

An ensemble English Document Image Dewarping using Retinex and Generative Adversarial Neural Networks

Soumya B J, Vasudev T, Sharath Kumar Y H

Department of Information Science and Engineering,
Maharaja Institute of Technology Mysore,
Karnataka, 571477, India.

Abstract-Document images obtained from bound or folded documents mainly undergo a specific image degradation known as document warping, making it difficult to read and significantly diminish the performance of an Optical Character Recognition (OCR) system. This paper attempts to develop a comprehensive framework for dewarping and restoring highly distorted document images using the Retinex theory and Generative Adversarial Networks (GANs). To the best of our knowledge, we could not find much research conducted on this topic using Retinex and GANs for ensembling. It is observed that the proposed method can perform better in tasks, like illumination correction, binarization, edge detection, and dewarping in order to reconstruct an improved version of the distorted text. Further, to demonstrate efficacy of the method in restoring distorted text lines in document images against their original state, we compare it to state-of-the-art algorithms on the widely-used CBDAR 2007 and personal dataset. With SSIM of 0.83 and PSNR of 65.21dB, the proposed approach is found better at dewarping highly distorted English document images. The results on a wide range of distorted document images show that the proposed methodology can be extended to address other document enhancing tasks.

Keywords: Document dewarping, Distortion correction, GANs, Document enhancement, Retinex Theory

1. INTRODUCTION

Optical Character Recognition (OCR) technology has made significant strides in recent years, enabling the automation of document processing by converting documents into formats that computer systems can understand. The rapid advancement in this field can be attributed to the proliferation of public databases, which have facilitated the development of more robust and accurate OCR systems [1-3]. However, the presence of distortions, such as warping, folding, or damage caused by use, poses challenges to the reliability of day to day OCR technology. Additionally, inadequate scanning conditions, such as those encountered when using smartphone cameras, can further degrade document quality, leading to issues such as shadows, blurs, and variations in lighting conditions.

Document image restoration is a crucial task in computer vision and image processing. As document images are often subject to various forms of degradation [4,5], it is essential to develop effective methods to restore them and improve the accuracy and

reliability of document analysis and processing. Existing techniques, however, are limited in their applicability, as they depend on specific hardware, environmental factors, or assumptions. In addition, the presence of stampings, cartoons, and annotations on some papers make the recovery process extremely difficult. Few instances of documents with varying degrees of distortion are shown in figure 1. When such input documents are subjected to OCR, the performance of OCR system drops considerably, and such degraded documents require an enhancement technique to improve the quality of input image. This requirement has motivated us to investigate an approach to enhance the quality of input image and make it more suitable for OCR system.



Figure 1: Document images with varying degrees of distortion

Dewarping is an essential pre-processing step in document analysis and OCR systems, as it rectifies the geometric distortions introduced in the document images due to factors like folding, binding, crumpling, or camera lens distortions. To address the challenges in document warping, a new framework is proposed that combines Retinex theory [11], and Generative Adversarial Networks (GANs)[22]. The framework uses Multi-Scale Retinex theory to enhance poorly exposed images and the application of Contrast Limited Adaptive Histogram Equalization (CLAHE) for image enhancement. Retinex is an image enhancement technique originally developed for improving the visual quality of images by addressing issues related to illumination and reflectance. In the context of OCR document dewarping, Retinex can provide certain advantages, which are discussed subsequently.

Retinex enhances the local contrast of an image by normalizing the illumination levels, resulting in improved contrast and visibility of textual content in document images, particularly when the original image suffers from uneven lighting conditions or low contrast. Moreover, it is capable of adapting to various illumination conditions, making it suitable for processing document images with diverse lighting scenarios. This adaptability enables OCR systems to handle a wide range of document images with different levels of distortion and lighting conditions, ensuring more accurate and reliable text recognition across various contexts.

Incorporating Retinex in document dewarping process contributes to enhance the image quality and improves OCR performance through overcoming issues related to illumination, contrast, and noise.

Convolutional networks are then employed to fix enhanced documents, with GAN outputs for pixel-displacement decoding and foreground/background segmentation. The proposed model regulates the smoothness of displacements, resulting in an effective method to fix distorted document images. This approach offers an effective method for restoring distorted document images, and overcoming the limitations of traditional techniques.

This work is organized into five sections. In section 1, the introduction to the study is highlighted. Section 2 is on related works on document image dewarping. The core methodology is presented in Section 3. In section 4, the results of the experiments are discussed. Section 5, concludes the work with scope for future avenues.

2. RELATED WORKS

In this section, the highlights of the traditional and machine learning-based approaches on document image dewarping are discussed. Traditional techniques hinge on auxiliary hardware, distinct features, or visual representations for estimating the 3D shape of documents within images, which unfortunately curtails their practicality in the face of hardware restrictions[6-9]. These traditional methods rely heavily on specific visual cues, creating a limiting factor in their overall utility. On the contrary, certain methods offer an avenue for simplifying the reconstruction process, reducing the complexity by confining the exposure area[10,11]. Thus, providing a workaround to the limitations imposed by the traditional methodologies.

2.1 Poorly illuminated image recovery and binarization

Document images obtained mainly from cameras are with poor illumination

and normally require enhancement of images. Enhancement of such documents use thresholding-based document binarization methods [12 - 16] and are designed with the appropriate global threshold as filters. Pixels are assigned to either the text or the degradation based on the values of the thresholds. Moghaddam and Cheriet[17,18] suggested a variational model to fix bleed-through in deteriorated two-sided document images. This method was also used for the problem of scene text binarization [19]. Similarly, Xiong et al.[20] used a mathematical morphological approach to estimate the background and remove it from the image. After the document image has been enhanced, Laplacian energy-based segmentation is used to assign classes to the individual pixels. The major shortcoming of these traditional approaches is that their results are highly conditional and cannot be extended for general cases. For instance, background complexity or uneven brightness were two such common causes for failure.

These studies were focused on restoring warped document images, rather than converting them into binary format. However, to assess their effectiveness, relevant literatures employed deep learning solutions for document image restoration.

2.2 Generative Adversarial Networks approaches to dewarped images

The proliferation of vast datasets has paved the way for numerous deep learning models, such as ResNet50 and VGG19 [21-22]. One notable model, Generative Adversarial Networks (GANs), was initially proposed by Goodfellow et al. [23] and has attracted substantial research interest for various image enhancement and image-to-image transformation applications. In studies [24-25], GANs

were utilized to refine the semantic segmentation of distorted documents. For image super-resolution, Ledig et al. [26] presented SRGAN, a conditional GAN capable of producing near-photorealistic reconstructions even at exceptionally high upscaling factors. CycleGAN, introduced in [27], addresses multiple image-to-image translation tasks involving imperfect data. The "pix2pix-HD" model, featured in [28], focuses on high-resolution photorealistic image-to-image translation tasks, such as 2048x1024. Additionally, the authors in [29] developed an unsupervised image translation technique using a "DualGAN", which consists of two GANs trained concurrently. These developments highlight the adaptability and potential of GANs in tackling a wide array of image processing challenges. However, it does not thoroughly detail the architecture, the number of layers and the size of images present in the dataset. Further, there is no mention about potential challenges of employing GANs to an unseen data, an approach that would ensure model generalization.

2.3 Deep learning-based dewarping rectification techniques

Leveraging a convolutional neural network (CNN), Das et al. [30] developed a method for detecting page creases, subsequently dividing them into blocks for correction. To address document distortion and camera tilt, Xing et al. [31] employed a CNN-based approach. Ramanna et al. [32] used a conditional GAN (cGAN) trained on a pix2pix-HD network [27] to correct document images affected by curling or geometric distortion. However, this technique was primarily effective for minor deformations and images with a solid color background. Ma et al. [33] introduced a stacked U-Net for predicting forward mapping for warping, trained from

scratch. However, due to the significant differences in their contributed dataset from other heterogeneous datasets, the model performed poorly on unseen images. To overcome the limitations of synthetic dataset training, Das et al. [34] utilized real-world documents and rendering software that encompassed various types of document image ground truth, ultimately proposing dewarping and refinement networks to rectify geometric and shading errors in document images. To capture the realistic intricacies of distorted document images, Li et al. [35] generated a training dataset in 3D space and processed it using a rendering engine. More recently, Rahman et al. [36] suggested the application of Retinex theory to enhance illumination in underexposed images. These advancements demonstrate deep learning techniques' continuous evolution and potential for document image dewarping and enhancement.

Although attempts have been made using specific examples of deep learning methods for document dewarping, the limitations of these techniques are only briefly discussed. For instance, the technique proposed by Ramanna et al. [37] is primarily effective for minor deformations and images with a solid color background. Additionally, the model proposed by Ma et al. [33] underperforms on unseen images due to the significant differences between their contributed dataset and other heterogeneous datasets. Thus, a robust approach that addresses these challenges incorporating a wider variety of images for training dataset, and appropriate fine-tuning that leverage pre-existing models trained on large datasets of 3500 document images is proposed.

3. METHODOLOGY

The methodology of the proposed work along with the dataset used for the study, are discussed in detail in the subsequent subsections.

3.1 Dataset Description

Two datasets are used in this work to evaluate the effectiveness of the proposed technique. The first dataset, CBDAR2007 [37], was initially introduced as part of a page dewarping dataset. It consists of 102 camera-captured photographs of documents, along with their corresponding ASCII text ground-truth. This dataset was developed to aid the research community in addressing complex document image processing challenges and to enable the comparison of different approaches.

In addition to the CBDAR2007 dataset, we have created another dataset comprising 250 camera-captured images. These images were specifically used to test the robustness of the proposed technique in handling various types of degradation, including poor exposure and distortion. With these two datasets, the proposed work is thoroughly evaluated for performance and compared with other state-of-the-art results.

3.2 Proposed Methodology

The proposed approach is an ensemble of Retinex and Generative Adversarial Neural Networks models that aims to improve the process of English document image dewarping. The network consists of two main components. The first is Retinex-based pre-processing [11] and, the second one is Generative Adversarial .

Table 1: Parameters for image augmentation on the CBDAR2007 dataset

Method	Default	Augmented
Rescale	-	1./255
Zoom range	-	0.25
x-Shift, y-Shift	None	[-0.1, +0.1]
x-Scale, y-Scale	None	[0.75, 1.25]
Adjusted image	Varies	128 x 128

The rescale factor is applied to the pixel values of images. In the augmented setting, pixel values are divided by 255, converting them into a range between 0 and 1. This normalization helps to improve the neural network's performance by scaling input values to a consistent range. The Zoom range parameter defines the range of random zoom levels applied to images. In the default setting, no zooming is applied followed by a random zoom level between -25% and +25%. This augmentation helps the neural network in learning to recognize objects and features at different scales. The x-Shift and y-Shift parameters represent the range of random horizontal (x) and vertical (y) translations applied to the images that randomly shifted the value between -10% and +10% in both directions. This augmentation helps the neural network learn to recognize objects and features with varying positions in the input image.

The x-Scale and y-Scale parameters refer to the range of random horizontal (x) and vertical (y) scaling applied to the images. The Adjusted image parameter represents the final image size after preprocessing. The images are arbitrarily resized to 128 x 128 pixels. This resizing ensures that all images have a consistent size before being fed into the neural network.

Subsequently, a Multi-scale Retinex (MSR)[36] is employed for illumination correction. The formula for brightness correction known as the “luma” is indicated in equation 1.

$$\text{Brightness} = \frac{\sqrt{0.241 * C_R^2 + 0.691 * C_G^2 + 0.068 * C_B^2}}{\text{Width} * \text{Height}} \quad (1)$$

Where C = Image coordinates in decimals

R, G, B represents the Red, Green, and Blue color channels.

Width, Height corresponds to the image dimensions

Neural Network [23]. The integration of these two components forms an ensemble technique that enhances the dewarping process.

Retinex is a color image enhancement technique, based on the premise that the observed image can be decomposed into the product of illumination and reflectance. In the context of document image dewarping, the Retinex algorithm is used to correct uneven illumination and improve the contrast of the text, making it more legible and easier for the GAN to process.

The proposed method includes preprocessing the input images to eliminate noise using a median filtering technique. To ensure uniformity with the layers of models pre-trained networks, these images are then resized to the standard 128 x 128 pixels. The specific parameters used in this process are detailed in Table 1. The process culminates in the dewarping of the images, thereby readying them for further analysis. The justification for choosing these parameters is to enhance the training data by creating diverse and augmented images, helping the model to generalize better and improve its performance on real-world distorted and degraded document images.

The flow diagram for dewarping process employed in the work is shown in Fig. 2.

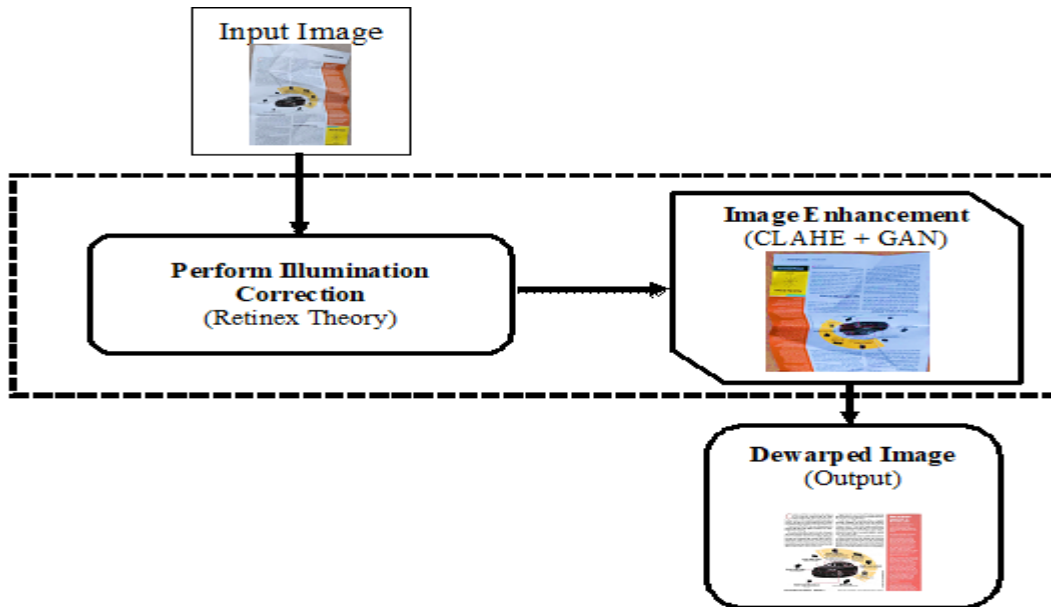


Figure 2: The block diagram of the proposed methodology

Image enhancement using CLAHE[44] and GAN involves two separate techniques combined to improve an image's quality, specifically in terms of contrast. CLAHE is an image enhancement technique that improves the local contrast in images, particularly useful for images with uneven lighting or poor contrast. CLAHE works by independently dividing the input image into small, non-overlapping regions (tiles) and applying histogram equalization to each tile. The algorithm then interpolates the histograms between adjacent tiles to eliminate artificial boundaries introduced by the independent processing of each tile. This results in a more natural-looking enhanced image. A contrast limit is also applied during histogram equalization to prevent excessive noise amplification, hence the name "Contrast Limited."

GAN are a class of deep learning models that consist of two primary components: a generator and a discriminator. In this paper, GANs are used to improve the quality of images by generating high-quality, refined versions of the input images. The generator learns to generate dewarped images from the input warped images, while the discriminator tries to differentiate between real

dewarped images and those generated by the generator. This competition leads the generator to produce better quality dewarped images over time. The generator is trained using a combination of adversarial loss (to ensure the generated images are indistinguishable from real images) and content loss (to maintain the content and structure of the input image).

3.2.1 The Generator

It is common practise to employ auto-encoder models while performing the task of image-to-image translation [38-40]. Most of these models are made up of a series of convolutional layers, collectively called an encoder, that carry out downscaling until a given layer. After that, the up-sampling and convolutional layers of the decoder take over. A great deal of unnecessary redundancy is often exchanged, which wastes time and resources. As a result, we use skip connections organised in the fashion of the U-Net model[41]. Adding skip connections for every two layers helps to recover images with less degradation, and using them during deep model training helps avoid gradient disappearing and explosion issues. Additional training time is saved by adding batch normalization layers. The generator utilised in this investigation is shown in Fig. 3; its

architecture is similar to that introduced in for biomedical picture segmentation[41].

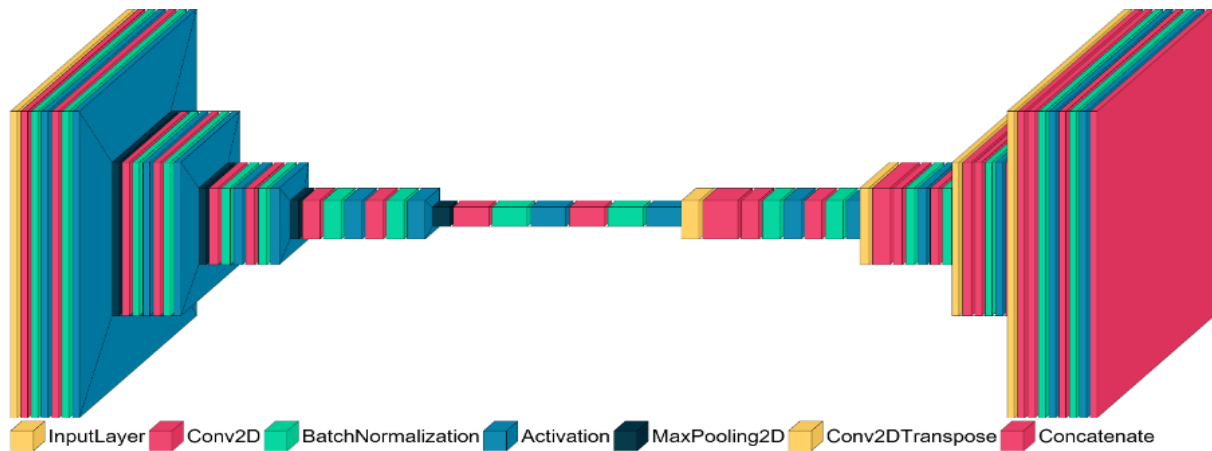


Figure 3: Block diagram of the proposed generator model

3.2.2 The Discriminator

The discriminator employed in this work is a 6-layer Convolutional Neural Network (CNN) that generates a 2x2 probability matrix, indicating the likelihood of the generated image being real, as illustrated in Figure 4. This discriminator is designed to accept both dewarped and unaltered images, either as ground truth or cleaned by the generator. By concatenating these images, a tensor of shape 128x128x2 is formed. This tensor is then processed through the model, resulting in a 16x16x1 matrix at the final layer.

If the cleaned image accurately represents the real scenario, the probabilities within this matrix should approach 1. Conversely, if the image is generated by the generator, the probabilities should be close to zero. A squashing function (sigmoid) is employed as the activation function for the last hidden layer to achieve this. Once the training is complete, the discriminator is no longer required. Instead, when presented, the generative network alone is relied upon to enhance low-quality images.



Figure 4: The discriminator block diagram as a decoder

By combining CLAHE and GAN, an image enhancement pipeline that first improves the local contrast and visibility using

CLAHE is achieved. This combination leads to more visually pleasing images with better contrast, detail, and overall quality, which is

particularly beneficial for applications such as OCR, where image quality directly impacts the accuracy and reliability of the extracted information.

By utilising the proposed network, we forecast the displacement and the categories (foreground or background) at the pixel level.

Next, the background of the input image is removed, and the foreground pixels are then interpolated into the rectified image following the expected displacements. Since the architecture is based on a generator-discriminator model, we provide a description of the process through an algorithm.

The pseudocode for the proposed method is highlighted in Algorithm 1

Algorithm 1:	Strategy for warped image correction
1:	Load input images: Resize (128 x 128)
2:	Perform illumination correction: <i>F_n, Retinex theory</i>
3:	For $n = 1, k$ do:
4:	Apply GANs for all images, $Img \in [1, k]$
5:	Apply CLAHE
6:	ProposedNet, K do:
7:	Label (P^i_j) \leftarrow Label (P^i) in F_n vector space
8:	If Label (P^i_j) = K , then
9:	Label (P^i) \leftarrow max(Label (P^i_j))
10:	else
11:	Label (P^i) \leftarrow Null
12:	endFor
13:	return dewarped image

The Algorithm 1 outlines a strategy for correcting warped document images by employing a combination of preprocessing, image enhancement, and dewarping techniques. It starts by resizing the input images and correcting their illumination using Retinex theory. Next, it applies a GAN to refine the images and improve their quality, followed by CLAHE to enhance the local contrast. Finally, the proposed model is

utilized to dewarp the images by adjusting the labels of the points in the vector space. The resulting output is a set of corrected and enhanced images that are more suitable for further processing, such as OCR document analysis, and information extraction.

Overall, the dewarping process for the proposed network is illustrated in figure 5 in the form of a flow diagram.

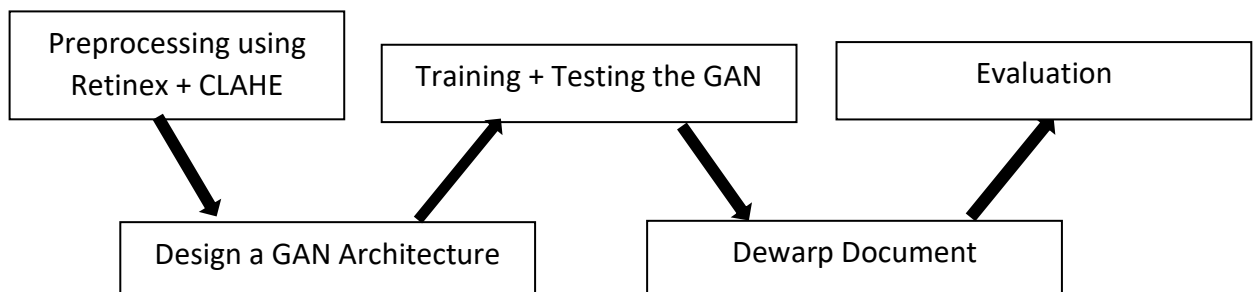


Figure 5: The block diagram of the proposed model

The process begins with pre-processing of the scanned document images, involving resizing, grayscale conversion, and

labeling the images for training, validation, and testing. A suitable Generative Adversarial Network architecture, such as a conditional

GAN or CycleGAN, is then chosen to dewarp the scanned documents, involving designing a generator network for creating dewarped images and a discriminator network for differentiating between the generated and original images. The GAN, is specifically trained using the labeled images with parameters fine-tuned for satisfactory results. Following training, the GAN dewarps the scanned documents by processing them through the generator network, with the dewarped images subsequently saved. Finally, the results are evaluated using metrics like mean squared error (MSE), structural similarity index matrix (SSIM), and peak signal-to-noise ratio (PSNR), with these results compared to the original images and ground truth labels.

3.3 The training process and experimental setup

The proposed model was trained and tested using CBDAR2007 dataset, consisting of camera-captured document images and their corresponding ASCII text ground-truth. The dataset is randomly partitioned into 70%, for training and 30% for testing. In addition, the model was evaluated on a private dataset that we have created.

A discriminator component that takes in both the created and degraded images was used to distinguish between real and generated images. The discriminator forces the generator to produce outputs that are indistinguishable from genuine images. Further, an Adaptive Momentum (ADAM) optimizer [43] with a learning rate of 0.00001 was used to train the model. The proposed approach was implemented on a NVIDIA 1060 GPU with a memory of 16GB, which enabled the analysis of 20 input images in a batch. As a result, the approach was able to process an image in approximately 0.8 and 0.7 seconds for training and test phases, respectively.

4 RESULTS AND DISCUSSION

The results of the proposed technique to various warped English documents are presented in figure 8 (a-d), and figures 9 (a-d). These images were chosen to represent various degrees of illumination and deformations.

Figures-6 through figures-10 depicts the results of each stage in the methodologies.

4.1 Illumination Correction Results

The result of illumination correction is presented in figure 6. The results are evaluated on CBDAR2007 dataset, described in [37] and includes examples of various forms of degradation.

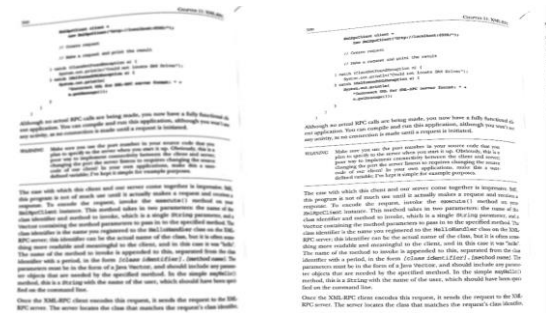


Figure 6: (a) Input Image (Poorly illuminated)

Figure 6: (b) Output Image (After illumination correction)

4.2 Image Enhancement Correction Results

To enhance the image quality further, GAN was applied, followed by an adaptive histogram equalization that stretches the intensity values. The results are in figure 7.

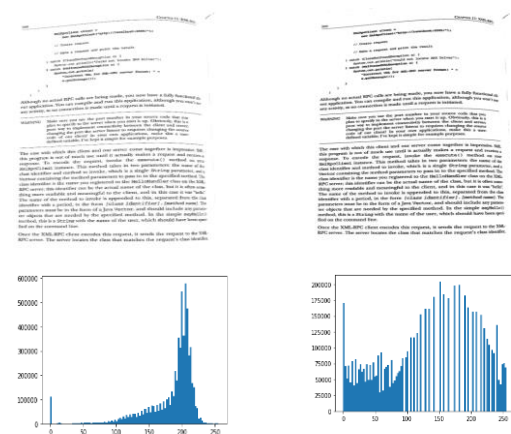


Figure 7: (a) Image enhancement (GAN)

Figure 7: (b) Image enhancement (GAN, followed by CLAHE)

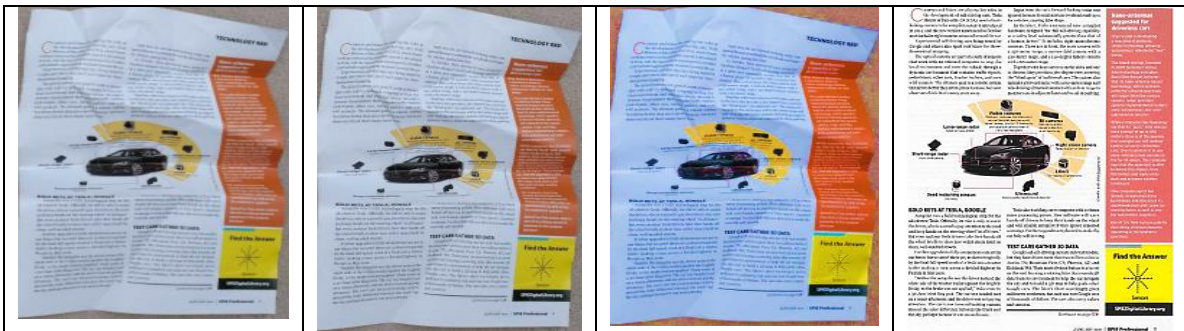


Figure 8: (a) Input image

(b)After illumination correction

(c) Effects of GAN + CLAHE

(d) Dewarped image

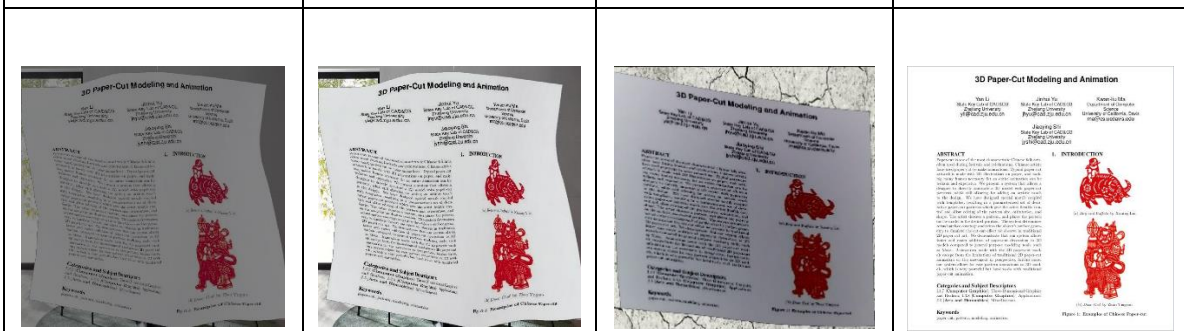


Figure 9: (a) Input image

(b) After illumination correction

(c) Effects of GAN + CLAHE

(d) Dewarped image

4.4 Dewarping results on a private dataset

Experiments are conducted on a personal dataset to test the robustness of the proposed methodology. The results are presented in figure 10.



Figure 10: (a) Poorly illuminated images with warped horizontal and diagonal lines

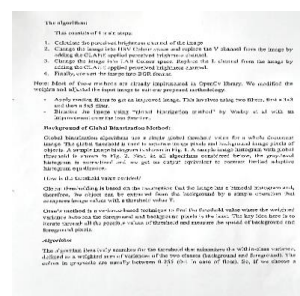


Figure 10: (b) Image enhancement & dewarping using GAN, and the proposed network

world challenges that OCR systems often encounter.

The dataset contained a variety of poorly illuminated images with warped horizontal and diagonal lines, simulating real-

Figure 10(a) showcases the original images from the dataset, demonstrating the complexities arising from poor illumination and distortions that hinder the OCR process.

These images exemplify the difficulties in accurately recognizing and extracting text information due to uneven lighting, document warping, and other degradations.

Figure 10(b) displays the results of the proposed methodology for image document enhancement. The output images exhibit significant improvements in both illumination and geometric corrections. The combination of Retinex theory for illumination correction, GANs for dewarping has effectively mitigated the issues presented in the original images.

The experimental results provide insightful evidence that the proposed methodology effectively addresses the challenges posed by document distortions and poor illumination. By significantly enhancing the quality of the images, the approach not only improves the legibility of the text but also increases the reliability and accuracy of OCR systems in extracting valuable information from distorted and degraded document images.

4.5 Comparative Analysis

Comparative analysis on methods proposed by [33,34,43], respectively are presented in Table 2. For effective representations, we computed the mean squared error (MSE), structural similarity index matrix(SSIM), and peak signal to noise ratio (PSNR) according to equations 2, 3 and 4.

Structural Similarity Index Matrix(SSIM)

measures the similarity between two images. Given an input windows (M x N), the measure between both is calculated as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

Where:

μ_x, μ_y represents the sample mean of x and y;

σ_{xy} = the covariance of x and y

σ_x^2 = the variance of x;

σ_y^2 = the variance of y;

Mean Square Error (MSE): This measures the sum of all the squared differences in a given image, computed as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y - y)^2 \quad (3)$$

Where:

Y = Actual value

y = predicted value

n = number of images

Peak Signal to Noise Ratio (PSNR): This metric is defined via the MSE and expressed as:

$$PSNR \text{ (in dB)} = 10 * \log_{10} \frac{MAX}{MSE} \quad (4)$$

Where:

MAX = the maximum pixel values in an image

MSE = the Mean Squared Error

Table 2: Comparison with other methods

Author	Methods	SSIM	MSE	PSNR (dB)
Ma et al.[32]	DocUNet	0.41	NA	NA
Das et al.[33]	DewarpNet	0.47	NA	NA
Pratikakis et al.[42]	Adaptive Binarization	NA	NA	17.53
Ours	Proposed	0.83	0.00057	65.21

The table 2 presents a comparison of different methods for document image restoration and binarization, evaluated using three metrics: Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR) in decibels (dB).

Ma et al.'s DocUNet achieves the lowest SSIM score of 0.41, indicating poor image quality compared to the other methods evaluated. However, the MSE and PSNR values are not available for comparison. Das et al.'s DewarpNet performs better than DocUNet with an SSIM score of 0.47, but again, MSE and PSNR values are not available for comparison. Pratikakis et al.'s AdaptiveBinarization has the highest PSNR value of 17.53 dB, indicating high image quality with low noise levels. However, the SSIM and MSE values are not available for comparison.

In comparison to these methods, our proposed method achieves the highest SSIM score of 0.83, indicating superior image quality compared to the other methods. Additionally, our method also achieves the lowest MSE value of 0.00057 and the highest PSNR value of 65.21 dB, indicating that our method is robust in handling image degradation and producing high-quality results.

Overall, the comparative analysis shows that the proposed method performs better than other methods evaluated in terms of image quality and robustness to image degradation.

5. CONCLUSION

This paper presents a technique that employs a Retinex theory and Generative Adversarial Network ensemble to rectify distorted document images. The proposed approach ensures the local and global pixels of the rectified image are enhanced through a background information-based function. The proposed methodology demonstrates a significant improvement in OCR performance compared to the reference implementation. The method displays the capability of learning and generalization from incomplete and warped images on the validation dataset (CBDAR2007). However, further work is required to improve the dewarped images' boundary regions and the calculation speed. Future research will involve using a wide range of noisy documents, each presenting unique challenges, such as distortion and occlusions. Overall, this work represents a promising step forward in the development of robust image processing algorithms for document analysis and OCR applications.

REFERENCES

- [1] M. Athimethphat, "A Review on Global Binarization Algorithms for Degraded Document Images," 2011.
- [2] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed, "DeepDeSRT: Deep Learning for Detection and Structure Recognition of Tables in Document Images," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Nov. 2017, pp. 1162–1167. doi: 10.1109/ICDAR.2017.192.
- [3] X. Li, J. Liu, and S. Zhang, "Text Recognition in Natural Scenes: A Review," in *2020 International Conference on Culture-oriented Science & Technology (ICCST)*, Beijing, China, Oct. 2020, pp. 154–159. doi: 10.1109/ICCST50977.2020.00036.
- [4] S. Bako, S. Darabi, E. Shechtman, J. Wang, K. Sunkavalli, and P. Sen, "Removing Shadows from Images of Documents," in *Computer Vision – ACCV 2016*, vol. 10113, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds. Cham: Springer International Publishing, 2017, pp. 173–183. doi: 10.1007/978-3-319-54187-7_12.
- [5] S. Murali, V. K. Govindan, and S. Kalady, "Single image shadow removal by optimization using non-shadow anchor values," *Comp. Visual Media*, vol. 5, no. 3, pp. 311–324, Sep. 2019, doi: 10.1007/s41095-019-0148-x.
- [6] Li Zhang, Yu Zhang, and Chew Lim Tan, "An Improved Physically-Based Method for Geometric Restoration of Distorted Document Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 728–734, Apr. 2008, doi: 10.1109/TPAMI.2007.70831.
- [7] B. Ge *et al.*, "An Active Filter Method to Eliminate DC-Side Low-Frequency Power for Single-Phase Quasi-Z Source Inverter," *IEEE Trans. Ind. Electron.*, pp. 1–1, 2016, doi: 10.1109/TIE.2016.2551680.
- [8] F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos, "Shape from shading for the digitization of curved documents," *Machine Vision and Applications*, vol. 18, no. 5, pp. 301–316, Oct. 2007, doi: 10.1007/s00138-006-0062-y.
- [9] Y. Tian and S. G. Narasimhan, "Rectification and 3D reconstruction of curved document images," in *CVPR 2011*, Colorado Springs, CO, USA, Jun. 2011, pp. 377–384. doi: 10.1109/CVPR.2011.5995540.
- [10] J. Liang, D. DeMenthon, and D. Doermann, "Geometric Rectification of Camera-Captured Document Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 591–605, Apr. 2008, doi: 10.1109/TPAMI.2007.70724.
- [11] A. Pugliese, S. Pomes, S. Ferilli, and D. Redavid, "A Novel Model-based Dewarping Technique for Advanced Digital Library Systems," *Procedia Computer Science*, vol. 38, pp. 108–115, 2014, doi: 10.1016/j.procs.2014.10.018.
- [12] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
- [13] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317–327, Mar. 2006, doi: 10.1016/j.patcog.2005.09.010.
- [14] Neerad Phansalkar, Sumit More, Ashish Sabale, and Madhuri Joshi, "Adaptive local thresholding for detection of nuclei in diversity stained cytology images," in *2011 International Conference on Communications and Signal Processing*, Kerala, India, Feb. 2011, pp. 218–220. doi: 10.1109/ICCSP.2011.5739305.
- [15] F. Z. Ait Bella, M. El Rhabi, A. Hakim, and A. Laghrib, "An innovative document image binarization approach driven by the non-local p-Laplacian," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, p. 50, Dec. 2022, doi: 10.1186/s13634-022-00883-2.
- [16] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation," *IEEE Trans. on Image Process.*, vol. 7, no. 6, pp. 918–921, Jun. 1998, doi: 10.1109/83.679444.
- [17] R. F. Moghaddam and M. Cheriet, "A Variational Approach to Degraded Document Enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1347–1361, Aug. 2010, doi: 10.1109/TPAMI.2009.141.
- [18] R. Hedjam, M. Cheriet, and M. Kalacska, "Constrained Energy Maximization and Self-Referencing Method for Invisible Ink Detection from Multispectral Historical Document Images," in *2014 22nd International Conference on Pattern Recognition*, Stockholm, Sweden, Aug. 2014, pp. 3026–3031. doi: 10.1109/ICPR.2014.522.
- [19] G.-W. Xie, F. Yin, X.-Y. Zhang, and C.-L. Liu, "Document Dewarping with Control Points".
- [20] W. Xiong, X. Jia, J. Xu, Z. Xiong, M. Liu, and J. Wang, "Historical document image binarization using background estimation and energy minimization," in *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, Aug. 2018, pp. 3716–3721. doi: 10.1109/ICPR.2018.8546099.
- [21] A. Victor Ikechukwu and S. Murali, "ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and

classification of Pneumonia from chest X-ray images,” *Global Transitions Proceedings*, vol. 2, no. 2, pp. 375–381, Nov. 2021, doi: 10.1016/j.gltp.2021.08.027.

[22] A. Victor Ikechukwu and M. S, “CX-Net: an efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis,” *Mach. Learn.: Sci. Technol.*, vol. 4, no. 2, p. 025021, Jun. 2023, doi: 10.1088/2632-2153/acd2a5.

[23] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” 2014, doi: 10.48550/ARXIV.1406.2661.

[24] A. Victor Ikechukwu and S. Murali, “i-Net: a deep CNN model for white blood cancer segmentation and classification,” *IJATEE*, vol. 9, no. 95, Oct. 2022, doi: 10.19101/IJATEE.2021.875564.

[25] C. Ledig *et al.*, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” 2016, doi: 10.48550/ARXIV.1609.04802.

[26] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” 2017, doi: 10.48550/ARXIV.1703.10593.

[27] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs,” 2017, doi: 10.48550/ARXIV.1711.11585.

[28] Z. Yi, H. Zhang, P. Tan, and M. Gong, “DualGAN: Unsupervised Dual Learning for Image-to-Image Translation,” 2017, doi: 10.48550/ARXIV.1704.02510.

[29] S. Das, G. Mishra, A. Sudharshana, and R. Shilkrot, “The Common Fold: Utilizing the Four-Fold to Dewarp Printed Documents from a Single Image,” in *Proceedings of the 2017 ACM Symposium on Document Engineering*, Valletta Malta, Aug. 2017, pp. 125–128. doi: 10.1145/3103010.3121030.

[30] Y. Xing, R. Li, L. Cheng, and Z. Wu, “Research on Curved Chinese Document Correction Based on Deep Neural Network,” in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, Dec. 2018, pp. 342–345. doi: 10.1109/ISCID.2018.10179.

[31] V. Ramanna, S. Bukhari, and A. Dengel, “Document Image Dewarping using Deep Learning,” in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, Prague, Czech Republic, 2019, pp. 524–531. doi: 10.5220/0007368405240531.

[32] K. Ma, Z. Shu, X. Bai, J. Wang, and D. Samaras, “DocUNet: Document Image Unwarping via a Stacked U-Net,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 4700–4709. doi: 10.1109/CVPR.2018.00494.

[33] S. Das, K. Ma, Z. Shu, D. Samaras, and R. Shilkrot, “DewarpNet: Single-Image Document Unwarping With Stacked 3D and 2D Regression Networks,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), Oct. 2019, pp. 131–140. doi: 10.1109/ICCV.2019.00022.

[34] X. Li, B. Zhang, J. Liao, and P. V. Sander, “Document rectification and illumination correction using a patch-based CNN,” *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–11, Dec. 2019, doi: 10.1145/3355089.3356563.

[35] Z. Rahman, D. J. Jobson, and G. A. Woodell, “Multi-scale retinex for color image enhancement,” in *Proceedings of 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, 1996, vol. 3, pp. 1003–1006. doi: 10.1109/ICIP.1996.560995.

[36] S. S. Bukhari, F. Shafait, and T. M. Breuel, “The IUPR Dataset of Camera-Captured Document Images,” in *Camera-Based Document Analysis and Recognition*, vol. 7139, M. Iwamura and F. Shafait, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 164–171. doi: 10.1007/978-3-642-29364-1_13.

[37] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *arXiv:1605.06211 [cs]*, May 2016, Accessed: Mar. 27, 2022. [Online]. Available: <http://arxiv.org/abs/1605.06211>

[38] J. Xie, L. Xu, and E. Chen, “Image Denoising and Inpainting with Deep Neural Networks”.

[39] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[40] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *arXiv*, May 18, 2015. Accessed: Nov. 10, 2022. [Online]. Available: <http://arxiv.org/abs/1505.04597>

[41] M. Ibrahim, M. Wagdy, F. S. AlHarithi, A. M. Qahtani, W. S. Elkilani, and S. Zarif, “An Efficient Method for Document Correction Based on Checkerboard Calibration Pattern,” *Applied Sciences*, vol. 12, no. 18, p. 9014, Sep. 2022, doi: 10.3390/app12189014.

- [42] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Nov. 2017, pp. 1395–1403. doi: 10.1109/ICDAR.2017.228.
- [43] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic Segmentation using Adversarial Networks," 2016, doi: 10.48550/ARXIV.1611.08408.

