracy out of all bands after transformation for PCA and ICA. From the quantitative results and visual comparison we can see the performance of PCA and ICA is worse then our method on data created by the optical density model. Although ICA demonstrated high OCR accuracy, that can be explained by the resilience of our classification algorithm to partial character vanishing, it does not hold the same improvement among other metrics. For the *min* model data we can see that our method demonstrates significant improvement for the multi-spectral data (labeled as "ours, 1–2 bands"). However, it shows much lower performance when we use for reconstruction only the first band. This result is explained by low contrast of the under-text, which to the some extant demonstrates the limitation of our method in the case of low under-text-to-background contrast for one channel data (the reconstruction of the second band showed 65.71% accuracy with only slight improvement in contrast.) We investigate further this issue in the next experiment where we perform the reconstruction at different contrast levels.

Overall the difference in performance between the traditional methods and our method can be largely seen in the overlapping text regions in Fig. 11. This is because the PCA and ICA methods could not separate a non-linear mixture of signals, which is manifested in these areas.

Palimpsest text image quality degradation Palimpsest manuscripts are usually very old documents and as a result, can be severely damaged. Also, the residual content of the ink in the parchment may significantly vary. All these factors lead to the uneven level of contrast of the text in the page and high "noisiness" of the image. To estimate the performance of our framework at low signal level regions of the manuscript, we change the contrast level of the MNIST palimpsest "under-text". We create the palimpsest in a similar manner as was done with the *min* model, but only for 1-channel data. Additionally, we added Gaussian noise with $\mu = 0$, $\sigma^2 = 0.1$ on top of the normalized MNIST digit images before overlaying them with foreground text. In this way, we decrease the image Signal-to-Noise Ratio (SNR) along with the contrast. We ran the experiment for four levels of contrast: 15%, 30%, 45%, and 60%, and samples are shown in Fig. 12. The formula for the intensity contrast change is the same as in [1].



(a) *min* model



(b) Optical density model

Fig. 11 Example of synthetic palimpsest dataset reconstruction: **a** is reconstruction of dataset created with *min* operation; **b** is reconstruction of dataset created with Optical density model [38]. We show two samples for the first model and one sample for the second model. The rows that are enclosed in red rectangles are ground truth of "under-text" and "over-text" maps

Fig. 12 MNIST handwritten digits palimpsest with added noise at different level of contrast: 60%, 45%, 30%, 15%

[2] code available at https://github.com/igul222/improved_wgan_training.

Table 2 Character recognition and reconstruction results for palimpsest MNIST handwritten digits at different level of contrast

Contrast	Acc. (%)	PSNR	Fscore	MI
15	58.00	12.470	0.937	0.159
30	87.64	15.715	0.963	0.296
45	91.24	17.122	0.970	0.341
60	92.24	17.882	0.973	0.359

Table 3 Reconstruction results for Archimedes palimpsest

Method	Accuracy
Before unmixing	30.0
Without background	41.5
With background	52.5
Background fine-tuned	59.5

Table 2 shows that performance rapidly drops at the contrast level of 15%.

4.4 Results on the Archimedes palimpsest

Background pre-training and fine-tuning. We performed an ablation study on one-channel images of the Archimedes palimpsest to see how our model performs for different background settings: without background generator, with background generator trained on "clean" manuscript, and with the background generator fine-tuned on the palimpsest. The results are presented in Table 3. For the Archimedes palimpsest we used Deep Convolutional Generative Adversarial Networks (DCGAN) since we can use for it original anti-shorcut loss. We used the same training parameters for all the experiments. The character classification accuracy for the unprocessed image improved from 30 to 59.5% after the reconstruction with fine-tuned background. For visual comparison, we illustrate on Fig. 13 samples from the original palimpsest and the results after reconstruction for three experiment settings: (1) without background, (2) with background, and (3) with fine-tuned background generators. From the figure, we can see that in all three cases the generated under-text is less accurate than for the synthetic palimpsest reconstruction. This discrepancy can be explained by several issues with dataset, and generator architecture that is not directly fit for generating continuous text. The first issue is that Minuscule Greek which is used in Greek 960 manuscript and Archimedes palimpsest is a continuous script, meaning that characters are connected to each other. For this reason, it would be extremely difficult to isolate one character at a time for training or reconstruction. Therefore, the generative model not only needs to correctly generate the character in the center of the patch but also the surrounding characters.

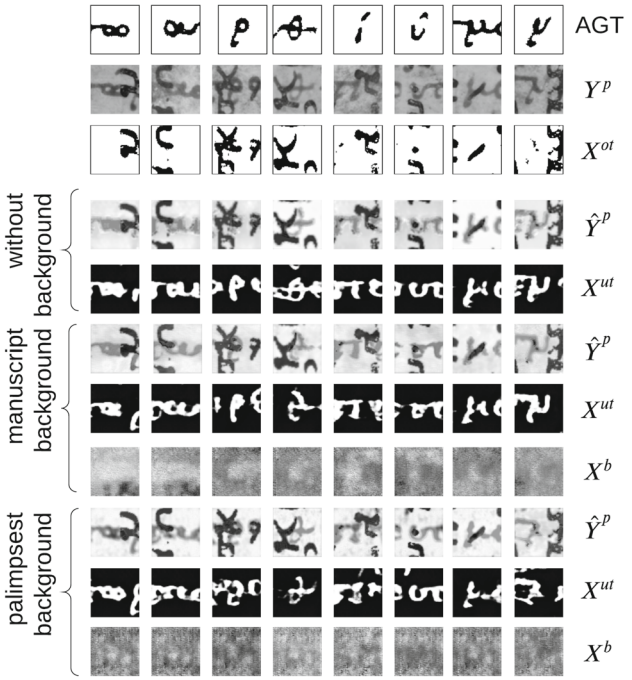


Fig. 13 Archimedes palimpsest reconstruction results. Top line called Approximate Ground Truth (AGT) is shown for the reference as a substitute of ground truth It represents under-text characters from unobscured regions that are closest to the potential ground-truth. Second row is the original palimpsest image. The third row is the a foreground text mask. The nine lower rows correspond to reconstruction and separated layers for three training setting: without background, with background, and with background generator tuned on the palimpsest. For each case, we show three outputs: reconstructed palimpsest, separated under-text, and separated background

However, this should not be a problem for the noncontinuous scripts, where all the characters are disconnected from one another, such as Uncial script. For continuous scripts, the problem can be solved by using a more advanced generator architecture for continuous handwritten text, such as [10].

Also, another issue is the difference between the source dataset for training the under-text generator and the final palimpsest dataset. The source dataset was made from a set of pages written in the same script but with different scribes. Therefore, some characters vary in style and differ from the palimpsest dataset characters. This fact brings two main disadvantages: first, our under-text reconstruction is no longer constrained only to the set of characters met in the palimpsest; second, some character styles are less represented than others. For example, if we look at the recognition rate of η 40/40/30%, γ 40/20/30% κ 20/10/10%, λ 40/40/10% for three ablation experiments, we can see that these characters have consistently low reconstruction. This tendency was constant over all our experiments. We provide a few samples of these characters from the palimpsest and the under-text dataset in Fig. 14a. On the other hand, such characters as μ: 90/100/60%, ν: 70/70/70%, ω: 80/90/60%, π: 90/60/60%,

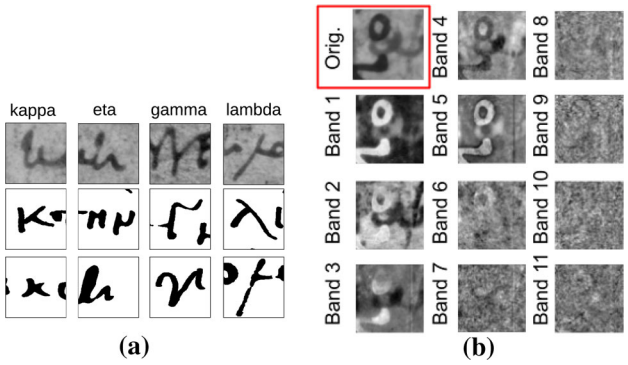


Fig. 14 **a** Samples of characters that have the lowest recognition rate after reconstruction. At the top row are characters from the Archimedes palimpsest. On the middle and bottom rows are characters from the manuscript used for under-text generator training. The middle row represents the widespread different variant, and the bottom the closest in appearance variant. **b** Original greyscale image of the character from Archimedes dataset versus the PCA transformed bands of multispectral image

ρ: 100/100/70% have consistently high reconstruction accuracy. This coincides with our observations between two datasets: characters with styles that were more represented in the source dataset, such as μ, ν, ω, etc., receive better reconstruction. In our experiments for the training of under-text GAN, we used a dataset of characters extracted by a semiautomatic method. The characters were binarized and then extracted with a connected component labeling algorithm automatically. Only then were they sorted out by a human, based on the simple criteria, whether the character is at the center of the patch or not. It is evident that involving the specialist in the language in the process of creating the under-text dataset would provide significant improvement to the accuracy of reconstruction.

Overall, from the ablation study, we can see that the background plays an important role in signal reconstruction. Moreover, tuning the background to a specific document brought an even more boost to the accuracy. The fine-tuned generator produces samples that are much closer to the background of the palimpsest document compared to the background generator trained on the first non-palimpsested manuscript. We also try to fine-tune the text generator. Unfortunately, we have only 1000 samples of palimpsest characters which are not enough for generator fine-tuning. Thus the generator collapsed into generating text together with background.

Comparison with ICA and PCA To compare our dataset against standard multispectral methods of manuscript processing, we created a multispectral Archimedes dataset. To make it the most comparable with the greyscale Archimedes dataset, we filtered the Archimedes palimpsest pages to find characters from the previous dataset to create their multispectral counterparts. In the end, we were able to detect 161 characters that are represented by 11 bands from 365 to 870 nm. We normalized the spectral image cubes to achieve the best performance. The accuracy for the PCA, ICA was calculated for each transformed band for the original and inverted variant. Here we show the best-achieved accuracy: PCA − 42.86%, ICA − 32.92%. We also considered the case when the camera settings were adjusted differently to each page, which can lead to different gains for different bands. In this case, we performed transformation separately for characters of the same page. To calculate the accuracy, we would select the transformed band for each page that gives the highest accuracy and then accumulate the result for overall accuracy. This way, we achieved 46.58% for PCA and 42.24% accuracy for ICA. We demonstrate the result in Fig. 14b. Although this is not a direct comparison, the OCR accuracy on corresponding 161 characters from the original greyscale Archimedes dataset is 53.42%. Similar to the MNIST dataset palimpsest, we saw the main difference in the reconstruction in overlapping areas between over-text and under-text. The PCA and ICA methods are not able to extract under-text regions in the places of overlap. For the multispectral dataset, our framework showed an accuracy of 47.83%. It is lower than for the original greyscale data, but the original dataset, although consist of one-band images, does not correspond to any band in a multispectral dataset. The greyscale image in the original Archimedes palimpsest dataset is a result of series of transformations. It is created by selecting few PCA transformed bands to form RGB image that, in the end, is converted to a greyscale image. These operations produce the image with higher contrast and lower noise. Also, the drop in performance can be explained by the need to reconstruct not only spatial but also spectral information. This made us use different loss functions and different mixing network architecture. For the mixing network, we used 1D deconvolution layers similar to one used in MNIST multispectral dataset. And for the loss, we used l2 loss with spectral cosine similarity loss from [23]. Further work is needed to find optimal architecture and loss for the reconstruction of noisy multispectral images.

5 Discussion

The proposed method is the first unsupervised deep learning method created for the separation of palimpsest text layers. From the provided experiment we have seen that it also demonstrates sufficient improvement compared to traditional palimpsest text separation techniques to be considered as a stand-alone method for the text separation problem. However, there are a number of limitations to this method that may create a question of its practicality if applied to the whole text page and not only on a character level that we need to clarify.

The main objective of our research was to incorporate spatial priors with a non-linear mixing model. That was implemented by a deep learning framework since it possesses both qualities of being able to learn high dimensional image abstractions and the non-linearity of the mixing process. The idea of learning the mixing model directly as was done in [38] was prohibited by our insight from the palimpsested document image acquisition process and some data properties which cause the variation in the mixing process. For example, the way the spectral signature of the ink may vary on the same piece of parchment, depending on the nature of damage such as parchment mold. On the other hand, the spatial properties of the writing for the same text are consistent regardless of the parchment damage or the imaging settings. This framework satisfies all of these requirements.

However, usage of deep generative networks as an image prior enforces constraints on the size of the reconstructed patch and also raises a concern of generating possible image artifacts, a known problem for GAN networks. Therefore, one of the directions of our future work is to investigate a more suited image prior. Such a prior can potentially be derived from autoregressive networks that are not limited in the generated image size. Also, as mentioned before we do not explicitly estimate the probability of the generated image as is required from the signal prior, but rather generate the image from the certain distribution. However, due to the poor distribution matching or mode collapse Generative networks may hallucinate text images that may not correspond to the ground-truth. Therefore, performing the unmixing stage first, and then assessing the likelihood of the separated under-text, is a more preferred sequence of action then generating text and mixing to match it to the original.

Another and more important course of future work is investigating a mixing function. Potentially, if we find the correct mixing function we can find an inverse operation, and then text layer separation can be performed in a more efficient fashion. Such an approach was proposed in the work [38], where the mixing process was represented by a physically inspired model for the superposition of two inks. The inverse operation was easily found with simple mathematical transformations. However, that model was based on the assumption of almost even text background, where all the values are estimated relative to its mean value. In very old documents such as palimpsested manuscripts, the background variation can be very significant and requires an adaptive model, which can be provided by approximating the mixing process with a neural network.

6 Conclusion

We propose the first known attempt to use an unsupervised deep neural network for palimpsest text separation on historical manuscripts. We cast the palimpsest document recreation problem as a separation problem, where both the under-text and background layers are modeled by deep generative networks and further mixed with a foreground text mask to simulate the original palimpsest images. The under-text image is optimized by comparing the generated image to the original palimpsest image. We consider a proposed framework as a foundation for future applications of Deep Learning techniques for the historical palimpsest manuscript text reconstruction. We construct the framework in such a way that each piece can be changed or adjusted to the palimpsest specification. We tested our framework on synthetic and real palimpsests. We improved character recognition by 50% after under-text separation on the synthetic dataset and over 29.5% on the real dataset. Our results provide insights into such problems of palimpsest text reconstruction as low signal-to-noise ratio and under-text dataset bias. Overall, we believe that this approach can facilitate palimpsest text recognition for scholars and automatic text transcription.

Acknowledgements The authors would like to thank Roger J. Easton and the Holy Monastery of Saint Catherine, His Eminence Archbishop Damianos, and the Monastery's Holy Council for providing digital copies of manuscripts from Mount Sinai. Also we would like to thank Rochester Institute of Technology Research Computing center for providing computational resources for some of the experiments [31].

Funding This research is funded under the Xerox Chair in Imaging Science held by Dr. Messinger.

Data availability The materials can be accessed from the permission of the authors.

Code availability The code is released at https://github.com/as3297/Palimpsest-Recostruction-Model.

Declarations

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

References

1. Akbarinia, A., Gegenfurtner, K.R.: How is contrast encoded in deep neural networks? arXiv preprint arXiv:1809.01438 (2018)
2. Anirudh, R., Thiagarajan, J.J., Kailkhura, B., Bremer, T.: An unsupervised approach to solving inverse problems using generative adversarial networks. arXiv preprint arXiv:1805.07281 (2018)
3. Arandjelović, R., Zisserman, A.: Object discovery with a copy-pasting gan. arXiv preprint arXiv:1905.11369 (2019)
4. Asim, M., Shamshad, F., Ahmed, A.: Blind image deconvolution using deep generative priors. In: 30th British Machine Vision Conference (2019)
5. Bora, A., Jalal, A., Price, E., Dimakis, A.G.: Compressed sensing using generative models. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 537–546. JMLR. org (2017)

6. Bora, A., Price, E., Dimakis, A.G.: Ambientgan: generative models from lossy measurements. ICLR **2**, 5 (2018)
7. Chollet, F., et al.: Keras. https://keras.io (2015)
8. Easton, R.L., Christens-Barry, W.A., Knox, K.T.: Spectral image processing and analysis of the archimedes palimpsest. In: 2011 19th European Signal Processing Conference, pp. 1440–1444. IEEE (2011)
9. Easton, R.L., Knox, K.T., Christens-Barry, W.A., Boydston, K., Toth, M.B., Emery, D., Noel, W.: Standardized system for multispectral imaging of palimpsests. In: Computer Vision and Image Analysis of Art, vol. 7531, p. 75310D. International Society for Optics and Photonics (2010)
10. Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., Litman, R.: Scrabblegan: semi-supervised varying length handwritten text generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4324–4333 (2020)
11. Gandelsman, Y., Shocher, A., Irani, M.: Double-dip: unsupervised image decomposition via coupled deep-image-priors. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 6, p. 2 (2019)
12. Glaser, L., Deckers, D.: The basics of fast-scanning XRF element mapping for iron-gall ink palimpsests. Manuscr. Cult. **7**(PUBDB–2015–06320), 104–112 (2014)
13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
15. Hanif, M., Tonazzini, A., Savino, P., Salerno, E.: Sparse representation based inpainting for the restoration of document images affected by bleed-through. In: Multidisciplinary Digital Publishing Institute Proceedings, vol. 2, p. 93 (2018)
16. He, S., Schomaker, L.: Deepotsu: document enhancement and binarization using iterative deep learning. Pattern Recognit. **91**, 379–390 (2019)
17. Hollaus, F., Gau, M., Sablatnig, R., Christens-Barry, W.A., Miklas, H.: Readability enhancement and palimpsest decipherment of historical manuscripts. Kodikologie und Paläographie im Digitalen Zeitalter 3: Codicology and Palaeography in the Digital Age, vol. 3, p. 31 (2015)
18. Jacobs, B., Momoniat, E.: A novel approach to text binarization via a diffusion-based model. Appl. Math. Comput. **225**, 446–460 (2013)
19. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
20. Kong, Q., Xu, Y., Wang, W., Jackson, P.J., Plumbley, M.D.: Single-channel signal separation and deconvolution with generative adversarial networks. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19) (2019)
21. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
22. Lettner, M., Sablatnig, R.: Multispectral imaging for analyzing ancient manuscripts. In: 2009 17th European Signal Processing Conference, pp. 1200–1204. IEEE (2009)
23. Li, J., Cui, R., Li, Y., Li, B., Du, Q., Ge, C.: Multitemporal hyperspectral image super-resolution through 3d generative adversarial network. In: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), pp. 1–4. IEEE (2019)
24. Li, S.C.X., Jiang, B., Marlin, B.: Misgan: learning from incomplete data with generative adversarial networks. In: ICLR (2019)
25. Lunz, S., Öktem, O., Schönlieb, C.B.: Adversarial regularizers in inverse problems. In: NIPS, pp. 8507–8516 (2018)
26. Mindermann, S.: Hyperspectral imaging for readability enhancement of historic manuscripts. Ph.D. thesis, Technical University of Munich (2018)
27. Moghaddam, R.F., Cheriet, M.: Low quality document image modeling and enhancement. Int. J. Doc. Anal. Recognit. (IJDAR) **11**(4), 183–201 (2009)
28. Pajot, A., de Bezenac, E., Gallinari, P.: Unsupervised adversarial image reconstruction. In: ICLR (2019)
29. Rapantzikos, K., Balas, C.: Hyperspectral imaging: potential in non-destructive analysis of palimpsests. In: IEEE International Conference on Image Processing 2005, vol. 2, pp. II–618. IEEE (2005)
30. Rick Chang, J., Li, C.L., Poczos, B., Vijaya Kumar, B., Sankaranarayanan, A.C.: One network to solve them all–solving linear inverse problems using deep projection models. In: CVPR, pp. 5888–5897 (2017)
31. Rochester Institute of Technology: Research computing services (2019). https://doi.org/10.34788/0S3G-QD15. https://www.rit.edu/researchcomputing/
32. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)
33. Salerno, E., Tonazzini, A., Bedini, L.: Digital image analysis to enhance underwritten text in the Archimedes palimpsest. Int. J. Doc. Anal. Recognit. (IJDAR) **9**(2–4), 79–87 (2007)
34. "Sinai Greek 960" sinai.library.ucla.edu, a publication of St. Catherine's Monastery of the Sinai in collaboration with EMEL and UCLA (2019). https://sinai.library.ucla.edu/. Accessed 5 June (2019)
35. Soltani, M., Jain, S., Sambasivan, A.: Learning generative models of structured signals from their superposition using gans with application to denoising and demixing. arXiv preprint arXiv:1902.04664 (2019)
36. Starynska, A., Easton Jr, R.L., Messinger, D.: Methods of data augmentation for palimpsest character recognition with deep neural network. In: Proceedings of the 4th International Workshop on Historical Document Imaging and Processing, pp. 54–58. ACM (2017)
37. Tonazzini, A., Bedini, L., Salerno, E.: A Markov model for blind image separation by a mean-field EM algorithm. IEEE Trans. Image Process. **15**(2), 473–482 (2006)
38. Tonazzini, A., Savino, P., Salerno, E.: A non-stationary density model to separate overlapped texts in degraded documents. Signal Image Video Process. **9**(1), 155–164 (2015)
39. Trier, O.D., Jain, A.K.: Goal-directed evaluation of binarization methods. IEEE Trans. Pattern Anal. Mach. Intell. **17**(12), 1191–1201 (1995)
40. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
41. Valdiviezo-N, J.C., Urcid, G.: Multispectral images segmentation of ancient documents with lattice memories. In: Digital Image Processing and Analysis, p. DMD6. Optical Society of America (2010)
42. Vatican palimpsests. https://spotlight.vatlib.it/palimpsests/
43. Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T.: scikit-image: image processing in python. PeerJ **2**, e453 (2014)
44. Walvoord, D.J., Easton, R.L.: Digital transcription of the Archimedes palimpsest [applications corner]. IEEE Signal Process. Mag. **25**(4), 100–104 (2008). https://doi.org/10.1109/MSP.2008.924960
45. Wu, K., Otoo, E., Shoshani, A.: Optimizing connected component labeling algorithms. In: Medical Imaging 2005: Image Processing,

vol. 5747, pp. 1965–1977. International Society for Optics and Photonics (2005)
46. Yeh, R.A., Chen, C., Yian Lim, T., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with deep generative models. In: CVPR, pp. 5485–5493 (2017)