Removal of Historical Document Degradations using Conditional GANs

Veeru Dumpala[1], Sheela Raju Kurupathi[1], Syed Saqib Bukhari[2] and Andreas Dengel[1,2]
[1] *University of Kaiserslautern, Germany*
[2] *German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany*

Keywords: Historical Documents, Degradations, Document Binarization, Conditional GANs.

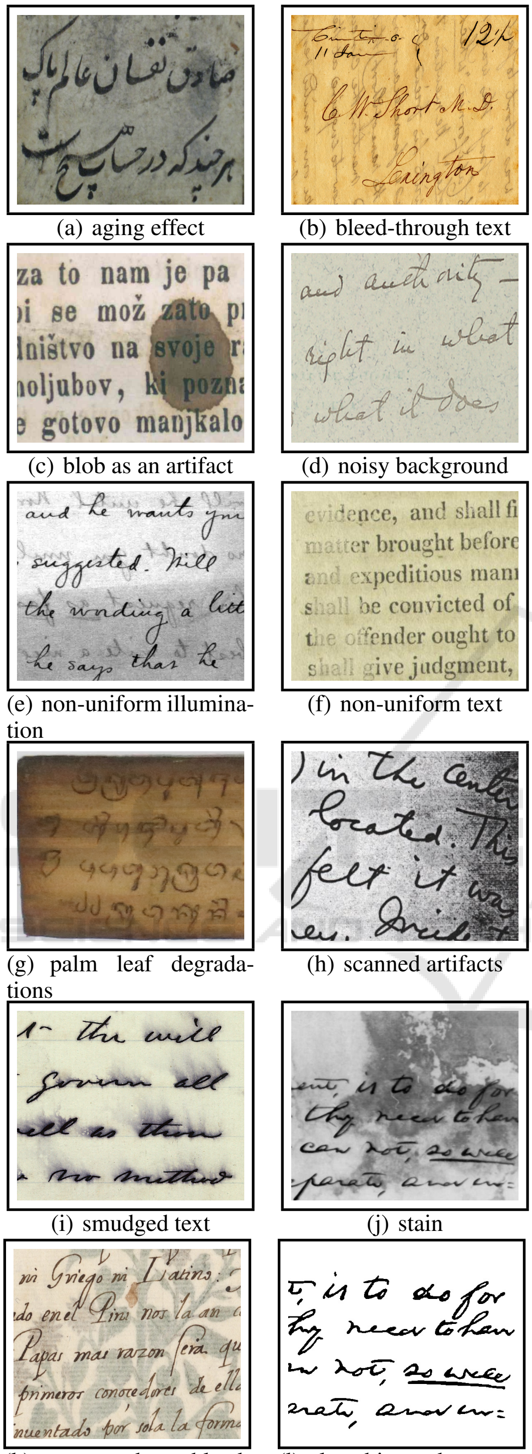
Abstract: One of the most crucial problem in document analysis and OCR pipeline is document binarization. Many traditional algorithms over the past few decades like Sauvola, Niblack, Otsu etc,. were used for binarization which gave insufficient results for historical texts with degradations. Recently many attempts have been made to solve binarization using deep learning approaches like Autoencoders, FCNs. However, these models do not generalize well to real world historical document images qualitatively. In this paper, we propose a model based on conditional GAN, well known for its high-resolution image synthesis. Here, the proposed model is used for image manipulation task which can remove different degradations in historical documents like stains, bleed-through and non-uniform shadings. The performance of the proposed model outperforms recent state-of-the-art models for document image binarization. We support our claims by benchmarking the proposed model on publicly available PHIBC 2012, DIBCO (2009-2017) and Palm Leaf datasets. The main objective of this paper is to illuminate the advantages of generative modeling and adversarial training for document image binarization in supervised setting which shows good generalization capabilities on different inter/intra class domain document images.

1 INTRODUCTION

Nowadays documents could be seen widely in many areas of our daily life and take the form of journals, manuscripts, invoices, quotes, contracts, certificates etc. Many document analysis pipelines for OCR (Bukhari et al., 2017), (Jenckel et al., 2016), (Breuel et al., 2013), (Breuel, 2008) require binarization as an initial step for pre-processing document images. These resulting binarized images will be further used by rest of the document analysis pipeline to transform the degraded document image into digital text. Binarization means separation of pixel intensity values into either black as a foreground or white as a background. There exists lot of challenges when generating the cleaner version of handwritten or machine-printed historical degraded documents like noise, non-uniform illumination, stains, non-uniform shadings etc (See Figure 1). Therefore, in order to extract the text from these noisy document images it is very important to differentiate the background from foreground text. In cleaned and scanned document images, it is very simple to achieve this but when we have noise in the documents, separating the background from the foreground pixels is really critical to achieve. To be successful in document binarization, one has to clean the historical artifacts while preserving the most meaningful content of the document image which can be seen as an ill-posed problem in document analysis. In this paper, we show that the proposed model learns the historical degradations and removes the noise while preserving most of the relevant information.

The most commonly used binarization techniques can be classified as global (Level Otsu, 1979), (Tensmeyer and Martinez, 2017), local (Niblack, 1986), (Mitianoudis and Papamarkos, 2015) and hybrid (Biswas et al., 2014), (Zemouri et al., 2014) thresholding. Global thresholding methods use a single threshold value for the entire document image. Local thresholding methods unlike global thresholding divide the image into blocks and use a local threshold value for each block of pixels. Hybrid thresholding methods use the combination of both local and global thresholding methods preserving the advantages of these methods. Examples of all these methods include Nick, Otsu, Sauvola, Niblack, Bradley, Bernsen, Local Adaptive thresholding etc. Although such techniques work well for normal degradations they fail in some cases of historical document degradations. Each method has its own pros and cons, we cannot claim that single technique is best suitable for all degraded

Dumpala, V., Kurupathi, S., Bukhari, S. and Dengel, A.
Removal of Historical Document Degradations using Conditional GANs.
DOI: 10.5220/0007367701450154
In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)*, pages 145-154
ISBN: 978-989-758-351-3



(a) aging effect (b) bleed-through text (c) blob as an artifact (d) noisy background (e) non-uniform illumination (f) non-uniform text (g) palm leaf degradations (h) scanned artifacts (i) smudged text (j) stain (k) textured bleed-through (l) clean binary document

Figure 1: Examples of machine-printed and handwritten text historical document degradations of DIBCO(b,c,d,e,f,h,i,j,k,l), PHIBC(a) and Palm Leaf(g) datasets, where (*l*) shows the clean binarized document image.

documents. There are also other techniques such as fuzzy logic based (Farahmand et al., 2017), gradient based (Pardhi and Kharat, 2017), RNNs (Westphal et al., 2018) etc. Selecting an optimal method for different degradations is still an open issue.

With the advent of deep learning approaches based on Convolution Neural Networks (CNNs) many problems related to binarization have been easily tackled using different architecture designs and training procedures (Tensmeyer and Martinez, 2017), (Ayyalasomayajula et al., 2018). One such emerging model which is being vastly studied and researched in the recent years is Generative Adversarial Network (GAN) (Goodfellow et al., 2014). These GANs have been applied in large image domains for solving different tasks like image generation, image manipulation, semantic manipulation etc,. Even though there are quiet few attempts in the recent years, most of the research on GANs is only restricted to natural images (Wang et al., 2018), (Zhu et al., 2017) and (Kim et al., 2017). In this paper, we show the usage and advantages of conditional GANs for document image binarization problem which can be treated or seen as high-resolution document image manipulation task. Major problem in applying deep learning methods to solve document binarization is to acquire clean ground truth of degraded documents. This could be solved mostly by using well-known document image degradation techniques like Ocrodeg (Breuel, 2018) and applying these degradations on already available UW-III (Phillips, 1996) and UNLV (Rice et al., 1996) clean datasets. Although using synthetic dataset could vastly solve acquiring ground truths of historical degraded documents, this in turn can make the learning model vulnerable to overfit to synthetically generated datasets and cannot generalize well with real world historical documents.

In this paper, we show that the proposed model can also binarize well using synthetic datasets for training when applied on unseen real-world document images. The proposed conditional generative adversarial model tries to learn the document degradations by mapping 1024x1024x3 degraded color images to 1024x1024x1 gray scale images. Further, while testing we binarize the resulting grayscale image using default 127 pixel value as a threshold to get binarized image. Based on recent approaches using global-to-local binarization techniques (Biswas et al., 2014), we make use of multi-resolution generator architecture of pix2pixHD model (Wang et al., 2018) for document image binarization. Here it has to be noted that, the proposed model can be easily trained end-to-end unlike other global-to-local binarization techniques. The main idea behind using the output of the proposed model to map to 1-channel instead of binary is that, the grayscale representation space allows the model to learn robustly and make decision based on the confidence values rather than just pixel-classification of the historical degraded image into foreground or background. Recent approaches using grayscale representation of the degraded documents had shown to improve the document image binarization on historical documents (Calvo-Zaragoza and Gallego, 2018), (Peng et al., 2017), (Hedjam et al., 2015). Unlike the original semantic manipulation model, our final proposed model also uses F-measure as an error function (Pastor-Pellicer et al., 2013). Finally, we show our binarization results on publicly available historical degraded datasets DIBCO-2017 (Pratikakis et al., 2017), PHIBC2012 (Ayatollahi and Nafchi, 2013) and Palm Leaf (Burie et al., 2016) which depicts that the proposed method for document image binarization outperforms recent state-of-the-art methods both quantitatively and qualitatively.

2 RELATED WORK

Over the years, various methods have been proposed and researched widely for document image binarization problem. From the perspective of this paper, they can be classified into two classes: data-driven based and heuristic based approaches. Though non-data-driven approaches work well for normal degraded documents they fail to achieve good binarization results on highly historically degraded documents. This made the document analysis community to focus on data driven approaches for document image binarization.

Majority of global and local/adaptive thresholding methods have been proposed over past few decades to solve binarization problem. Otsu (Level Otsu, 1979) is one such popular global method from image processing community, where it calculates single optimal threshold value to convert grayscale image to binary image. Sauvola (Sauvola and Pietikäinen, 2000) is also one such local method where it takes the context of local neighborhood for binarizing documents. Despite that they provide good results for normal document but they fail to supply justifiable output if a document contains degradations. These methods cannot even be acceptable in complex historical degradations scenario such as smudges, bleed-throughs, non-uniform shadings, stains etc. This made the researchers of document analysis and recognition community to focus on local thresholding approaches. They can vary from simple window-based techniques to pixel level classification.

From the progress of deep learning on different Computer Vision tasks, successful methods have been acquired and adapted for documents. SAE (Calvo-Zaragoza and Gallego, 2018) uses convolutional auto-encoder where the output activations indicates the likelihood of a pixel to be either foreground or background. (Peng et al., 2017) is another such encoder-decoder network. PDNet (Ayyalasomayajula et al., 2018) is based on network architecture that uses FCNs with an unrolled primal-dual network. Though these models outperform other hand-crafted or non-data driven models, the results from these networks are still qualitatively low. Going from global to local, these models lose the global information which could sometimes be useful to make predictions of a pixel into foreground or background (See Figure 4). This could be seen as the Global-to-Local generalization problem. For the past few years, researchers proposed global-to-local binarization approaches that make use of pixel information both globally and locally to threshold the degraded documents. Despite these binarization algorithms use local and global features they are still far from generalizing well to different inter class domain degraded documents. With the success of high-resolution image synthesis using conditional generative adversarial networks (Wang et al., 2018), (Gulrajani et al., 2017), we make our proposed model to use similar network architecture for document image binarization. Here, the primary goal of generator is to provide binary result of the input degraded document. We also show that, our model trained on synthetic datasets could even generalize well to real world historical document datasets. The main idea behind using GANs for document image binarization task is that the generator does not see the binary version of the corresponding degraded document rather learns to differentiate between good and bad binarization images with the help of multi-scale discriminators. Incorporating additional losses like F-Measure loss (Pastor-Pellicer et al., 2013) in the GAN objective function will also provide better gradient flow and faster convergence during training.

With the help of DIBCO contests (Gatos et al., 2009), (Ntirogiannis et al., 2014) and (Pratikakis et al., 2017), we had the chance to benchmark many different binarization approaches on a single scale using widely accepted evaluation metrics like Precision, Recall, F-Measure, pseudo-F-Measure, Peak Signal-to-Noise Ratio (PSNR) and Distance Reciprocal Distortion (DRD). As we can see from the recent approaches, one measure alone will not provide information about how well the binarization algorithm works. By this we can say that the defined models should not only provide better quantitative results but also qualitative as this could improve the performance of overall document image analysis pipeline systems.

3 DATASETS

As we know that, the dataset plays a critical role in training the deep learning models which in turn influences these overall performance of the models. The vital task of training deep models is to create sufficient amount of training data so that, the model could learn efficiently. In this paper, we present various synthetically generated, publicly available and real-world datasets like UW-III, PHIBC, DIBCO and Palm Leaf which vary in their sizes (ranging from 300x400 to 2500x3300), fonts, styles and have concrete degradations which may be due to aging effects, bleed-throughs and physical damages because of corrosion and fire. One of the challenges of historical document image binarization is to gather the ground truth. But, this could be solved partly by using publicly available libraries such as Ocrodeg (Breuel, 2018) or by simply applying alpha-channel blending on clean document images. For creating synthetic dataset (See Figure 2), we have used UW-III dataset that contains 1600 document images which are clean from the perspective of Optical Character Recognition (OCR) pipeline. We have applied various Ocrodeg degradations and alpha-channel blending for creating bleed-through degraded images. The total of 1500 corresponding images are used for training and the rest for testing. Even with manually generated degraded documents, we could not attain few of the most challenging historical degradations such as non-uniform shading, smudges and uneven pen strokes etc. Therefore, we used publicly available datasets of DIBCO from 2009 to 2016, PHIBC2012, Palm Leaf for training and benchmarked the results on DIBCO-2017 dataset.

4 DEEP LEARNING MODELS

Today, deep learning models are used to solve various problems in the fields of Computer Vision, Image Processing, Robotics, Social Networking, Astronomy etc. Most commonly used deep learning architectures are AlexNet, ResNet, Google Net etc., We have used Generative Adversial Network architecture to solve our problem related to document degradation. GANs (Goodfellow et al., 2014) are special class of artificial intelligence algorithms which consists of mainly two neural networks Generator and Discriminator. Generator network is responsible for generating the synthetic instances from random noise or conditioned on the input image. Discriminator is used for evaluating the synthetic instances by minimizing the loss to its original input. Both generator and discriminator networks compete with each other to minimize the losses such that synthesized data is as similar as real data. We can say that these networks model and learn to mimic the data distribution.

We have seen many applications of generative adversarial networks in which one of them is to take semantic label maps as input and generate the photo-realistic images. For our problem, we have chosen Image-to-Image translation GANs which take the input from one domain and translate it into another domain. So, the degraded documents are converted to clean binarized documents.

4.1 Conditional GANs

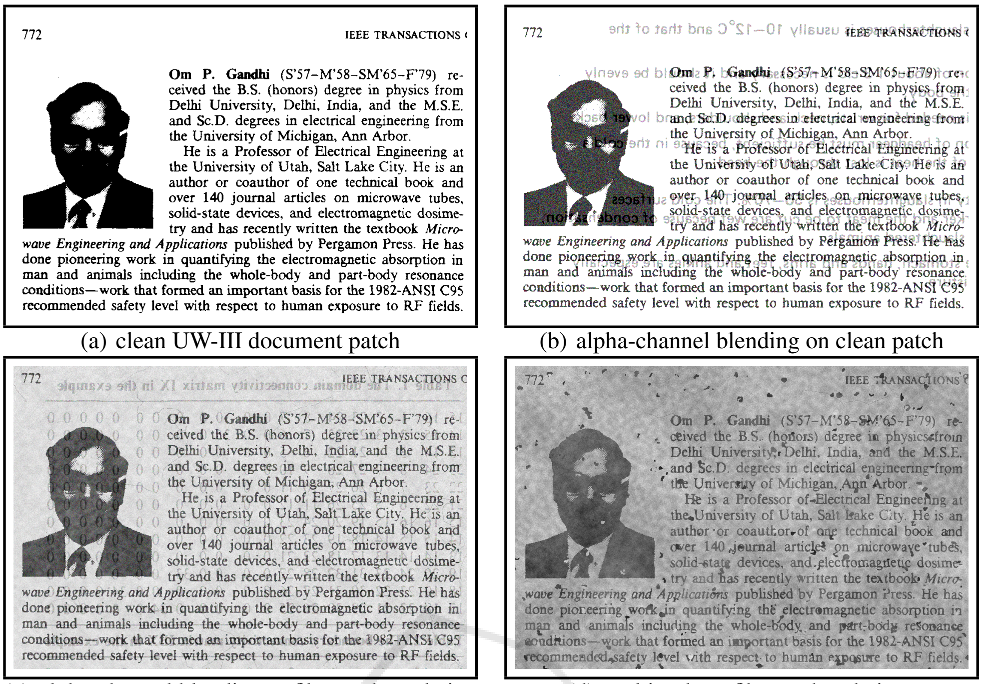
The objective of Conditional GANs is to model the conditional distribution of real images given the degraded images. pix2pix (Isola et al., 2017) method uses U-Net as the generator G and a patch-based fully convolutional network as the discriminator D. We need to have the corresponding clean and degraded document image pairs in the supervised setting. But the main drawback of this model was that it couldn't generate high-resolution images with good quality, it looses the finer details and also the training was unstable. So, to overcome the disadvantages of pix2pix model we have used modified model pix2pixHD (Wang et al., 2018) which consists of a coarse-to- fine generator, a multi-scale discriminator, and a robust adversarial learning objective function.

4.2 Proposed Model

With the inspiration from high resolution image synthesis, we propose document image binarization technique which uses the previously stated architecture as the baseline model.

4.2.1 Architectural Details

The building block of proposed binarization framework for generating cleaned version of degraded documents is an auto-encoder with residual blocks. As stated by the baseline model, we make use of coarse-to-fine generator architecture to achieve better binarization results even on high resolution degraded document images. From the evaluation point of view, we make a statement that some of the publicly available datasets can only obtain best accuracy when the local receptive field of model is 256x256. Because of this problem, there were more generalization errors


772 IEEE TRANSACTIONS (

Om P. Gandhi (S'57–M'58–SM'65–F'79) received the B.S. (honors) degree in physics from Delhi University, Delhi, India, and the M.S.E. and Sc.D. degrees in electrical engineering from the University of Michigan, Ann Arbor.

He is a Professor of Electrical Engineering at the University of Utah, Salt Lake City. He is an author or coauthor of one technical book and over 140 journal articles on microwave tubes, solid-state devices, and electromagnetic dosimetry and has recently written the textbook *Microwave Engineering and Applications* published by Pergamon Press. He has done pioneering work in quantifying the electromagnetic absorption in man and animals including the whole-body and part-body resonance conditions—work that formed an important basis for the 1982-ANSI C95 recommended safety level with respect to human exposure to RF fields.

(a) clean UW-III document patch
(b) alpha-channel blending on clean patch
(c) alpha-channel blending + fibrous degradation
(d) multiscale + fibrous degrdations

Figure 2: Examples of synthetically generated dataset from UW-III (Phillips, 1996), alpha-channel blending and Ocrodeg (Breuel, 2018) document degradations.

in the recent state-of-the-art methods when tested on other inter domain datasets. However, we can solve this problem by using incremental training.

We propose conditional generative adversarial network that contain generator and discriminator modules. The generator module can be further classified into 2 sub-networks $\{G1\}$ (See Figure 3) and $\{G2\}$ (See Figure 4) where both are based on auto-encoder architecture with residual blocks (Johnson et al., 2016) where one works on top of the other except that the local receptive field of two sub-networks varies from 512 and 1024 respectively. This topology of generator for document image binarization is considered to not only work with lower resolution documents but also on high resolution historically degraded images.

Similar to generator module, the discriminator module consists of multi-scale discriminators where each discriminator works at different scale. For generator sub-module $\{G1\}$, we use 2 discriminators where one works at model resolution and other works at by a factor of 2. The whole generator i.e., $\{G1, G2\}$ for generative adversarial framework uses 3-scale discriminators by factor of 2 and 4. We still downsample the real and synthesized cleaned versions of degraded documents before giving them to multi-scale discriminators to differentiate between real and fake degradations-free document images. This allows the generator to learn the historical document degradations at different scales and efficiently generate better binary image.

4.2.2 Model Extension

The proposed model based on conditional GAN framework has input and output resolutions 1024x1024x3 and 1024x1024x1 respectively. But still, the model output is a gray-scale channel image. So, we perform the global-thresholding on the resulted gray-scale image with fixed global value of 127. This is to make sure that the our model learns to differentiate between foreground and background pixel more robustly.

4.2.3 F-Measure as Loss Function $\mathcal{L}_{f-m}$

Recently with advances in Generative Adversarial Framework, we see many attempts in defining the GAN value objective function with different losses. Here, the different loss functions are used to tackle different optimization problems. The baseline model pix2pixHD not only uses GAN loss but also feature matching loss and VGG loss. In the proposed work, as we are using the GAN framework for document image binarization so we replace the VGG loss with f-measure error function, which previously shown to obtain better binarization results on challenging

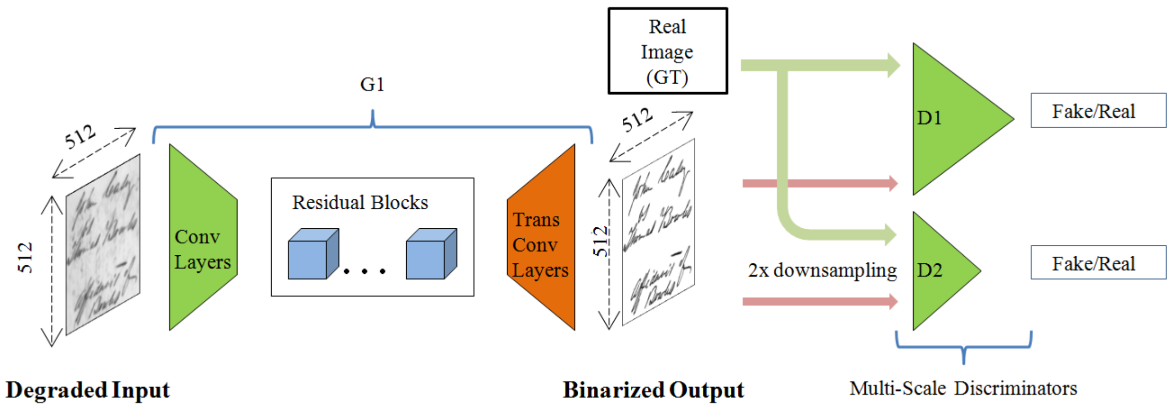


Figure 3: Global network with Convolution Layers, Residual Blocks and Transpose-Convolution Layers as Generator $\{G1\}$ and Multi-Scale Discriminators D1 and D2, which work at 512 X 512 and 256 X 256 respectively.

datasets (Pastor-Pellicer et al., 2013). We also show the evaluation results with and without $\mathcal{L}_{f-m}$ in GAN scenario.

Overall objective function of the proposed model can be described as below:

$$\min_{G}((\max_{D_i,..D_k}\sum_{i=1}^{k}\mathcal{L}_{GAN}(G,D_i))+\lambda\sum_{i=1}^{k}\mathcal{L}_{FM}(G,D_i)+ \sum_{i=1}^{k}\mathcal{L}_{f-m}(G,D_i)) \quad (1)$$

where $\mathcal{L}_{FM}$ is a feature matching loss and $\mathcal{L}_{f-m}$ is a f-measure loss.

4.2.4 Training Procedure

Initially, we train $\{G1\}$ and $\{G2\}$ with the help of respective multi-scale discriminators in the defined order of their resolutions and fine-tune all the networks accordingly. By the help of this multi-resolution pipeline, the proposed document image binarization model works well on wide-range of historically degraded documents.

5 EXPERIMENTS

In this section, we show the evaluation results of the proposed model with different configurations and compare them with other state-of-the-art methods for document image binarization. In 5.2.1, we provide the model performance with and without f-measure as error function. In 5.2.2, we show the test results of obtained generator with varying input resolutions. In 5.2.3, and 5.2.4, we present both quantitative and qualitative evaluation results on various publicly available datasets and also on the synthetically generated dataset.

5.1 Evaluation Metrics

For benchmarking the proposed model performance, we incorporate widely-known evaluation metrics for document binarization from previous document image binarization contests (DIBCO) like F-Measure (harmonic mean of Precision and Recall), DRD (measures the visual distortion of binary document images), pseudo-F-Measure (harmonic mean of pseudo-Precision and pseudo-Recall which uses weighted-distances to GT contours) and PSNR (computes peak signal-to-noise ratio between GT and predicted binary document images). Though these measures are widely accepted, we like to also show the qualitative results of generator with adversarial training that are visually appealing compared to present state-of-the-art approaches.

5.2 Quantitative and Qualitative Results

5.2.1 With and Without $\mathcal{L}_{f-m}$

As we are working on historical document images where the final goal of the proposed model is to output a clean version or binary image without any degradations, we make use of f-measure error function and compare it in the generative adversarial framework. From the Table 1, it is clear that the accuracy drops by 1-2% when GAN loss and Feature matching loss alone considered. This explains about the importance of $\mathcal{L}_{f-m}$ when building deep learning architectures for document image binarization where the sole task could be generalized to obtain relevant (for foreground) and non-relevant (for background) pixels from the degraded documents. We also provide comparison of f-measure accuracies on various de-

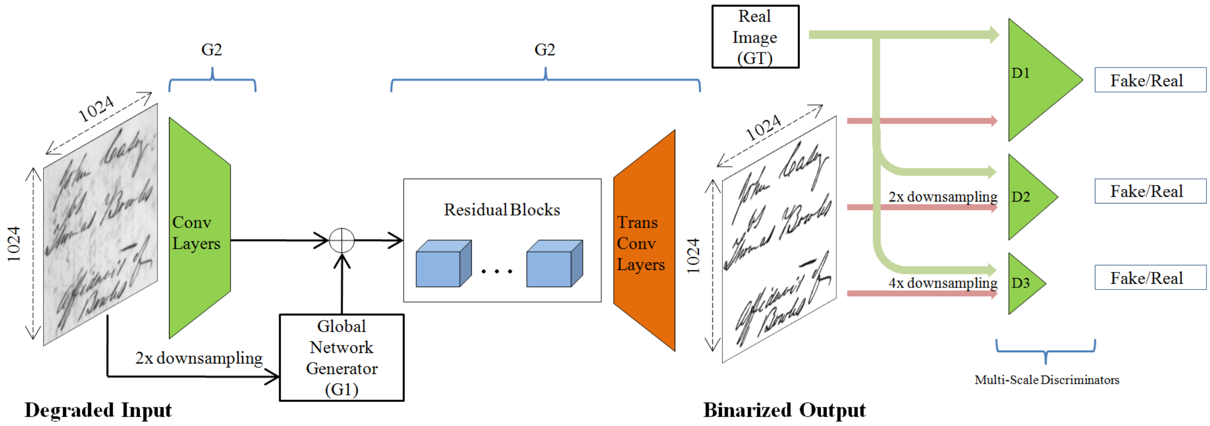


Figure 4: Local enhancing network with Convolution Layers, element-wise sum between intermediate G2 feature maps and G1 last feature maps, Residual Blocks and Transpose-Convolution Layers as Generator $\{G1, G2\}$ and Multi-Scale Discriminators D1, D2 and D3, which work at 1024 X 1024, 512 X 512 and 256 X 256 respectively.

fined datasets (Calvo-Zaragoza and Gallego, 2018) in the Table 2. In the case of DIBCO datasets (D14 and D16), we use all the images from the rest of the contest editions as training set. PL-I and PL-II are already provided with train and test partitions. While in the case of PHIBC, we randomly split the corpora into train and test partitions, with 80% and 20% respectively. It is clear that the model trained on PHIBC, PL-I and PL-II with $\mathcal{L}_{f-m}$ generalizes well with D14 and D16 test sets. This shows the better generalization advantage of adversarial training over other state-of-the-art end-to-end learning approaches. For the Palm-Leaf datasets, we show $\{G1\}$ and $\{G1, G2\}$ results separately because the aspect ratio of the PL-I and PL-II datasets varies substantially compared to other datasets.

Table 1: Comparison of our proposed model with and without $\mathcal{L}_{f-m}$ on DIBCO and PHIBC datasets based on F-Measure accuracy.

Trained on	Tested on DIBCO-2017
DIBCO 2009-2016 w and w/o $\mathcal{L}_{f-m}$	89.2 / 87.1
PHIBC w and w/o $\mathcal{L}_{f-m}$	83.8 /82.77

5.2.2 $\{G1\}$ and $\{G1, G2\}$ Evaluation Results

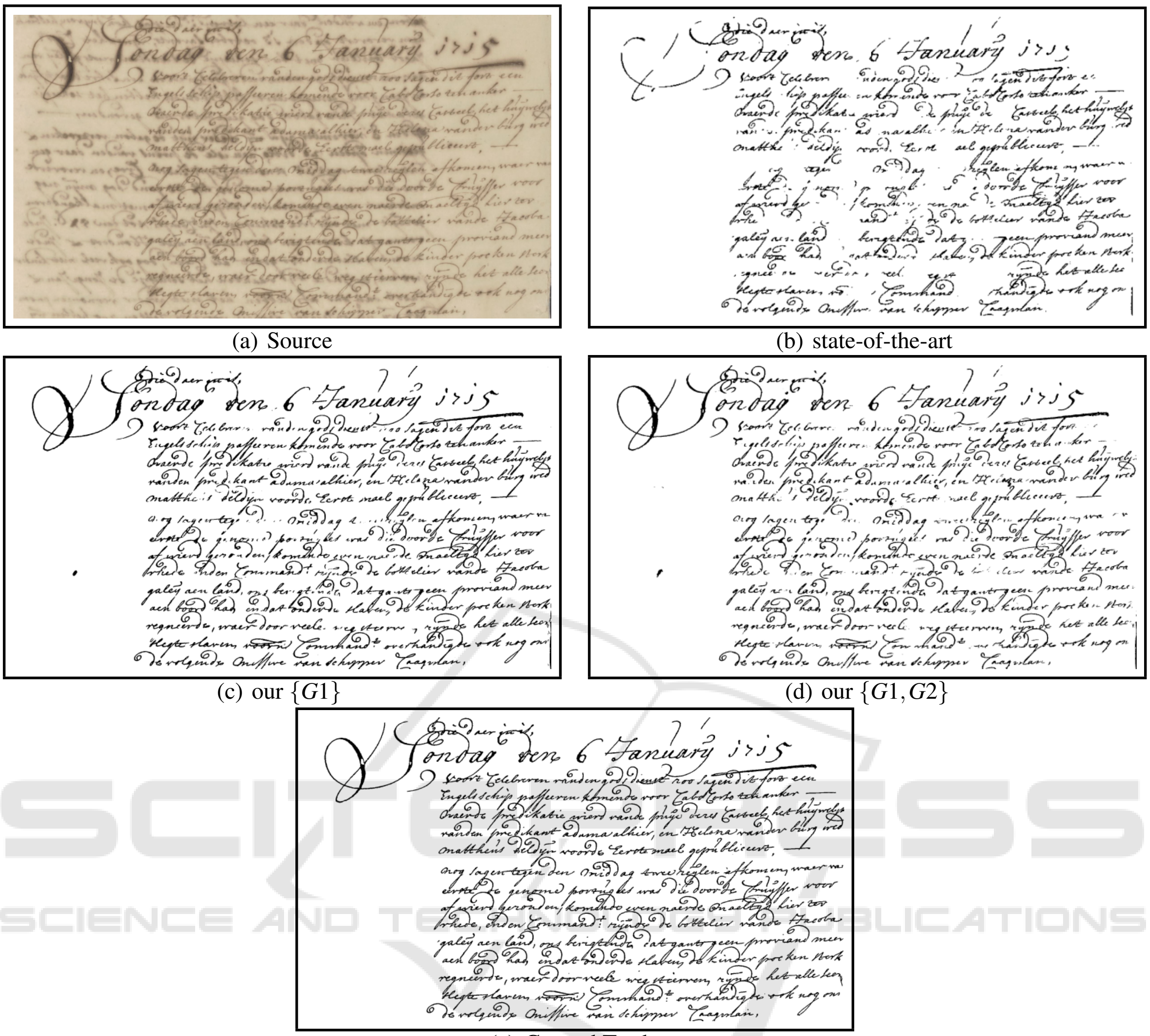
From Table 3, we can see that the proposed model alone with $\{G1\}$ outperforms the DIBCO 2017 benchmark challenge winner (Pratikakis et al., 2017) which is based on U-Net convolutional architecture by 0.5% in accuracy. Here, the training set consists of previous DIBCO contests datasets from 2009 to 2016 and tested on DIBCO 2017. From Figure 5 and Figure 6, it should also be noted that the obtained results not only outperform the other method quantitatively but also qualitatively. However, increasing the model size by fine-tuning with G2 made the model prediction to drop the f-measure accuracy by 1.42%. But the qualitative results are still better in comparison.

5.2.3 On Synthetically Generated Data and Manually Collected Dataset

Here, we compare the proposed model which is trained on synthetically generated degraded documents using libraries like Ocrodeg on UW-III and UNLV datasets with percentile based method (Afzal et al., 2013), which is a robust non-deep learning-based approach for document image binarization. We even achieve 80.7% f-measure accuracy on the most challenging DIBCO 2017 binarization dataset (see Table 4). This implies that the model trained in adversarial manner provides less generalization error and avoids over-fitting to the training set i.e., in this case to synthetically generated degraded dataset. It should be noted that the model used for the evaluation do not incorporate $\mathcal{L}_{f-m}$ while training. But with the inclusion to full objective function, we obtained f-measure of 81.2%.

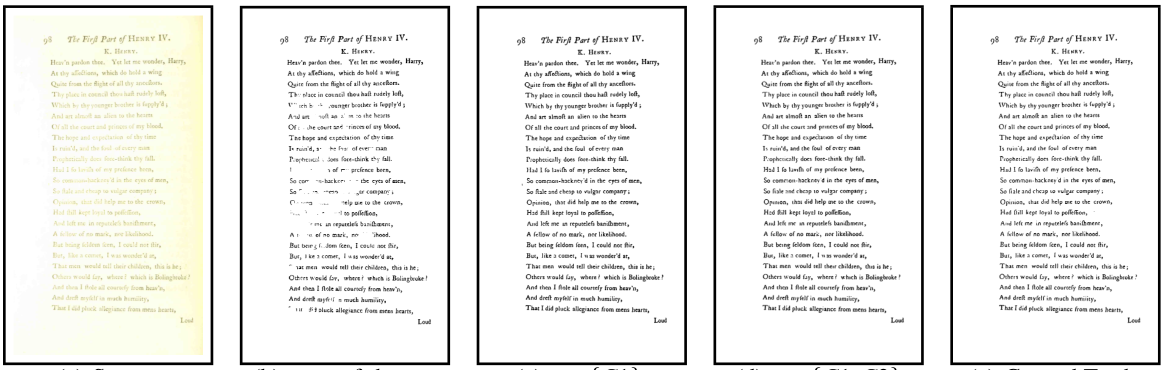
6 CONCLUSION

In this paper, we presented the conditional generative model that exploits the power of deep neural networks for removing various challenging degradations from historical documents like stains, bleed-through, non-uniform shadings etc,. to obtain high-resolution binarized result. Here, we also observed that integrating $\mathcal{L}_{f-m}$ to the objective function enhanced the learning behaviour of the model. Without much data augmentation (we used only horizontal flips), the proposed



(a) Source (b) state-of-the-art (c) our $\{G1\}$ (d) our $\{G1, G2\}$ (e) Ground Truth

Figure 5: Qualitative Results of the proposed model which is trained on DIBCO 2009-2016 and evaluated on DIBCO-2017 Handwritten Text outperforms recent state-of-the-art (Pratikakis et al., 2017).



(a) Source (b) state-of-the-art (c) our $\{G1\}$ (d) our $\{G1, G2\}$ (e) Ground Truth

Figure 6: Qualitative Results of the proposed model which is trained on DIBCO 2009-2016 and evaluated on DIBCO-2017 Machine-Printed Text outperforms recent state-of-the-art (Pratikakis et al., 2017).

Table 2: Proposed model trained and tested on defined datasets (Calvo-Zaragoza and Gallego, 2018). Asterisk (*) indicating better performance over other method.

Train	**Test**					
	D14	D16	PL-I	PL-II	PHIBC	Avg
D14	92.4	-	49.72	49.41	89.32	70.21
D16	-	85.2	46.09	45.78	89.11*	66.55
PL-I	55.46*	50.05*	49.81	49.9	60.66	53.18
PL-I (only $\{G1\}$)	57.56*	46.5*	57.6	57.6	37.67	51.39
PL-II	57.7*	54.59*	53.06	53.2	54.48	54.61
PL-II (only $\{G1\}$)	60.28*	55.77*	57.18	57.27	48.60	55.82
PHIBC	90.18*	85.81*	52.4*	52.53*	87.73	73.73

Table 3: Quantitative results of $\{G1\}$ and $\{G1, G2\}$ which are trained on DIBCO 2009-2016 and evaluated on DIBCO-2017 outperforms recent state-of-the-art (Pratikakis et al., 2017) in document image binarization.

Metric	**only $\{G1\}$**	$\{G1, G2\}$	**state-of-the-art**
F-Measure	91.53*	90.11	91.04
pseudo-F-Measure	94.2*	91.72	92.86
DRD	2.820*	3.803	3.40
PSNR	18.241*	17.544	18.28

Table 4: F-Measure results of our proposed model trained on synthetic dataset with and without $\mathcal{L}_{f-m}$ and compared to percentile based method.

Method	**Tested on DIBCO 2017**
Ours + w and w/o $\mathcal{L}_{f-m}$	81.2 / 79.9
(Afzal et al., 2013)	82.3

model converges and also generalizes well to never seen data. We showed that our model outperforms recent state-of-the-art for document image binarization by providing benchmark results on publicly available and manually acquired datasets. From our experiments, we conclude that the previous state-of-the-art models which were trained end-to-end without coarse-to-fine architecture are prone to a problem which we state that as Global-to-Local generalization problem. We also presented results which depict that our model has qualitative improvement over other methods. We have exhibited the pros of generative modeling with adversarial training for document image binarization in supervised and incremental setting which provides good generalization capabilities on different inter and intra class domain document images. In the future work we would like to further improve the model efficiency by optimizing its architecture. Although significant work has been done on using deep learning architectures for document binarization there is still lot to be explored where these models can be adapted for solving several handwritten or machine-printed document analysis and recognition problems.

REFERENCES

Afzal, M. Z., Krämer, M., Bukhari, S. S., Yousefi, M. R., Shafait, F., and Breuel, T. M. (2013). Robust binarization of stereo and monocular document images using percentile filter. In *International Workshop on Camera-Based Document Analysis and Recognition*, pages 139–149. Springer.

Ayatollahi, S. M. and Nafchi, H. Z. (2013). Persian heritage image binarization competition (phibc 2012). *arXiv preprint arXiv:1306.6263*.

Ayyalasomayajula, K. R., Malmberg, F., and Brun, A. (2018). Pdnet: Semantic segmentation integrated with a primal-dual network for document binarization. *Pattern Recognition Letters*.

Biswas, B., Bhattacharya, U., and Chaudhuri, B. B. (2014). A global-to-local approach to binarization of degraded document images. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 3008–3013. IEEE.

Breuel, T. M. (2008). The ocropus open source ocr system. In *Document Recognition and Retrieval XV*, volume 6815, page 68150F. International Society for Optics and Photonics.

Breuel, T. M. (2018). Document image degradation for data augmentation for handwriting recognition and ocr applications. https://github.com/NVlabs/ocrodeg. Accessed: 2018-05-30.

Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A., and Shafait, F. (2013). High-performance ocr for printed english and fraktur using lstm networks. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 683–687. IEEE.

Bukhari, S. S., Kadi, A., Jouneh, M. A., Mir, F. M., and Dengel, A. (2017). anyocr: An open-source ocr system for historical archives. In *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, volume 1, pages 305–310. IEEE.

Burie, J.-C., Coustaty, M., Hadi, S., Kesiman, M. W. A., Ogier, J.-M., Paulus, E., Sok, K., Sunarya, I. M. G., and Valy, D. (2016). Icfhr2016 competition on the analysis of handwritten text in images of balinese palm leaf manuscripts. In *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*, pages 596–601. IEEE.

Calvo-Zaragoza, J. and Gallego, A.-J. (2018). A selectional auto-encoder approach for document image binarization. *Pattern Recognition*.

Farahmand, A., Sarrafzadeh, H., and Shanbehzadeh, J. (2017). Noise removal and binarization of scanned document images using clustering of features.

Gatos, B., Ntirogiannis, K., and Pratikakis, I. (2009). Icdar 2009 document image binarization contest (dibco 2009). In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 1375–1382. IEEE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5767–5777.

Hedjam, R., Nafchi, H. Z., Kalacska, M., and Cheriet, M. (2015). Influence of color-to-gray conversion on the performance of document image binarization: toward a novel optimization problem. *IEEE transactions on image processing*, 24(11):3637–3651.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv preprint*.

Jenckel, M., Bukhari, S. S., and Dengel, A. (2016). anyocr: A sequence learning based ocr system for unlabeled historical documents. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 4035–4040. IEEE.

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.

Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192*.

Level Otsu, N. (1979). A threshold selection method from gray-level histogram. *IEEE Trans. Syst. Man Cybern*, 9(1):62–66.

Mitianoudis, N. and Papamarkos, N. (2015). Document image binarization using local features and gaussian mixture modeling. *Image and Vision Computing*, 38:33–51.

Niblack, W. (1986). *An introduction to digital image processing*, volume 34. Prentice-Hall Englewood Cliffs.

Ntirogiannis, K., Gatos, B., and Pratikakis, I. (2014). Icfhr2014 competition on handwritten document image binarization (h-dibco 2014). In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 809–813. IEEE.

Pardhi, S. and Kharat, G. (2017). An improved binarization method for degraded document. In *National Conference MOMENTUM*, volume 17.

Pastor-Pellicer, J., Zamora-Martínez, F., España-Boquera, S., and Castro-Bleda, M. J. (2013). F-measure as the error function to train neural networks. In *International Work-Conference on Artificial Neural Networks*, pages 376–384. Springer.

Peng, X., Cao, H., and Natarajan, P. (2017). Using convolutional encoder-decoder for document image binarization. In *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, volume 1, pages 708–713. IEEE.

Phillips, I. (1996). Users reference manual for the uw english/technical document image database iii. *UW-III English/Technical Document Image Database Manual*.

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)*, pages 1395–1403. IEEE.

Rice, S. V., Jenkins, F. R., and Nartker, T. A. (1996). *The fifth annual test of OCR accuracy*. Information Science Research Institute.

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

Tensmeyer, C. and Martinez, T. (2017). Document image binarization with fully convolutional neural networks. In *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, volume 1, pages 99–104. IEEE.

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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 5.

Westphal, F., Lavesson, N., and Grahn, H. (2018). Document image binarization using recurrent neural networks. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 263–268. IEEE.

Zemouri, E., Chibani, Y., and Brik, Y. (2014). Enhancement of historical document images by combining global and local binarization technique. *International Journal of Information and Electronics Engineering*, 4(1):1.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*.