Pattern Recognition 123 (2022) 108370

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Enhance to read better: A Multi-Task Adversarial Network for Handwritten Document Image Enhancement

Sana Khamekhem Jemni[a,c,1], Mohamed Ali Souibgui[b,1,*], Yousri Kessentini[c,d], Alicia Fornés[b]

[a] MIR@CL: Multimedia, InfoRmation systems and Advanced Computing Laboratory, Tunisia
[b] Computer Vision Center, Computer Science Department, Universitat Autónoma de Barcelona, Spain
[c] Digital Research Center of Sfax, B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia
[d] SM@RTS: Laboratory of Signals, systeMs, aRtificial Intelligence and neTworkS, Tunisia

ARTICLE INFO

Article history:
Received 13 May 2021
Revised 13 August 2021
Accepted 10 October 2021
Available online 11 October 2021

Keywords:
Handwritten document image binarization
Document enhancement
Handwriting text recognition
Generative adversarial networks
Recurrent neural networks

ABSTRACT

Handwritten document images can be highly affected by degradation for different reasons: Paper ageing, daily-life scenarios (wrinkles, dust, etc.), bad scanning process and so on. These artifacts raise many readability issues for current Handwritten Text Recognition (HTR) algorithms and severely devalue their efficiency. In this paper, we propose an end to end architecture based on Generative Adversarial Networks (GANs) to recover the degraded documents into a *clean* and *readable* form. Unlike the most well-known document binarization methods, which try to improve the visual quality of the degraded document, the proposed architecture integrates a handwritten text recognizer that promotes the generated document image to be more readable. To the best of our knowledge, this is the first work to use the text information while binarizing handwritten documents. Extensive experiments conducted on degraded Arabic and Latin handwritten documents demonstrate the usefulness of integrating the recognizer within the GAN architecture, which improves both the visual quality and the readability of the degraded document images. Moreover, we outperform the state of the art in H-DIBCO challenges, after fine tuning our pretrained model with synthetically degraded Latin handwritten images, on this task.

1. Introduction

Handwritten document analysis is an active and important field in computer vision and pattern recognition community. With the recent developments in machine learning [1], processing handwritten documents is reaching a good accuracy, especially in the application of Handwritten Text Recognition (HTR). HTR is the crucial part towards automatically understanding a document, which facilitates the access to various automatic applications such as: information extraction, search, indexation, validation, etc.

One of the problems that HTR systems are facing is the degradation of an inputted document, which significantly decreases the reading performance, reflecting on its utility. Indeed, many degradation scenarios can be attached to a handwritten document, especially the historical ones. Degradation includes background noise, corrupted text, dust, wrinkles and historical effects just to name a few related to the condition of the document itself [2]. The bad scanning process can also produce problems (shadows, blur, light distortion, angle, etc.) [3,4]. Moreover, some documents contain watermarks or stamps inserted for security reasons, those can cover the text and obstruct the HTR engine [5]. Some degradation examples are presented in Fig. 1, as it can be seen, cleaning the document before passing it to the HTR stage should be done.

This cleaning task, called document enhancement, includes different recovering techniques, to reverse the degradation effect, for example: Binarization, dewarping, deblurring, watermark removal, etc. Classic recovery techniques were integrating image processing algorithms to be used as a filter that separates the degradation from the text. However, those methods are failing in removing the high degradation. Also, their parameters are usually set depending on the quality state of the addressed document to produce an optimal result. Thus, a manual intervention is needed in some cases to adjust the parameters, which is quite costly.

Given this, some modern document recovery techniques are appearing, using machine learning tools. Those are training deep learning models, mainly Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), to learn the parameters for a direct mapping of any degraded document image into a clean binary version (without a restriction on degradation level) [6–8]. Similar to those, we proposed in [5] a document enhance-

* Corresponding author.
E-mail address: msouibgui@cvc.uab.cat (M.A. Souibgui).
[1] Those authors were equally contributed to this paper

https://doi.org/10.1016/j.patcog.2021.108370
0031-3203/

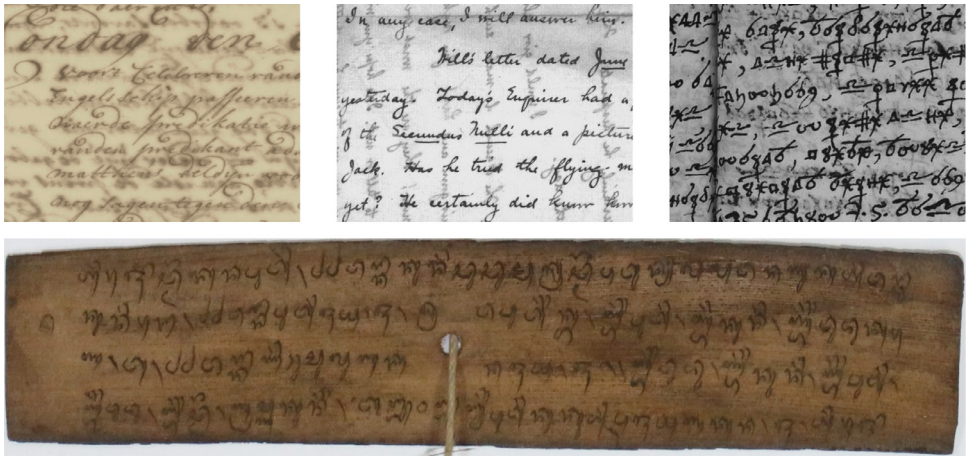

Fig. 1. Examples of the degradation that can be appeared in handwritten text images.

ment model called DE-GAN. However, and despite the high accuracy that we achieved in various enhancement tasks (Binarization, cleaning, watermark removal, deblurring), an important evaluation was not done. In fact, the goal of document enhancement is to provide a cleaner version of the image which is highly beneficial for HTR engines. But, in the mentioned approaches (including ours), the evaluation was conducted using only the visual similarity measurement between the recovered image and the Ground Truth (GT) clean version with some metrics that depend on pixels values. Thus, a HTR evaluation (means, passing the images to a HTR engine and compare the recognized text with the GT) is missing, for a better validation of the developed approaches. Also, these models are generally trained using only the images, while ignoring the text. As result, a model can easily evolve to deteriorate the text while cleaning the degraded image.

Motivated by those challenges, we propose a new method consists in an improved version of our previously developed DE-GAN, which was designed to recover the handwritten document to a clean version while ensuring its readability. Our approach is a deep learning model based on GANs that learns its parameters not only from the handwritten images pairs (degraded + GT), but also from the associated GT text. For this aim, we propose to add a recognizer that is trained jointly in a GAN model to assess the readability of the recovered document image. Hence, the model shall learn the best mapping of the degraded image to be as clean as possible while keeping its text readable. To accomplish this, and since the used datasets for document binarization does not (or rarely) contain the text information, we used two publicly available handwriting text images datasets (KHATT for Arabic script and IAM for Latin script) that are originally used for HTR to create degraded versions from the GT clean text lines images. The contribution of this paper can be summarized as follows:

- To the best of our knowledge, this is the first work that integrates a recognition stage in a document binarization model. Thus, the degraded handwritten document will be recovered while maximising its readability, simultaneously. This is done by combining the GAN and the Connectionist Temporal Classification (CTC) losses functions: We eliminate the noise while preserving the handwritten text strokes.
- We demonstrate that training the recognizer progressively (on images ordered from the degraded domain to the clean versions), improves the recognition performance.
- The proposed model is simple, and flexible to restore different forms of degradation, independently of the document language. This was shown by the experiments conducted on two created datasets namely degraded-IAM (Latin script) and degraded-KHATT (Arabic script).
- We achieved the SOTA performance in handwritten document binarization according to H-DIBCO benchmarks.

The rest of this paper is organized as follows. Section 2 provides a review of prior works on document enhancement, in particular for document recovery and binarization. Then we present our proposed model in Section 3. After that, experimental results and comparisons with recent methods will be described in Section 4. Finally, a conclusion with a future direction is given in Section 5.

2. Related work

This work aims to recover images that contain hard degradation by removing the background noise, while keeping its readability by HTR systems as accurate as possible. This application called document enhancement, is generally a preprocessing stage that produces an enhanced version of the document, in order to improve the recognition results of HTR engines [9].

Early methods known by global binarization, aimed to find a single threshold value for the whole document. A more sophisticated approaches, named local binarization, determine a different threshold value for each pixel. These threshold values are further used to classify the image pixels into foreground (black) or background (white) [10,11]. Nowadays, thresholding based methods are still evolving, for instance a global threshold selection method was introduced in [12] basing on fuzzy expert systems (FESs). In this method the image contrast is enhanced in a first step. Then, a pixel-counting algorithm is used with another FES for thresholds adjustment as a range, before choosing the right value from the middle of that produced list. Moreover, a support vector machine (SVM) based approach was proposed in [13]. Given the local features of the gray level images, degraded regions were classified according to SVM to be binarized according to one of four different threshold values. The main drawback of these thresholding methods is the sensitivity to the document condition, they usually fail to restore highly degraded images [2], since it is hard to obtain a good filter in those scenarios.

Later, energy based methods were introduced to track the text while binarizing the image. In [14], the ink was considered as a target to be maximized by an energy function, while minimizing the degraded background. Similarly, the background is estimated and subtracted from the degraded image using a mathematical morphology in [15]. However, the results using those hand crafted image processing algorithms were unsatisfactory.

Recently, deep learning architectures were used to tackle this problem by training their weights directly from raw data. In [16], the problem was formulated as pixels classification depending on sequences. Hence, a 2D Long Short-Term Memory (LSTM) was used to predict each pixel value whether belonging to the text or the degradation given a sequence of its neighbours. This process is, of course, time consuming. Thus, instead of classifying each pixel separately, images were mapped in an end to end fashion from the degraded version into the clean one using the Convolutional Neural Networks (CNNs). These architectures, called auto-encoders, lead to recent improvements in image denoising [17] and more particularly documents binarization [18–20] or deblurring problems [21]. This kind of applications are now called image-to-image translation, since the goal is to start from a degraded image and learn a mapping function that translate it into a clean domain. Following these approaches, [7] proposed an auto-encoder architecture that performs a cascade of pre-trained U-Net models [22] to learn the binarization with less data. Also, [23] proposed a neural network to learn the enhancement/binarizatoin in an iterative fashion.

Other deep learning approaches modeled the problem as a generation task, where the goal is to generate a clean version of the image by conditioning on the degraded one. This process was carried out using GANs architectures, composed of a generative model that produce a clean version of the image and a discriminator to assess the binarization result. Thus, and motivated by other approaches where GANs significantly surpasses autoencoders in image-to-image tasks [24], some approaches applying this method were introduced. In [5], a conditional GAN approach was developed for document enhancement and achieved good results in recovering handwritten documents with several backgrounds degradation scenarios, it was also used for optical documents deblurring and dense watermarks removal. A similar cGAN based method was also proposed in [8], where the binarization performed by the generator was done in two stages, learning the pixels in different scales then combining the results to provide the final output. In [25], a strokes preservation method was developed using a GAN model, this was done by learning the text boundary in an auxiliary task for a better document binarization, especially with weak or ambiguous strokes. Another GAN's based method was proposed in [26] using a two networks frameworks, for document binarization. The first one was conditioning on the clean image to generate a degraded one, while the other network reconstructed the clean version conditioning on the degraded image. Thus, an unpaired data training was performed leading to good results when using the second network to binarize images. Similarly, [6] treated the problem as two stages: The first stage was devoted to augment the data by creating degraded handwritten images using a GAN model, while the second stage exploited the generated images to train an inverse problem binarization model.

To summarize, deep learning based methods are now significantly surpassing the classic or modern image processing handcrafted algorithms for the handwritten document binarization task. However, the only limitation that can be noticed is the "*image only*" based training. Because, the usual benchmarks for testing the binarization performance do not include a text recognition evaluation (a GT truth text of the degraded documents). Thus, those methods could easily delete some parts of the handwritten text while binarizing the image, without noticing. It is to note that we addressed this problem in a previous work [3], but for printed documents domain and using the Tesseract OCR engine to evaluate the produced text. Where the character error rate of the OCR on the generated image is inputted as an additional input channel to the discriminator. Contrary, we are proposing in this study, is a handwritten text recognizer that is jointly trained in the GAN architecture to maintain the text, while cleaning the degraded image. Thus, it is more flexible to be used for different handwritten languages and writing styles.

3. Proposed method

We treat recovering a clean version from a handwritten degraded document as an image-to-image translation task using a generative model. Our GAN architecture is composed of three main parts, as shown in Fig. 2: A generator, a discriminator and a handwritten text recognizer. Since we are using the text information, the patches that are used during training should be in a readable form by a HTR, after binarization. Thus, the model is designed to be working at handwritten line images level. During training, the generator is conditioning on the degraded line image to generate the clean version. The generated image is passed to the discriminator to assess it as real (looks clean) or fake (looks degraded), for ensuring a realistic visually recovery. The image is also passed also to the HTR model to read it and compare the recognized text to the GT, hence, to maintain its readability while recovering it. The discriminator, as well as the recognizer, passed their feedback about the generated image through the adversarial loss. Noting that another additional Binary Cross Entropy (BCE) loss is integrated in the generator, for a faster convergence. In this way, the generator parameters are learned to produce a handwritten image that is as clean as possible, while keeping the text quality. In what follows, we explain the three components presented in our architecture with more details.

3.1. Generator

Since we are doing an image-to-image translation process, the generator is designed as an auto-encoder model. We employ the U-net [22] for this task, in which the inputted image is encoded through a sequence of convolution layers with a down-sampling to reach a specific layer. After that, the image is decoded with a sequence of up convolution layers with an up-sampling. The model involves some skip connections after each two successive layers to recover images with lower deterioration, since the goal is to keep the text while removing the degradation. Thus, skip connections can help the decoder in maintaining the text features while producing the image. Figure 3 shows further details about the used generator. As can be seen, it is composed of 23 convolutional layers, with Dropout regularization and batch normalization layers. The output of this model is a single channel (in gray level) image, assumed to be the cleaned version of the inputted degraded image.

3.2. Discriminator

The discriminator is another Fully Convolutional Network (FCN) that produces an assessment of the generated image in term of visual similarity (pixel level) with the GT (real) images. The model was designed to take a degraded image with its clean version and output the class "real" if the clean version is the real GT, or assign the class "fake" if the clean image was produced by the generator. Both input images, which have of course the same size, are concatenated in an $H \times W \times 2$ shape. Then, the obtained volume is propagated in the discriminator model detailed in Fig. 4, to end up in the last layer as a form of $H/16 \times W/16 \times 1$ matrix. During training, this matrix contains values that are equal to 1 in case of inputting the GT as a clean image, and equal to 0 in case of inputting the generator based enhanced image.

3.3. Handwritten text line recognizer

The used handwritten recognizer is a Convolutional Recurrent Neural Network (CRNN) model, following the architecture pre-

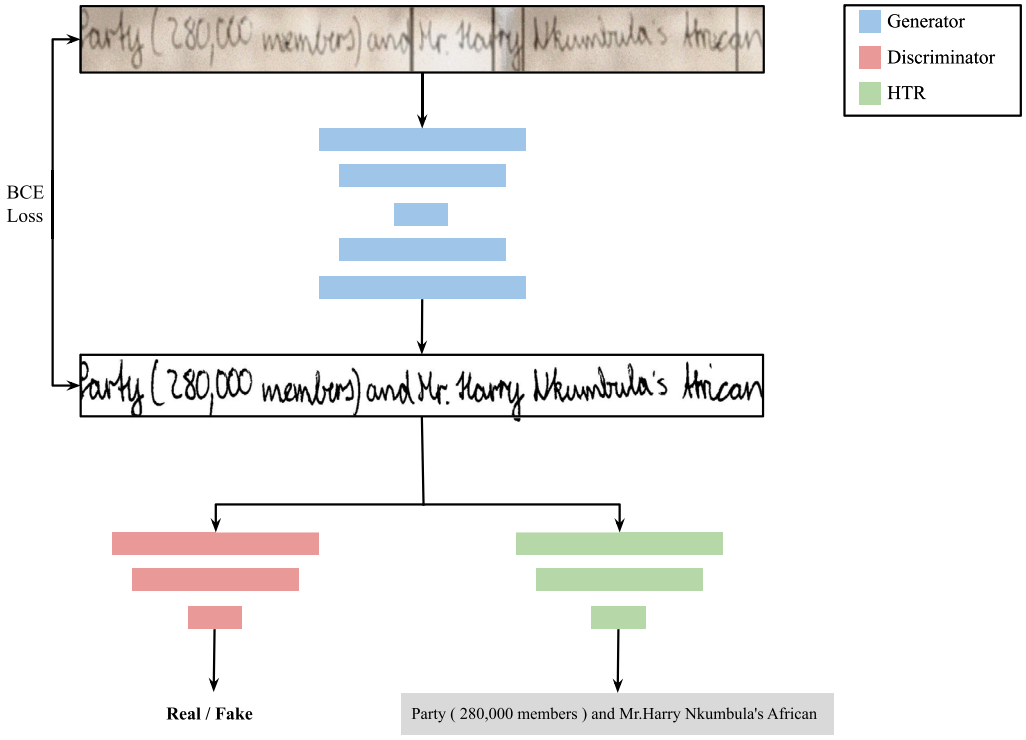

Fig. 2. Proposed architecture for document binarization.

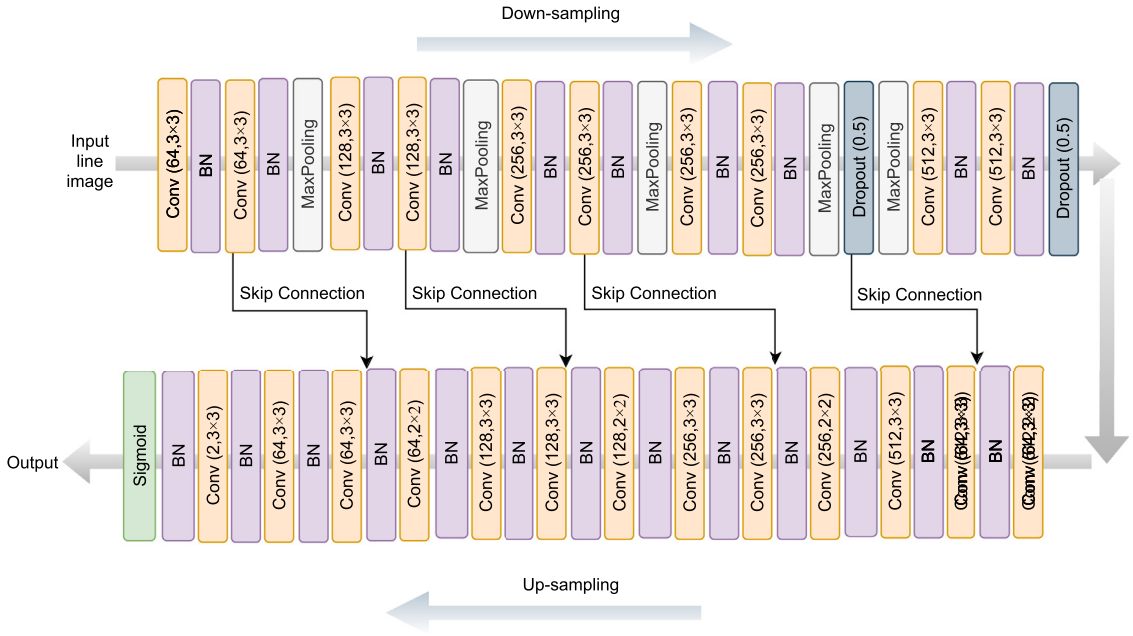

Fig. 3. Generator's architecture design used in this study.

sented recently in [27] and considered among the best HTR architectures. Noting that, any other HTR can be also used for this task, for instance: [28–30]. The model architecture is detailed in Fig. 5. After enhancing the image by the generator it is inputted to an encoding stage that uses convolutional and gated convolutional layers, with integrated regularization techniques. The encoded image is passed later to the decoding stage, consists in two bidirectional Gated Recurrent Unit (GRU) layers. Finally, the CTC is used to decode the feature frames into text characters. The CTC layer is having the size of the character set plus one additional symbol corresponding to the blank symbol. During training, the recognizer could be fitted with two types of clean images, forming the following two scenarios:

- **S1:** The Recognizer is trained at each iteration with the GT images that are related to the degraded batch images inputted to the generator. The GT images are used with the associated GT text transcription for training in this process.
- **S2:** The Recognizer is trained using the images enhanced by the generator at each iteration with the associated GT text. The intuition behind, is that we may obtain a better recognition con-

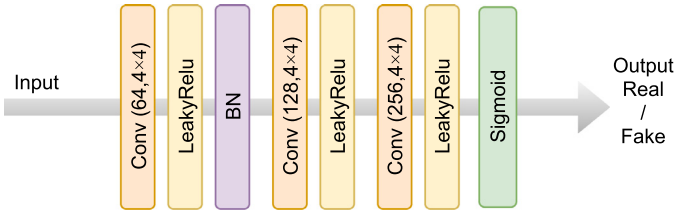

Fig. 4. Discriminator's architecture used in this study.

vergence that is going progressively from the degraded domain to the clean domain.

3.4. Training process

The different components presented above were trained jointly. The generator G, which is a function having the parameters θ_G, is conditioned on the degraded image I_d to provide its cleaned image that should be as close as possible to the GT image I_{gt}. This image is passed to be validated by the discriminator D and handwritten recognizer R, with parameters θ_D and θ_R, respectively. D is giving an assessment of the cleaned image about its cleanliness to be Real or Fake, $P(\text{Real}) = D_{\theta_D}(G_{\theta_G}(I_d))$. This adversarial training process of G and D can be formalized by:

$$\mathcal{L}_{adv}(\theta_G, \theta_D) = \mathbb{E}_{I_d, I_{gt}} \log[D_{\theta_D}(I_d, I_{gt})] + \mathbb{E}_{I_d} \log[1 - D_{\theta_D}(I_d, G_{\theta_G}(I_d))] \tag{1}$$

R is recognizing the generated image to maintain its readability with the CTC decoder, $CTC(t, R_{\theta_R}(G_{\theta_G}(I_d)))$, where t is the GT text. Note that it is trained with a clean version of the image (whether using S1 or S2, presented above), at each same iteration: $CTC(t, R_{\theta_R}(I_{gt}))$. Also, for a faster convergence, a simple BCE loss is used in the generator between the cleaned images and the GT ones, $BCE(\theta_G)$. Thus, the generator is being affected by three factors to produce its generation. The whole architecture is formalized as:

$$\begin{aligned}\mathcal{L}(\theta_G, \theta_D, \theta_R) = & \; min_{\theta_G} max_{\theta_D} L_{adv}(\theta_G, \theta_D) \\ & + \lambda (E_{t, I_d} CTC(t, R_{\theta_R}(G_{\theta_G}(I_d))) + \beta BCE(\theta_G)\end{aligned} \tag{2}$$

Where λ and β are the weights balancing the components intervention to produce the final generated image. During our experiments, we set λ to 1 and β to 10. For training, we used Adam's optimizer for the generator and discriminator components, while using the RMSProp for the handwritten text recognizer.

4. Experiments and results

We provide in this Section the experiments that were done to validate the effectiveness of our proposed method. First, we start by presenting the metrics and datasets used in our evaluation.

4.1. Metrics

Following the usual approaches for handwritten document image binarization [31], we use the same metrics to validate the cleaned images. Those metrics which compare the images visual similarity with the GT clean ones, are: Peak signal-to-noise ratio (PSNR), F-Measure (FM), pseudo-F-measure (Fps) and Distance reciprocal distortion metric (DRD). In addition, since we are using the text information to validate our model, we utilise as well the HTR metrics for comparing the recognized text to the GT one. These metrics are based on the Levenshtein distance [32], and they consists in the Character Error Rate (CER) and the Word Error Rate (WER) measures.

4.2. Handwritten text databases

Usual handwritten document binarization databases do not contain text information [2,33,34]. Thus, we opt to create synthetically degraded images from the databases used in HTR tasks in order to exploit the GT text provided within these datasets. We address in this study two different alphabets: Arabic and Latin. From each, we took the most used database for handwritten text line image recognition: KHATT [35] and IAM [36], to add degradation. We call the created datasets, degraded-KHATT and degraded-IAM.

4.2.1. Degraded-KHATT

The KHATT dataset was developed for Arabic manuscript recognition, contains texts lines images with their associated GT texts. In our experiments, we used 6161 lines for training and a set of 1861 lines for testing, while a set of 940 lines was used for validation as it was done in [28]. Then, we added random distortions as shown in Fig. 6 to obtain the degraded-KHATT dataset. To accomplish this, we insert different background images containing some flaws or artifacts. These background images are extracted mainly from public historical documents such as Nabuco, Bickley diary and Persian datasets. We have also applied different distortion operations, especially, dilation, erosion and blurs using random kernel sizes (2×2 and 3×3 for dilation, 2×2, 3×3 and 4×4 for erosion and from 1×1 to 15×15 for blurring). We inserted also random vertical lines having random widths in order to simulate the noise that can occur in old historical documents.

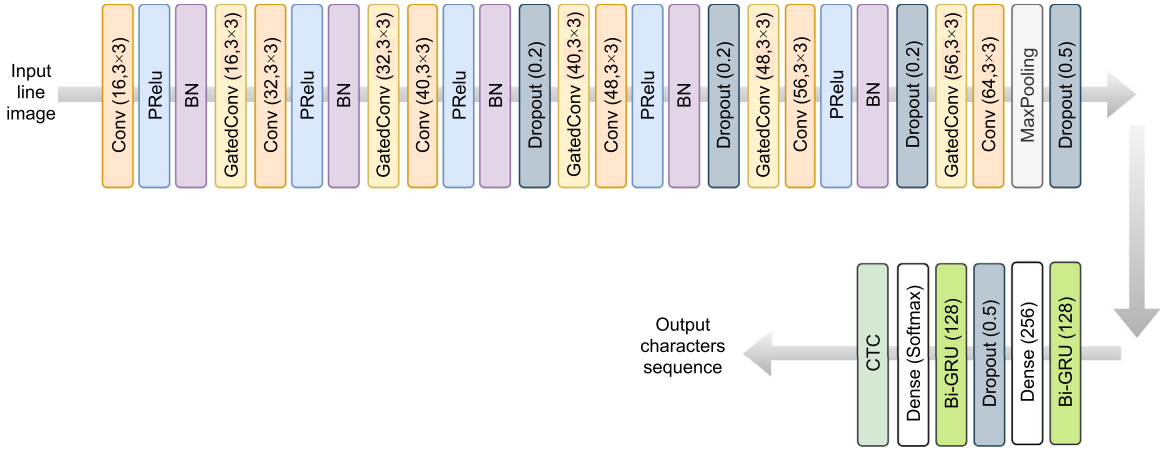

Fig. 5. Workflow of the CNN-Bi-GRU recognizer's architecture.

Fig. 6. Examples of distorted line images of the degraded-KHATT database used in this study, images are presented in gray level.

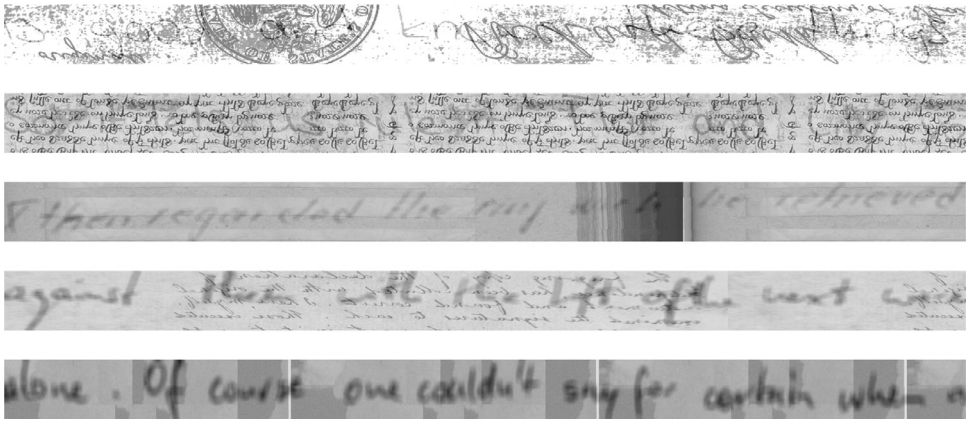

Fig. 7. Examples of distorted line images of the degraded-IAM database used in this study, images are presented in gray level.

4.2.2. Degraded-IAM

The IAM dataset was proposed for handwritten Latin script text recognition. It contains 8962 line images taken from the Lancaster-Oslo/Bergen (LOB) corpus. To insert degradation, we used same as in KHATT, 6161, 940, 1861 line images for training, validation and testing, respectively. We add dense backgrounds to simulate real historical deteriorated images same as it was done for the degraded-KHATT presented above. Examples of the obtained degraded-IAM are illustrated in Fig. 7.

4.3. Results

4.3.1. Arabic handwritten texts images recovery

For Arabic, we fed the proposed model with the training set of the created degraded-KHATT database. As stated above, the generator is trained to map the degraded image into a clean version, which will be evaluated by the discriminator and the recognizer. It is to note that in our experiments (for Arabic and Latin manuscripts), we used a high degradation for a meaningful evaluation in the hard scenarios. Also, we separate the background types between the training and testing sets (i.e there is no intersection in the background noise between the two sets).

Table 1 illustrates the obtained results of the performed image binarization methods on the test set of the degraded-KHATT database. Reminding that we proposed two binarization scenarios sharing the same GAN architecture and integrating a CRNN recognizer, S1 and S2, stated before. The key difference between the two scenarios is the data fed to the recognizer during the GAN training stage. In scenario S1, the recognizer (we call it CRNN1) is fed with ground truth clean images, while it is fed with generated images (cleaned by the generator) in the second scenario (called CRNN2).

As it can be seen, contrary to the previous related approaches, we evaluate the image in its visual quality (Binarization performance) and readability (recognition performance) at the same time. For readability, we tested each of our both scenarios S1 and S2 (Reco. CRNN1 and Reco. CRNN2) using the two recognizers (CRNN1 and CRNN2).

To compare our approach, we used a simple GAN architecture as a baseline. The architecture contains the same generator and discriminator of our architecture, but, without using a recognizer. We compare also with the method presented in [3] for printed text recovery, where an OCR is used during training as a part of the discriminator. However, since we are doing a handwritten text recognition (not optical text). We modify it by training a HTR having the same architecture as [27] to use it as a part of the discriminator, more details are given in [3].

Table 1
Image binarization results for the *test set* (degraded-KHATT database). (A → B): The CRNN is trained on images from the domain A and tested on images from the domain B. Deg.: Degraded images. Reco.: Recognition performance.

	Binarization Performance (Visual Quality)				Reco. CRNN1%		Reco. CRNN2%	
Method	PSNR	FM	Fps	DRD	CER	WER	CER	WER
CRNN [27] (GT → GT)	ND	ND	ND	ND	12.04	32.39	-	-
CRNN [27] (Deg. → Deg.)	4.80	25.45	25.70	107.22	30.34	54.44	-	-
CRNN [27] (GT → Deg.)	4.80	25.45	25.70	107.22	91.18	100	-	-
Baseline cGAN	**15.52**	75.01	75.11	**6.05**	29.24	53.68	-	-
cGAN [3]	15.10	75.56	75.75	11.78	28.84	54.37	-	-
Ours (S1)	15.45	**77.45**	**77.60**	7.97	27.03	52.84	**24.33**	**47.67**
Ours (S2)	15.44	74.52	74.62	6.18	27.90	53.49	25.31	48.48

Out of the results, we can see that using the GT images, a trained HTR engine based on CRNN [27] is reaching a CER of 12.04% and 32.39% as a WER, this is considered as our upper bound for recognition. Using the same model to recognize the degraded images, we obtain a poor performance of 91.18% in CER, obviously because the model is trained on the clean data. If we train the model on the degraded train set, it results in 30.34% of CER and 54.44% as WER. This experiment is done to verify later if we can surpass this performance (as a baseline) by cleaning the images then reading them, instead of training a model on the degraded domain.

The different binarization approaches, as it can be noticed, are enhancing the visual quality and the readability of the degraded lines. However, we can see that the baseline cGAN which is not taking the text into consideration while cleaning the image, is producing a result having a better visual quality in term of PSNR and DRD, but a worse readability compared to the methods integrating the text information. For the recognizer based methods, it is clear that our recovery method (S1) is leading to the best performance in terms of having a good visual enhancement while conserving the text readability. Since by recognizing the produced images, we get a CER of 27.03% and a WER of 52.84% when using the recognizer of S1 and a CER of 24.33% and a WER of 47.67% when using the S2 recognizer. This proves that using the text information during binarizing the images is useful. Also, we notice that the progressive learning of a HTR (training in an order from the degraded images to their clean versions) in a multitask framework, is better for the recognition task. However, using the HTR pretrained with the clean GT images (as a separate task) during enhancing the document, is better for the binarization performance (visually).

To illustrate our method's effectiveness, we show in Figs. 8 and 9 some qualitative results of recovering the distorted lines, ranging from easy distortions to the hard ones. We can see that our method is the most successful in producing clean images especially in cases of highly degraded ones. In fact, it can even recover the vanished handwritten text strokes.

Furthermore, as we stated above and since different weights can be used in the recognizer loss level to control the enhancement, we perform an ablation study to evaluate the right trade of between the visual quality and the readability during enhancing. In other words, the effectiveness of the weight λ presented in Eq. (2). This is done by varying the weight λ, then training the model with that setting and finally measuring the visual quality and readability (using the recognizer of S1) at each time to have the right option. The obtained results are shown in Table 2, where the experiments were carried out using the first scenario (S1) on a set of the Degraded-KHATT database, and ended up by selecting the setting of λ to be 1.

4.3.2. Latin handwritten texts images recovery

For Latin manuscript, we performed the same experiments using the degraded-IAM dataset. The obtained results are presented

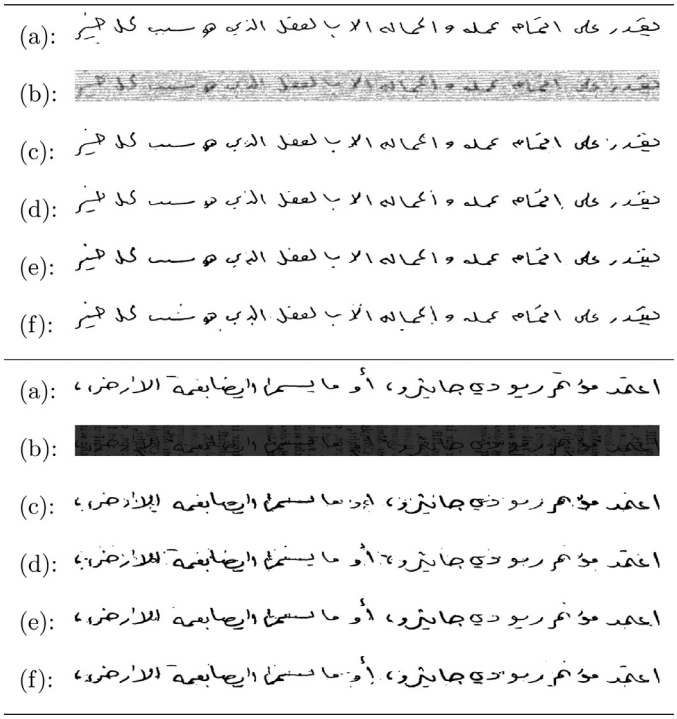

Fig. 8. Results of our proposed method for recovering degraded lines images. (a): GT, (b): Distorted, (c): Baseline cGAN, (d): cGAN [3], (e): Ours S1, (f): Ours S2.

Table 2
Impact of the recognizer weight on the final generated image.

	Binarization Performance (Visual Quality)	Reco. performance	
λ	PSNR	CER%	WER%
0.5	**17.94**	11.98	31.07
1	17.88	**11.74**	**31.05**
5	17.71	12.66	32.44
10	17.07	15.22	36.74
20	16.32	19.72	40.84

in Table 3. As it can be seen, training a CRNN [27] on the degraded images leads to 40.34% as a CER and a WER of 74.05%, with an obvious poor visual binarization quality since there was not a performed binarization with this way (using directly the degraded version). Contrary, by cleaning the image and passing it to the recognizer, better results were obtained. Here, same as the previous experiment, we are comparing our method to the basic GAN (without a recognizer), to validate the use of text information in our current method and our proposed one in [3]. It can be noticed that our method surpasses both GAN methods in the visual quality, and achieves the best text recognition rate compared to the other options. By using the recognizer trained in S1, we boost the CER

Table 3
Image binarization results for the *test set* (degraded-IAM database). (A → B): The CRNN is trained on images from the domain A and tested on images from the domain B. Deg.: Degraded images. Reco.: Recognition performance.

	Binarization Performance (Visual Quality)				Reco. CRNN1%		Reco. CRNN2%	
Method	PSNR	FM	Fps	DRD	CER	WER	CER	WER
CRNN [27] (GT → GT)	ND	ND	ND	ND	11.92	36.07	-	-
CRNN [27] (Deg. → Deg.)	6.01	26.13	26.12	70.81	40.34	74.05	-	-
CRNN [27] (GT → Deg.)	6.01	26.13	26.12	70.81	90.46	99.50	-	-
Baseline cGAN	14.99	75.44	75.01	5.91	31.51	60.95	-	-
cGAN [3]	15.86	80.89	80.83	5.00	27.55	58.08	-	-
Ours (S1)	**15.97**	**81.69**	**81.55**	**4.83**	26.05	56.07	**21.98**	**49.74**
Ours (S2)	15.87	81.12	81.16	5.09	27.48	58.35	23.07	51.15

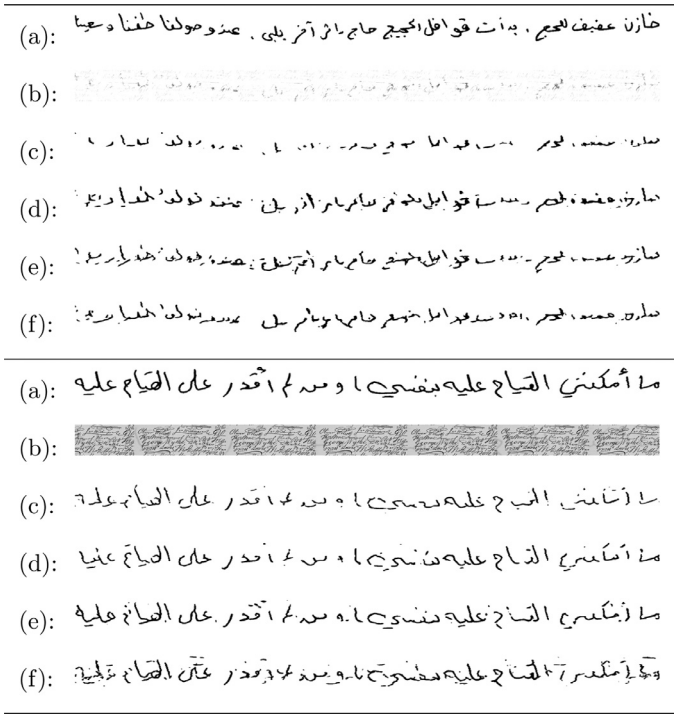

Fig. 9. Results of our proposed method for recovering extremely degraded lines images. (a): GT, (b): Distorted, (c): Baseline cGAN, (d): cGAN [3], (e): Ours S1, (f): Ours S2.

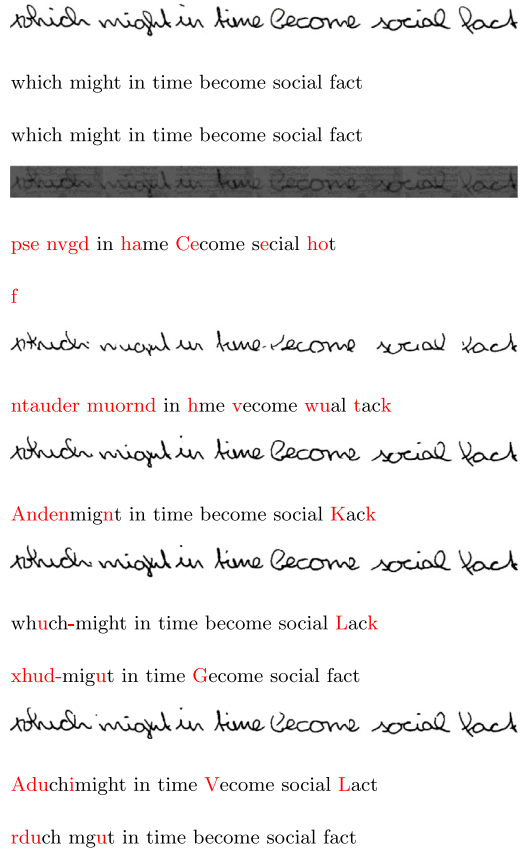

Fig. 10. Results of fixing a degraded handwritten line image. Errors made by the CRNN reading engine are shown in character level with the red color. R (GT): recognition by the CRNN [27] trained on clean images, R (D): recognition by the CRNN [27] trained on degraded images (better viewed in color),R (Generated): recognition by CRNN [27] trained on generated images (S2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by 1.50% compared to [3], 5.46% compared to the basic GAN and 14.29% compared to reading handwritten images in the distorted domain. Moreover, using the recognizer trained in S2 we can even improve the CER result by 4.07%.

Furthermore, we show some qualitative results in Figs. 10, 11 and 12, to visualize the performances of the different methods. Of course, reading the degraded image by a model trained on the GT clean images is not a suitable option. Also, training a model on degraded images is not improving the recognition, especially in the hard scenarios. That is why, enhancing the image and then reading it is the better solution. As it can be seen, our method is better in this practice especially than the baseline cGAN (without a recognizer), because ours is a text conservative method. Hence, it maps the image to a clean but readable domain, while the basic GAN is mapping the image to a visually clean version, without taking the text into consideration (see Figs. 11 and 12).

For the sake of more confirmation and to prove that our model is independent from the used recognizer, we took a different state-of-the-art HTR that is Puigcerver's model [37] trained on the GT images of IAM and KHATT original datasets. Then, we carried a binarization stage to our degraded databases using different methods (including ours) and measure the final recognition performance. As it can be seen from Table 4, our proposed binarization method enhances the performance of the recognizer compared to the use of images binarized by the classic methods [10,11] or the recent cGAN's based one. Also, we can confirm the efficiency of our proposed binarization method compared to the baseline cGAN which did not integrate the text readability information.

4.3.3. H-DIBCO Competitions

After demonstrating the suitability of our proposed method in recovering clean and readable images from highly degraded ones. In what follows, we validate it in H-DIBCO competition on handwritten document binarization, using H-DIBCO 2012 [31], H-DIBCO 2016, DIBCO 2017 [39] and H-DIBCO 2018 [2]. Since our model was designed to enhance line images with a size of 128 × 1024, we bi-

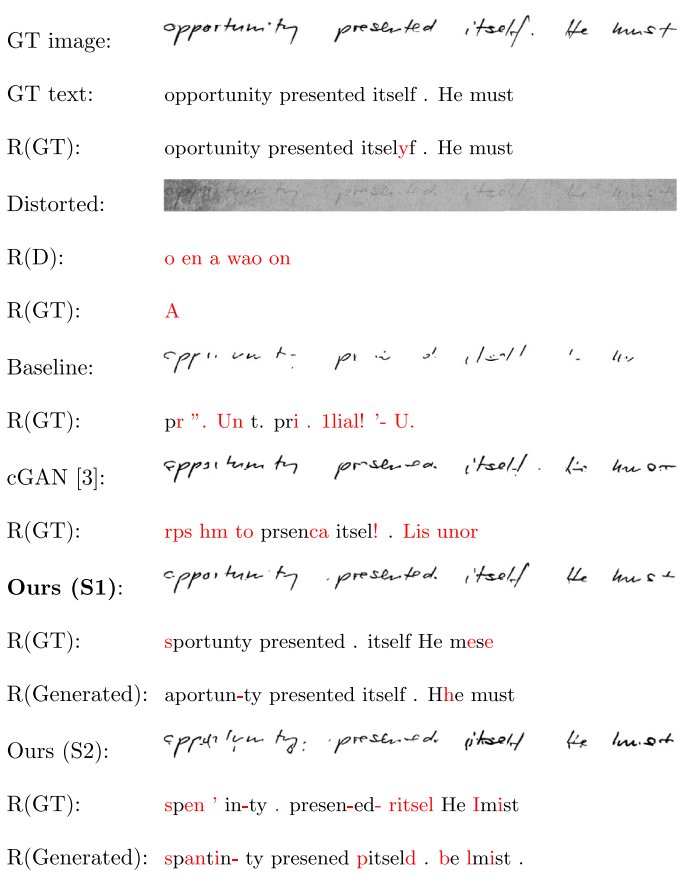

Fig. 11. Results of fixing a highly degraded handwritten line image. Errors made by the CRNN reading engine are shown in character level with the red color. R (GT): recognition by the CRNN [27] trained on clean images, R (D): recognition by the CRNN [27] trained on degraded images, R (Generated): recognition by CRNN [27] trained on generated images (S2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

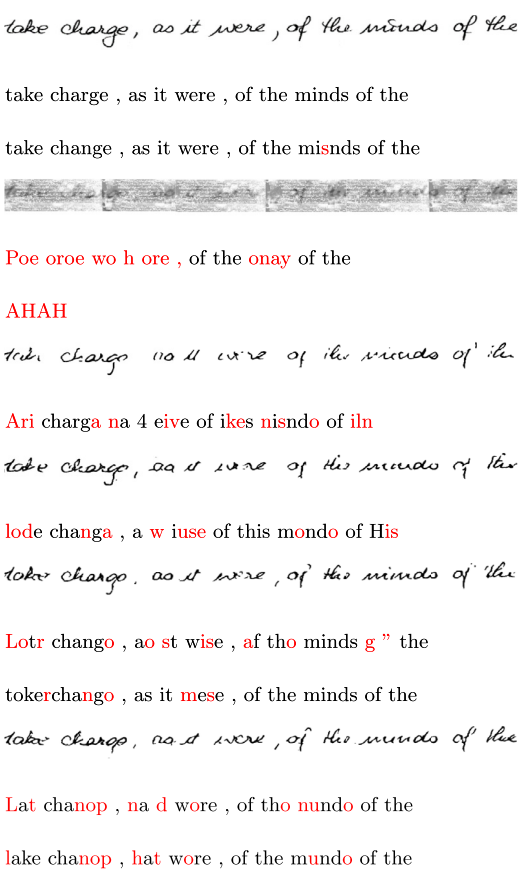

Fig. 12. Results of fixing an extremely degraded handwritten line image. Errors made by the CRNN reading engine are shown in character level with the red color. R (GT): recognition by CRNN [27] trained on clean images, R (D): recognition by the CRNN [27] trained on degraded images, R (Generated): recognition by the CRNN [27] trained on generated images (S2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Impact of the proposed binarization method (scenario S1) on the recognition performance by a HTR system.

Dataset	Binarization Method	CER%	WER%
degraded-KHATT	Otsu [10]	54.28	85.42
	Sauvola [11]	58.42	99.57
	cGAN [3]	28.32	53.96
	Baseline cGAN	28.61	53.73
	Ours (S1)	**26.57**	**52.31**
degraded-IAM	Otsu [10]	62.62	81.96
	Sauvola [11]	72.48	98.00
	cGAN [3]	27.21	58.18
	Baseline cGAN	31.31	61.79
	Ours (S1)	**25.79**	**56.43**

Table 5
Comparative results of our proposed method on *H-DIBCO 2012* Dataset for document binarization. Avg = (PSNR + FM + Fps + (100 - DRD)) / 4.

Method	PSNR	FM	Fps	DRD	Avg
Otsu [10]	15.03	80.18	82.65	26.46	62.85
Sauvola et al. [11]	16.71	82.89	87.95	6.59	70.24
Guo et al. [38]	17.86	86.40	89.00	4.67	72.14
Zhao et al. [8]	21.91	94.96	96.15	1.55	77.86
Competition winner [31]	21.80	89.47	90.18	3.44	74.50
Kang et al. [7]	21.37	95.16	**96.44**	**1.13**	**77.96**
Ours (S1)	16.29	79.25	85.96	7.33	68.54
Ours (S1) + Fine-tuning	**22.00**	**95.18**	94.63	1.62	77.54

narize H-DIBCO images in form of patches having the same dimensions. We compare with the recent state of the art approaches, the winners of the different competitions [31] and the classic binarization methods [10,11]. To clean H-DIBCO images, and since they are formed of Latin script text, we used our pretrained model on the developed degraded-IAM dataset. Two scenarios were investigated: Using the model directly to clean the images, or fine tuning it with a similar distribution before using it. For fine tuning, we used the other DIBCO and H-DIBCO versions [2] and the Palm-Leaf dataset [33]. It is to note also that since the DIBCO datasets are not holding the text information, we removed the recognizer component during fine tuning process, we have frozen also the batch normalisation layers of the generator and we trained the architecture for one only epoch to keep the learned knowledge of the degraded-IAM. During cleaning, we feed our model with the original degraded image in two forms: A normal condition and a vertically flipped version. Thus, we produce two instances of the recovered images. The flipped image is, then, re-flipped again to the normal condition. After that, a voting method is used to produce the final binarized image, by assigning zero to the pixel value (black) if it is indeed black in the two produced images by our model. We found that this led to a better result instead of using just one image condition.

We start our experiments with H-DIBCO 2012, the obtained results are given in Table 5. As it can be seen, our model leads to competitive results to the state of the art approaches, with a superiority in two metrics (PSNR and FM). However, we can see that

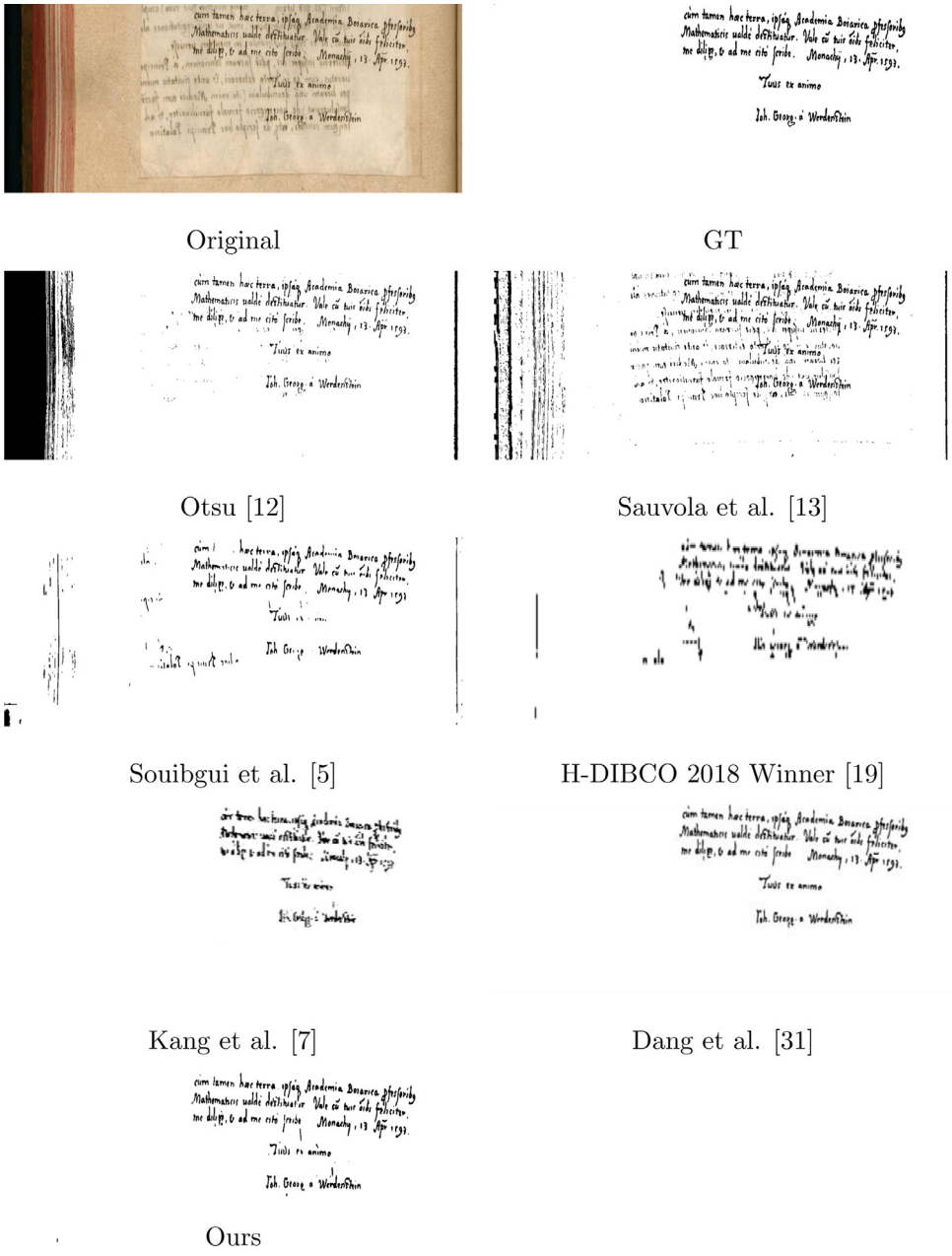

Fig. 13. Results of our method in binarization of some samples from the H-DIBCO 2018 dataset. Images in columns are: Left: original image, Middle: GT image, Right: Binarized image using our proposed method.

the model proposed in [7] is better in term of F_{ps} and DRD. Then, we tested our method on a more recent dataset which is H-DIBCO 2016 [40]. As presented in Table 6, our model gives the state of the art compared to all the methods in the three metrics PSNR, FM and DRD and in the overall average.

Next, we tested with DIBCO 2017, which contains a mix of handwritten and printed degraded documents images. The results in Table 7 show that our model is not superior in this dataset, but it is competitive with the best approaches. We note that our model performance was affected by the type of binarized documents in this dataset, which contain several printed documents, while our model is designed essentially for the handwritten text. Finally, we tested with the most recent H-DIBCO 2018 [2]. The results are shown in Table 8, where we compare with the most recent state of the art results, the winner of the H-DIBCO 2018 competition and the classic binarization methods. The performance of our model is superior than the different approaches in term of PSNR, FM, F_{ps} and average score.

Out of the obtained results in the different datasets, we can say that the classic thresholding methods [10,11] have a moderate performance compared to the recent deep learning approaches. Also, we can notice that if our model using only degraded-IAM for training, does not reach a satisfactory result because there is a domain gap between the training and testing data. However, fine tuning our model with the similar datasets leads to the best performance compared to all the state of the art methods in H-DIBCO 2016 and H-DIBCO 2018. While having a competitive result with the best approach in H-DIBCO 2012 that is [7], where we obtain a superior PSNR and FM scores. We can conclude also that our model is more suitable for binarizing the handwritten images, since it was pre-trained on the developed degraded-IAM dataset before the fine tuning stage.

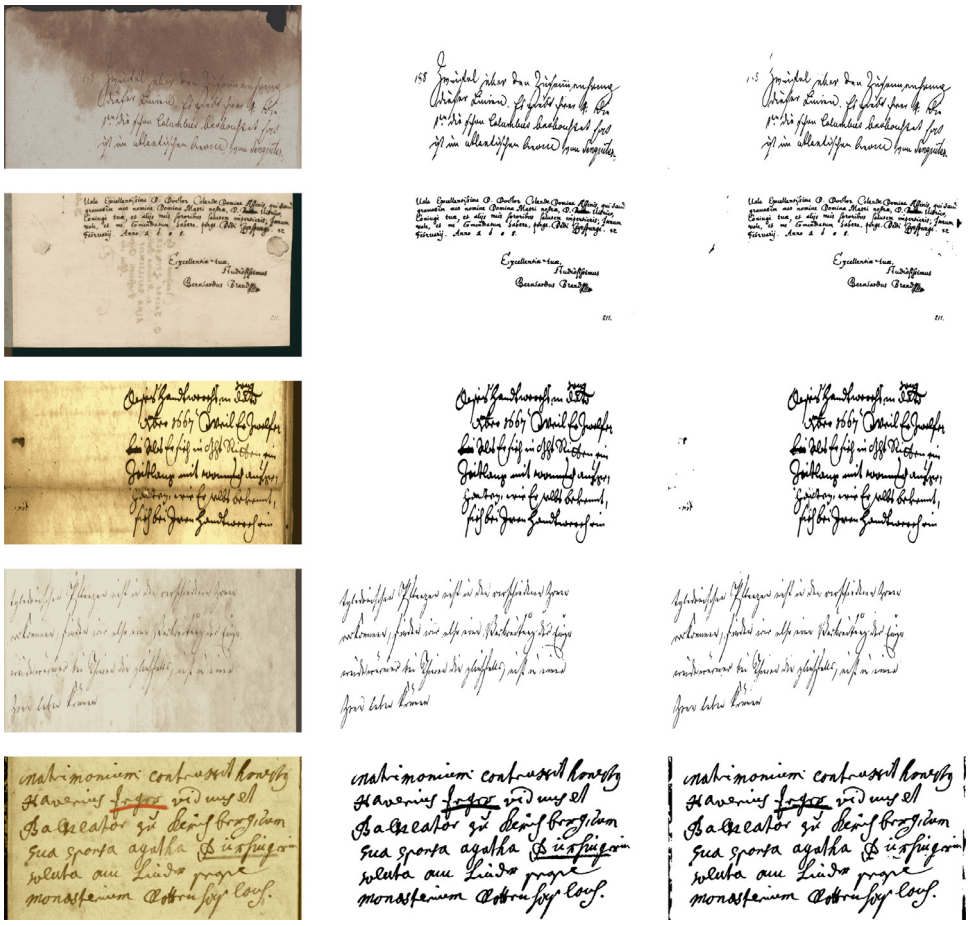

Fig. 14. Results of the different enhancement on the sample 4, from H-DIBCO 2018 Dataset.

Table 6
Comparative results of our proposed method on *H-DIBCO 2016* Dataset for document binarization. Avg = (PSNR + FM + Fps + (100 - DRD)) / 4.

Method	PSNR	FM	Fps	DRD	Avg
Otsu [10]	17.80	86.61	88.67	7.46	71.40
Sauvola et al. [11]	16.42	82.52	86.85	5.56	70.05
Vo et al. [41]	19.01	90.10	93.57	3.58	74.77
Guo et al. [38]	18.42	88.51	90.46	4.13	73.31
He and Schomaker [23]	19.60	91.40	94.30	2.90	75.6
Zhao et al. [8]	19.64	91.66	**94.58**	2.82	75.76
Competition winner [40]	18.11	87.61	91.28	5.21	72.94
Bera et al. [42]	18.94	90.43	91.66	3.51	74.38
Kang et al. [7]	19.18	93.09	94.85	3.03	76.02
Ours (S1)	14.26	69.52	78.01	12.11	62.42
Ours (S1) + Fine-tuning	**21.85**	**94.95**	94.55	**1.56**	**77.44**

Table 7
Comparative results of our proposed method on *DIBCO 2017* Dataset for document binarization. Avg = (PSNR + FM + Fps + (100 - DRD)) / 4.

Method	PSNR	FM	Fps	DRD	Avg
Otsu [10]	13.85	77.73	77.89	15.54	63.48
Sauvola et al. [11]	14.25	77.11	84.1	8.85	66.65
Zhao et al. [8]	17.83	90.73	92.58	3.58	74.39
Competition winner [39]	**18.28**	91.04	92.86	3.40	**74.69**
Kang et al. [7]	15.85	**91.57**	**93.55**	**2.92**	74.51
Bera et al. [42]	15.45	83.38	89.43	6.71	70.38
Ours (S1)	13.54	71.13	80.39	9.60	63.86
Ours (S1) + Fine-tuning	17.45	89.8	89.95	4.03	73.29

Table 8
Results for all methods on *H-DIBCO 2018* Dataset for handwritten document binarization. Avg = (PSNR + FM + Fps + (100 - DRD)) / 4.

Method	PSNR	FM	Fps	DRD	Avg
Otsu [10]	9.74	51.45	53.05	59.07	38.79
Sauvola et al. [11]	13.78	67.81	74.08	17.69	59.50
Adak et al. [2]	14.62	73.45	75.94	26.24	59.44
Souibgui et al. [5]	16.16	77.59	85.74	7.93	67.89
Tamrin et al. [6]	17.04	83.08	88.46	5.09	70.87
Zhao et al. [8]	18.37	87.73	90.6	4.58	73.03
Competition winner [2]	19.11	88.34	90.24	4.92	73.19
Akbari et al. [20]	19.17	89.05	93.65	4.80	74.26
Kang et al. [7]	19.39	89.71	91.62	**2.51**	74.55
Dang et al. [25]	19.81	91.26	93.97	3.42	75.40
Bera et al. [42]	15.31	76.84	83.58	9.58	66.53
Ours (S1)	13.88	65.06	73.46	12.86	59.89
Ours (S1) + Fine-tuning	**20.18**	**92.41**	**94.35**	2.60	**76.08**

Finally, we show some qualitative results about the binarization performance in Fig. 14 that demonstrates our method superiority compared to the other ones in this task. Also, we provide the binarization result of other images from the H-DIBCO 2018 dataset in Figure 13 where we obtained images that are very close to the GT. Moreover, Fig. 15 shows an example where our method can even complete the missing pixels (that are not existent in the GT image), to provide a more readable text.

4.3.4. Dataset selection for the fine-tuning stage

Selecting the right dataset for fine-tuning will improve the binarization performance. Thus, in this section, we study the impact of the different datasets on the binarization process carried out on the H-DIBCO 2016. For all our experiments, we tested using

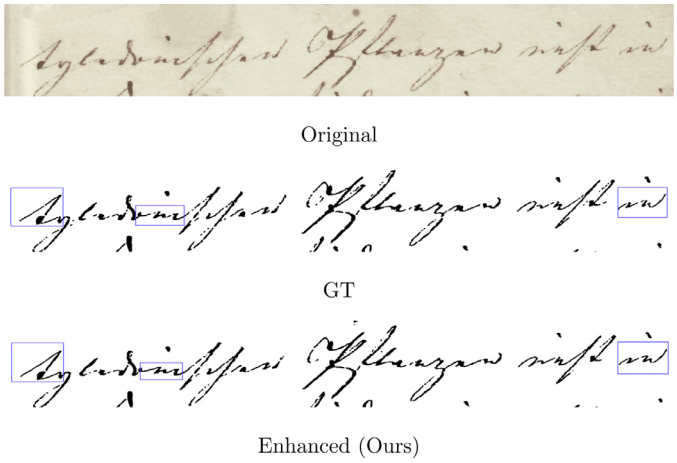

Fig. 15. Qualitative results of our proposed method evaluated on an a part taken from a sample from H-DIBCO 2018 Dataset (Pixels restoration).

Table 9
Impact of the fine-tuning data selection on the binarization performance on H-DIBCO **2016** Dataset.

H-DIBCO	DIBCO	Nabuco	Bickley-diary	Palm-Leaf	PNSR
					14.26
√					18.25
	√				18.08
√	√				18.10
√	√	√			16.89
√	√		√		17.98
√	√			√	**21.85**
		√	√	√	14.78
√	√	√	√	√	17.63

the other variations of DIBCO and H-DIBCO (from 2009 to 2018), except DIBCO 2019 since it has different distributions in term of degradation and document types. Also, we tested including similar datasets which were developed for binarization task, namely, Palm-Leaf [33], Nabuco [43] and Bickley-diary [44]. Images contained in these datasets suffer from different kinds of degradation, such as water stains, ink bleed-through, and significant foreground text intensity. As it can be seen from Table 9, using our model trained only on degraded-IAM leads to poor results, thus, a fine-tuning stage is necessary. Using the H-DIBCO images for fine-tuning improves the performance with a slight superiority than using the DIBCO ones. This can be explained by the type of text, because our model is pretrained to binarize the handwritten text. However, using H-DIBCO and DIBCO at the same time is a better option. Because, DIBCO contains a useful types of degradation that can be learned to be cleaned by our model even with a printed text. Also, adding other datasets is sometimes useful, but in other times deteriorates the performance. This can be noticed when adding the Palm-Leaf dataset which improves the binarization, while adding the Nabuco or the Bickley-diary is leading the model to learn a non suitable parameters for H-DIBCO 2016. This can be justified by the similar domain (degradation and text types) between H-DIBCO 2016 and Palm-Leaf distributions, while it is different with the other datasets.

5. Conclusion

In this paper, we proposed an architecture for handwritten document binarization based on GANs. Our method recovers the degraded images while conserving its readability by integrating a HTR to evaluate the enhanced image in addition to the discriminator. To the best of our knowledge, this is the first approach that includes textual information when performing the recovery process of handwritten documents. Experimental results proved the effectiveness of the proposed model in cleaning extremely degraded documents. We proved also that training a HTR model progressively on the images binarized by the generator at each iteration leads to a better performance in CER and WER. We used in this paper a CRNN as a recognizer, but it is worth to note that by using other HTR architectures we may obtain a better recognition performances, since our method is flexible to integrate different ones. Moreover, we obtain the best performance compared to the state of the art in H-DIBCO benchmarks of degraded documents binarization.

Training supervised approaches using paired data (degraded image with its clean version and GT text) is costly, since it is really hard to obtain this kind of resources. That is why most of the datasets are synthetically made. As a future work, we intend to explore the use of unpaired data. In this way, the real images can be easily obtained (degraded images, and clean ones). Thus, our model could be trained to produce images that are for the discriminator as much real as possible (similar to the clean real documents) and for the recognizer as much readable and real as possible (the recognized text must have sense).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been partially supported by the Swedish Research Council (grant 2018-06074, DECRYPT), the Spanish project RTI2018-095645-B-C21, the Ramon y Cajal Fellowship RYC-2014-16831 and the CERCA Program/Generalitat de Catalunya.

References

[1] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep Learning, vol. 1, MIT press Cambridge, 2016.
[2] I. Pratikakis, K. Zagori, P. Kaddas, B. Gatos, ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018), in: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018, pp. 489–493, doi:10.1109/ICFHR-2018.2018.00091.
[3] M.A. Souibgui, Y. Kessentini, A. Fornés, A conditional GAN based approach for distorted camera captured documents recovery, in: Mediterranean Conference on Pattern Recognition and Artificial Intelligence, Springer, 2020.
[4] B. Wang, C.L.P. Chen, An effective background estimation method for shadows removal of document images, in: 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 3611–3615, doi:10.1109/ICIP.2019.8803486.
[5] M.A. Souibgui, Y. Kessentini, DE-GAN: a conditional generative adversarial network for document enhancement, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[6] M.O. Tamrin, M. El-Amine Ech-Cherif, M. Cheriet, A two-stage unsupervised deep learning framework for degradation removal in ancient documents, in: Pattern Recognition. ICPR International Workshops and Challenges, Springer International Publishing, 2021, pp. 292–303.
[7] S. Kang, B.K. Iwana, S. Uchida, Complex image processing with less data document image binarization by integrating multiple pretrained U-Net modules, Pattern Recognit. 109 (2021) 107577.
[8] J. Zhao, C. Shi, F. Jia, Y. Wang, B. Xiao, Document image binarization with cascaded generators of conditional generative adversarial networks, Pattern Recognit. 96 (2019) 106968.
[9] S.K. Jemni, Y. Kessentini, S. Kanoun, Out of vocabulary word detection and recovery in arabic handwritten text recognition, Pattern Recognit. 93 (2019) 507–520.
[10] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[11] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, Pattern Recognit. 33 (2) (2000) 225–236.
[12] M. Annabestani, M. Saadatmand-Tarzjan, A new threshold selection method based on fuzzy expert systems for separating text from the background of document images, Iran. J. Sci. Technol.Trans. Electr. Eng. 43 (1) (2019) 219–231.
[13] W. Xiong, J. Xu, Z. Xiong, J. Wang, M. Liu, Degraded historical document image binarization using local features and support vector machine (SVM), Optik 164 (2018) 218–223.

[14] R. Hedjam, M. Cheriet, M. Kalacska, Constrained energy maximization and self-referencing method for invisible ink detection from multispectral historical document images, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 3026–3031.
[15] W. Xiong, X. Jia, J. Xu, Z. Xiong, M. Liu, J. Wang, Historical document image binarization using background estimation and energy minimization, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 3716–3721.
[16] M.Z. Afzal, J. Pastor-Pellicer, F. Shafait, T.M. Breuel, A. Dengel, M. Liwicki, Document image binarization using LSTM: a sequence learning approach, in: Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing, 2015, pp. 79–84.
[17] X.-J. Mao, C. Shen, Y.-B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2810–2818.
[18] K.G. Lore, A. Akintayo, S. Sarkar, LLNet: a deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61 (2017) 650–662.
[19] J. Calvo-Zaragoza, A.-J. Gallego, A selectional auto-encoder approach for document image binarization, Pattern Recognit. 86 (2019) 37–47.
[20] Y. Akbari, S. Al-Maadeed, K. Adam, Binarization of degraded document images using convolutional neural networks and wavelet-based multichannel images, IEEE Access 8 (2020) 153517–153534, doi:10.1109/ACCESS.2020.3017783.
[21] M. Hradiš, J. Kotera, P. Zemcık, F. Šroubek, Convolutional neural networks for direct text deblurring, in: Proceedings of BMVC, vol. 10, 2015.
[22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[23] H. Sheng, S. Lambert, DeepOtsu: document enhancement and binarization using iterative deep learning, Pattern Recognit. 91 (2019) 379–390, doi:10.1016/j.patcog.2019.01.025.
[24] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[25] Q.-V. Dang, G.-S. Lee, Document image binarization with stroke boundary feature guided network, IEEE Access 9 (2021) 36924–36936, doi:10.1109/ACCESS.2021.3062904.
[26] A.K. Bhunia, A.K. Bhunia, A. Sain, P.P. Roy, Improving document binarization via adversarial noise-texture augmentation, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, 2019, pp. 2721–2725.
[27] A.F.S. Neto, B.L.D. Bezerra, A.H. Toselli, E.B. Lima, HTR-Flor: a deep learning system for offline handwritten text recognition, in: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020, pp. 54–61, doi:10.1109/SIBGRAPI51738.2020.00016.
[28] T. Bluche, R. Messina, Gated convolutional recurrent neural networks for multilingual handwriting recognition, in: Proceeding of International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2017, pp. 646–651.
[29] S.K. Jemni, Y. Kessentini, S. Kanoun, J.-M. Ogier, Offline arabic handwriting recognition using BLSTMs combination, in: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), IEEE, 2018, pp. 31–36.
[30] L. Kang, P. Riba, M. Villegas, A. Fornés, M. Rusiñol, Candidate fusion: integrating language modelling into a sequence-to-sequence handwritten word recognition architecture, Pattern Recognit. 112 (2021) 107790.
[31] I. Pratikakis, B. Gatos, K. Ntirogiannis, ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012), in: Proceedings of the International Conference on Frontiers in Handwriting Recognition, IEEE, 2012, pp. 817–822.
[32] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, in: Soviet Physics Doklady, vol. 10, Soviet Union, 1966, pp. 707–710.
[33] J.-C. Burie, M. Coustaty, S. Hadi, M.W.A. Kesiman, J.-M. Ogier, E. Paulus, K. Sok, I.M.G. Sunarya, D. Valy, ICFHR2016 competition on the analysis of handwritten text in images of balinese palm leaf manuscripts, in: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2016, pp. 596–601.
[34] S.M. Ayatollahi, H.Z. Nafchi, Persian heritage image binarization competition (PHIBC 2012), in: 2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA), IEEE, 2013, pp. 1–4.
[35] S.A. Mahmoud, I. Ahmad, W.G. Al-Khatib, M. Alshayeb, M.T. Parvez, V. Märgner, G.A. Fink, KHATT: an open arabic offline handwritten text database, Pattern Recognit. 47 (3) (2014) 1096–1112.
[36] U.-V. Marti, H. Bunke, The IAM-database: an english sentence database for offline handwriting recognition, Int. J. Doc. Anal. Recogn. 5 (1) (2002) 39–46.
[37] J. Puigcerver, Are multidimensional recurrent layers really necessary for handwritten text recognition? in: 2017 IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2017, pp. 67–72.
[38] J. Guo, C. He, X. Zhang, Nonlinear edge-preserving diffusion with adaptive source for document images binarization, Appl. Math. and Comput. 351 (2019) 8–22, doi:10.1016/j.amc.2019.01.021.
[39] I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, ICDAR 2017 competition on document image binarization (DIBCO 2017), in: 2017 International Conference on Document Analysis and Recognition, IEEE, 2017, pp. 1395–1403, doi:10.1109/icdar.2017.228.
[40] I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, ICFHR 2016 handwritten document image binarization contest (H-DIBCO 2016), in: 2016 International Conference on Frontiers in Handwriting Recognition, IEEE, 2016, pp. 619–623, doi:10.1109/icfhr.2016.0118.
[41] Q.N. Vo, S.H. Kim, H.J. Yang, G. Lee, Binarization of degraded document images based on hierarchical deep supervised network, Pattern Recognit. 74 (2018) 568–586, doi:10.1016/j.patcog.2017.08.025.
[42] S.K. Bera, S. Ghosh, S. Bhowmik, et al., A non-parametric binarization method based on ensemble of clustering algorithms, Multimed. Tools Appl. 80 (2021) 76537673, doi:10.1007/s11042-020-09836-z.
[43] R.D. Lins, Nabuco-two decades of processing historical documents in Latin America, J. Univers. Comput. Sci. 17 (2011) 151161.
[44] A.F. Bickley, Bickley diary dataset, Private Journal (unpublished), donated to the Methodist Church Archives, Singapore by Erin Bickley(1926).

Dr. Sana Khamekhem Jemni: graduated in Computer Science engineering from the National Engineering School of Sfax (ENIS) in 2009 and received her Ph.D. degree in the field of Pattern Recognition from the University of Sfax, Tunisia in 2020. She was a postdoctoral researcher at CRNS in 2020. Her current research interests include handwritten Document Image Analysis and Recognition, Arabic and Latin optical character recognition, information extraction, data fusion and Deep Learning. She is the author and co-author of several papers and she has been a reviewer for some journals.

Mr. Mohamed Ali Souibgui: Received his bachelor and master degrees in computer science from the University of Monastir, Tunisia, in 2015 and 2018, respectively. He is currently working towards the PhD degree in the Computer Vision Center, Autonomous University of Barcelona, Spain. His research interests include machine learning, computer vision and document analysis.

Dr. Yousri Kessentini: graduated in Computer Science engineering from the National Engineering School of Sfax (ENIS) in 2003 and received his Ph.D. degree in the field of pattern recognition from the University of Rouen, France in 2009. He was a postdoctoral researcher at ITESOFT company and LITIS laboratory from 2011 to 2013. Currently, he is Assistant Professor at CRNS and the head of the DeepVision research team. His main research areas concern deep learning, document processing and recognition, data fusion, and computer vision. He is certified as an official instructor and ambassador from the NVIDIA Deep Learning Institute. He has coordinated several research projects in partnership with industry.

Dr. Alicia Fornes: received the Ph.D. degree in 2009 from the Universitat Autonoma de Barcelona (UAB). Her Ph.D. work on writer identification of old music scores received the best thesis award 2009.2010 by the AERFAI (Spanish Branch of the IAPR). She received the IAPR/ICDAR Young Investigator Award in 2017 for outstanding contributions in the recognition of handwriting, text and graphics, with high impact to the field of Digital Humanities. She is currently a Senior Research Fellow at the Computer Vision Center (CVC) and the UAB.