Document Image Dewarping using Deep Learning

Vijaya Kumar Bajjer Ramanna[1], Saqib Bukhari[2] and Andreas Dengel[1,2]
[1]*Informatik, Technical University of Kaiserlautern, Kaiserslautern, Germany*
[2]*Department of Knowledge Management, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Kaiserlsutern, Germany*

Keywords: Document Image Dewarping, Deep Learning, Geometric Distortion, Page Curl, Line Curl.

Abstract: The distorted images have been a major problem for Optical Character Recognition (OCR). In order to perform OCR on distorted images, dewarping has become a principal preprocessing step. This paper presents a new document dewarping method that removes curl and geometric distortion of modern and historical documents. Finally, the proposed method is evaluated and compared to the existing Computer Vision based method. Most of the traditional dewarping algorithms are created based on the text line feature extraction and segmentation. However, textual content extraction and segmentation can be sophisticated. Hence, the new technique is proposed, which doesn't need any complicated methods to process the text lines. The proposed method is based on Deep Learning and it can be applied on all type of text documents and also documents with images and graphics. Moreover, there is no preprocessing required to apply this method on warped images. In the proposed system, the document distortion problem is treated as an image-to-image translation. The new method is implemented using a very powerful pix2pixhd network by utilizing Conditional Generative Adversarial Networks (CGAN). The network is trained on UW3 dataset by supplying distorted document as an input and cleaned image as the target. The generated images from the proposed method are cleanly dewarped and they are of high-resolution. Furthermore, these images can be used to perform OCR.

1 INTRODUCTION

These days, most of the libraries and companies want to convert their old documents and books to digital form. Evidently, these days Optical Character Recognition (OCR) (Mori et al., 1999) is extensively used for this purpose. OCR performs well on the undistorted document images. The hard copies of input documents are scanned using cameras or flatbed scanners. These scanned images are passed as inputs to OCR. However, these scanned documents are distorted in most of the cases and they are not in convenient form to apply the recognizer. Also, OCR doesn't give a good accuracy for distorted images. Hence, these images are not appropriate to apply OCR and even not for archiving because of the existing document distortion.

There could be several reasons for document dewarping, sometimes the instruments used for scanning like camera and flatbed scanner are the main basis. Camera-based scanned images are generally suffer from perspective distortion because of the varied angle of camera position and geometrical distortion due to the curvature of the unfolded book surface. For non-planner documents like books, the page curl is added as an additional distortion to the scanned images. Apart from that, the distortion is also caused by document aging, for example, historical documents.

In recent days, document image dewarping research community is continuously growing and several algorithms have been developed to dewarp distorted document images (Liang et al., 2005). Most of the algorithms are based on conventional aligned text lines and few of them also utilize line segmentation in addition to that. Some of these algorithms dependent on only single camera images to apply dewarping. Most of the algorithms are for textual documents but few recent techniques work on both text and image documents. However, textual based algorithms need some preprocessing of the input images before using any of these dewarping methods. The majority of the algorithms just focus on only one problem, so we would need to apply two methods to remove multiple distortions (Ex perspective and page curl).

In contrast, our model doesn't rely on text line alignment or segmentation. Our approach won't even need any preprocessing steps for the input images. It works well on text and non-text parts of a document

Bajjer Ramanna, V., Bukhari, S. and Dengel, A.
Document Image Dewarping using Deep Learning.
DOI: 10.5220/0007368405240531
In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)*, pages 524-531
ISBN: 978-989-758-351-3

like images, tables, graphics etc. The dewarped image is a high resolution and it can be used for OCR. Unlike other methods, we can use images from multiple scanners or cameras as input.

Recently, Deep Learning has gained success in many fields. The powerful matrix multiplication of Deep Neural Networks with the combination of GPU architecture has wide application. The research in document images is getting extensive benefits from Deep Learning. The proposed method is based on the most advanced image-to-image translation Deep Learning network, which has the ability to synthesize high resolution dewarped image by retaining the quality of text or image/graphics in the input image. The document distortion problem is formulated as an image to image translation problem to avoid the dependency on line/image/graphics alignment in the input image. In our experiment, we the benefits from the state-of-the-art image-to-image model pix2pixhd (Wang et al., 2017). This network is trained on the high-resolution UW3 dataset by using the distorted image as input and the normal image as the target.

2 RELATED WORK

Document image dewarping is an essential preprocessing step for OCR to get better character recognition. There is a wide list of developed methods (Stamatopoulos et al., 2008)(Bukhari et al., 2009)(Ulges et al., 2005) for document dewarping. But, the most popular approach is to use depth-measuring equipments like structured light or laser scanner to measure the depth. However, Depth Measuring equipments are certainly not appropriate for common users because of the need for special hardware and calibration of the hardware. Also, In case of non-linear curled pages, these depth measuring instruments work efficiently only on multiple images but for common users, it might be too inconvenient to take so many pictures (Koo et al., 2009). This limits the use of depth measuring methods. Generally, people tend to focus on easy and comfortable devices like mobile cameras with a high resolution and flatbed scanners.

The majority of the currently available methods utilize prior document layout information to remove page curl and/or perspective distortion. These methods work by keeping text lines as basis as they are the salient features of a document. The dewarping method using an estimation of curled text lines (Ulges et al., 2005) takes only single camera image as an input. It's a line-by-line dewarping approach, which uses local text line extraction and depth extraction. But, these methods made an assumption that the line spacing is constant and the input documents must have a single column. Also, this method takes nearly 10 secs to process one image. Another method (Kim et al., 2015) which does not work directly by using text lines instead they use connected components like discrete representations of text-lines and text-blocks. Nevertheless, this method is not suitable if the text has multiple skews and each skew has different angles. In general, there are lots of disadvantages with line based methods as it is hard to extract text lines for complex layouts.

In another research work (Wu and Agam, 2002), they reconstruct the target mesh by using a single input document image to remove perspective and geometric distortion. To create the target mesh they calculated the starting point and the orientation of curled text lines. This method is language independent and doesn't depend on structural models, however, works on the basis of an assumption that all the text lines in the input documents are straight. So, this method cannot be used to remove page curl and/or line curl.

3 OUR APPROACH

In the proposed method, the document dewarping problem is considered as the image-to-image translation. The goal is to translate the input image from one domain to other domain by taking input distorted image and the target clean image. The proposed model has generative-adversarial networks to learn the style transfer from input distorted images to clean images. For the experiment, the training is performed on UW3 dataset. We prepared the synthetic distorted UW3 dataset using OCRODEG package (NVlabs, 2018) by inducing random page and/or line curl distortion. The generated input dataset looks similar to the originally distorted test images. The experiment is conducted by training the proposed network using synthetic UW3 dataset prepared and the evaluation is done on the UW3 dataset and additionally, another originally distorted dataset.

In our approach, we used the powerful image-to-image translation network (Wang et al., 2017), which translates the image from one domain to another domain. The model gets the benefit by using the adversarial loss instead of L1 loss since L1 loss produce blurry images. The adversarial training helps the discriminator to learn to better distinguish between generated dewarped images from the generator and the real images. The discriminator uses the trainable loss which brings additional benefits to the results with the ability to better distinguish images at the target. The advantage of multi-scale generator, discriminator and the new objective function makes it easier to generate high-resolution images with realistic textures. The generator has a global network which is responsible for style transfer and a local enhancer which helps to generate high-resolution images.

3.1 Pix2pixhd Network

There are a couple of methods to solve image-to-image translation problem. For example, pix2pix image translation network with conditional Generative Adversarial Network (CGAN) (Isola et al., 2017) is one of them. Initially, we made an attempt to apply this network for our problem but the restriction of the network to use the image size of 256x256 limits the benefit for our situation because we have high-resolution document images. Since the model has Generative networks, if the input images are resized then it's difficult for the network to reconstruct the characters. Nevertheless, pix2pixhd (Wang et al., 2017) network accepts high-resolution input images and furthermore, the network generates high-resolution images with the help of GANs and by considering adversarial loss instead of the L1 loss. GANs produce natural image distribution by forcing the generated samples to be indistinguishable from natural images. This pix2pixhd model fine-tunes the generator and multi-scale discriminator, which helps to generate images conditionally at high resolution.

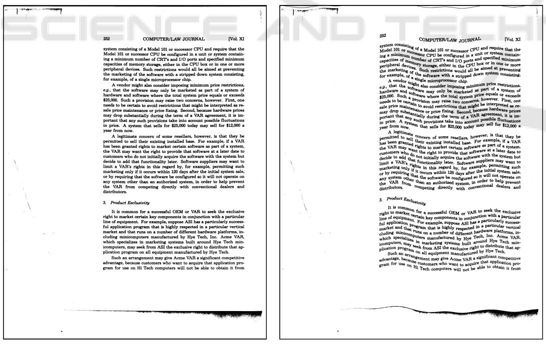


Figure 1: Synthetic data preparation using OCRODEG package; (a) Original clean image, (b) After applying synthetic distortion.

3.2 Dataset Preparation

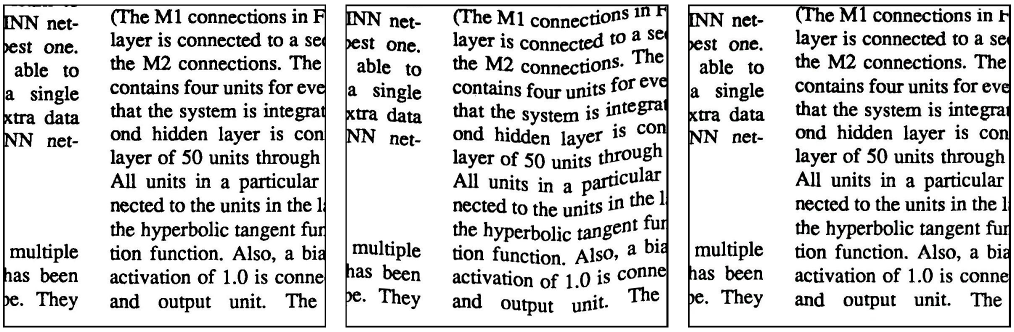
The new method does not require any preprocessing on the input images, which makes the dataset preparation quite simple. UW3 dataset images which are in 'tif' format are converted into the required 'jpg' format and stored in the directory named 'train_B' (ground truths). No re-sizing has been done on the input images since the model can be trained on high-resolution images. As mentioned in the previous section, the distorted input images are generated by using the package OCRODEG. This package let us to generate random noise and the images are distorted using this noise. Figure (1) shows the synthetic data generated using this package. Distorted input dataset has the same name, size, and format as the ground truth and they are saved in a directory 'train_A'. We have used two test sets, the UW3 dataset which was not included for training and a new dataset from the historical domain; OCR-D images (Ocr-d.de, 2018). The synthetic UW3 test set is prepared using OCRODEG in a similar way we did to the train set but OCR-D images have distortions already in the original images.

3.3 Implementation

The state-of-the-art method for the image-to-image translation problem is high-resolution photo-realistic image-to-image translation (Wang et al., 2017), the pytorch implementation of the same (NVIDIA, 2018) has been used for this research work. The model is trained by changing the network parameters to get better result. We have used 1500 UW3 raw images without resizing, but during training, images are cropped with a finesize of 768 (the network can also be trained with higher resolution) and loaded with a loadsize of 1024 to reduce the hardware requirement for training. The number of label channels used is 0 and the maximum number of samples allowed per dataset is 200.

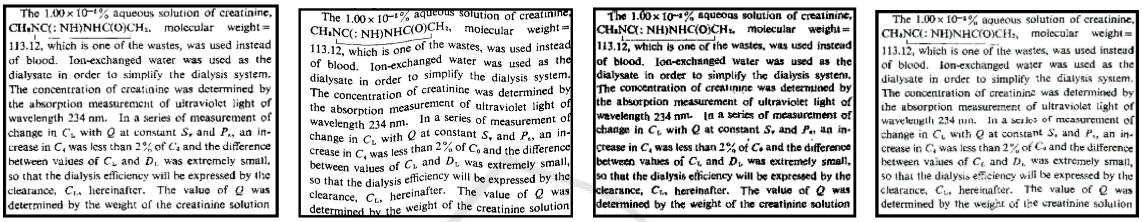
Pix2pix baseline is a CGAN. It is a supervised learning framework with two networks; a generator to map the input images to real images and a discriminator to distinguish between real and generated images. In pix2pixhd model, the generator is a combination of two networks; global generator and local enhancer. Global generator is responsible for the image style transfer. It consists of three components; a convolutional front-end, a set of residual blocks and a transposal convolutional backend. It works at a resolution of 1024x512. In our experiment we have 2 local enhancers, that means the output resolution will be doubled (2048x1024). The Local Enhancer is responsible for a high-resolution output image and it also contains the same three components. During training, the global generator is trained before the local enhancer network and they are fine tuned together later.

Multiscale discriminators have been used to avoid over-fitting caused due to deeper network, which is necessary to produce high-resolution images. In our experiment, we have used 3 scale discriminators with similar network structures but different scales, which discriminates the synthesized image and real image in 3 different scales. Basically, this network has



(a) (b) (c)

Figure 2: Training results of UW3 images using our method (a) Original Image, (b) Distorted input image, (c) Dewarped image using our DL method.



(a) (b) (c) (d)

Figure 3: Comparing test results of Computer Vision and Deep Learning methods. (a) Original Image (b) Warped image (c) Dewarped image using CV method, (d) Dewarped image using our DL method. As one can see from the patch (c) the lines are not straight.

been adopted to guide the generator with both global overview and minute details of an image by including larger and smaller scales of the network.

During training, the overall objective of the loss function includes the GAN loss and feature loss. This features loss is extracted from each layer of the discriminator and it will be matched with different scales of real and synthesized image features. Additional feature loss stabilizes the training since we are using different scales.

Training: The powerful deep learning pix2pixhd network is trained on 1500 UW3 dataset of average size (3300,2500). However, the model is learned on the patches of size (1024, 768) by cropping the input distorted image. We considered batch size of one to train the network to learn one to one mapping between the distorted and the clean image. We used GPU with size of 12 GB for training. The input training data is very generalized and it contains images with mix of text, images and tables.The entire training is carried out for over 250 epochs. As we use the distorted images in place of random noise as input to GANs, the network has already learnt the layout of the document image in the first few epochs. The reason for learning over 200 epochs is to train the network to learn and generate minute features of each character. Testing is performed on two unseen dataset; UW3 dataset Images which are not included in training and OCR-D historical images. The input warped image has been split into patches and we applied the dewarping method on these patches. The final dewarped image has higher resolution (3300,2500), with a very good quality of characters and images of the document. These resulted images can be used to apply OCR.

4 PERFORMANCE EVALUATION

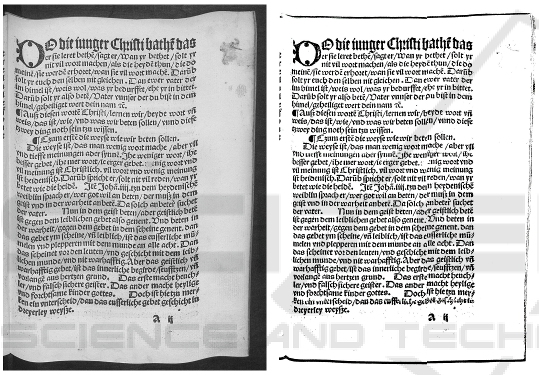
The qualitative analysis of the results is done by comparing results of our method to ground truth images. The proposed network trained on UW3 dataset is very generalized and it works great on varied types of documents; Documents with a single column and multi-column (Figure 1), text, images (Figure 5), equations and tables. The model has been evaluated against all these document elements. The training results are demonstrated in Figure (1), these are UW3 patches generated during the training at Epoch 225. The input distorted image (Figure 2b) is a patch from two column document image with line level distortion on two columns. The generated image (Figure 2c) looks similar to the original clean image (Figure 2a), all the text lines are straight and the characters still holds the same resolution as the ground truth image.

Table 1: This table shows the average values for Pixel-based image evaluation for both Deep Learning (DL) and Computer Vision (CV) methods. These results are calculated between the UW3 ground truth test image and synthesized image from DL and CV method.

Evaluation Method	Computer Vision	Deep Learning
Pixel Accuracy	79.64%	93.25%
Mean Accuracy	61.04%	84.66%
Frequency Weighted IU	67.74%	88.53%
SSIM	61.17%	72.88%
HaarPSI	98.00%	98.99%

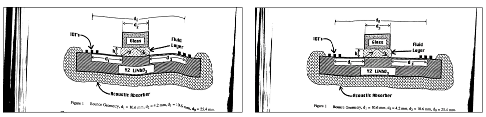
Table 2: Tesseract and ABBYY OCR evaluation results on the UW3 dataset. The average Character Error Rate (CER) is computed between the ground truth text file against the text file generated by applying Tesseract OCR on the ground truth image, input warped image and dewarped images using existing and proposed methods.

Image Type	Tesseract	ABBYY
Clean Image	3.78%	9.69%
Warped Image	17.1%	40.45%
Dewarped using CV method	5.07%	13.32%
Dewarped using DL method	4.39%	10.92%



(a) (b)

Figure 4: Testing results from our method on OCR-D historical data. (a) Distorted input image, (b) Dewarped image from DL method.



(a) (b)

Figure 5: Evaluation results on a document with image (UW3 dataset) (a) Distorted input image, (b) Dewarped image from DL method.

The performance evaluation is also conducted on unseen OCR-D dataset by using the trained model to remove distortion from these images. OCR-D dataset is from a completely different domain and it differs a lot from the UW3 dataset, which has a different language (German) and font style (Futura). Besides the fact that the model was trained on UW3 dataset, yet the performance is also great on the historical OCR-D dataset. Similar to the training process, no preprocessing has been done on this dataset except that the images are binarized. The reason for binarization is that the input images are blurred because of document aging and it is hard to distinguish the results. The binarization method used is from (Afzal et al., 2013). Dewarping results from our method are independent of binarization. From Figure (4), it is visually evident that the proposed method successfully removed distortions from the historical document image. The image (Figure 4a) has geometrical distortions; the line curls and page curls caused mostly by document aging and also a non-linear surface of the book. The dewarped image using our method has effectively removed the page curl and the lines are mostly straight.

5 EVALUATION

Nowadays, there are numerous possibilities to assess two images to check the similarity between them. The majority of them are quantitative analysis and some of them are based on the human perceptual analysis. In our experiment, we have used diverse evaluation methods based on our data. As pixel-based evaluation methods are fundamental and common for image comparison, we have computed Pixel Accuracy, Mean Accuracy and Frequency Weighted IU (Long et al., 2015) between the ground truth image and the dewarped image. The numerals are tabulated in the column Deep Learning in Table (1). These metrics compare two images at the pixel level and provide similarity score between the two images. our experiment results reveal very good similarity (93.25% pixel accuracy) between the dewarped image and the original image. However, pixel-based methods estimate absolute errors, which may not be adequate to assess two images quality, so we have additionally considered the Structural Similarity Index (SSIM) (Wang et al., 2004). The idea behind this strategy is that the inter-dependencies between pixels in space, carry very significant visual structural information. The calculated SSIM (kersner, 2018) between the ground truth image and the synthesized image shows 72.88% (Table 1) of similarity.

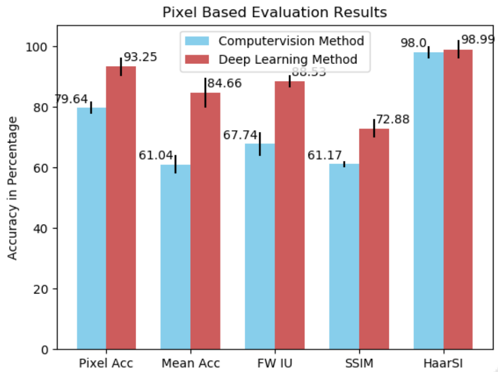


Figure 6: Pixel Accuracy comparison between Computervision method and Deep Learning method.

Additionally, another state-of-the-art method for image comparison is Haar Perceptual Similarity Index (HaarPSI)(Reisenhofer et al., 2018). In the experiment, it has been proved that this method works better than other currently available similarity measures so far and also outcomes from this method are highly correlated with the human opinion scores. Once again, HaarPSI reported outstanding similarity (98%) between the synthesized dewarped image and ground truth image. Based on these metrics we can infer that the dewarped image is almost similar to the ground truth image. So, the proposed method is successful in removing the distortion from the input image and the generated image looks close to the ground truth image.

Another interesting evaluation method we have used is Optical Character Recognition (OCR). We have considered two OCR engines, which are freely available on the Internet to evaluate the results of our method against the existing Computer Vision method and to the ground truth images. To evaluate the recognizer results, we have computed character error rate between text files the synthesized images from our method and the ground truth images. This method computes edit distance between the ground truth and recognizer output image. The outcome of this evaluation method is the character error in percentage between the two files.

1. Tesseract (Smith, 2007) is an open source OCR engine, which supports a variety of languages. Until now, it is one of the best open source OCR. This OCR has been applied on three images; input distorted image, ground truth, and dewarped image. The obtained results are compared with the ground truth text file by computing character error rate. Tesseract OCR results (Table 2 and Figure 7) show better error score for the dewarped image when compared to the distorted image. This proves the increase in recognizer accuracy after using our proposed method and also, shows the quality of the Deep Learning based dewarping method. Tesseract evaluation is also performed on OCR-D historical images. The recognizer results in Table (3) explains the improvement in the recognizer accuracy after applying our dewarping method.

2. Abbyy (Finereaderonline.com, 2018) is an OCR engine developed by a Russian company ABBYY, which is used to convert document image to editable electronics format. For the evaluation purpose, Abbyy OCR is applied to the results of the existing method (Bukhari et al., 2009) and our method. The generated text files are compared to the ground truth text file using a character comparison method. The recognizer results from the Table (2) and Figure (7) also gave the best results for our method. The error percentage for the results of our method is less than the result of the warped image.

The Deep Learning based dewarping method works well both on text-based documents as well as on documents with images. However, these two OCR engines work on the basis of document layout. Moreover, the ground truth text files of the images used for the experiment include only non-image content ie, only the textual part of the document. Hence, for the sake of evaluation, we have cleaned (removed image content from document) the images before applying OCR. Nevertheless, if we look at the error percentage from the Character Comparison method, the dewarped images have less error when compare to the warped images. This quantitatively proves that our method increases the percentage of character recognition.

Finally, in order to better evaluate our method, we compared our results with the existing document image dewarping method, which is developed based on a computer vision technique (Bukhari et al., 2009). This method works only on textual documents but fails to extend its application to documents with images. Figure (3) conveys the qualitative comparison of dewarping results for the current and existing method. The dewarped results from Computervision method (Figure 3c) still have line curls in the middle but the results from our method (Figure 3d) looks similar to the original image (Figure 3a). In order to make a quantitative analysis, the results of the two methods have been compared in two stages. Firstly,

Table 3: Tesseract OCR on OCR-D images(Error % is shown in table). The edit distance is computed between Tesseract OCR output text files of the warped input image and dewarped image (proposed method).

Image File ID	Warped Input Image	Dewarped image
estor_rechtsgelehrsamkeit02_1758_0119	21.84%	21.09%
luther_auszlegunge_1520_0003	66.91%	56.20%
luther_auszlegunge_1520_0029	34.23%	24.98%
Average	40.99%	34.09%

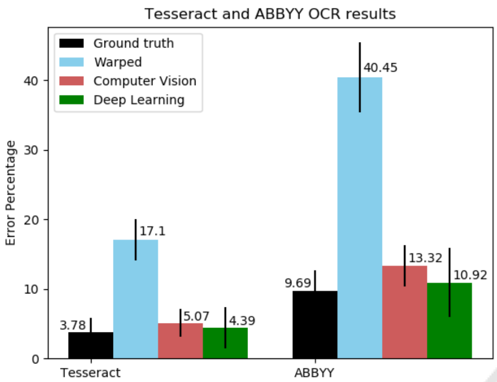
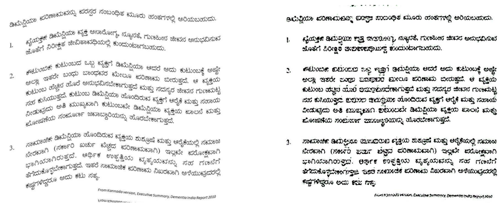


Figure 7: Comparison of OCR results. Tesseract and ABBYY OCR are applied on clean image, warped image, and resulting image of CV and DL methods. The bar graph shows the Character error rate in percentage.



(a) (b)

Figure 8: Testing on a image with low resolution, from different language (Kannada) and domain (Source: google images) (a) Distorted input image, (b) Dewarped image from DL method.

we have considered the pixel-based metrics (Table 1 and Figure 6) to evaluate which method is better in dewarping the distorted image and focus on to generate the resulting image close to the original image. For all of the pixel-based evaluation metrics, out method has an average of 14% higher accuracy when compared to the Computervision based approach. This infers that the DL based approach generates un-distorted images resembles the original image. Secondly, in order to evaluate the dewarping method with respect to the recognizer accuracy, Tesseract and Abbyy OCR have been applied to the results of the two methods. The numerals are tabulated in Table (2), DL based approach increased the recognizer accuracy by 1-2%. Besides, the work in (Bukhari et al., 2009) proved that the Computer Vision method works better than any other Document Dewarping methods. In our experiment, from qualitative and quantitative analysis of the results we can conclude that the DL method is good dewarping method compared to Computervision Based method. Hence, our proposed method is better than other existing document dewarping method to remove page and line curl from the distorted document.

Until now we have discussed the best part of the proposed approach, but the method also fails in some cases. As discussed, the input image passed to the model is always high resolution. If the input distorted image is resized then it is difficult for the generator to produce good quality characters. The model corrects only the style of the images but fails to produce a good quality image. As seen from the results (Figure 8), the model removed the curls from the text and to generated straight text lines but fail to produce the local features of the characters. The future work can concentrate on training the model with resized input data so that the model can be applied for low resolution or resized images.

6 CONCLUSION

In this paper, we have proposed Document Image Dewarping Method using the advanced Deep Learning Image to Image Translation technique. The outcomes are evaluated using the state-of-the-art evaluation methods and also OCR has been applied on the resulting images. The experimental results show that the proposed method works great on text document images, documents with graphics and images, and also on the historical documents. The main advantage of this method is that it doesn't require any sophisticated text line or layout extraction and moreover, this approach doesn't need any preprocessing to be done on the input images. Currently, this method can be applied to dewarp historical documents, geometrically degraded images, images with page and line curl. The future work can extend our method to include also perspective distortion problems.

ACKNOWLEDGEMENTS

This work was supported by my Supervisor, Special thanks to him for his guidance to finish this work smoother and successfully. The test dataset was borrowed from OCR-d.de, special gratitude for providing the dataset to evaluate our work. Our appreciation for the author for sharing the code of existing computer vision based method, which helped us in great extent to evaluate our method effectively.

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.

Bukhari, S. S., Shafait, F., and Breuel, T. M. (2009). Dewarping of document images using coupled-snakes. In *Proceedings of Third International Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain*, pages 34–41. Citeseer.

Finereaderonline.com (2018). Ocr online - text recognition & pdf conversion service — abbyy finereader online. *https://finereaderonline.com/en-us*. (Accessed on: 10-09-2018).

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

kersner, M. (2018). martinkersner/py_img_seg_eval. *https://github.com/martinkersner/py_img_seg_eval*. (Accessed on: 10-08-2018).

Kim, B. S., Koo, H. I., and Cho, N. I. (2015). Document dewarping via text-line based optimization. *Pattern Recognition*, 48(11):3600–3614.

Koo, H. I., Kim, J., and Cho, N. I. (2009). Composition of a dewarped and enhanced document image from two view images. *IEEE Transactions on Image Processing*, 18(7):1551–1562.

Liang, J., Doermann, D., and Li, H. (2005). Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition (IJDAR)*, 7(2-3):84–104.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Mori, S., Nishida, H., and Yamada, H. (1999). *Optical character recognition*. John Wiley & Sons, Inc.

NVIDIA (2018). Nvidia/pix2pixhd. *https://github.com/NVIDIA/pix2pixHD*.

NVlabs (2018). Nvlabs/ocrodeg. *https://github.com/NVlabs/ocrodeg*.

Ocr-d.de (2018). Koordinierte förderinitiative zur weiterentwicklung von verfahren der optical character recognition (ocr). *http://ocr-d.de/*.

Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T. (2018). A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43.

Smith, R. (2007). An overview of the tesseract ocr engine. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 629–633. IEEE.

Stamatopoulos, N., Gatos, B., Pratikakis, I., and Perantonis, S. J. (2008). A two-step dewarping of camera document images. In *Document Analysis Systems, 2008. DAS'08. The Eighth IAPR International Workshop on*, pages 209–216. IEEE.

Ulges, A., Lampert, C. H., and Breuel, T. M. (2005). Document image dewarping using robust estimation of curled text lines. In *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, pages 1001–1005. IEEE.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2017). High-resolution image synthesis and semantic manipulation with conditional gans. *arXiv preprint arXiv:1711.11585*.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Wu, C. and Agam, G. (2002). Document image de-warping for text/graphics recognition. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 348–357. Springer.