# Document Enhancement System Using Auto-encoders

**Mehrdad J. Gangeh**
Ernst & Young (EY) AI Lab, USA
Mehrdad.J.Gangeh@ey.com

**Sunil R. Tiyyagura***
Ernst & Young (EY) GDS India LLP
Sunil.Tiyyagura@gds.ey.com

**Sridhar V. Dasaratha**
Ernst & Young (EY) GDS India LLP
Sridhar.Dasaratha@gds.ey.com

**Hamid Motahari**
Ernst & Young (EY) AI Lab, USA
Hamid.Motahari@ey.com

**Nigel P. Duffy**
Ernst & Young (EY) AI Lab, USA
Nigel.P.Duffy@ey.com

## Abstract

The conversion of scanned documents to digital forms is performed using an Optical Character Recognition (OCR) software. This work focuses on improving the quality of scanned documents in order to improve the OCR output. We create an end-to-end document enhancement pipeline which takes in a set of noisy documents and produces clean ones. We train a blind model (auto-encoders) that works on different noise levels of scanned text documents. Results are shown for blurring and watermark noise removal from noisy scanned documents.

## 1 Introduction

Scanned documents are stored as images and need to be processed by an Optical Character Recognition (OCR) software to extract the text contents into a digital format such as an ASCII text file. This is an active research area and there are many tools in the market that process a scanned document and extract the content in a digital format. The success with extraction of digital output is heavily dependent on the quality of the scanned document. In practice, however, there is some noise associated with the scanned document. Typical noises seen in scanned documents are blurring, watermarks, fading, and salt & pepper.

With the rise of deep learning adoption in computer vision tasks, there are many neural network models available for image denoising and restoration [1]. However, most of the literature focuses on pictures (e.g., images from natural scenes) but not text documents, and the techniques used are not directly applicable due to very different nature of text document images.

## 2 Methodology

### 2.1 Model - Network Architecture

In this study, we have adapted the neural network architecture described in [2] called REDNET (Residual Encoder-Decoder Network). The main advantage of this method is having symmetric skip connections between a convolutional layer and the corresponding deconvolutional layer. Another advantage over fully convolutional network is that pooling and un-pooling, which tend to eliminate

*First and second authors have equal contributions.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

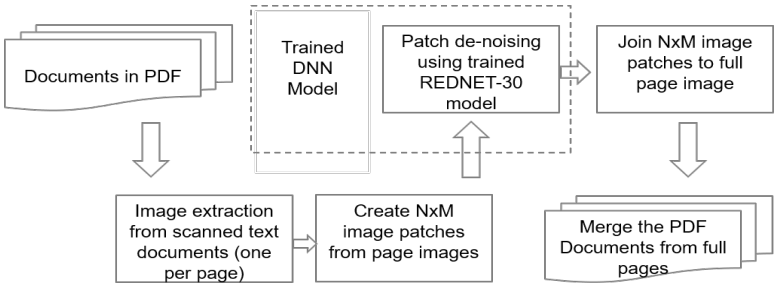



Figure 1: Block diagram of the image enhancement system.

image details, is avoided for low-level image task such as image restoration, resulting in higher resolution outputs. The key differences of this work from [3] is the use of larger dataset and training a blind model.

In this research, a REDNET with 15 convolutional and 15 deconvolutional layers was designed, including 8 symmetric skip connections between alternate convolutional layers and the mirrored deconvolutional layers. The filter size in the convolutional and deconvolutional layers was $3\times3$ except the final decoding convolutional layer where it was set to a filter size of $5\times5$. The number of filters used in the first 3 convolutional layers and their mirrored deconvolutional layers was 64. All other layers had 128 filters. ReLU activation function was employed in all the layers except the final decoding convolutional layer where a linear activation function was used. Also, Adam optimizer was used with a base learning rate of $10^{-4}$. Figure 1 illustrates the steps followed in the proposed system that denoises input scanned documents with multiple pages using a trained model.

### 2.2 Patch- versus Image-level Training

In contrast to images used from natural scenes, document page images are usually large (typically $2200\times1700$ or larger). It is not practical to directly submit such a large image to the REDNET (or any CNN-based network). There are two main alternative solutions: 1) down-sampling the page, or 2) extracting patches from the page and submitting them to the network. The latter has two main advantages: first, down-sampling degrades the quality of the image and therefore, deteriorate the performance of the noise removal network. Second, extracting patches from the images, significantly augment the dataset. In this research, therefore, patch-level training was adopted.

## 3 Experiments and Results

### 3.1 Datasets

In a noise-to-clean training strategy, noisy input and clean target output pairs are needed for the training of the network. The dataset preparation based on this strategy is explained below.

**Blurring Noise Type:** We use both Gaussian blur and box blur noise kernels to introduce noise on clean image patches. The kernel sizes for Gaussian and box blur noise are between $1\times1$ to $21\times21$. Each kernel size is applied on equal portion of patches from the whole dataset.

**Watermark Removal:** As explained in Subsection 2.2, patch-level training was adopted in this research. One major problem with extracting patches from the original watermarked documents as input to the REDNET is that the most of the patches contain no watermarks (about 90% of patches). A remedy to this problem is to divide each document page to a grid of, e.g., $4\times2$, and synthetically adding a watermark to each grid. This significantly increases the number of patches containing watermarks. An additional advantage of this method for generating noise-clean pairs is that since the watermarks are synthetically added to documents, the clean pair is naturally available. It is worthwhile to highlight that the synthetically added watermarks have variations in text, orientation, font, opacity, size , and color.

Table 1 presents a summary of the experimental setup, including the details about the data used in the training of the network.

Table 1: Experimental setup and the details of the datasets used in the experiments.

| Name | Blur | Watermark |
|---|---|---|
| No. of Documents | 116 | 410 |
| No. of Pages | 984 | 5652 |
| Page Size (pixels) | 3500×2500 | 2200×1700 |
| Patch Size (pixels) | 250×250 | 220×170 |
| No. of Patches per Page | 140 | 100 |
| Training Set Size | 137,760 | 565,200 |
| No. of Channels per Patch | 1 (Gray-scale) | 4 (RGBA) |
| Loss Function | $\ell_2$ norm | $\ell_1$ norm |
| Network | REDNET 30 | REDNET 30 |
| Epochs | 20 | 10 |

Table 2: Comparison between the performance of the OCR with and without watermark removal. The numbers shown in the table are the percentage of the OCR error compared with the ground truth. The results are averaged over 9 pages.

| No. of Pages | OCR on Watermarked Page | OCR on Cleaned Page |
|---|---|---|
| 9 | 27.0% | 8.6% |

### 3.2 Evaluation

Peak signal-to-noise ratio (PSNR) metric was used to measure improvement on the validation set. In addition, the OCR performance (in terms of the number of correct words converted by the OCR) on watermarked and cleaned documents were compared with the ground truth to quantitatively measure the effectiveness of noise removal by the trained model (Table 2). ABBYY FineReader v12 was used in this evaluation as the OCR engine.

### 3.3 Results

Figure 2 depicts the results of deblurring for different noise levels using the trained REDNET. As can be observed from the results, the trained network performs an excellent job in deblurring the patches even in the presence of very large blurring kernels. We have also tested the OCR improvement on 10 real scanned contracts with a small amount of blur noise. We measured an improvement of 0.7% in the number of valid words after cleaning the documents using the trained deblurring model.

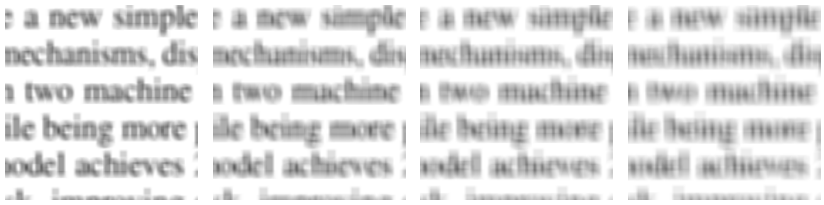
Figure 3 presents the results of watermark removal on a sample test document page with synthetically added watermarks as well as on a real watermarked document page. As the results indicate, the trained network is able to completely remove the watermarks without any visual distortion on the original text. To further investigate the effect of watermark removal on the final OCR quality, we compared the OCR accuracy in Table 2 on nine sample document pages with and without watermark removal. As can be observed from these results, the OCR performance is improved by a large margin after watermark removal using the trained REDNET.

Finally, the trained deblurring model for the setup described in Table 1 resulted in a PSNR of 34.52 dB after 20 epochs for 8 bit gray-level images in a validation set of about 10K patches. Similarly, the trained watermark removal model resulted in a PSNR of 50.24 dB on a validation set of 23.6K RGBA color patches after 10 epochs.

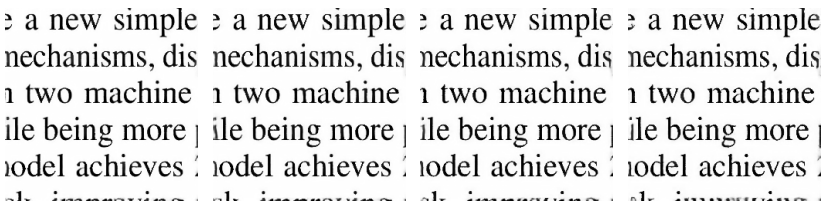
## 4 Conclusion and Future Work

The designed REDNET was successfully tested on deblurring document images with various levels of intensity as well as removing both gray-level and color watermarks from text image documents.

Currently, research on designing a unified network that can remove all noise types from text documents is ongoing.

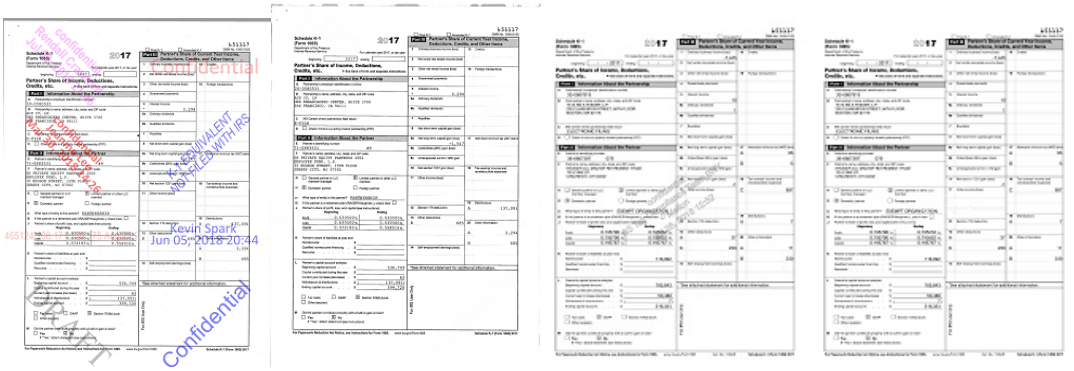


(a) 9x9 blur noise (b) 11x11 blur noise (c) 13x13 blur noise (d) 15x15 blur noise



(e) 9x9 cleaned image (f) 11x11 cleaned image (g) 13x13 cleaned image (h) 15x15 cleaned image

Figure 2: The results of cleaning different blur noise levels using a trained model. The top row (*a, b, c &d*) are input noisy patches and the bottom row (*e, f, g & h*) are cleaned patches.



(a) Original (b) Watermark Removed (c) Original (d) Watermark Removed

Figure 3: The results of watermark removal on sample test document pages with synthetically added watermarks (*(a) & (b)*) and real watermarked document (*(c) & (d)*). The images are shown in low resolutions. Best to be seen enlarged.

## References


[1] M.T. McCann, K.H. Jin, and M. Unser. A review of convolutional neural networks for inverse problems in imaging. *ArXiv*, abs/1710.04011, 2017.

[2] X. Mao, C. Shen, and Y.-B. Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems*, pages 2802–2810. 2016.

[3] G. Zhao, J. Liu, J. Jiang, H. Guan, and J. Wen. Skip-connected deep convolutional autoencoder for restoration of document images. In *24th International Conference on Pattern Recognition (ICPR)*, pages 2935–2940, Aug 2018.