Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 4, Issue 5; October-December, 2017; pp. 335-346

http://www.krishisanskriti.org/Publication.html

Document Image Denoising Using Convolutional Denoising Autoencoder

Shabana Nargis Rasool[a]

[a]*Department Of Computer Science, Islamic University of Science and Technology Kashmir India, shabana.nargis@islamicuniversity.edu.in*

Abstract

Dirty documents obscure valuable information and can make it difficult for the OCR system to accurately recognize the characters. In this research paper, we employed a deep-learning solution to denoise dirty documents and uncover their hidden treasures. Our approach employed a Convolutional Denoising Autoencoder (DAE) to remove noise from an input. The Autoencoder (AE) was trained on a dataset of noisy and clean documents to learn the relationship between the two and minimize the reconstruction error. We also explore the usage of data augmentation on the training dataset to improve performance. Finally, the performance of these approaches was compared with a traditional image-processing technique (Median Filter). Our results demonstrate that the Convolutional DAE outperformed the "Median Filter" in terms of denoising accuracy, achieving a Structural similarity (SSIM) index of 0.9626 and signal-to-Noise Ratio (PSNR) of 23.680. Additionally, we discovered that the usage of data augmentation on the training dataset further improves the performance of the model, resulting in an SSIM index of 0.9731 and PSNR of 25.248. These findings suggest that the application of deep learning techniques with data augmentation can be an effective approach for denoising document images.

Keywords: Convolutional Denoising Autoencoders, Deep Learning, Autoencoders, Data Augmentation

1. Introduction

Dirty documents, whether in the form of old manuscripts or degraded digital files, often contain valuable information that is obscured by various types of noise such as ink smudging, water damage, stains, and scanning artifacts. To preserve and access this information, it is important to denoise these documents and restore their clarity. Traditionally, denoising has been approached using methods such as image processing techniques [1, 2], thresholding [3], and filtering [4]. However, these methods are limited in their ability to remove complex noise patterns and preserve the structure and content of the documents.

In recent years, deep learning techniques, specifically Denoising Autoencoders have shown promising results in removing noise from dirty documents. Denoising Autoencoders were introduced by [5] as an extension to classic Autoencoders. Autoencoders, a type of neural network for unsupervised learning, are trained to reconstruct the input data from a lower dimensional representation, known as the bottleneck. During training, the AE learns to remove noise from the input and reconstruct the original document. Once trained, the AE can be used to denoise new, unseen dirty documents. In comparison to traditional methods, deep learning approaches offer several advantages for denoising dirty documents. Firstly, deep learning models are able to learn and remove complex noise patterns, leading to higher-quality denoised documents. Secondly, deep learning models can be trained on a large and diverse dataset, allowing them to generalize well to new, unseen documents. Finally, deep learning models can be easily adapted to new types of noise and document structures, making them highly flexible and scalable.

In this research study, Convolutional Denoising Autoencoder was employed for Denoising document images. We also studied the impact of data augmentation on the performance of Autoencoders for denoising dirty documents. Data augmentation involves synthesizing new training samples from the existing data to increase the size of the training set and improve the robustness of the model. We then compared the denoising performance of Autoencoders trained with and without data augmentation with the traditional image processing technique such as “Median Filter”. Our results show that the use of data augmentation can significantly improve the performance of Autoencoders for denoising dirty documents, leading to clearer and more readable denoised documents. This highlights the importance of data augmentation in improving the robustness and generalization ability of deep learning models, especially in the context of denoising dirty documents.

In the following sections, we describe the methodology and experiments used in our study, present the results and analysis, and conclude with a discussion of the implications and future work.

2. Literature Review

Document images can be affected by several types of document noise due to various factors such as handling of the paper, poor quality of scanning devices, environmental conditions, and more [6]. These noises can include things like smudges, creases, shadows, and distortions, which can make it difficult to accurately extract information from the image. In some cases, these noises can even make the text in the image unreadable. To address these issues, various image-processing techniques including Spatial Filtering methods and Transform Domain Filtering methods have been developed to improve the quality of the document image [7]. Linear Filters such as mean filters are simple to implement and are effective in removing random noise from images. However, they have some limitations in terms of preserving the fine details and edges in the image, which can lead to blurring and loss of information [8]. To address these limitations, the Wiener Filtering [9] technique can be used, which aims to remove noise while preserving the underlying signals. However, the Wiener Filtering technique also has its own limitations, such as the requirement for information about the noise spectra and the actual image, and its effectiveness in preserving fine details and edges is limited to smooth underlying signals [3, 10]. In some cases, other techniques such as Non-Linear Filters or Adaptive Filters may be needed to address the specific types of noise present in the document image.

Non-Linear Filters such as Morphology and Median Filters are popular techniques for removing noise from images. Morphological operations, such as erosion and dilation, are effective in removing salt-and-pepper noise [11], but they are not well-suited for non-natural images, such as document images, and can lead to other problems such as over-smoothing or loss of information. Therefore, researchers [12, 4, 13] have been exploring ways to enhance the performance of morphological techniques, but they still have some limitations that need to be addressed [14].

Transform Domain Filtering (TDF) is a popular approach for image processing and computer vision applications, including OCR. TDF methods are categorized into Non-Adaptive Transform and Data-Adaptive Transform. Non-Adaptive Transform filters are those that operate in a fixed manner and do not change based on the input image data. Examples of Non-Adaptive Transform filters include low-pass filters, high-pass filters, and Fast Fourier Transform. Non-Adaptive Transform filters have some limitations, one of which is that they can be sluggish and highly dependent on the cut-off frequency and performance of the filter function. The choice of the cut-off frequency can have a significant impact on the performance of the filter, and finding the optimal cut-off frequency can be challenging. Another limitation of Non-Adaptive Transform filters is that they generate artificial frequencies [15]. These limitations of Non-Adaptive Transform filters have motivated the development of Data-Adaptive Transform filters, which adjust their parameters based on the input image data. These filters are designed to provide more accurate results by adapting to the specific characteristics of the image. Therefore, Data-Adaptive approaches have proven to be a better option for image denoising, both for natural and non-natural images [16]. However, Data-Adaptive Transform filters are generally more computationally expensive and high-memory demanding than Non-Adaptive Transform filters [17], and finding the optimal parameters can also be challenging.

Recently, researchers in the field of image denoising have reported promising results using Support Vector Regression [17]. This Machine Learning approach has proven to be effective for both natural and non-natural images. In [18] Convolutional Neural Network has been applied as the image denoising technique. It was discovered that with only a limited number of training pictures, it is possible to get performance that is comparable to or even superior to the most recent advances in the field, which are based on wavelets and Markov random fields.

3. Preliminaries

3.1. Denoising Autoencoder

Denoising Autoencoder is a type of Neural Network (NN) architecture that is used for removing noise from images. This AE is trained to reconstruct the original, clean signal from a corrupted version of it, in which some part of the input has been masked or added with random noise. The idea behind a DAE is to first encode the corrupted input into a lower-dimensional representation, and then decode it back to the original shape. During the training process, the AE learns to separate the meaningful information from the noise, and the weights of the network are optimized so that the output is as close as possible to the original signal. Mathematically, let the original image be represented by a vector x and the corrupted image be represented by a vector x'. The goal of the DAE is to learn a mapping from x' to x that minimizes some reconstruction loss.

$$\min_{f,g} \mathcal{L}(x, g(f(x'))) \tag{1}$$

where f and g represent the encoder and decoder, respectively, and $\mathcal{L}$ represents the reconstruction loss. A common choice for $\mathcal{L}$ is the "Mean Squared Error":

$$\mathcal{L}(x, x') = \frac{1}{n}\sum_{i=1}^{n}(x_i - x'_i)^2 \tag{2}$$

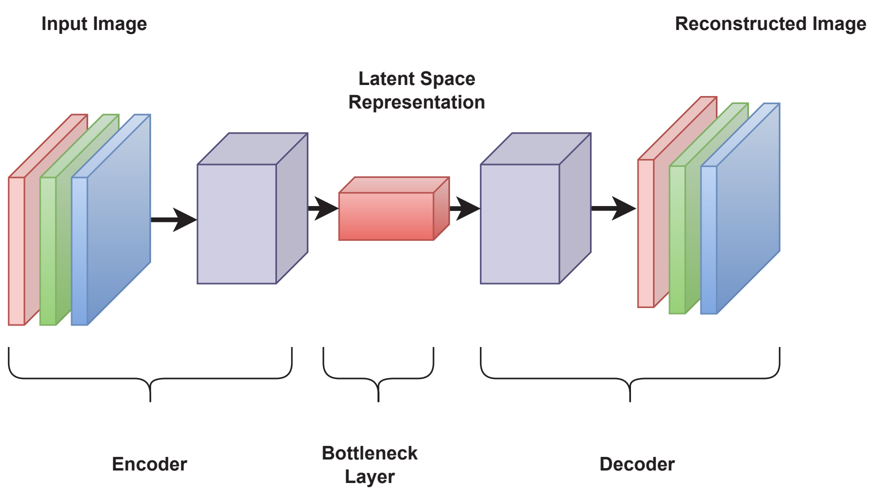


Figure 1: Convolutional Denoising Autoencoder Architecture

The DAE can be trained using gradient-based optimization algorithms, such as Stochastic Gradient Descent, to find the parameters of f and g that minimize the reconstruction loss.

3.2. Convolutional Denoising Autoencoder

A Convolutional DAE is a type of NN that is used to reconstruct an original image from a corrupted version of that image. In this case, the encoder and decoder use convolutional layers instead of fully connected layers. Figure 1 represents the architecture of Convolutional Denoising Autoencoder.

```
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 420, 540, 1)]     0

 conv2d (Conv2D)             (None, 420, 540, 32)      320

 max_pooling2d (MaxPooling2D  (None, 210, 270, 32)     0
 )

 dropout (Dropout)           (None, 210, 270, 32)      0

 conv2d_1 (Conv2D)           (None, 210, 270, 32)      9248

 max_pooling2d_1 (MaxPooling  (None, 105, 135, 32)     0
 2D)

 conv2d_2 (Conv2D)           (None, 105, 135, 32)      9248

 up_sampling2d (UpSampling2D  (None, 210, 270, 32)     0
 )

 dropout_1 (Dropout)         (None, 210, 270, 32)      0

 conv2d_3 (Conv2D)           (None, 210, 270, 32)      9248

 up_sampling2d_1 (UpSampling  (None, 420, 540, 32)     0
 2D)

 conv2d_4 (Conv2D)           (None, 420, 540, 1)       289

=================================================================
```

Figure 2: Convolutional Denoising Autoencoder model Summary

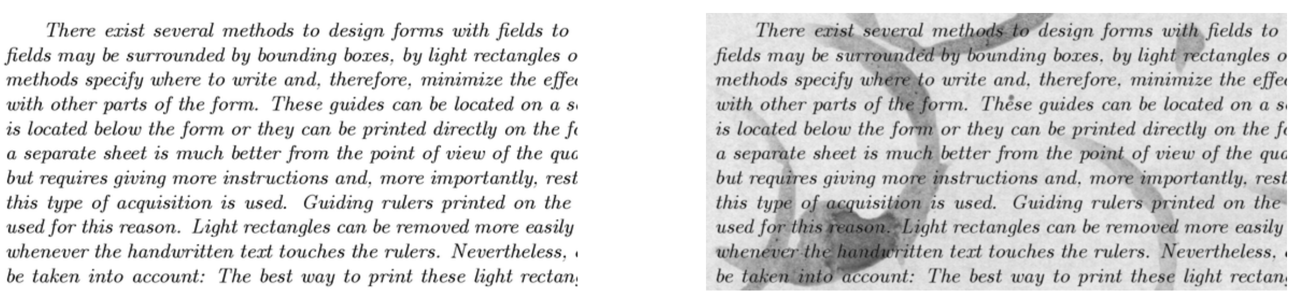
Let the original image be represented by a tensor $X \in \mathbb{R}^{h\times w\times c}$, where h, w, and c are the height, width, and number of channels of the image, respectively. The corrupted image is represented by a tensor $X' \in \mathbb{R}^{h\times w\times c}$. The goal of the Convolutional DAE is to learn a mapping from X' to X that minimizes some reconstruction loss.

$$\mathcal{L}(X, X') = \frac{1}{hwc}\sum_{i=1}^{h}\sum_{j=1}^{w}\sum_{k=1}^{c}(X_{i,j,k} - X'_{i,j,k})^2 \tag{3}$$

In a Convolutional DAE, the weights of the convolutional filters are shared across all spatial locations in the input. This allows the network to learn spatial hierarchies, where local spatial information is combined to form higher-level abstract representations. By preserving the spatiality of the input, the network is able to capture local patterns and relationships in the data, which is particularly useful for image-processing tasks. The shared weights also make the network more computationally efficient, since the number of parameters is reduced compared to a fully connected AE. Figure 2 represents the model that has been adopted in our experimentation for denoising document images.

4. Dataset

The Document Imaging dataset was collected from Kaggle, an online platform that hosts data science and machine learning competitions. The dataset contains two sets of images: a training set and a test set. The training set includes noisy document images as well as the same images without the added noise. The purpose of the noise is to simulate real-world artifacts, such as smudges, stains, or other visual interference that can make it difficult for optical character recognition (OCR) systems to accurately read and interpret the text. Figure 3a depicts the example of a clean and Figure 3b represents the example of a noisy image from the dataset.


There exist several methods to design forms with fields to fields may be surrounded by bounding boxes, by light rectangles o methods specify where to write and, therefore, minimize the effe with other parts of the form. These guides can be located on a s is located below the form or they can be printed directly on the fo a separate sheet is much better from the point of view of the qua but requires giving more instructions and, more importantly, rest this type of acquisition is used. Guiding rulers printed on the used for this reason. Light rectangles can be removed more easily whenever the handwritten text touches the rulers. Nevertheless, be taken into account: The best way to print these light rectan

(a) Clean Image (b) Noisy Image

Figure 3: Examples of Document Imaging Dataset

5. Experimentation Results

The experimentation was carried out on a "Dirty Document Imaging" dataset collected from the Kaggle competition and was divided into a training set and a validation set. The training set contained 90% of the data and the validation set contained 10% of the data. The documents contained various types of noise, such as Coffee stains, wrinkles, and dog-eared pages. A Convolutional DAE model was implemented using the Keras framework. The model consisted of an encoder that learned a compact representation of the document, and a decoder that reconstructed the denoised document from the compact representation. The model was trained on the training set using "Mean Squared Error" as the loss function. The model was trained for 500 epochs, with the best model selected based on the validation loss. The "Early Stopping" mechanism was employed to prevent overfitting by interrupting the training process when the model's performance on the validation set starts to degrade. The "monitor" argument was set to "val_loss", which means that the model's validation loss will be monitored for changes. The "mode" argument was set to "min", which means that the training process will be stopped when the validation loss stops decreasing and starts to increase. The patience argument was set to 10, which means that the training process will be stopped after 10 epochs have passed without improvement in the validation loss. Figure 4 represents the training validation loss function of Convolutional Denoising Autoencoder without applying data augmentation techniques on the training dataset.

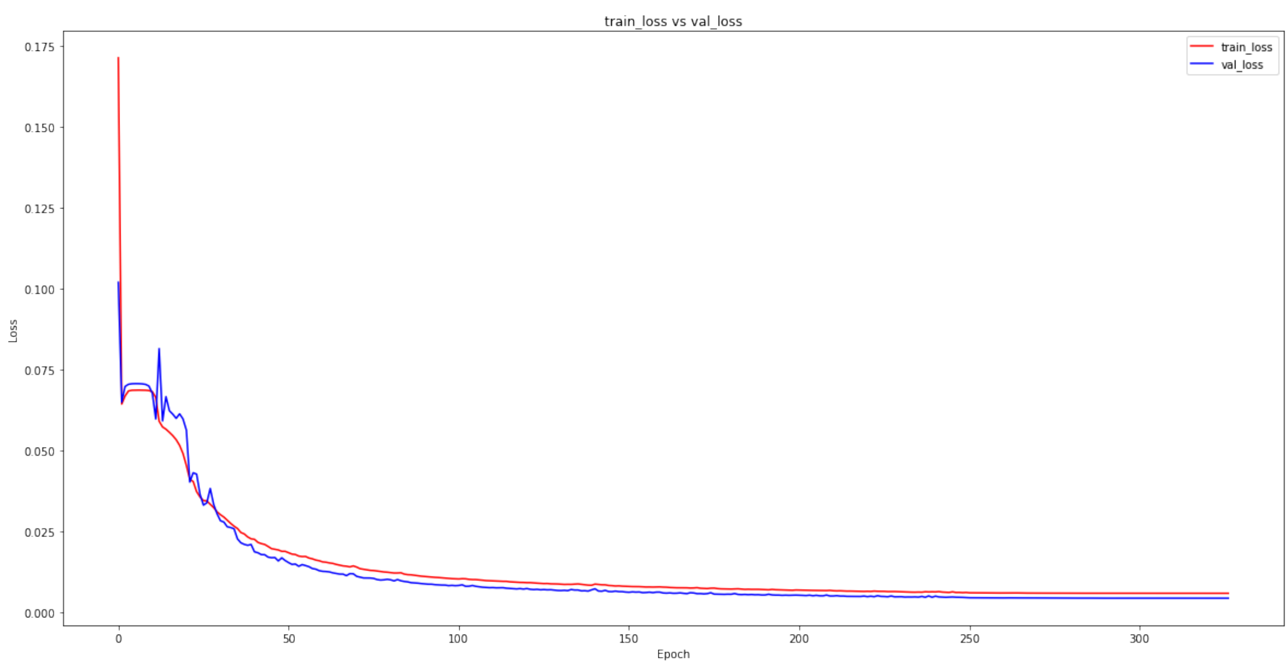


Figure 4: Training Validation Loss of Convolutional Denoising Autoencoder Without Augmentation

To improve the generalization ability of the models, data augmentation techniques were applied to the training set. The techniques used included horizontal and vertical flipping and rotation at different angles. The training dataset increased from 129 to 774. Figure 5 represents the training validation

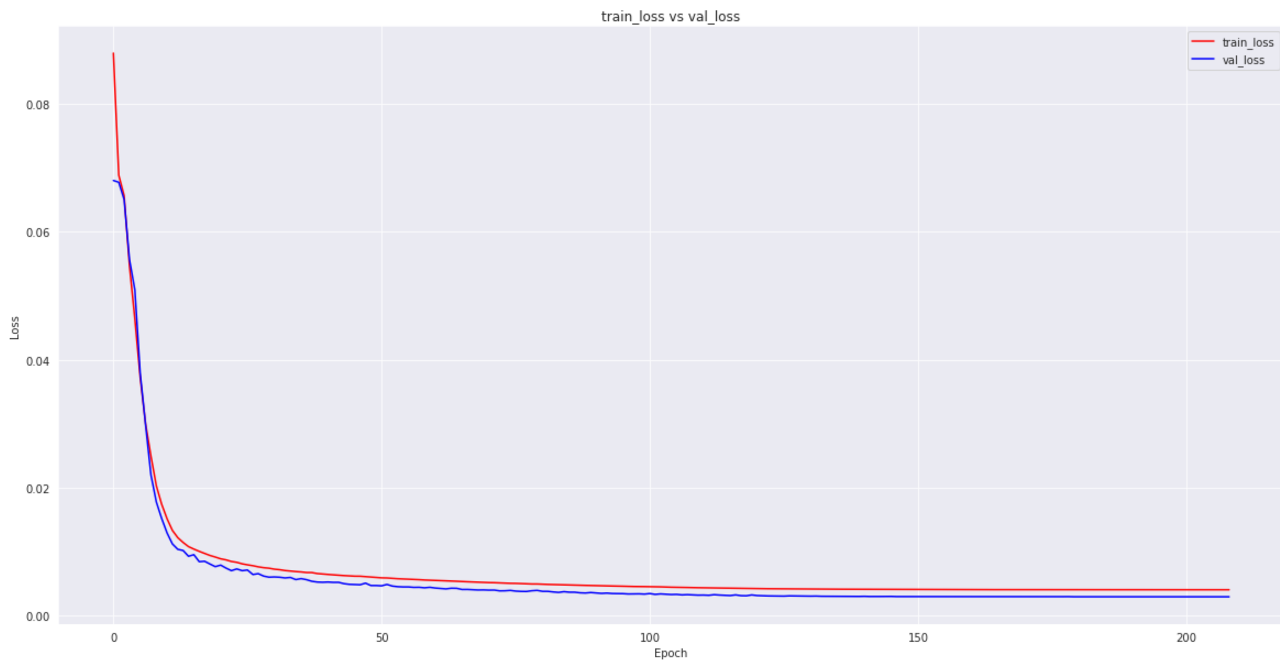


Figure 5: Training Validation Loss of Convolutional Denoising Autoencoder after Applying Data Augmentation

loss of the Convolutional Denoising Autoencoder after training the model by the augmented dataset. The trained model was evaluated on the test set using two metrics: Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The SSIM score measures the similarity between the denoised document and the clean reference document, while the PSNR score measures the signal-to-noise ratio of the denoised document.

The results of the Convolutional DAE model were compared with a baseline model that used traditional image-denoising techniques, such as Median Filtering. Figure 6 demonstrates the comparison of the Convolutional DAE model with and without augmentation and the Median Filtering technique. The results of the experiment showed that the use of data augmentation improved both the SSIM score and the PSNR score of the Convolutional DAE model. Without data augmentation, the model achieved an SSIM score of 0.9626 and a PSNR of 23.680. With data augmentation, the model achieved an SSIM score of 0.9731 and a PSNR of 25.248. Table 1 presents the SSIM index, PSNR of Convolutional DAE, and "Median Filter" models when evaluated on the test set.

Table 1: Comparison Using SSIM and PSNR for Different Models

Model	SSIM	PSNR
Convolutional DAE (Without Augmentation)	0.9626	23.680
Convolutional DAE (With Augmentation)	**0.9731**	**25.248**
Median Filter	0.8719	16.473

These results demonstrate the effectiveness of data augmentation in improving the performance of deep learning models for denoising tasks. These results suggest that Convolutional DAE is a promising approach for denoising dirty documents and that data augmentation can be an effective method to improve the performance of these models.

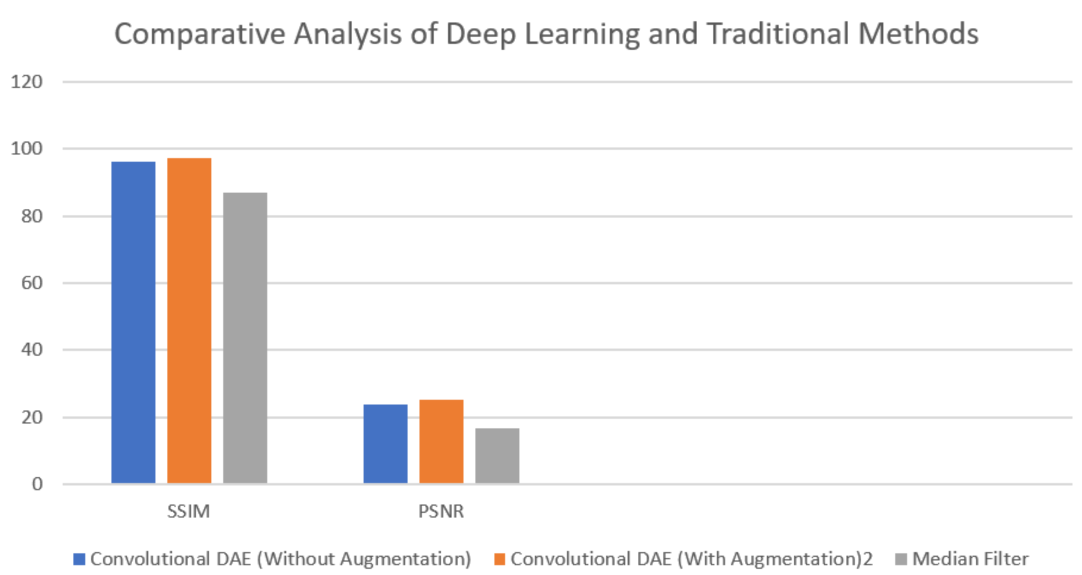


Figure 6: Comparative analysis of Convolutional DAE model and Traditional Model

6. conclusion

In this study, Convolutional Denoising Autoencoder (DAE) was implemented for denoising dirty documents. we also investigated the impact of data augmentation on the performance of Autoencoder for denoising dirty documents. Without data augmentation, the Convolutional Autoencoder model achieved an SSIM score of 0.9626 and a PSNR of 23.680. With data augmentation, the model achieved an SSIM score of 0.97 and a PSNR of 25.248. The results of this experiment suggest that Convolutional Autoencoders are a promising approach for denoising dirty documents and that data augmentation can be an effective method to improve the performance of these models. Further research is needed to explore the impact of using different types of deep learning architectures, such as Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs) on document denoising performance.

References

[1] R. R. Coifman, D. L. Donoho, Translation-invariant de-noising, Springer, 1995.

[2] J. Portilla, V. Strela, M. J. Wainwright, E. P. Simoncelli, Image denoising using scale mixtures of gaussians in the wavelet domain, IEEE Transactions on Image processing 12 (2003) 1338–1351.

[3] X. Li, G.-p. Yan, X.-m. Li, L. Chen, Image denoise based on soft-threshold and edge enhancement, in: Second Workshop on Digital Media and its Application in Museum & Heritages (DMAMH 2007), IEEE, 2007, pp. 53–56.

[4] S.-S. Wang, C.-H. Wu, A new impulse detection and filtering method for removal of wide range impulse noises, Pattern Recognition 42 (2009) 2194–2202.

[5] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proceedings of the 25th international conference on Machine learning, 2008, pp. 1096–1103.

[6] T. Saba, A. Rehman, G. Sulong, An intelligent approach to image denoising, Journal Of theoretical and Applied information technology 17 (2010) 32–36.

[7] M. C. Motwani, M. C. Gadiya, R. C. Motwani, F. C. Harris, Survey of image denoising techniques, in: Proceedings of GSPX, volume 27, Proceedings of GSPX, 2004, pp. 27–30.

[8] L. Fan, L. Fan, C. L. Tan, Binarizing document image using coplanar prefilter, in: Proceedings of sixth international conference on document analysis and recognition, IEEE, 2001, pp. 34–38.

[9] H. Farahanirad, J. Shanbehzadeh, M. Pedram, A. Sarrafzadeh, A hybrid edge detection algorithm for salt and-pepper noise, proceedings of the IMECS (2011).

[10] T. Saba, A. Rehman, M. Elarbi-Boudihir, Methods and strategies on off-line cursive touched characters segmentation: a directional review, Artificial Intelligence Review 42 (2014) 1047–1066.

[11] C. Cong-Ping, W. Jian, Q. Wu, D. Xiao-gang, A new adaptive weight algorithm for salt and pepper noise removal, in: 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), IEEE, 2011, pp. 26–29.

[12] Q. Zheng, T. Kanungo, Morphological degradation models and their use in document image restoration, in: Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205), volume 1, IEEE, 2001, pp. 193–196.

[13] A. Mehrotra, K. K. Singh, M. Nigam, K. Pal, A novel algorithm for impulse noise removal and edge detection, International Journal of computer applications 38 (2012) 30–34.

[14] A. Buades, B. Coll, J.-M. Morel, A review of image denoising algorithms, with a new one, Multiscale modeling & simulation 4 (2005) 490–530.

[15] H. S. Al-Khaffaf, A. Z. Talib, R. A. Salam, Removing salt-and-pepper noise from binary images of engineering drawings, in: 2008 19th International Conference on Pattern Recognition, IEEE, 2008, pp. 1–4.

[16] M. Mundher, D. Muhamad, A. Rehman, T. Saba, F. Kausar, Digital watermarking for images security using discrete slantlet transform, Applied Mathematics & Information Sciences 8 (2014) 2823.

[17] M. Mahmoudi, G. Sapiro, Fast image and video denoising via nonlocal means of similar neighborhoods, IEEE signal processing letters 12 (2005) 839–842.

[18] V. Jain, S. Seung, Natural image denoising with convolutional networks, Advances in neural information processing systems 21 (2008).