



# A Two-Stage Unsupervised Deep Learning Framework for Degradation Removal in Ancient Documents

Milad Omrani Tamrin<sup>(✉)</sup>, Mohammed El-Amine Ech-Cherif,  
and Mohamed Cheriet

École de technologie supérieure, 1100 Notre-Dame St W, Montreal, Qc, Canada  
{milad.omrani-tamrin.2,mohammed-el-amine.ech-cherif.1}@ens.etsmtl.ca,  
mohamed.cheriet@etsmtl.ca

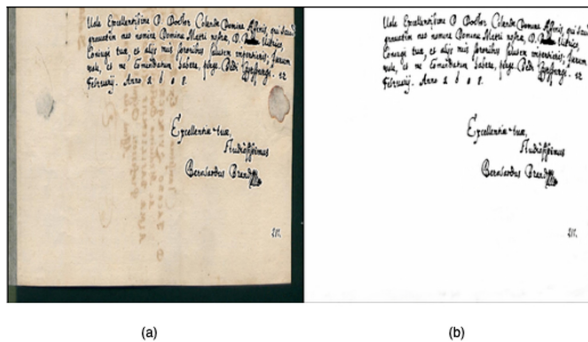
**Abstract.** Processing historical documents is a complicated task in computer vision due to the presence of degradation, which decreases the performance of Machine Learning models. Recently, Deep Learning (DL) models have achieved state-of-the-art accomplishments in processing historical documents. However, these performances do not match the results obtained in other computer vision tasks, and the reason is that such models require large datasets to perform well. In the case of historical documents, only small datasets are available, making it hard for DL models to capture the degradation. In this paper, we propose a framework to overcome issues by following a two-stage approach. Stage-I is devoted to data augmentation. A Generative Adversarial Network (GAN), trained on degraded documents, generates synthesized new training document images. In stage-II, the document images generated in stage-I, are improved using an inverse problem model with a deep neural network structure. Our approach enhances the quality of the generated document images and removes degradation. Our results show that the proposed framework is well suited for binarization tasks. Our model was trained on the 2014 and 2016 DIBCO datasets and tested on the 2018 DIBCO dataset. The obtained results are promising and competitive with the state-of-the-art.

**Keywords:** Historical documents · Degradation · Low-scale data set · Generative adversarial networks · Restoration

## 1 Introduction

Motivated by discovering hidden information in ancient manuscripts, the analysis of historical documents is considered an important and active research area in image processing/understanding, which has been the subject of much recent research. In traditional applications of historical document image processing, input data distributions are assumed to be noise-free, whereas due to various degradations such as shading, non-uniform shapes, bleed-through background,

and warping impacts in some corrupted images need to be removed and are considered as a pre-processing step [9]. Numerous techniques have been put forward to deal with naturally degraded images during the past few decades, and promising results have been obtained [5]. Deep learning models and mainly convolutional neural networks have recently outperformed traditional image processing techniques on numerous tasks ranging from computer vision, speech recognition, and times series forecasting. The latter models' tremendous success is due mainly to their reliance on computed or learned features from an extensive collection of images rather than handcrafted features obtained from the raw image pixels. Moreover, in both supervised and unsupervised learning settings, convolutional neural networks must be trained on massive data sets to achieve a better generalization error on unseen data. For image processing tasks, data augmentation is customarily and successfully used to provide arbitrarily large data sets for training convolutional neural networks. It is important to note that data augmentation must operate on clean images, i.e., not corrupted by degradation. We do not have much data for the deep network to capture the degradation process. The degradation will also damage the generated data and, therefore, does not improve the model's accuracy [11]. For the ancient image analysis tasks, data augmentation must not be restricted to only elementary geometric transformations. Instead, it requires geared towards reproducing the artifacts that the old document has been subjected to, such as ageing, spilled ink, stamps hindering some essential parts of the image, etc. [8]. The latter task requires advanced mathematical modelling and is beyond the scope of the present work. To overcome the lack of dataset in ancient manuscripts, GAN [7] provides a new perspective of synthesizing documents. We aim to leverage the deep learning paradigm to extract binarized images from the generated historical documents using the recently developed deep image prior [19], as shown in Fig. 1. This paper proposes a two-stage method. The first stage aims to generate realistic-looking high-quality documents by training the state-of-the-art generative models, Deep



**Fig. 1.** Restored degraded document designed by cleaning foreground from noisy background. a) original degraded document b) Restored image using proposed neural network based approach with an inverse function

Convolutional GAN (DC-GAN), on the DIBCO datasets. In the second stage, we adapt a deep image prior to the generated images, to produce binarized images to be evaluated on the 2016 version of DIBCO datasets. The contribution of this paper is three-fold:

- We propose a modified DC-GAN structure to synthesize more realistic data from ancient document images. The model generates high-resolution images.
- We adapt deep image prior to the generated images and develop a new loss function to perform image binarization.
- We validate the binarized images on the DIBCO datasets and we obtain competitive results in the 2014 and 2016 versions.

Section 2 presents in generative models and depth restoration using deep neural networks. Section 3 describes how we will compare them. Finally, we will review in Sect. 4, the results regarding the different measurements are discussed.

## 2 Related Work

Several methods have proposed to generate historical documents, while the corruption process in the case of large-scale damage becomes a complicated task to in-paint the lost area. Here we review some related works on augmenting historical documents. A Deep Learning algorithm [7] is proposed to generate an artificial dataset. Recently, GANs have become more interesting for synthesizing documents images. By having vanilla GAN, many approaches were introduced to synthesize document images such as GAN-CLS [17], where the proposed method consists of two neural networks. Generator  $G$  generates fake images, where the Discriminator  $D$  plays a discriminating role between  $G$  output and real images, including an auxiliary condition on text description  $\varphi(t)$ . However, a significant problem of such a method for document image augmentation would be a lack of substantial training data. The model requires a large number of images, including their text descriptions. Another version of GAN related to synthesis data for digital documents is [2] that the author proposed a style-GAN to synthesize alphabets that could predict the missed alphabets. However, the input needs labels, which are time-consuming and complex tasks. Despite the argument that generative models could be an advantage to overcome limited datasets, such a model can also increase damaged images' resolution. However, it is necessary to add a technique that can understand the available samples' underlying characteristics by considering the low quality of training images, as shown in Fig. 2. For multiple decades, inverse problems have been the subject of many studies in image restoration. Their success heavily depends on designing an excellent prior term to uncover the degraded images. The prior is usually hand-crafted based on specific observations made on a dataset. Creating a prior is often a difficult task as it is hard to model the degradation distribution. In the context of DL, the prior is learned from training a ConvNet on a vast dataset [3]. Most of the proposed methods using deep learning models only perform as good as the available datasets. The solution is tied up to the image space. In [19],



Fig. 2. Training samples from DIBCO degraded dataset [15]

the authors have shown that the structure of a ConvNet contains a great deal of information, and a prior can be learned within the weights of the architecture. In other words, exploring ConvNet weights’ space can result in recovering a clean image from a degraded image without the need to have a considerable dataset. Moreover, Processing Document Image Binarization (DIB) in historical image documents suffers in different challenges due to the nature of the old manuscripts that leads to degraded image either by faded or stains ink, bleed-through, document ages, documents quality and many other factors that may affect the historical documents. Therefore, the degradation manuscripts increase the challenging binarization process task since it requires classifying the foreground from background pixels as a pre-processing stage. That being said, the initial methods [13] used for classifying document image pixels (foreground vs background) are based on different single and multiple threshold values.

### 3 Work Methodology

#### 3.1 Stage I - Data Augmentation Framework

In this section, we will introduce DC-GAN, and provide technical information regarding the generative model.

**Deep Convolutional GAN (DC-GAN).** As of the original GAN’s general idea, in Deep Convolutional GAN (DC-GAN), the augmentation process is similar to the unique GAN but specifically concentrates on deep fully-connected networks. The model uses an adversarial game to solve generalization tasks. The generator is liable to create synthetic instances from random noises, and the discriminator tries to distinguish between fake and real images. By this adversarial process, the generator attempts to improve its weights and also generate images. The Convolutional-transpose layers try to do the feature extraction task by finding the correlated areas of images. The authors in [16] proposed that DC-GAN precisely fits for unsupervised-learning, whereas the original idea of GAN more relies on the public domain. Following the Eq. 1 in DCGAN, where the  $G$  utilizes the transposed technique to apply up-sampling of image size and allow to transfer the random noises into the shape of the input images. In  $D(x)$ , the ConvNet tries to find the correlated area of images.  $G(z)$  represents real data; the  $D(x)$  is also used to distinguish the difference between generated images versus real data using a classifier. The  $x$  is the samples of images from the actual dataset, and also the distribution of data is represented by  $P_{data(x)}$ .  $z$  is also a sample from the generator with the distribution of  $P(z)$ .

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{x} \sim P_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim P_{noise}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \tag{1}$$

The objective of training consists of two processes. In the first step, the discriminator updates parameters by maximizing the expected log-likelihood, and in the second step, while the discriminator parameters are updated, the generator generates fake images. The architecture used is given in Table 1. Hence, the input size of each image is  $(3 \times 128 \times 128)$ , the learning rate is considered 0.0002, batch-size is 256, and the number of epochs is 25k. To evaluate the performance of the generation effect of modified DC-GAN, we perform a quantities evaluation index called the Frechet Inception Distance network (FID) [10].

**Table 1.** Architecture of generative model used

Summary of DC-GAN		
Layer(type)	Output	Connected to
$input_1$ layer	None, 128, 512	
Lambda layer	512, $eps = 1e-05$	$input_1$
$input_2$ layer	512, $eps = 1e-05$	$input_2$ , lambda
Gan model	None, 0, 1	$input_1$ , lambda
yFake	None, 1	gan(1, 0)
yReal	None, 1	gan(1, 1)
Total Sample: 200	6 layers	Nb epochs: 25k

### 3.2 Stage II Convolutional Neural Network-Based Document Binarization

Several deep learning models have achieved state-of-the-art performance on binarization for degraded document analysis and printed machinery text [4, 20]. In our training, to get promising results from generative models, the first step is that the enhancement task is performed to improve the quality of degraded document images. However, for this process, it is necessary to train a learning model that requires many data. Indeed, there is a lack of big datasets to train a learning model when it comes to historical documents. To overcome the limitation, we explore a way that can allow us to enhance the quality of our images, using inverse problems. Inverse problems have been widely studied in document images but without promising results. Previously, the problem was formulated, and the goal was to look for the prior (inverse image).

In our approach, we will use the structure of a neural network proposed in [19]. Convolutional networks have become a popular tool for image generation and restoration. Generally, their excellent performance is credited to their ability to learn realistic image priors from many example images. In this stage, we adapt and extend the original deep image prior method to historical documents. We show that the structure of a generator ConvNet is sufficient to capture any information about the degradation of historical documents without any learning involved. To do so, we define a ConvNet architecture (U-Net) that is untrained. The network is then used as handcrafted prior to performing the text’s binarization from the background and hence removing the degradation.

In image restoration problems the goal is to recover the original image  $x$  having a corrupted image  $x_0$ .

Such problems are often formulated as an optimization task:

$$\min_x E(x; x_0) + R(x) \quad (2)$$

where  $E(x; x_0)$  is a data term and  $R(x)$  is an image prior.

The data term  $E(x; x_0)$  is usually easy to design for a wide range of problems, such as super-resolution, denoising, inpainting, while image prior  $R(x)$  is a challenging one. Today’s trend is to capture the prior  $R(x)$  with a ConvNet by training it using a large number of examples.

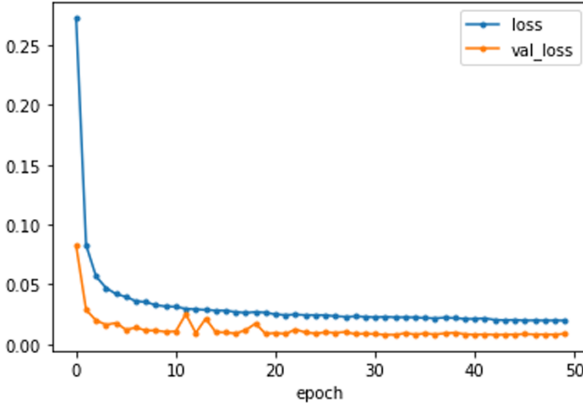
It is noticed, that for a surjective  $g : \theta \mapsto x$  the following procedure, in theory, is equivalent to 2:

$$\min_{\theta} E(g(\theta); x_0) + R(g(\theta))$$

In practice,  $g$  dramatically changes how the image space is searched by an optimization method. Furthermore, by selecting a “good” (possibly injective) mapping  $g$ , we could get rid of the prior term. We define  $g(\theta)$  as  $f_{\theta}(z)$ , where  $f$  is a ConvNet (U-Net) with parameters  $\theta$  and  $z$  is a fixed input, leading to the formulation:

$$\min_{\theta} E(f_{\theta}(z); x_0)$$

Here, the network  $f_\theta$  is initialized randomly and input is filled with noise and fixed. Figure 3 depicts the learning of the proposed networks. Moreover, the reduction of losses for training and validations proves that the model has improved and eliminate the noises from generated images.



**Fig. 3.** Training and validation Loss functions error getting minimized as the training epochs increase.

In other words, instead of searching for the answer in the image space we now search for it in the space of the neural network’s weights. We emphasize that only a degraded document image  $x_0$  is used in the binarization process. The architecture is shown in Fig. 4. The whole process is presented in Algorithm 1.

---

**Algorithm 1:** Deep GAN Perior

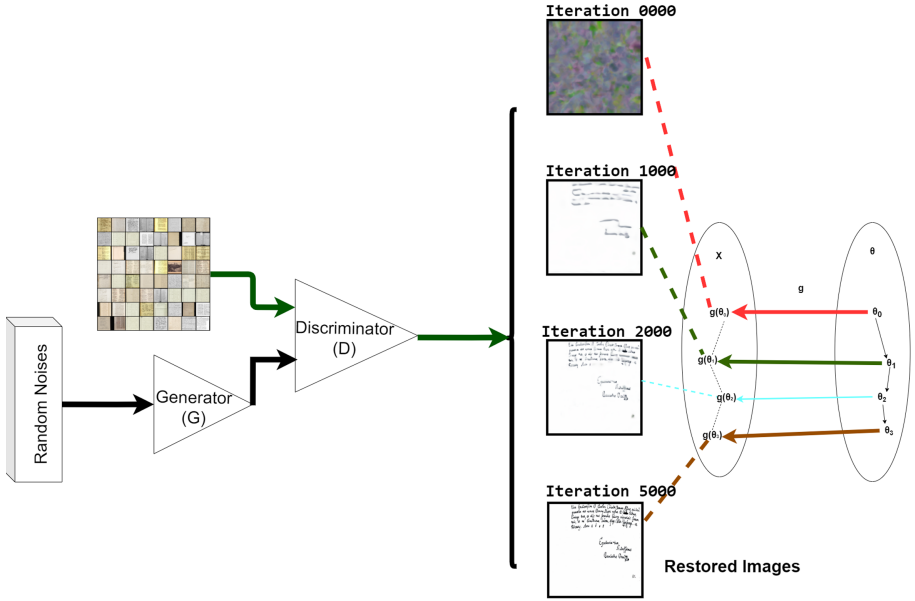
---

```

for number of training do
  for n steps do
    sample  $\{x^{(i)}\}_{i=1}^{i=m} \sim P_{data}$  a real sample
    sample  $\{z^{(i)}\}_{i=1}^{i=m} \sim P_z$  a restored sample
    update the discriminator by ascending its stochastic gradient
  end for
  sample  $\{z^{(i)}\}_{i=1}^{i=m} \sim P_z$  a restored sample
  Update the generator by ascending its stochastic gradient
end for
initialize generated document  $z^{(i)}$ 
take  $x^{(i)}$  min by  $\arg \min_{\theta} E(f_{\theta}(z))\hat{x}^{(i)}$ 
apply the gradient
result  $x^* = f_{\theta^*}(z^{(i)})$ 

```

---



**Fig. 4.** The proposed framework in two-fold: Stage-I on the left generates new synthetic images using DC-GAN, and stage-II on the right, removes degradation and perform binarization from generated images.

### 3.3 Datasets

To train and validate our developed methods, we used the most common image binarization datasets in handwriting documents, namely 2014 H-DIBCO [16] (Document Image Binarization Competition), 2016 H-DIBCO [17] and 2018 H-DIBCO [18], organized by ICFHR (International Conference on Frontiers in Handwriting Recognition) 2014, ICFHR 2016 and CFHR 2018 respectively. These benchmark datasets have been extensively used to train and validate the results of binarization algorithms in historical handwritten documents. The 2014 and 2016 H-DIBCO datasets are used to train our models, and the 2018 H-DIBCO is used to validate our results.

## 4 Result and Analysis

To evaluate our method, we adopt the benchmark historical handwritten dataset DIBCO described in Sect. 3.3. Moreover, we have tested the results for denoised images to understand the effectiveness of the proposed model. The document images in the dataset suffer from degradation. Furthermore, to further assess the performance of the method, we employ the four commonly used metrics to evaluate competitors in the DIBCO contest, namely F-Measure (FM), pseudo-FMeasure (Fps), PSNR, and Distance Reciprocal Distortion (DRD) [12].



The model learns well the general idea of historical document augmentation that can be noted in Fig. 5. By having the degraded samples, it is clear that the new documents have been enhanced, rather than using original generator methods. To evaluate the synthesized images in the recognition accuracy, we apply FID to measure the quality of generated images. FID computes the KL-divergence between real images distributions and generated image distributions. Table 2 shows the FID implies that both distributions are close to the real images.

**Table 2.** FID evaluation on generated distribution over real images distributions

Model	FID-5k	FID-20k
Real images	9.33	7.4
DC-GAN	36.73	6.45

Furthermore, the proposed method’s output could remove degradation and increase the accuracy of CNN, resulting in better classifications in document analysis. The encouraging results we obtained motivate more the effectiveness of data augmentation and the challenge of limited data and degraded in ancient documents taken from the basic GAN. A consequence of underlying GAN leads us to get more in-depth by using deep convolutional networks in the generator. Simultaneously, we were not convinced that we would get the same results from the underlying GAN. Our goal was to improve the quality of augmented document images. During this process, it was noted that DC-GAN provided better performance in the use of basic GAN. To also include, the PyTorch [14] framework was used during the discoveries.

**Table 3.** FM, Fps, PSNR and DRD evaluation and comparison with DIBCO 2018 winners

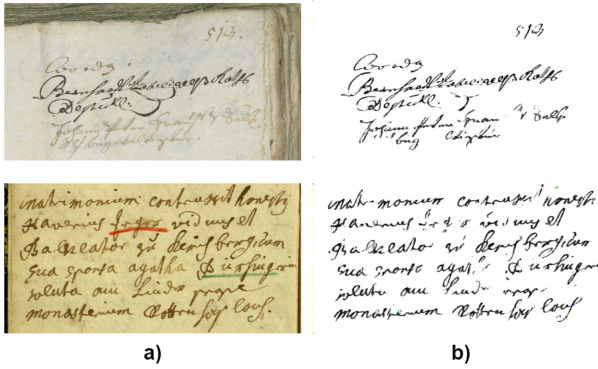
Method	FM	Fps	PSNR	DRD
Adak et al. [1]	73.45	75.94	14.62	26.24
Gattal et al. [6]	64.52	68.29	13.57	16.67
Saddami et al. [18]	46.35	51.39	11.79	24.56
Proposed method	<b>83.08</b>	<b>88.46</b>	<b>17.04</b>	<b>5.09</b>

The results obtained in this research paper are attributed to the design of good generative models and adapting newly discovered inverse-problem algorithms based on ConvNets. We constructed two new custom DC-GANs architecture. These architectures’ choice seems to work best because very deep networks are known to learn more features. In our paper, this seemed to be accurate and helped us get excellent results. Due to our efforts, we could generate realistic-looking synthetic document images by training our proposed DC-GANs. The



**Fig. 5.** Samples generated by our proposed method from the degraded documents augmentation by Stage-I

generator presents different transformations such as cropping to capture the documents' characters and resize the samples to normalize the data to meet the requirements for synthesizing high-quality images. Furthermore, to improve the augmentation task with unlabeled data, we alter the G and D networks for  $128 * 128$  size, including the extra Conv and pooling layers. To perform efficient binarization, we adapted and extended the original deep image prior algorithm to the problem. The unique deep image prior was developed to work on generated images. However, to the best of our knowledge, no one intended to work with ancient historical documents. The performance of previously developed state-of-the-art algorithms in binarization heavily relies on the data. Table 3 shows either state-of-the-art results or competitive results with the best binarization algorithms. The metrics shown in Table 3 are FM, Fps, PSNR and DRD. Our method outperforms all the algorithms that were used in the DIBCO 2018 competition. In this work, we showed that the space of a ConvNet contains valuable information about clustering the degraded ancient documents into two clusters, both background and foreground. As shown in Fig. 6, the proposed method shows promising results with removing noise from degraded document images.



**Fig. 6.** a) samples generated by our proposed DC-GAN b) binarized images produced by our method.

## 5 Conclusion

In this paper, a combined deep generative - image binarization model has been implemented and trained on degraded documents datasets. Our algorithm consists of two main steps. At first, an augmentation task is performed on unlabeled data to generate new synthetic samples for training. The second step is to remove the noise from generated images by taking the generators’ minimum error and removing the degradations. Experimental results have shown that the method was able to generate new realistic historical image documents. We performed binarization on the 2018 DIBCO dataset to validate our approach. The obtained results demonstrate that our method gets very close results to the 2018 DIBCO contest winner and vastly surpasses the other participants. Despite competitive results, there is room to improve our model by exploring different ConvNet architectures. We believe that the choice of the structure in the binarization task profoundly impacts our method’s performance. In future work, we will explore other ConvNet architectures as hyper-parameter model selection.

**Acknowledgement.** The authors thank the NSERC Discovery held by Prof. Cheriet for their financial support.

## References

1. Adak, C., Chaudhuri, B.B., Blumenstein, M.: A study on idiosyncratic handwriting with impact on writer identification. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 193–198. IEEE (2018)
2. Azadi, S., Fisher, M., Kim, V.G., Wang, Z., Shechtman, E., Darrell, T.: Multi-content GAN for few-shot font style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7564–7573 (2018)

3. Bui, Q.A., Mollard, D., Tabbone, S.: Automatic synthetic document image generation using generative adversarial networks: application in mobile-captured document analysis. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 393–400. IEEE (2019)
4. Calvo-Zaragoza, J., Gallego, A.J.: A selectional auto-encoder approach for document image binarization. *Pattern Recogn.* **86**, 37–47 (2019)
5. Dumpala, V., Kurupathi, S.R., Bukhari, S.S., Dengel, A.: Removal of historical document degradations using conditional GANs. In: ICPRAM, pp. 145–154 (2019)
6. Gattal, A., Abbas, F., Laouar, M.R.: Automatic parameter tuning of k-means algorithm for document binarization. In: Proceedings of the 7th International Conference on Software Engineering and New Technologies, pp. 1–4 (2018)
7. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Hedjam, R., Cheriet, M.: Historical document image restoration using multispectral imaging system. *Pattern Recogn.* **46**(8), 2297–2312 (2013)
10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems, pp. 6626–6637 (2017)
11. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
12. Lu, H., Kot, A.C., Shi, Y.Q.: Distance-reciprocal distortion measure for binary document images. *IEEE Signal Process. Lett.* **11**(2), 228–231 (2004)
13. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)
14. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)
15. Pratikakis, I., Zagori, K., Kaddas, P., Gatos, B.: ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018). In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 489–493 (2018). <https://doi.org/10.1109/ICFHR-2018.2018.00091>
16. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) (2015)
17. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. arXiv preprint [arXiv:1605.05396](https://arxiv.org/abs/1605.05396) (2016)
18. Saddami, K., Afrah, P., Mutiawani, V., Arnia, F.: A new adaptive thresholding technique for binarizing ancient document. In: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 57–61. IEEE (2018)
19. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
20. Vo, Q.N., Kim, S.H., Yang, H.J., Lee, G.: Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recogn.* **74**, 568–586 (2018)