# DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement

Mohamed Ali Souibgui [ID] and Yousri Kessentini [ID]

**Abstract**—Documents often exhibit various forms of degradation, which make it hard to be read and substantially deteriorate the performance of an OCR system. In this paper, we propose an effective end-to-end framework named document enhancement generative adversarial networks (DE-GAN) that uses the conditional GANs (cGANs) to restore severely degraded document images. To the best of our knowledge, this practice has not been studied within the context of generative adversarial deep networks. We demonstrate that, in different tasks (document clean up, binarization, deblurring and watermark removal), DE-GAN can produce an enhanced version of the degraded document with a high quality. In addition, our approach provides consistent improvements compared to state-of-the-art methods over the widely used DIBCO 2013, DIBCO 2017, and H-DIBCO 2018 datasets, proving its ability to restore a degraded document image to its ideal condition. The obtained results on a wide variety of degradation reveal the flexibility of the proposed model to be exploited in other document enhancement problems.

**Index Terms**—Document analysis, document enhancement, degraded document binarization, watermark removal, deep learning, generative adversarial networks

✦

## 1 INTRODUCTION

AUTOMATIC document processing consists in the transformation into a form that is comprehensible by a computer vision system or by a human. Thanks to the development of several public databases, document processing has made a great progress in recent years [1], [2]. However, this processing is not always effective when documents are degraded. Lot of damages could be done to a document paper. For example: Wrinkles, dust, coffee or food stains, faded sun spots and lot of real-life scenarios [3]. Degradation could be presented also in the scanned documents because of the bad conditions of digitization like using the smart-phones cameras (shadow [4], blur [5], variation of light conditions, distortion, etc.). Moreover, some documents could contain watermarks, stamps or annotations. The recovery is even harder when certain types of these later take the text place for instance in cases where the stains color is the same or darker than the document font color (Fig. 1 shows some examples). Hence, an approach to recover a clean version of the degraded document is needed.

In this study, we are focusing on two document enhancement problems. Degraded documents recovery, i.e., to produce a clean (grayscale or binary) version of the document given any type of degradation, and watermark removal. The faced obstacles are as follows: Overlaps of noise or watermarks with the text, dense watermarks, intense dirt or degradation can cover the entire text and reading it becomes very hard, there is no prior knowledge about the degradation or the watermark that should be removed, etc. An ideal system should be good in performing two tasks simultaneously, removing the noise and the watermarks as well as retaining the text quality in the document images.

Recently, a great success is made by deep neural networks in natural images generation and restoration, especially deep convolutional neural networks (auto-encoders and variational auto-encoders (VAE)) [6], [7], [8] and generative adversarial networks (GANs) [9], [10]. GANs, which were introduced in [11], are now considered as the ideal solution for image generation problems [10] with a high quality, stability and variation compared to the auto-encoders. Generative models gained more attention because of the ability of capturing high dimensional probability distributions, imputation of missing data and dealing with multimodal outputs. Despite that, document analysis research community is not benefiting enough from those approaches, yet. Using them is very limited, for instance, in font translation [12], handwritten profiling [13] and staff-line removal from a music score images [14], where a promising results were found.

In [9], Isola *et al.* show that conditional generative adversarial networks (cGANs), a variation of GANs, performs good in image-to-image translation (labels to facade, day to night, edges to photo, BW to color, etc.). While GANs learn a generative model of data, conditional GANs (cGANs) learn a conditional generative model, where it conditions on an input image and generate a corresponding output image. Since document enhancement follows the same process, means, we want to preserve the text and remove the damage in a conditioned image, cGANs shall be the suitable solution, and this is what motivated us for this study.

---

- *Mohamed Ali Souibgui is with the Computer Vision Center, and Computer Science Department, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. E-mail: msouibgui@cvc.uab.es.*
- *Yousri Kessentini is with the Digital Research Center of Sfax, Sfax 3021, Tunisia , and also with the MIRACL Laboratory, University of Sfax, Sfax 3029, Tunisia. E-mail: yousri.kessentini@crns.rnrt.tn.*
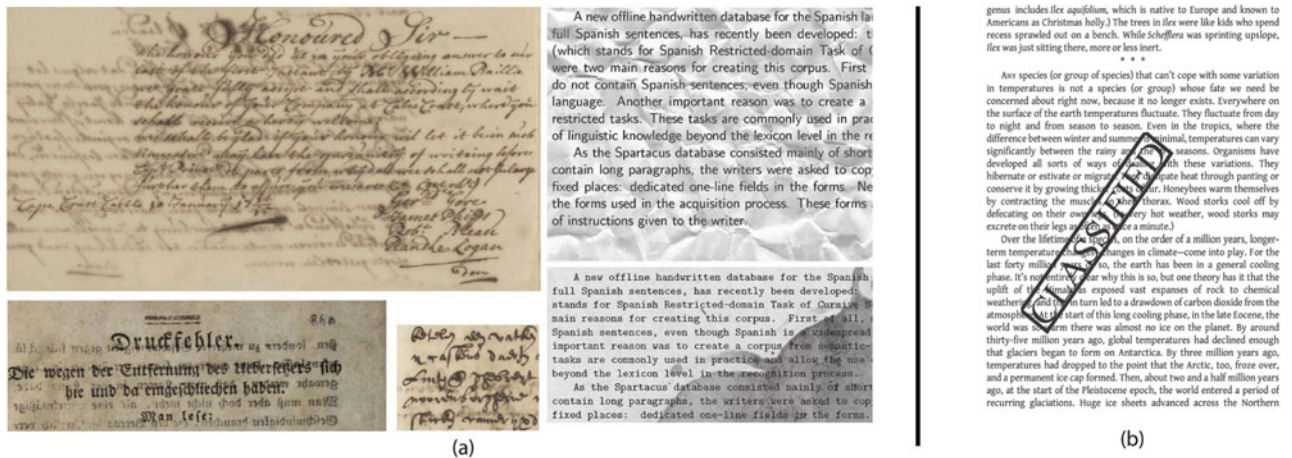
Fig. 1. Examples of the documents used in this study: (a): Degraded documents, (b): A document with dense watermark.

The main contributions of this paper are: As primary, to the best of our knowledge, this is the first occurrence of GANs, conditional GANs specifically, in a framework that addresses different document enhancement problems (clean up, binarization and watermark removal). Second, we used a simple but flexible architecture that could be exploited to tackle any document degradation problem. Third, we introduce a new document enhancement problem consists in dense watermarks and stamps removal. Finally, we experimentally prove that our approach achieves a higher performance compared to the state-of-the-art methods in degraded document binarization.

The rest of the paper is organized as follows. In Section 2, a summary of previous works on document enhancement, especially for document clean up and binarization and watermark removal in documents as well as in natural images. We review also some related works using the GANs in image-to-image translation. Then, we provide our proposed approach in Section 3. Some experimental results and comparison with traditional and recent methods are described in Section 4. Finally, a conclusion with some future research directions are presented in Section 5.

## 2 RELATED WORKS

In this work, we focus on the enhancement of degraded document images by addressing different kinds of degradation which are document clean up, binarization, and watermark removal. From a document analyst viewpoint, this recovery of a clean version from the degraded document falls in the research field called document enhancement. Where we can find, in addition to those three, many other ways to enhance a document. For instance: unshadowing [4], [15], super-resolution [16], deblurring [5], dewarping [17], etc. In what follows, we cover some related works to our addressed problems.

### 2.1 Degraded Document Recovering and Binarization

Dirty document cleaning is related to document image binarization, especially the degraded ones. Where the goal is to produce a binary but clean document. That is why, dealing with these two problems was almost the same. The idea is to classify the pixels of the document as one of two categories: degradation or text. Afterward, assigning zeros to the text pixels and ones for the degradation will generate a binary clean image. While generating a gray-scale or colored image is done by preserving the same value for the text pixels.

Classic document binarization methods [18], [19], [20], [21], [22], [23] were based on thresholding, many algorithms were developed for the goal of finding the suitable global or local threshold(s) to apply as a filter. According to the threshold(s), pixels are classified to be belonging on the text (zero) or the degradation (one). Lelore *et al.* [24] presented an algorithm called FAIR, based on edge detection to localize the text in a degraded document image. A global threshold selection method was proposed in [25], basing on fuzzy expert systems (FESs), the image contrast is enhanced. Then, the range of the threshold value is adjusted by another FES and a pixel-counting algorithm. Finally, the threshold value is obtained as the middle value of that range. A machine learning based approach was proposed in [26], the goal was the determination of the binarization threshold in each image region given a three-dimensional feature vector formed by the distribution of the gray level pixel values. The support vector machine (SVM) was used to classify each region into one of four different threshold values. An other and similar SVM based approach was introduced in [27]. The main drawbacks of these classic methods is that the results are highly sensitive to document condition. With a complex image background or a non uniform intensity, problems occurred.

Later, evolved techniques were proposed. Moghaddam *et al.* [28] proposed a variational model to remove the bleed-through from the degraded double-sided document images. For the cases where the verso side of the document is not available, a modified variational model is also introduced. By transferring the model to the wavelet domain and using the hard wavelet shrinkage, the interference patterns was removed. Other energy based methods were also introduced. In [29], authors considered the ink as a target and tried to detect it by maximizing an energy function. This technique was applied also for scene text binarization [30], which is a similar task. Similarly, Xiong *et al.* [31] estimated the background and subtracted it from the image by a mathematical morphology. Then the Laplacian energy based

segmentation is performed on the enhanced document image to classify the pixels. Although these sophisticated image processing techniques, document binarization results are still unsatisfactory. For this reason, some deep learning frameworks were recently used to tackle this problem. The goal here is not to train a model for predicting a threshold, it is to directly separate the foreground text from the background noise given, of course, a considerable amount of paired data (degraded images and their binarized versions). These deep learning based models lead to better results compared to the other hand-crafted methods. Several End-to-end frameworks, based on fully convolutional neural-network (encoder-decoder way), was used to binarize and enhance the document image [32], [33], [34]. Afzal *et al.* [35] formulated the binarization of document images as a sequence learning problem. Hence, the 2D Long Short-Term Memory (LSTM) was employed to process a 2D sequence of pixels for the classification of each pixel as text or background. In [36], an other Fully convolutional network was trained with a combined Pseudo F-measure and F-measure loss function for the task of document image binarization. A method that inspires from the two previous approaches, i.e., a recurrent neural network based algorithm using grid LSTM cells for image binarization, as well as a pseudo F-Measure based weighted loss function could be found in [37]. Vo *et al.* [38] proposed a hierarchical deep supervised network (DSN) architecture to predict the text pixels. They claimed that their architecture incorporates side layers to decrease the learning time, while taking a lot of training data.

It is to note that in this paper our object is not to binarize the document images but to clean the degraded ones and preserve them in their basic grey or colored level. But, we will test our approach in this problem for the purpose of comparison with the state-of-the art approaches.

## 2.2 Watermark Removal

Watermark removal is also related to classical document binarization or image matting, where the goal is to decompose a single image into background and foreground knowing that this time the text is in the background while the watermark is in the foreground. But, this problem was not proposed in document processing. In fact, the appeared works that deal with watermark removal was in natural images processing. In [39], authors used an image inpainting algorithms to remove the watermark. Before that, a statistical method was used to detect the watermark region. Dekel *et al.* [40] proposed to estimate the outline sketch and alpha matte of the same watermark from a batch of different images. Two watermarks were used in this study, the goal was testing the effectiveness of a single visible watermark to protect a large set of images. Wu *et al.* [41] used the generative adversarial networks [11] to remove watermark from faces images used in a biometric system. Cheng *et al.* [42] proposed a method based on convolutional neural networks (CNN). First, object detection algorithms were used to detect the watermark region in natural images and then pass it to an other model to remove the watermark. In our study, we investigate for the first time the problem of watermark removal in document images, this leads us to compare our approach with some results obtained on natural images for the same purpose.

## 2.3 Generative Adversarial Networks for Image-to-Image Transform

As mentioned above, GANs are now achieving impressing results in image generation and translation. In this paragraph, we investigate the use of this mechanism in related problems to document processing and enhancement. This shall gives intuitions to document analysis community about exploiting GANs for these tasks. In [43], it was demonstrated that GANs lead to improvements in semantic segmentation. Ledig *et al.* presented SRGAN [44], a Generative Adversarial Network for image Super-Resolution. Through it, they achieved a photo-realistic reconstructions for large upscaling factors ($4\times$). In [44], conditional GANs were used for several image-to-image translation tasks (these tasks are related to document enhancement), given a paired data. This work was extended to [45], where CycleGAN, a GAN that uses impaired data, was proposed as a solution. An other model called "pix2pix-HD" and deals with high-resolution (e.g., 2048x1024) photorealistic image-to-image translation tasks was appeared in [46]. Furthermore, an unsupervised method for image-to-image translation was proposed in [47], where authors train two GANs, or "DualGAN" as they denoted. In their architecture, the primal GAN learns to translate images from a domain $U$ to a domain $V$, while the dual GAN learns to invert the task. The closed loop architecture allows images from each domain to be translated and then reconstructed. Hence, a loss function that accounts for the reconstruction error of images can be used to train the translators.

## 3 PROPOSED APPROACH

We consider the problems of document enhancement as an image to image translation task where the goal is to *generate* clean document images given the degraded ones. Since GANs have outperformed auto-encoders in generating high fidelity samples and while we are using paired data, we propose to use a cGAN. We called our model *DE-GAN* (for Document Enhancement conditional Generative Adversarial Network). GANs were initially proposed in [47] and consist in two neural networks, a generator $G$ and a discriminator $D$ characterized by the parameters $\varphi_G$ and $\varphi_D$, respectively. The generator have the goal of learning a mapping from a random noise vector $z$ to an image $y$, $G_{\varphi_G}: z \rightarrow y$. While the discriminator has the function of distinguishing between the image generated by $G$ and the ground truth one. Hence, given $y$, $D$ should be able to tell if it is *fake* or *real* by outputing a probability value, $D_{\varphi_D}: y \rightarrow P(real)$. Those two networks compete against each other in a min-max game, in other words if one wins the other loses. The generator aims to cheat the discriminator by producing a close image to the ground truth, however, the discriminator will improve his prediction of the image being fake, and this is what is called the adversarial learning. cGANs follow the same process, except that, they introduced an additional parameter $x$. Which is the conditioned image. Here, the generator is learning the mapping from an observed image $x$ and a random noise vector $z$, to $y$, $G_{\varphi_G}: \{x, z\} \rightarrow y$ and the discriminator is looking, also, to the conditioned image which makes his process as: $D_{\varphi_D}: \{x, y\} \rightarrow P(real)$.

In our situation, the generator will generate a clean image denoted by $I^C$ given the degraded (or watermarked) one
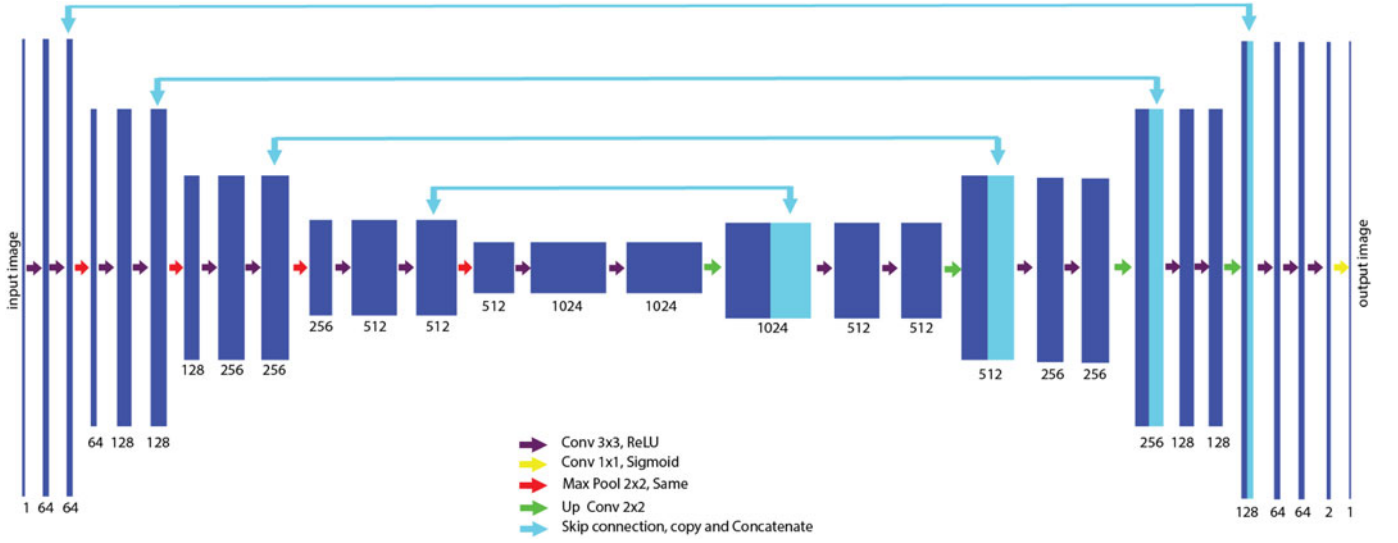
Fig. 2. The generator follows the U-net architecture [48]. Each box corresponds to a feature map. The number of channels is denoted on bottom of the box. The arrows denote the different operations.

which we will denote $I^W$. The generator aims, of course, to produce an image that is very close to the ground truth image denoted by $I^{GT}$. The training of cGANs for this task is done by the following adversarial loss function:

$$L_{GAN}(\varphi_G, \varphi_D) = \mathbb{E}_{I^W, I^{GT}} \log \left[ D_{\varphi_D}(I^W, I^{GT}) \right]$$
$$+ \mathbb{E}_{I^W} \log \left[ 1 - D_{\varphi_D}(I^W, G_{\varphi_G}(I^W)) \right]. \tag{1}$$

Using this function, the generator should produce, after several epochs, a similar image to the ground truth, i.e., the watermark and the degradation will be removed and this may fool the discriminator. But, it is not guaranteed that the text will be preserved in a good condition. To overcome this, we employ an additional log loss function between the generated image and the ground truth, for the purpose of forcing the model to generate images that have the same text as the ground truth. It is to note also that this additional loss boosts the training speed, the added function is

$$L_{log}(\varphi_G) = \mathbb{E}_{I^{GT}, I^W} [-(I^{GT} \log(G_{\varphi_G}(I^W))$$
$$+ ((1 - I^{GT}) \log(1 - G_{\varphi_G}(I^W)))]. \tag{2}$$

Thus, the proposed loss of our network, denoted by $L_{net}$ becomes:

$$L_{net}(\varphi_G, \varphi_D) = min_{\varphi_G} max_{\varphi_D} L_{GAN}(\varphi_G, \varphi_D) + \lambda L_{log}(\varphi_G). \tag{3}$$

Where, $\lambda$ is a hyper-parameter that was set to 100 for text cleaning and 500 in watermark removal and document binarization. The architecture of generator and discriminator networks are described in the next sections.

### 3.1 Generator

The generator is performing an image-to-image translation task. Usually, auto-encoder models are used for this problem [49], [50], [51]. These models consist, mostly, in a sequence of convolutional layers called encoder which perform down-sampling until a particular layer. Then, the process is reversed

to a sequence of up-sampling and convolutional layers called decoder. There are two disadvantages of using an encoder-decoder model for the proposed problem: First, due to down-sampling (pooling), lot of information is lost and the model will have difficulties to recover them later when predicting an image with the same size as the input. Second, image information flow pass through all the layers, including the bottleneck. Thus, sometimes, a huge amount of unwanted redundant features (inputs and outputs are sharing a lot of same pixels) are exchanged. Which leads to energy and time loss. For this reason, we employ skip connections following the structure of a the model called U-net [48]. Skip connections are added every two layers to recover images with less deterioration, it is to note also that skip connections are used when training a very deep model to prevent the gradient vanishing and exploding problems. Some batch normalization layers are also added to accelerate the training. The architecture of the generator used in this study is illustrated in Fig. 2, it is similar to [48] where it was introduced for the purpose of biomedical image segmentation.

### 3.2 Discriminator

The defined discriminator is a simple Fully Convolutional Network (FCN), composed of 6 convolutional layers, that output a 2D matrix containing probabilities of the generated image being real. This model is presented in Fig. 3. As shown, the discriminator receives two input images which are the degraded image and its clean version (ground truth or cleaned by the generator). Those images were concatenated together in a $256 \times 256 \times 2$ shape tensor. Then, the obtained volume propagated in the model to end up in a $16 \times 16 \times 1$ matrix in the last layer. This matrix contains probabilities that should be, to the discriminator, close to 1 if the clean image represents the ground truth. If it is generated by the generator the probabilities should be close to 0. Therefore, the last layer takes a sigmoid as an activation function. After completing training, this discriminator is no longer used. Given a degraded image, we only use the generative network to enhance it. But, this discriminator shall force the generator during training to produce better results.

Fig. 3. The discriminator architecture.



Fig. 4. The proposed DE-GAN.

TABLE 1
The Obtained Results of Document Cleaning
Using Noisy Office Database [3]

| Model | SSIM | PSNR |
|---|---|---|
| FCN (U-net) | 0.9970 | 36.02 |
| **DE-GAN** | **0.9986** | **38.12** |

## 3.3 Training Process

Training our DE-GAN was as follows, we took patches from the degraded images of size $256 \times 256$ and fed it as an input to the generator. The produced images are fed to the discriminator with the ground truth patches and the degraded ones. Then, as presented in Equation (3), the discriminator starts forcing the generator to produce outputs that cannot be distinguished from "real" images, while doing his best at detecting the generator's "fakes". This training is illustrated in Fig. 4 and it is done using Adam with a learning rate of $1e^{-4}$ as an optimizer.

## 4 EXPERIMENTS AND RESULTS

### 4.1 Document Cleaning and Binarization

We begin our experiments with document cleaning. For this task, the Noisy Office Database which contains different types of degradation, and presented in [3], is used. We defined 112



Fig. 5. Cleaning degraded documents by DE-GAN.

TABLE 2
Results of Image Binarization on DIBCO 2013 Database

| Model | PSNR | F-measure | $F_{ps}$ | DRD |
|---|---|---|---|---|
| Otsu [18] | 16.6 | 83.9 | 86.5 | 11.0 |
| Niblack [20] | 13.6 | 72.8 | 72.2 | 13.6 |
| Sauvola et al.[19] | 16.9 | 85.0 | 89.8 | 7.6 |
| Gatos et al. [59] | 17.1 | 83.4 | 87.0 | 9.5 |
| Su et al. [60] | 19.6 | 87.7 | 88.3 | 4.2 |
| Tensmeyer et al. [36] | 20.7 | 93.1 | 96.8 | 2.2 |
| Xiong et al. [31] | 21.3 | 93.5 | 94.4 | 2.7 |
| Vo et al. [38] | 21.4 | 94.4 | 96.0 | 1.8 |
| Howe [61] | 21.3 | 91.3 | 91.7 | 3.2 |
| **DE-GAN** | **24.9** | **99.5** | **99.7** | **1.1** |

images for training and 32 for testing. From the 112 training images, a set of overlapped patches of size $256 \times 256$ pixels was extracted. This has generated 1,356 pairs of patches that were fed to our model. This first test intend to demonstrate the adversarial training effect. Thus, we train another model which is a simple FCN which is the U-net presented in Fig. 2. A validation set of 15 percent from the training images was used in this model. The results obtained by both models are presented in Table 1. As could be interpreted, the result of the encoder-decoder network (U-net) are acceptable for denoising and cleaning tasks. But, our DE-GAN is further improving the results. Which expose the reason of using an adversarial training for these types of problems. For more comparison, we have participated to the kaggle competition on denoising dirty documents,[1] we obtain a root mean squared error score of 0.01952. This makes our method as one of the best approaches in the leaderboard.

In order to give an idea of the cleaning made by our model, some examples are given in Fig. 5 which demonstrate the ability of recovering a very close document to the ground truth.

Nevertheless, we will compare our approach with state-of-the-art results in the document binarization problem. We take the DIBCO 2013 Dataset [52] for testing. While training

---

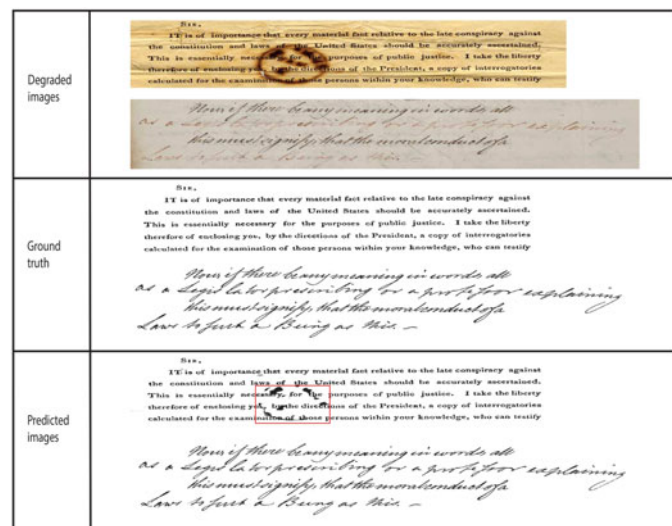1. https://www.kaggle.com/c/denoising-dirty-documents/

Fig. 6. Binarization of degraded documents by DE-GAN, the result is satisfactory, except in some parts that were highly dense (the red boxes in the predicted images row).
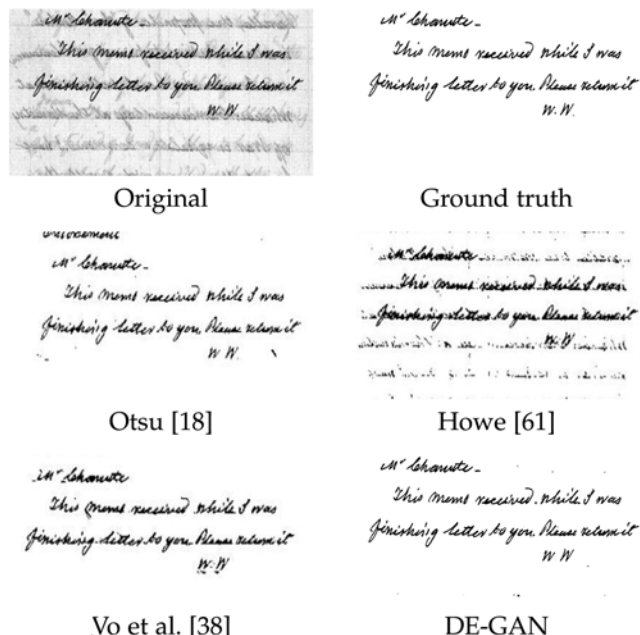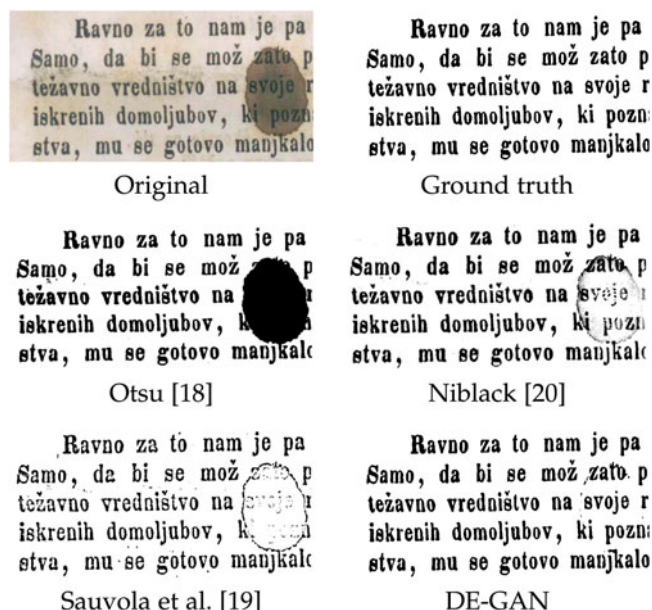


Fig. 7. Qualitative binarization results produced by different methods of a part from the sample (PR5), which is included in DIBCO 2013 dataset.



Fig. 8. Qualitative binarization results produced by different methods of of a part from the sample (HW5), which is included in DIBCO 2013 dataset.

TABLE 3
Results of Image Binarization on DIBCO 2017 Database, a Comparison With DIBCO 2017 Competitors Approaches

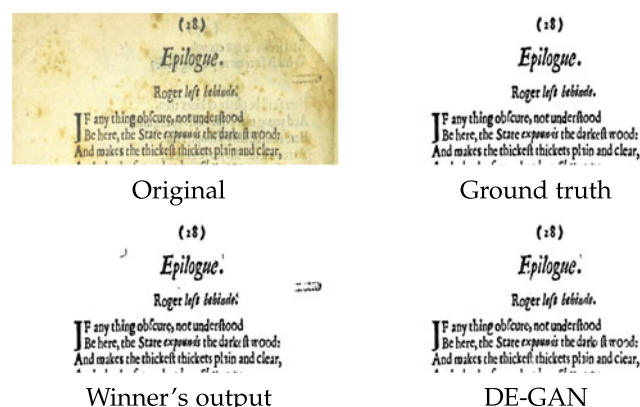| Model | PSNR | F-measure | $F_{ps}$ | DRD | Rank in the competition |
|---|---|---|---|---|---|
| 10 [58] | 18.28 | 91.04 | 92.86 | 3.40 | 1 |
| 17a [58] | 17.58 | 89.67 | 91.03 | 4.35 | 2 |
| 12 [58] | 17.61 | 89.42 | 91.52 | 3.56 | 3 |
| 1b [58] | 17.53 | 86.05 | 90.25 | 4.52 | 4 |
| 1a [58] | 17.07 | 83.76 | 90.35 | 4.33 | 5 |
| **DE-GAN** | **18.74** | **97.91** | **98.23** | **3.01** | - |



Fig. 9. Qualitative binarization results produced by different of the sample 16 in DIBCO 2017 dataset, here we compare DE-GAN with the winner's approach.

our model was done with different versions of DIBCO Databases [53], [54], [55], [56], [57], [58]. Same as the previous test, a set of 6,824 training pairs (patches of size $256 \times 256$) was taken from its 80 total images. The obtained results are compared with several approaches in Table 2.

Out of the results, we can say that DE-GAN is superior than the current state-of-the-art methods according to the following metrics [52]: Peak signal-to-noise ratio (PSNR), F-measure, pseudo-F-measure ($F_{ps}$) and Distance reciprocal distortion metric (DRD). Some examples of DIBCO 2013 images binarization by DE-GAN are presented in Fig. 6.

To reflect the results of the previous Table, an illustrative comparisons between those different methods could be found in Figs. 7 and 8. It is easy to visualize the superiority of our method over the classic methods, like those of [18], [19], [20], which fail to remove the background degradation from the document when it get very dense, because they are basing on thresholds that make the degraded pixels classified as a text, or classifying the text pixels as a damage to be removed. For the recent approaches [38], [61], they yield a better result than the classic ones and separate the text from the background successfully. However, our method gives a higher performance in terms of closeness to the ground truth image.

TABLE 4
Results of Image Binarization on DIBCO 2017 and DIBCO 2018 Databases, a Comparison With DIBCO 2018 Competitors Approaches

| Model | DIBCO 2018 | | | | DIBCO 2017 | | | | Rank in the competition |
|---|---|---|---|---|---|---|---|---|---|
| | PSNR | F-measure | $F_{ps}$ | DRD | PSNR | F-measure | $F_{ps}$ | DRD | |
| 1 [62] | **19.11** | **88.34** | **90.24** | **4.92** | 17.99 | 89.37 | 90.17 | 5.51 | 1 |
| 7 [62] | 14.62 | 73.45 | 75.94 | 26.24 | 15.72 | 84.36 | 87.34 | 7.56 | 2 |
| 2 [62] | 13.58 | 70.04 | 74.68 | 17.45 | 14.04 | 79.41 | 82.62 | 10.70 | 3 |
| 3b [62] | 13.57 | 64.52 | 68.29 | 16.67 | 15.28 | 82.43 | 86.74 | 6.97 | 4 |
| 6 [62] | 11.79 | 46.35 | 51.39 | 24.56 | 15.38 | 80.75 | 87.24 | 6.22 | 5 |
| DE-GAN | 16.16 | 77.59 | 85.74 | 7.93 | **18.74** | **97.91** | **98.23** | **3.01** | - |

Moreover, we tested DE-GAN on a recent DIBCO dataset, which is DIBCO 2017 [58]. We train our model on 6,098 patches from similar datasets [52], [53], [54], [55], [56], [57].



The comparison is done with the top 5 ranked approaches in ICDAR 2017 competition on document image Binarization [58]. 18 research groups have participated in the competition with 26 distinct algorithms. The results are presented in Table 3, where you can notice the superiority of our DE-GAN over the different methods. It is to note that most of these approaches are based on encoder-decoder models and the winner team was using a U-net with several data augmentation techniques. However, GANs were not exploited in this competition. An example to compare our output with the winner algorithm is given in Fig. 9.

In addition, we compared our model with the most recent approaches, presented in the H-DIBCO 2018 competition [62] that was held in ICFHR 2018 conference. The results are presented in Table 4. As shown, our approach has the best performance on DIBCO 2017 test set and gives the second best DRD, PSNR, F-measure and pseudo F-Measure on H-DIBCO 2018 test set. We note that the winner system in the competition integrates a lot of pre-processing and a post-processing steps in their algorithm, that make it more efficient for this particular H-DIBCO 2018 dataset. On the contrary, we are presenting a simple end-to-end model that shows a good ability in a several datasets and enhancements tasks without any additional processing step. Finally, for a more practical usage of the model, we tried to binarize some real (naturally degraded) documents as well, the degradation consists in stains and show-through. The obtained results are given in Fig. 10, the model is producing a better versions of the real images, which will certainly improves their recognition rate.

## 4.2 Watermark Removal

After testing our model in document cleaning and binarization, we will evaluate it on the problem of watermark removal. Dense watermarks (or stamps) can cause a huge deterioration in the *foreground* of the document, which make it hard to be read. However, this problem was not investigated by document analysis community. We decided to be the first that address it using DE-GAN. Hence, it was not possible to find a public dataset for testing. We created our own database which contains 1000 pairs (image of a document with a dense watermark and stamps and its clean version).

The used watermarks have random texts, sizes, colors, fonts, opacities and locations (see Fig. 11). As shown, these watermarks are sometimes covering the entire text making it unseen by the unaided eye. The code used to produce this data is available at GitHub,[2] for the same dataset used in
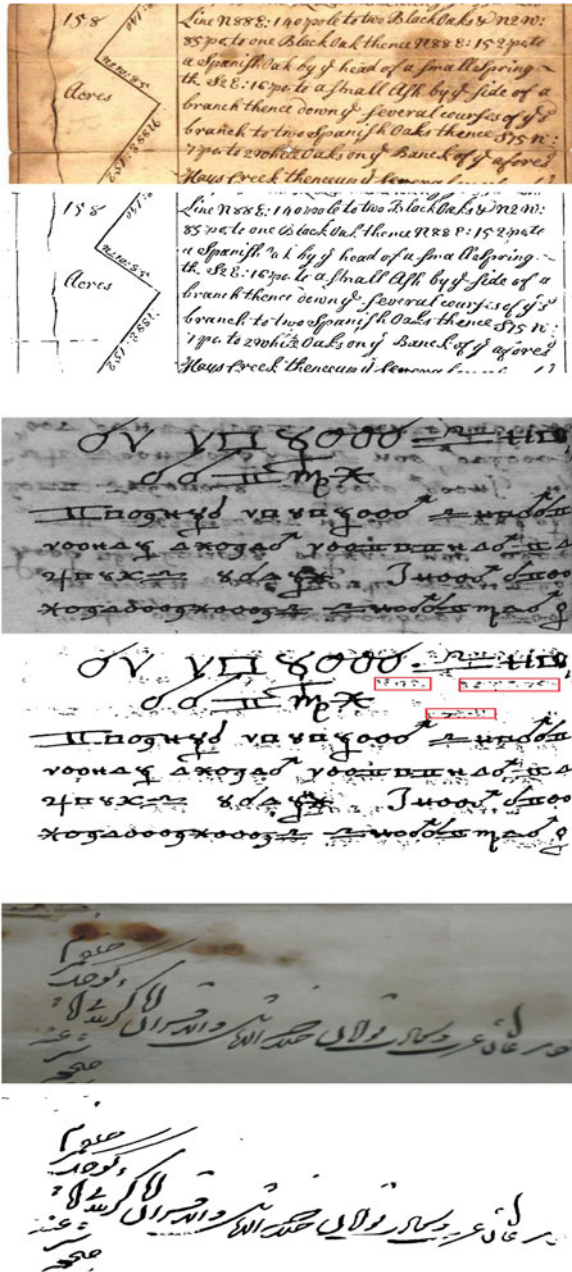
Fig. 10. Binarization of three historical degraded documents by DE-GAN, the binarized version is presented under each original image. Some parts are not well recovered as shown in the red boxes.

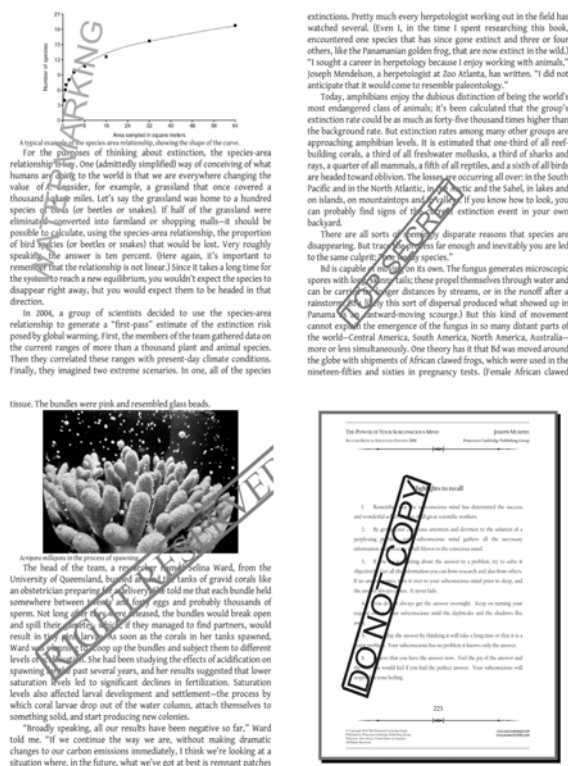2. https://github.com/dali92002/watermarking-documents/blob/master/Watermarking.ipynb

Fig. 11. 4 Samples from our developed dataset.



Fig. 12. Qualitative results for dense watermark removal. Above, a section from watermarked invoice. Below, it's enhanced version. Some parts of the text in the invoice was blurred due to privacy constraints. Because of different domains, synthetic versus real, we can see that some tiny parts of the watermark were not completely removed (red boxes).

our study the reader can contact the first author to obtain it. Training our DE-GAN was done, same as document cleaning, by using overlapped patches (7658 pairs of patches from 800 watermarked document images). While taking 200 documents for testing. Since, for the best of our knowledge, there is no approach in the literature that addresses this problem in documents. Comparing our obtained results was done with the approaches used in natural images watermark removal. The comparison results are presented in Table 5.

Despite that the watermarks used in our study were very dense and we believe that removing them is harder than the related approaches presented in Table 5. Our approach surpasses, by far, those in natural images. Fig. 13 shows some examples of watermark removal by DE-GAN, the produced images are preserving the text quality while removing the foreground watermarks. In addition, since the presented watermarked documents were synthetically made, it was interesting to apply DE-GAN to remove watermarks from a naturally degraded document. Fig. 12 shows that DE-GAN successfully removes a dense watermark from a document paper. As you can see, the watermark is completely removed, and the reader or the OCR system can easily read the enhanced document compared to the degraded one.

### TABLE 5
### Results of Watermark Removal

| Model | PSNR | SSIM |
|---|---|---|
| Dekel *et al.* [40] | 36.02 | 0.924 |
| Wu *et al.* [41] | 23.37 | 0.884 |
| Cheng *et al.* [42] | 30.86 | 0.914 |
| **DE-GAN** | **40.98** | **0.998** |

## 4.3 Comparison With Other GAN Models

As it is a fact that our model is inspired from the pix2pix model [9] (we are using a deeper generator and a different additional loss), it would be useful if we tried some other similar models that are based on GANs and dedicated to the same image-to-image translation problem. For this aim, cycleGAN [45] and pix2pix-HD [46] models are considered for the comparison. We evaluate these models on H-DIBCO 2018 dataset [62] with the same conditions and data used to train the DE-GAN. The quantitative and qualitative obtained results are presented in Table 6 and Fig. 14, respectively. Experimental results shows the superiority of DE-GAN compared to cycleGAN and pix2pix-HD in achieving higher PSNR, F-measure and Fps and a lower DRD. We note that the unsupervised training capabilities of CycleGAN are quite useful since paired data is harder to find in document enhancement applications. For pix2pix-HD, the results are promising, since the training samples that we used for training were few (the number of DIBCO samples is small if we split them to patches with size $512 \times 1024$, that's why we used some flips of images to augment the data). With more data, we believe that pix2pix-HD could perform much better.

## 4.4 Document Deblurring

The DE-GAN model presented in this paper is able to outperform many state-of-the-art approaches in different problems like binarization, denoising and watermark removal. To experimentally prove the efficiency and the flexibility of the proposed method, we evaluate it on a more challenging
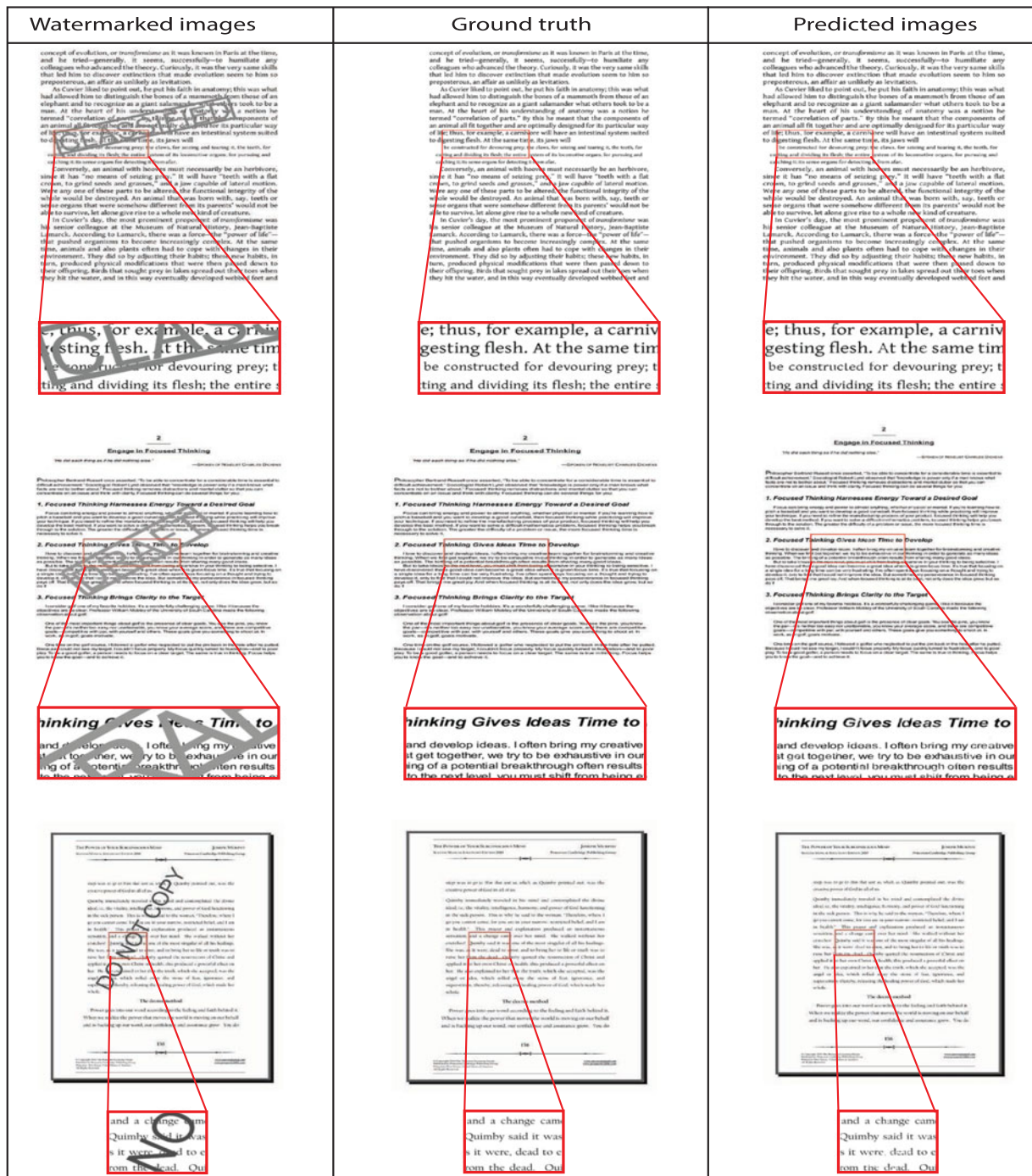
Fig. 13. Watermark removal by DE-GAN.

scenario, which is document deblurring. We use 4000 patches from the dataset developed in [63] to train our model, and 932 patches for testing. Noting that, in [63] a convolutional neural network architecture is proposed to address the problem.
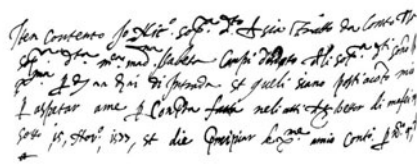
TABLE 6
Results of Image Binarization for DIBCO 2018 Database

| Model | PSNR | F-measure | $F_{ps}$ | DRD |
|---|---|---|---|---|
| cycleGAN | 11.00 | 56.33 | 58.07 | 30.07 |
| pix2pix-HD | 14.42 | 72.79 | 76.28 | 15.13 |
| **DE-GAN** | **16.16** | **77.59** | **85.74** | **7.93** |

Thus, we will compare the results with this CNN and pix2-pix-HD models trained on this selected data. The obtained results are presented in Table 7. We can see that GAN's models surpasses the CNN. This is much clear in the qualitative results of some patches presented in Fig. 15. We can also see that DE-GAN gives similar results to pix2pix-HD, however, it is more accurate for predicting some characters. For example, in the second patch row, third line, the word "kind" is correctly predicted by DE-GAN but it is predicted as "bind" by pix2pix-HD. We note that the used dataset is composed of 300x300px image patches, which can explain why pix2pix-HD does not give a better performance (it works generally with larger input patch with a size of 512x1024, or 1024x2048).
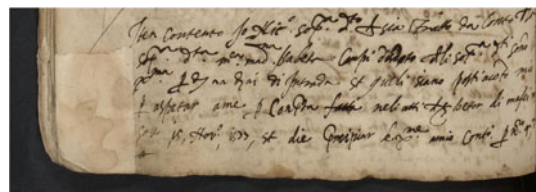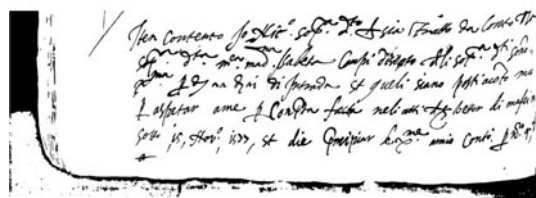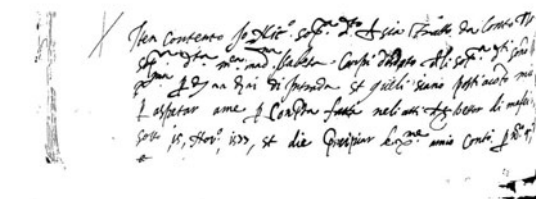
Original

Ground truth
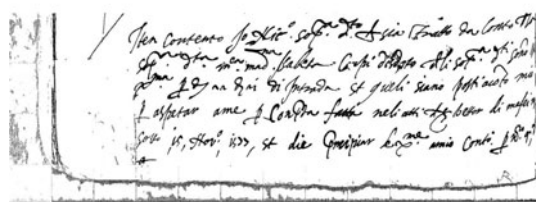
CycleGan

Pix2pix-HD

DE-GAN

Fig. 14. Qualitative binarization results produced by different models of the sample (9) from H-DIBCO 2018 dataset.



Original    GT    CNN    pix2pix-HD    DE-GAN
                  [63]        [46]

Fig. 15. Qualitative deblurring results of some patches produced by different methods.



For Nol expoundsytbe officers expound,

For Nol expounds, the officers expound,

term temperaturef ham Yhanges
in climate-come into play. For the

term temperature changes-changes
in climate-come into play. For the
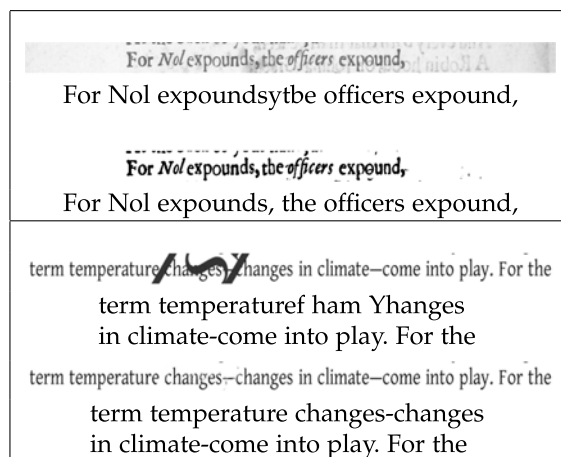
Fig. 16. Qualitative results for Tesseract recognition of some text lines.

## TABLE 7
The Obtained Results of Document Deblurring

| Method | PSNR |
| --- | --- |
| CNN [63] | 19.36 |
| pix2pix-HD [46] | 19.89 |
| **DE-GAN** | **20.37** |

### 4.5 OCR Evaluation

After the quantitatively and qualitatively evaluation of the resulted enhanced images presented previously, we compare in what follows the performance of OCR on degraded and enhanced documents. For this aim, we took a set of 4 images (2 degraded ones from DIBCO datasets, and 2 images with a dense watermark from our dataset).Then, we used Tesseract OCR [64] to recognize those image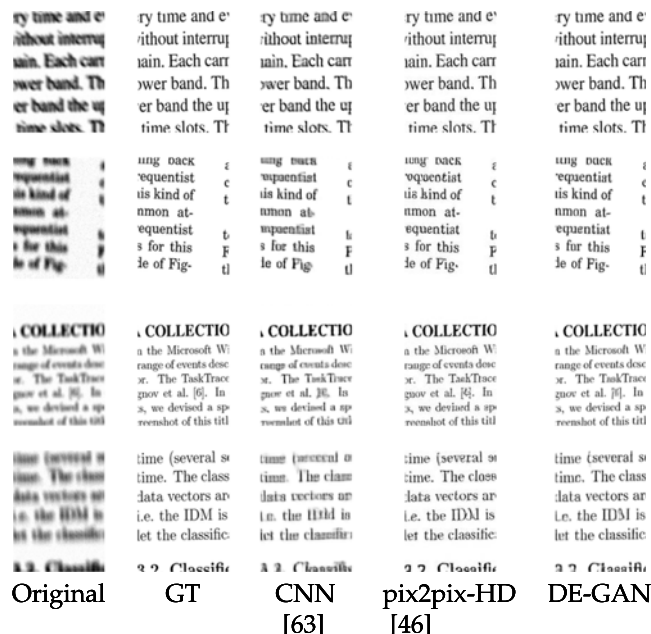s and their enhanced versions with DE-GAN. We found that the proposed enhancement method boosts the baseline OCR performance by a large margin, the character error rate is decreased from 0.37 for the degraded documents to 0.01 for the enhanced ones. Fig. 16 shows a tiny example of this process. In each row, you can find a line of a degraded document image and the text produced by the OCR system, then its enhanced version followed the OCR text.

## 5 CONCLUSION

In this paper we proposed a Document Enhancement Generative Adversarial Network named DE-GAN to restore severely degraded document images. DE-GAN is a modified version of the pix2pix with a deeper generator and a different additional loss (adversarial + log) to generate an enhanced document given the degraded version. To the best of our knowledge, this is the first application of GANs for studying document enhancement problem. Moreover, we present a new problem in document enhancement that

is dense watermark (or stamps) removal, hoping that it takes the attention of document analysis community. Extensive experiments show that DE-GAN achieved interesting results in different document enhancement tasks that outperform the fully convolutional networks, cycleGAN and pix2pix-HD models. Furthermore, we achieve improved results compared to many recent state-of-the-art methods on benchmarking datasets like DIBCO 2013, DIBCO 2017 and H-DIBCO 2018.

We showed that the proposed enhancement method boosts the baseline OCR performance by a large margin. Hence, as an immediate future work, we plan to add the OCR evaluation in the discriminator part. Thus, we can give the discriminator the ability of reading the text to decide if it is real or fake, which will force it to generate more readable images. We intend, also, to test the performance of the DE-GAN on mobile captured documents which present many problems like shadow, real blur, low resolution, distortion, etc.

## ACKNOWLEDGMENT

## REFERENCES

[1] K. Chellapilla, S. Puri, and P. Simard, "High performance convolutional neural networks for document processing," in *Proc. Int. Workshop Front. Handwriting Recognit.*, 2006. [Online]. Available: https://scholar.googleusercontent.com/scholar.bib?q=info:W8yn_wnlEl0J:scholar.google.com/&output=citation&scisdr=CgVFlI0mEP7SzqjSoHc:AAGBfm0AAAAAX1zXuHck0ZQCcjUmDlB9jOwrINeGsggP&scisig=AAGBfm0AAAAAX1zXuDmqymV0GuOewudd3gh3ULzjjwVr&scisf=4&ct=citation&cd=-1&hl=en

[2] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed, "DeepDeSRT: Deep learning for detection and structure recognition of tables in document images," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit.*, 2017, pp. 1162–1167.

[3] F. Zamora-Martinez, S. España-Boquera, and M. J. Castro-Bleda, "Behaviour-based clustering of neural networks applied to document enhancement," in *Proc. Int. Work-Conf. Artif. Neural Netw.*, 2007, pp. 144–151.

[4] S. Bako, S. Darabi, E. Shechtman, J. Wang, K. Sunkavalli, and P. Sen, "Removing shadows from images," in *Proc. Asian Conf. Comput. Vis.*, 2016, pp. 173–183.

[5] X. Chen, X. He, J. Yang, and Q. Wu, "An effective document image deblurring algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 369–376.

[6] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 2810–2818.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[8] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. 2nd Int. Conf. Learn. Representations*, 2014.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.

[10] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. 6th Int. Conf. Learn. Representations*, 2018.

[11] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[12] A. K. Bhunia *et al.*, "Word level font-to-font image translation using convolutional recurrent generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognit.*, 2018, pp. 3645–3650.

[13] A. Ghosh, B. Bhattacharya, and S. B. R. Chowdhury, "Handwriting profiling using generative adversarial networks," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4927–4928.

[14] A. Konwer *et al.*, "Staff line removal using generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognit.*, 2018, pp. 1103–1108.

[15] N. Kligler, S. Katz, and A. Tal, "Document enhancement using visibility detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2374–2382.

[16] R. K. Pandey and A. G. Ramakrishnan, "Language independent single document image super-resolution using CNN for improved recognition," *CoRR*, vol. abs/1701.08835, 2017. [Online]. Available: http://arxiv.org/abs/1701.08835

[17] C. L. Tan, L. Zhang, Z. Zhang, and T. Xia, "Restoring warped document images through 3D shape modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 195–208, Feb. 2006.

[18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[19] J. Sauvola and M. Pietik, "Adaptive document image binarization," *Pattern Recognit.*, vol. 33, pp. 225–236, 2000.

[20] W. Niblack, *An Introduction to Digital Image Processing*. Birkeroed, Denmark: Strandberg Publishing Company, 1985.

[21] N. Phansalkart, S. More, A. Sabale, and M. Joshi, "Adaptive local thresholding for detection of nuclei in diversity stained cytology images," in *Proc. Int. Conf. Commun. Signal Process.*, 2011, pp. 218–220.

[22] G. Chutani, T. Patnaik, and V. Dwivedi, "An improved approach for automatic denoising and binarization of degraded document images based on region localization," in *Proc. Int. Conf. Advances Comput. Commun. Inform.*, 2015, pp. 2272–2278.

[23] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 918–921, Jun. 1998.

[24] T. Lelore and F. Bouchara, "FAIR: A fast algorithm for document image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2039–2048, Aug. 2013.

[25] M. Annabestani and M. Saadatmand-Tarzjan, "A new threshold selection method based on fuzzy expert systems for separating text from the background of document images," *Iranian J. Sci. Technol. Trans. Elect. Eng.*, vol. 43, pp. 219–231, 2019.

[26] C.-H. Chou, W.-H. Lin, and F. Chang, "A binarization method with learning-built rules for document images produced by cameras," *Pattern Recognit.*, vol. 43, pp. 1518–1530, 2010.

[27] W. Xiong, J. Xu, Z. Xiong, J. Wang, and M. Liu, "Degraded historical document image binarization using local features and support vector machine (SVM)," *Optik*, vol. 164, pp. 218–223, 2018.

[28] R. F. Moghaddam and M. Cheriet, "A variational approach to degraded document enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1347–1361, Aug. 2010.

[29] R. Hedjam, M. Cheriet, and M. Kalacska, "Constrained energy maximization and self-referencing method for invisible ink detection from multispectral historical document images," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 3026–3031.

[30] S. Milyaev, O. Barinova, T. Novikova, P. Kohli, and V. Lempitsky, "Fast and accurate scene text understanding with image binarization and off-the-shelf OCR," *Int. J. Document Anal. Recognit.*, vol. 18, pp. 169–182, 2015.

[31] W. Xiong, X. Jia, J. Xu, Z. Xiong, M. Liu, and J. Wang, "Historical document image binarization using background estimation and energy minimization," in *Proc. 24th Int. Conf. Pattern Recognit.*, 2018, pp. 3716–3721.

[32] G. Meng, K. Yuan, Y. Wu, S. Xiang, and C. Pan, "Deep networks for degraded document image binarization through pyramid reconstruction," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit.*, 2017, pp. 727–732.

[33] J. Calvo-Zaragoza and A.-J. Gallego, "A selectional auto-encoder approach for document image binarization," *Pattern Recognit.*, vol. 86, pp. 37–47, 2019.

[34] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, 2017.

[35] M. Z. Afzal, J. Pastor-Pellicer, F. Shafait, T. M. Breuel, A. Dengel, and M. Liwicki, "Document image binarization using LSTM: A sequence learning approach," in *Proc. 3rd Int. Workshop Historical Document Imag. Process.*, 2015, pp. 79–84.

[36] C. Tensmeyer and T. Martinez, "Document image binarization with fully convolutional neural networks," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit.*, 2017, pp. 99–104.

[37] F. Westphal, N. Lavesson, and H. Grahn, "Document image binarization using recurrent neural networks," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst.*, 2018, pp. 263–268.

[38] Q. N. Vo, S. H. Kim, H. J. Yang, and G. Lee, "Binarization of degraded document images based on hierarchical deep supervised network," *Pattern Recognit.*, vol. 74, pp. 568–586, 2018.

[39] C. Xu, Y. Lu, and Y. Zhou, "An automatic visible watermark removal technique using image inpainting algorithms," in *Proc. 4th Int. Conf. Syst. Inform.*, 2017, pp. 1152–1157.

[40] T. Dekel, M. Rubinstein, C. Liu, and W. T. Freeman, "On the effectiveness of visible watermarks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6864–6872.

[41] J. Wu, H. Shi, S. Zhang, Z. Lei, Y. Yang, and S. Z. Li, "De-Mark GAN: Removing dense watermark with generative adversarial network," in *Proc. Int. Conf. Biometrics*, 2018, pp. 69–74.

[42] D. Cheng *et al.*, "Large-scale visible watermark detection and removal with deep convolutional networks," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 27–40.

[43] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *NIPS Workshop Adversarial Training*, Dec. 2016.

[44] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.

[45] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[46] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8798–8807.

[47] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2868–2876.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv.*, 2015, pp. 234–241.

[49] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[50] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 350–358.

[51] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[52] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2013 document image binarization contest (DIBCO 2013)," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, 2013, pp. 1471–1476.

[53] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 document image binarization contest (DIBCO 2009)," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, 2009, pp. 1375–1382.

[54] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2011 document image binarization contest (DIBCO 2011)," in *Proc. Int. Conf. Document Anal. Recognit.*, 2011, pp. 1506–1510.

[55] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012)," in *Proc. Int. Conf. Front. Handwriting Recognit.*, 2012, pp. 817–822.

[56] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "ICFHR2014 competition on handwritten document image binarization (H-DIBCO 2014)," in *Proc. 14th Int. Conf. Front. Handwriting Recognit.*, 2014, pp. 809–813.

[57] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICFHR 2016 competition on handwritten document image binarization (H-DIBCO 2016)," in *Proc. 15th Int. Conf. Front. Handwriting Recognit.*, 2016, pp. 619–623.

[58] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gato, "ICDAR2017 competition on document image binarization (DIBCO 2017)," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit.*, 2017, pp. 1395–1403.

[59] B. Gatos, I. Pratikakis, and S. Perantonis, "An adaptive binarization technique for low quality historical documents," in *Proc. Int. Workshop Document Anal. Syst.*, 2014, pp. 102–113.

[60] B. Su, S. Lu, and C. L. Tan, "Robust document image binarization technique for degraded document images," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1408–1417, Apr. 2013.

[61] N. Howe, "Document binarization with automatic parameter tuning," *Int. J. Document Anal. Recognit.*, vol. 16, pp. 247–258, 2013.

[62] I. Pratikakis, K. Zagoris, P. Kaddas, and B. Gatos, "ICFHR2018 competition on handwritten document image binarization (H-DIBCO 2018)," in *Proc. 16th Int. Conf. Front. Handwriting Recognit.*, 2018, pp. 489–493.

[63] M. Hradiš, J. Kotera, P. Zemčík, and F. Šroubek, "Convolutional neural networks for direct text deblurring," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–13.

[64] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. 9th Int. Conf. Document Anal. Recognit.*, 2007, pp. 629–633.

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

**Yousri Kessentini** received the graduate degree in computer science engineering from the National Engineering School of Sfax (ENIS), Sfax, Tunisia, in 2003, and the PhD degree in the field of pattern recognition from the University of Rouen, Rouen, France, in 2009. He was postdoctoral researcher with ITESOFT Company and LITIS Laboratory from 2011 to 2013. Currently, he is an assistant professor at CRNS and the head of the DeepVision Research Team. His main research interests include 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 coordinate several research projects in partnership with industry and the author and co-author of several papers.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.