EHDI: Enhancement of Historical Document Images via Generative Adversarial Network

Abir Fathallah[1,2] [a], Mounim A. El-Yacoubi[2] and Najoua Essoukri Ben Amara[3]

[1] *Université de Sousse, Institut Supérieur de l'Informatique et des Techniques de Communication, LATIS - Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*

[2] *Samovar, CNRS, Télécom SudParis, Institut Polytechnique de Paris, 9 rue Charles Fourier, 91011 Evry Cedex, France*

[3] *Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*

Keywords: Historical Documents, Document Enhancement, Degraded Documents, Generative Adversarial Networks.

Abstract: Images of historical documents are sensitive to the significant degradation over time. Due to this degradation, exploiting information contained in these documents has become a challenging task. Consequently, it is important to develop an efficient tool for the quality enhancement of such documents. To address this issue, we present in this paper a new modelknown as EHDI (Enhancement of Historical Document Images) which is based on generative adversarial networks. The task is considered as an image-to-image conversion process where our GAN model involves establishing a clean version of a degraded historical document. EHDI implies a global loss function that associates content, adversarial, perceptual and total variation losses to recover global image information and generate realistic local textures. Both quantitative and qualitative experiments demonstrate that our proposed EHDI outperforms significantly the state-of-the-art methods applied to the widespread DIBCO 2013, DIBCO 2017, and H-DIBCO 2018 datasets. Our suggested model is adaptable to other document enhancement problems, following the results across a wide range of degradations. Our code is available at https://github.com/Abir1803/EHDI.git.

1 INTRODUCTION

Historical Arabic Documents (HADs) are a valuable part of cultural heritage, but access to them is often limited due to inadequate storage conditions. To be understood automatically by machine vision, digital historical documents must be transcribed into a readable form, as they are not readily processed in their original form. These documents often suffer from various types of degradation. The restoration of historical documents can be complicated by the presence of watermarks, stamps, or annotations, especially when these forms of degradation occur in the text itself. This is particularly challenging when the stain color is similar to or more intense than the font color of the document. Document processing, which involves transcribing digital historical documents into a readable form, can be done using a computer vision tool or by a human being. In recent years, the development of various public databases has led to a significant expansion of document processing. The processing of historical documents is a very challenging task and it is not always efficient due to the poor quality of manuscripts. These documents can be affected by various types of damage, such as wrinkles, dust damage, nutrition stains, and discolored sunspots, which can hinder their processing efficiency (Zamora-Martínez et al., 2007). Degradation may also occur in the scanned documents due to poor scanning conditions related to the use of the smartphone camera (shadow (Finlayson et al., 2002), blurring (Chen et al., 2011), varying light conditions, warping, etc.). In addition, several documents are sometimes infested with stamps, watermarks, or annotations. In this paper, a new document enhancement model is evolved for enhancing degraded documents to provide a cleaned-up version. Specifically, we consider the document enhancement task as a GAN-based image-to-image converter process. This paper proposes a new GAN architecture specifically designed to improve the clarity of images of historical documents. In contrast to previous methods, this approach aims to simultaneously remove noise and watermarks while preserving the quality of the text. The ultimate goal is to create a system that is able to

[a] https://orcid.org/0000-0003-0433-1029

Fathallah, A., El-Yacoubi, M. and Ben Amara, N.
EHDI: Enhancement of Historical Document Images via Generative Adversarial Network.
DOI: 10.5220/0011662700003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 238-245
ISBN: 978-989-758-634-7; ISSN: 2184-4321

effectively enhance document images. An ideal system should be able to remove noise and watermarks while also maintaining the quality of the text in document images. The ability to perform both tasks simultaneously would be highly desirable. Deep neural networks, specifically deep convolutional neural networks (auto-encoders and variational auto-encoders (VAE)) (Mao et al., 2016; Dong et al., 2015), and generative adversarial networks (GANs) (Isola et al., 2017), have recently made significant progress in generating and restoring natural images.

The remainder of this paper is organized as follows. Section 2 provides an overview of relevant previous work. Section 3 presents the proposed approach for enhancing degraded documents. The experimental study is described in section 4. Finally, the conclusion and future directions are outlined in Section in section 5.

2 RELATED WORK

Document enhancement involves the perceptual quality improvement of document images and the removal of degradation effects and artifacts from the images to make them look as they originally did (Hedjam and Cheriet, 2013). In order to improve the quality of historical documents, the binarization of documents is the most widely applied technique. It aims to separate each pixel of text from the background (Sauvola and Pietikäinen, 2000). This technique reduces the amount of noise in document images. Traditional methods of document binarization (Otsu, 1979; Sauvola and Pietikäinen, 2000) are generally formulated using the thresholding technique. Hence, several approaches have evolved to determine the most optimal thresholds for applying it as a filter. Depending on the threshold(s), binary classification is performed to specify whether the pixels are part of the text or the degradation. In order to determine the binarization threshold, the authors in (Chou et al., 2010) proposed a method based on machine learning techniques. Each region of the image is defined as a three-dimensional feature vector obtained from the gray-level pixel value distribution. To classify each region into one of four different threshold values, the support vector machine (SVM) was employed. Generally, the major limitation of these traditional methods lies in the fact that the results are extremely dependent on the document conditions. Indeed, in the presence of a complex image background or with an atypical intensity, several problems appear. A variational model aimed at eliminating transparencies from degraded two-sided document images is introduced in (Moghaddam and Cheriet, 2009).

Recently, GANs have gained impressive achievements in both image generation and translation. In this section, we consider the application of GANs in document processing and enhancement issues. In terms of image segmentation, Ledig et al. reported SRGAN (Ledig et al., 2017), which provides a GANs for super-resolution of images. The authors employed conditional GANs for document enhancement tasks based on image-to-image translation. Dual GAN generator algorithms (Yi et al., 2017) intended for underwater image enhancement exploit two or more generators for predicting the enhanced image. The intention behind using two generators along with one discriminator or two generators with two discriminators is to either share features between the generators or to consider the prediction of one generator as an input to the other generator. The model is weakly supervised and avoids the need to use matched underwater images for training in that the underwater images can be considered in unknown locations, thus allowing for adversarial learning. Isola et al. (Isola et al., 2017) developed the GAN Pix2Pix for image-to-image translation using CGAN. An adversarial loss is used to train the GAN Pix2Pix model generator, thus promoting plausible image generation in the target domain.The discriminator assesses whether the generated image is a real transformation of the source image. There are various approaches to improving documents, but most of these approaches address a specific problem. For example, in a study published in (Souibgui and Kessentini, 2020), the authors examined the issue of documents that have been damaged due to watermarks or stamps and proposed a solution using conditional GANs to restore historical documents to their original, undamaged state. Similarly, the authors in (Jemni et al., 2022) developed a method using GANs that combines document binarization with a recognition stage. Another study (Gangeh et al., 2021) specifically addressed problems such as blurry text, salt and pepper noise, and watermarks by proposing a unified architecture that combines a deep network with a cycle-consistent GAN for the purpose of denoising document images.

As motioned above, there are some recent techniques for document enhancement that are emerging, involving the use of deep learning tools. Most of them employ Convolutional Neural Networks (CNNs) and GANs, in order to learn how to generate a clean binary version for any degraded document image (Souibgui and Kessentini, 2020; Tamrin et al., 2021). However, most authors employed a simple architecture with an optimization technique that is not well adapted to the complexity of historical document images. Essentially, the aim of document enhancement is to generate a significantly better and cleaner version of degraded document images, which is very beneficial for further processing tasks. To this end, we propose a robust GAN architecture that is trained and optimized following several loss functions in order to overcome the complexity of historical documents and to generate a quite clean image comparing with the existing methods.

3 PROPOSED METHOD

In this section, we present the main steps of our proposed approach called EHDI (Enhancement of Historical Document Images via GANs). It aims to generate a clear version of a degraded historical document using GANs. GANs are machine learning models that are used to learn the distribution of real data and generate images based on random noise. The goal of our model is to use GANs to improve the quality of historical documents. The main purpose of GANs, as previously discussed in studies such as (Marnissi et al., 2021; Souibgui and Kessentini, 2020; Jemni et al., 2022), is to learn the distribution of real data and create an output image from random noise. In our GAN model, the goal is to generate a clean version of a degraded historical document. This can be thought of as an image-to-image conversion process, where our model learns to map a degraded document image (represented as ”x”) to a clean document image (represented as ”y”). During the training process, the GAN takes a degraded document image as input and attempts to generate a cleaned version of it. On the other hand, the discriminator receives two inputs: the generated image and the ground truth, which is the known clean version of the degraded image. It then determines whether the generated image is a realistic representation or not based on the ground truth. As shown in Figure 1, our model consists of two main components: a generator (G) and a discriminator (D). The generator is trained to convert a degraded document image into a clean version, while the discriminator helps the generator to produce more realistic images by distinguishing between generated and real images.

3.1 Generator Architecture

The generator in our model is an image transformation network that generates the transformed image using the input image. It is designed as an autoencoder model, which consists of an encoder and a decoder. The input image is typically processed through a series of convolutional layers with downward sampling to reach a particular layer, and then decoded through a series of up-sampling and convolutional layers. Figure 1 illustrates the details of our suggested GAN architecture. Our generator network is based on the architecture proposed in (Wang et al., 2018), with each sub-network following the structure outlined in (Johnson et al., 2016).

3.2 Discriminator Architecture

The discriminator in our model is a simple fully convolutional network that receives two input images: the generated image and its ground truth. Its purpose is to determine whether the generated image is real or fake. As shown in Figure 1, our suggested discriminator architecture consists of five convolutional layers, followed by a normalization layer (except for the last layer) and a LeakyReLU activation function (except for the first and last layers). Inspired by PatchGAN (Isola et al., 2017), we use a 70×70 patch as input to our discriminative network, which determines whether local image patches are real or fake. The discriminator's goal is to identify whether the input patch in an image is genuine or synthetic.

3.3 Loss Functions of Proposed GAN

To effectively train our EHDI model, we include a content loss to penalize the distance between the generated and ground-truth images. The adversarial discriminator helps the generator to synthesize fine and specific details. The discriminator helps the generator create more accurate and specific details by identifying and penalizing deviations from the desired output. To further enhance the clarity and precision of these details, we also incorporate a combination of perceptual and Total Variation (TV) losses. Our objective loss is comprised of four losses: adversarial loss, content loss, perceptual loss, and TV loss. These losses are defined as follows:

- Adversarial loss: To encourage the generator to produce high-quality, accurate images, we use an adversarial loss. This loss function, defined in Eq.(1), helps ensure that the generated clean images $G(x)$ are as close as possible to the true clean images y.

$$\mathcal{L}_{adv} = \mathbb{E}_y[-log(D(G(x),y)] \quad (1)$$

- Content loss: To preserve the content information present in the ground truth image y in the generated image $G(x)$, we use a content loss in our improved GAN. This loss function, defined in Eq.(2), is a pixel-wise mean squared error that

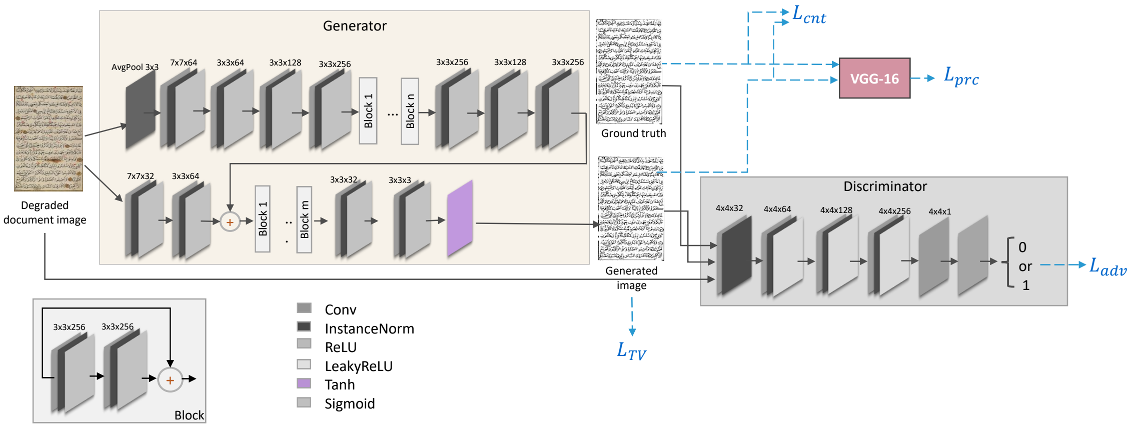


Figure 1: Proposed GAN architecture for our document enhancement model EHDI.

minimizes low-level content errors between the generated cleaned images and their corresponding ground truth images.

$$\mathcal{L}_{content} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} \parallel y_{i,j} - G(x)_{i,j} \parallel_1 \qquad (2)$$

where W and H refer to the height and width of the degraded image, respectively. The L_1 norm is denoted by $\parallel . \parallel_1$, and the pixel values of the ground truth image and the generated image are represented by $y_{i,j}$ and $G(x)_{i,j}$, respectively.

- Perceptual loss: To improve the perceptual quality of the generated results and correct any distorted textures caused by the adversarial loss, we use a perceptual loss function introduced in (Johnson et al., 2016). This loss function calculates the distance between the generated image and its ground truth based on high-level representations extracted from a pre-trained VGG-16 model.The perceptual loss is defined by Eq.(3).

$$\mathcal{L}_{prc} = \sum_k \frac{1}{C_k H_k W_k} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} \parallel \Phi_k(y)_{i,j} - \Phi_k(G(x))_{i,j} \parallel_1 \qquad (3)$$

where Φ_k represents the feature representations of the k^{th} maxpooling layer in the VGG-16 network, and $C_k H_k W_k$ represents the size of these feature representations.

- Total variation loss: To prevent over-pixelization and improve the spatial smoothness of the cleaned document images, we use the TV loss introduced in (Aly and Dubois, 2005). This loss function is defined in Eq.(4).

$$\mathcal{L}_{tv} = \frac{1}{WH} \sum |\nabla_x G(\tilde{y}) + \nabla_y G(\tilde{y})| \qquad (4)$$

where $|.|$ refers to the absolute value per element of the indicated input.

The loss function that optimizes the network parameters of the generator (G) is referred to as the global loss function $\mathcal{L}$ given by Eq.(5).

$$\mathcal{L} = \mathcal{L}_{cnt} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{prc}\mathcal{L}_{prc} + \lambda_{tv}\mathcal{L}_{tv} \qquad (5)$$

where λ_{adv}, λ_{prc} and λ_{tv} represent the weights that control the share of different losses in the full objective function.

4 EXPERIMENTS

In this section, we present the main experiments that were conducted to evaluate the effectiveness of our proposed approach.

4.1 Datasets

To train our document enhancement architecture, we used the Noisy Office database (Zamora-Martínez et al., 2007). For the evaluation phase, we considered the DIBCO 2013 (Pratikakis et al., 2013), DIBCO 2017 (Pratikakis et al., 2017), H-DIBCO 2018 (Pratikakis et al., 2018b) and (Pantke et al., 2014) datasets.

4.2 Experimental Setup

4.2.1 Evaluation Protocol

To evaluate the ability and the quality of our EHDI model, we have conducted a series of experiments on different datasets. We introduce qualitative and quantitative results to evaluate our EHDI. We select four performance assessments (Pratikakis et al., 2013): Peak signal-to-noise ratio (PSNR), pseudo-F-measure (F_{ps}), Distance reciprocal distortion metric (DRD) and F-measure.

4.2.2 Implementation Details

The learning process is optimized using the stochastic gradient descent algorithm with a learning rate of 10^{-3} and a batch size of 512. T. All parameter values were chosen based on empirical testing. The experiments were implemented using the PyTorch framework and the EHDI model was trained on an NVIDIA Quadro RTX 6000 GPU with 24 GB of RAM. To facilitate the training of the architecture, each image is resized to 1024 × 1024 pixels and a set of stacked patches of size 256 × 256 pixels are extracted. This results in a total of 2,304 patch pairs, which are used to train the EHDI model. During training, the weights of the different losses in the full objective function are set to $\lambda_{adv} = 0.3$, $\lambda_{prc} = 1$, and $\lambda_{tv} = 1$, respectively.

4.3 Results

In this section, we evaluate the performance of the proposed EHDI model and compare it to the current state of the art in document binarization. It is important to note that the model was only trained on the Noisy Office (Zamora-Martínez et al., 2007) database, while the evaluation was conducted on external databases not included in the training. The results of our EHDI model are presented in Table 1 and compared to other approaches on the DIBCO 2013 dataset. Figure 2 also shows a qualitative comparison of the results on the DIBCO 2013 dataset, where it can be seen that the EHDI model produces a cleaner image quality than DE-GAN, especially when the degradation is very dense. This is because DE-GAN may struggle to remove such degradation from the document background in these cases.

Table 1: Results of our proposed EHDI on DIBCO 2013 dataset.

Model	PSNR	F-measure	F_{ps}	DRD
(Otsu, 1979)	16.6	83.9	86.5	11.0
(Niblack, 1985)	13.6	72.8	72.2	13.6
(Sauvola and Pietikäinen, 2000)	16.9	85.0	89.8	7.6
(Gatos et al., 2004)	17.1	83.4	87.0	9.5
(Su et al., 2012)	19.6	87.7	88.3	4.2
(Tensmeyer and Martinez, 2017)	20.7	93.1	96.8	2.2
(Xiong et al., 2018)	21.3	93.5	94.4	2.7
(Vo et al., 2018)	21.4	94.4	96.0	1.8
(Howe, 2013)	21.3	91.3	91.7	3.2
(Souibgui and Kessentini, 2020)	24.9	99.5	99.7	1.1
EHDI	**26.8**	**99.9**	**99.9**	**0.97**

To demonstrate the improvement achieved by our EHDI model, Figure 3 presents an example of the generated document image that is very close to and even superior to the ground truth image.

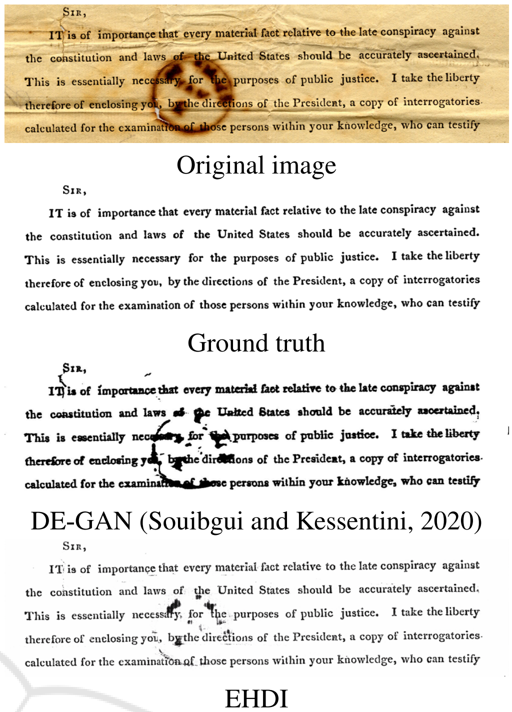


Figure 2: Example of degraded documents enhancement by our EHDI and DE-GAN on sample PR08 from DIBCO-2013 (Souibgui and Kessentini, 2020).

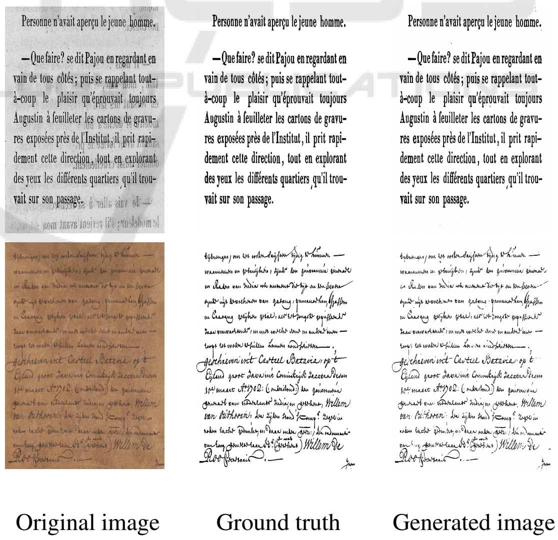
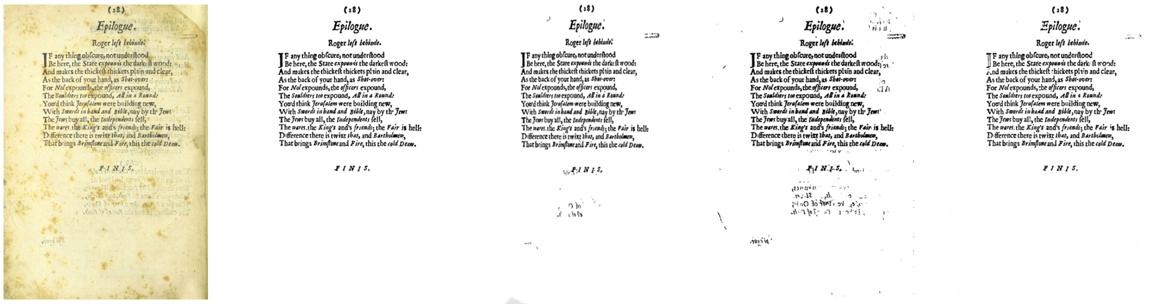


Figure 3: Example of enhancing degraded documents by our EHDI.

The proposed EHDI (enhanced document image) outperforms the state-of-the-art DE-GAN model (Souibgui and Kessentini, 2020). This is evident in the results of the 2017 DIBCO and 2018 H-DIBCO test sets, as shown in Table 2. An example of this superiority is shown in Figure 4, where EHDI outper-

Table 2: A comparative review of competitor approaches of DIBCO 2018 on DIBCO 2017 and DIBCO 2018 Datasets.

Model	DIBCO 2018				DIBCO2017			
	PSNR	F-measure	F_{ps}	DRD	PSNR	F-measure	F_{ps}	DRD
1 (Pratikakis et al., 2018a)	19.11	88.34	90.24	4.92	17.99	89.37	90.17	5.51
7 (Pratikakis et al., 2018a)	14.62	73.45	75.94	26.24	15.72	84.36	87.34	7.56
2 (Pratikakis et al., 2018a)	13.58	70.04	74.68	17.45	14.04	79.41	82.62	10.70
3b (Pratikakis et al., 2018a)	13.57	64.52	68.29	16.67	15.28	82.43	86.74	6.97
6 (Pratikakis et al., 2018a)	11.79	46.35	51.39	24.56	15.38	80.75	87.24	6.22
(Souibgui and Kessentini, 2020)	16.16	77.59	85.74	7.93	18.74	97.91	98.23	3.01
EHDI	**20.31**	**92.69**	**90.83**	**3.94**	**19.15**	**98.56**	**99.44**	**2.87**



Original image Ground truth Winner team DE-GAN EHDI

Figure 4: Qualitative binarization results on sample 16 in DIBCO 2017 dataset. Here, we compare the results of our proposed model with the winner's approach (Pratikakis et al., 2017) and DE-GAN (Souibgui and Kessentini, 2020).

Original image DE-GAN (Souibgui and Kessentini, 2020) EHDI

Figure 5: Example of qualitative results on HADARA dataset compared to DE-GAN enhancement approach (Souibgui and Kessentini, 2020).

forms the winner's method and DE-GAN on (sample 16) from DIBCO 2017. The winner's method used a U-net architecture and data augmentation techniques, while DE-GAN used a simple GAN network. Our model achieved better results due to the use of multiple loss functions that optimize the generator to produce images more closely aligned with the ground truth.

As shown in Figure 5, the visual quality of the enhanced document images on samples from the HADARA dataset (Pantke et al., 2014) is demonstrated. It is clear that EHDI consistently outperforms DE-GAN. To fairly evaluate our EHDI against state-of-the-art methods, we used the same sample as in the DE-GAN paper and compared EHDI to pix2pix-HD (Wang et al., 2018) and CycleGan (Zhu et al., 2017). The results, shown in Figure 6 and Table 3, demonstrate the superior performance of EHDI in terms of visual quality compared to cycleGAN, pix2pix-HD, and DE-GAN. In contrast to previous work in (Souibgui and Kessentini, 2020), which used three separate cGAN-based models for binarization, watermarking, and deblurring, this paper presents a single model that can handle all these tasks.

Table 3: Results of our proposed model on DIBCO 2018 dataset.

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
EHDI	**26.8**	**99.9**	**99.9**	**0.97**

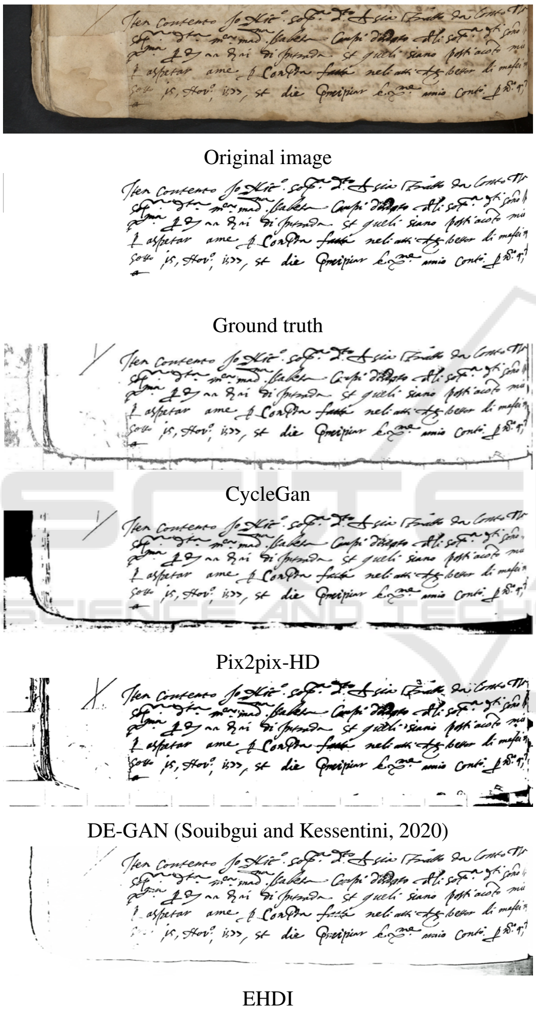


Figure 6: Example of qualitative enhancement results produced by different models of the sample (9) from the H-DIBCO 2018 dataset.

5 CONCLUSION

In this paper, we have put forward a conditional GAN as a means to generate clean document images from highly degraded images. Our suggested EHDI has been designed to handle different degradation tasks such as watermark removal and chemical degradation with the goal of producing hyper-clean document images and fine detail recovery performances.

Extensive experiments have shown the effectiveness of the proposed EHDI for cleaning extremely degraded documents. An interesting improvement for historical documents compared to many recent state-of-the-art methods on reference datasets.

Future work will include the adoption of the vision transformer techniques for a better document improvement process. In addition, we intend to add a recognition module to our framework to provide a comprehensive platform for processing historical documents.

REFERENCES

Aly, H. A. and Dubois, E. (2005). Image up-sampling using total-variation regularization with a new observation model. *IEEE Transactions on Image Processing*, 14(10):1647–1659.

Chen, X., He, X., Yang, J., and Wu, Q. (2011). An effective document image deblurring algorithm. In *CVPR 2011*, pages 369–376. IEEE.

Chou, C.-H., Lin, W.-H., and Chang, F. (2010). A binarization method with learning-built rules for document images produced by cameras. *Pattern Recognition*, 43(4):1518–1530.

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

Finlayson, G. D., Hordley, S. D., and Drew, M. S. (2002). Removing shadows from images. In *European conference on computer vision*, pages 823–836. Springer.

Gangeh, M. J., Plata, M., Motahari, H., and Duffy, N. P. (2021). End-to-end unsupervised document image blind denoising. *arXiv preprint arXiv:2105.09437*.

Gatos, B., Pratikakis, I., and Perantonis, S. J. (2004). An adaptive binarization technique for low quality historical documents. In *International Workshop on Document Analysis Systems*, pages 102–113. Springer.

Hedjam, R. and Cheriet, M. (2013). Historical document image restoration using multispectral imaging system. *Pattern Recognition*, 46(8):2297–2312.

Howe, N. R. (2013). Document binarization with automatic parameter tuning. *International journal on document analysis and recognition (ijdar)*, 16(3):247–258.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Jemni, S. K., Souibgui, M. A., Kessentini, Y., and Fornés, A. (2022). Enhance to read better: A multi-task adversarial network for handwritten document image enhancement. *Pattern Recognition*, 123:108370.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.

Mao, X., Shen, C., and Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems*, 29.

Marnissi, M. A., Fradi, H., Sahbani, A., and Essoukri Ben Amara, N. (2021). Thermal image enhancement using generative adversarial network for pedestrian detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6509–6516. IEEE.

Moghaddam, R. F. and Cheriet, M. (2009). A variational approach to degraded document enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1347–1361.

Niblack, W. (1985). *An introduction to digital image processing*. Strandberg Publishing Company.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Pantke, W., Dennhardt, M., Fecker, D., Märgner, V., and Fingscheidt, T. (2014). An historical handwritten arabic dataset for segmentation-free word spotting-hadara80p. In *2014 14th International Conference on Frontiers in Handwriting Recognition*, pages 15–20. IEEE.

Pratikakis, I., Gatos, B., and Ntirogiannis, K. (2013). Icdar 2013 document image binarization contest (dibco 2013). In *2013 12th International Conference on Document Analysis and Recognition*, pages 1471–1476. IEEE.

Pratikakis, I., Zagoris, K., Barlas, G., and Gatos, B. (2017). Icdar2017 competition on document image binarization (dibco 2017). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1395–1403. IEEE.

Pratikakis, I., Zagoris, K., Kaddas, P., and Gatos, B. (2018a). Icfhr 2018 competition on handwritten document image binarization (h-dibco 2018). *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 489–493.

Pratikakis, I., Zagoris, K., Kaddas, P., and Gatos, B. (2018b). Icfhr2018 competition on handwritten document image binarization contest (h-dibco 2018). In *International conference on frontiers in handwriting recognition (ICFHR). IEEE*, pages 1–1.

Sauvola, J. and Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern recognition*, 33(2):225–236.

Souibgui, M. A. and Kessentini, Y. (2020). De-gan: A conditional generative adversarial network for document enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Su, B., Lu, S., and Tan, C. L. (2012). Robust document image binarization technique for degraded document images. *IEEE transactions on image processing*, 22(4):1408–1417.

Tamrin, M. O., El-Amine Ech-Cherif, M., and Cheriet, M. (2021). A two-stage unsupervised deep learning framework for degradation removal in ancient documents. In *International Conference on Pattern Recognition*, pages 292–303. Springer.

Tensmeyer, C. and Martinez, T. (2017). Document image binarization with fully convolutional neural networks. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, volume 1, pages 99–104. IEEE.

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

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807.

Xiong, W., Jia, X., Xu, J., Xiong, Z., Liu, M., and Wang, J. (2018). Historical document image binarization using background estimation and energy minimization. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3716–3721. IEEE.

Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857.

Zamora-Martínez, F., España-Boquera, S., and Castro-Bleda, M. (2007). Behaviour-based clustering of neural networks applied to document enhancement. In *International Work-Conference on Artificial Neural Networks*, pages 144–151. Springer.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.