Proceedings of the 5th International Conference on Smart Systems and Inventive Technology (ICSSIT 2023)
IEEE Xplore Part Number: CFP23P17-ART; ISBN: 978-1-6654-7467-2

Medical Image Denoising using Convolutional Autoencoder with Shortcut Connections

Manas Gupta
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
gmanas813@gmail.com

Kushal Goel
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
kushal8601@gmail.com

Jatin Kansal
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
jatinkansal81@gmail.com

Anurag Goel
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
anurag@dtu.ac.in

2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) | 978-1-6654-7467-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSIT55814.2023.10061131

***Abstract*—In recent times, medical image analysis has gained immense attention among researchers for diagnosing and treating deadly diseases, e.g., Cancer. Denoising of Images is considered as a crucial process in medical image analysis. Several Deep Learning techniques are effectively used for image denoising including Autoencoders. In this paper, a convolutional autoencoders based approach with shortcut connections is proposed for medical image denoising. The proposed approach is evaluated by using three medical images datasets. The results demonstrated that the proposed approach outperforms the current cutting-edge methods of medical image denoising on all the three datasets.**

***Keywords*—Medical Image Denoising, Convolutional Neural Network, Convolutional Autoencoders, Shortcut Connections.**

I. INTRODUCTION

Noise refers to the distortions in an image. It can be in the form of brightness variations or color information in an image. Denoising is the technique used to eradicate these distortions.

Various techniques, including Ultrasound Imaging, Radiography, etc., are used to view the internal structure of the human body to diagnose various medical conditions. Such technologies are categorized as Medical Imaging and are vulnerable to noise. The vulnerability could arise from using distinct image acquisition methods to reduce patient exposure to radiation. The more the radiation exposure is reduced, the more the noise is increased. Other reasons could also result in a noisy image, like bit errors in transmission, statistical quantum fluctuation, etc. For medical analysis, medical image denoising is usually required to diagnose and treat diseases properly.

Being a classical problem, image denoising remains a popular topic of research. There are different image denoising approaches, such as Domain Transformation based models, Partial Differential Equation based models. The equation for the problem would be:

$$z = x + y \tag{1}$$

Here, x refers to the original image, y is the noise, and z is the image generated after the addition of noise in the original image. It is assumed that noise is generated from a distinct process, and using this assumption, y is calculated. Most of the models try to approximate x using z.

With the evolution of deep learning, the deep learning models perform better than the conventional denoising approaches. In deep learning models, Convolutional autoencoders are majorly implemented for image denoising task. Convolutional autoencoders consist of encoder-decoder network, which is built using Convolutional Neural Networks. An image denoising technique is used for medical dataset based on Convolutional autoencoders with shortcut connections.

The paper is divided into a number of sections. Related Work study is described in Section II. The planned approach is illustrated in Section III. The results of the work is shown in Section IV while in Section V conclusion is stated.

II. RELATED WORK

In [1], a review of various Convolutional Neural Network (CNN) based approaches for image denoising have been presented. In [1], it was shown that the CNN-based approaches gave a good performance. CNN models are designed with the information of noise in the dataset. In [2], the correlation between medical image denoising and medical image classification is explored using DenseNet-121 and CNN models.

In [2], it was concluded that image denoising significantly affects the performance of image classification. The approach proposed in [3] uses CNN for image denoising. It was found that performance can be equivalent to or higher than the methods established upon wavelets.

Autoencoder encodes an input into a latent space representation using the encoder network, and then the latent space reprsentation is converted back to the original input utilizing the decoder network. Through training, the model develops the ability to map various input images to precise locations in the latent space [4].

Vincent et al. [5] proposed Denoising Autoencoders, which are an extension to the traditional autoencoders, helping reconstruction of the input by introducing a noisy version. Later, the same authors proposed a stacked version of denoising autoencoders by stacking denoising autoencoders one over the other [6]. A Denoising Autoencoder based architecture is also proposed by Robinet et al. [7] for image denoising.

Convolutional encoding and decoding layers are added to a typical autoencoder architecture to create Convolutional autoencoders, proposed by Jonathan et al. [8]. Convolutional autoencoders are considered to be better for image processing as compared to classic autoencoders because they utilize the power of CNNs to exploit the contextual information in the image.

Lovedeep et al. [9] proposed Convolutional Neural Network based Denoising AutoEncoders (CNN DAE) and have shown that CNN DAE outperformed the median filter in terms of denoising performance on small datasets. It was suggested that these techniques could recover signals even in situations with high noise levels when the majority of denoising techniques would fall short. However, this straightforward network has problems recreating the original signal when the noise level is relatively high. However, this network is successful at partially generating actual images even when they are invisible to the human eye.

Prashanth et al. [10] have shown the effectiveness of Convolutional Autoencoder for image denoising. In Convolutional Autoencoders, convolutional layers are used instead of dense layers in standard autoencoders. In [10], it was demonstrated that the completely connected Autoencoder, that just uses dense layers, is not efficient to denoise the input; while Convolutional autoencoders output a nearly noise-free image. This is because, when mapping an image to a latent space, the convolutional layers extract and preserve the essential input features and eliminate noise.

From the current state-of-the-art technologies, it is observed that there is a need to reduce overfitting of the model. Deeper architectures sometimes prove to be less efficient because of memorizing the training data.

III. PROPOSED APPROACH

The approach suggested is based on Convolutional AutoEncoder with shortcut connections. Shortcut connections boost the performance of the model by reducing the depth of the architecture of the model. This helps in eliminating the problem of over training the model as well. The representation of a shortcut connection is shown in Fig. 1.

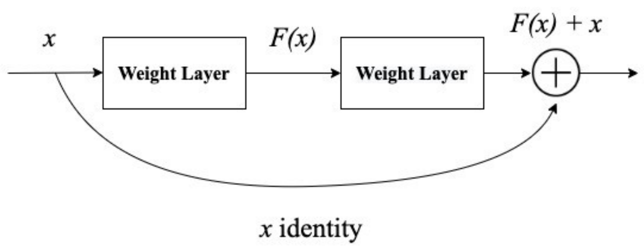

Fig. 1. Representation of Shortcut Connections

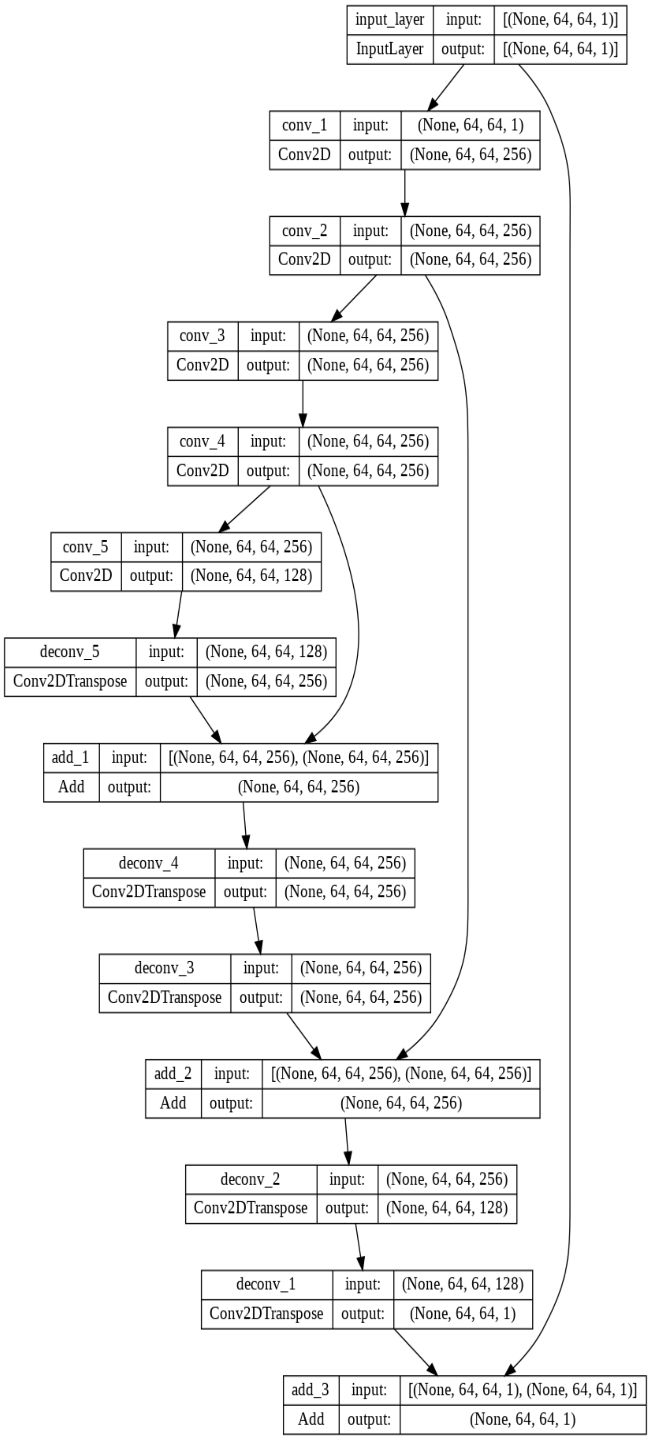

Fig. 2. Block Diagram Architecture of Proposed Approach

Fig. 2 illustrates the architecture of the proposed approach. The architecture includes an encoder network which is built using Convolutional layers and a decoder network consisting of Deconvolutional layers. Shortcut connections are introduced in the architecture. The goal of introducing the shortcut connections is to pass the features of an image from one level to another without interfering with in-between layers.

The proposed model takes an input image and fed it to a number of convolutional layers and transposed convolutional layers. The convolutional layers bring out the essential characteristics of the image and also help in removing noise from images. The kernel filters in the convolution layers processed over the input image to extract features and create a feature map. Deeper layer extracts more complex features from images.These feature maps store essential characteristics of the image needed to reconstruct it to the original form. In the proposed model, five convolutional layers have been implemented in which the first four layers use 256 convolutional kernel filters while the last layer uses 128 convolutional kernel filters. The proposed model also consists of five transposed convolutional layers (deconvolutional layers) used for the purpose of rebuilding the original image. The final output produced is a clean and clear image without noisebecause the shortcut connections eliminates the problem of vanishing gradient and deeper architecture.Also early stopping is used to avoid overfitting

The shortcut connections are implemented in between the deconvolutional layers. The shortcut connections are implemented with the help of add layer which adds the output from one layer to another layer directly skipping the layers in between.

In the proposed model architecture, pooling layers are not used because pooling may remove the essential information which becomes critical in the study of medical dataset. Using the shortcut connections, the feature maps extracted by the convolutional layers are passed directly to the deconvolutional layers by skipping some layers in between. This solves the problem of overfitting caused by the deep layer architecture of the model.

IV. EXPERIMENTS AND RESULTS

A. Datasets Used

As suggested in [14], large scale versatile images prove better than small scale similar images. The following datasets are used for performing experiments:

1) Chest X-Ray Dataset: In this dataset, there are total 247 Chest X-ray images [11]. The training set and the testing set conisists of 197 images and 50 images respectively. A sample image is shown in Fig. 3.

Fig. 3. Sample Image of Chest X-Ray Dataset

2) Dental X-Ray Dataset: In this dataset, there are total 120 images [12]. The training set and the testing set consists of 96 images and 24 images respectively. A sample image is shown in Fig. 4.

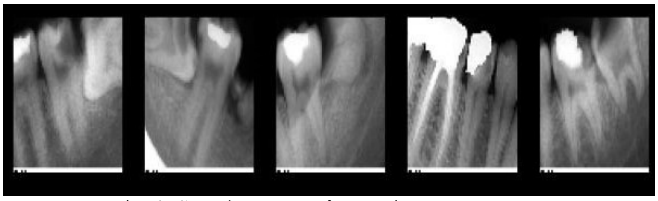

Fig. 4. Sample Image of Dental X-Ray Dataset

3) Covid CT Dataset: In this dataset, there are total 349 images of Computed Tomography (CT) scans of Covid patients as well as non-Covid patients [13]. For the training set and the testing set, 279 images and 70 images are used respectively. A sample image of this dataset is shown in Fig. 5.

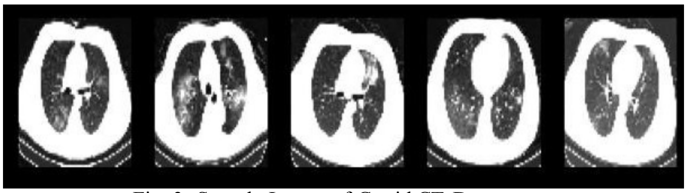

Fig. 3. Sample Image of Covid CT Dataset

B. Experimental Setup

The experiments have been performed on an Intel i5 processor running Ubuntu. The code has been written in Python language. For pre-processing, the images are loaded in grayscale mode with a target size of (64,64) and Gaussian noise (7%) is added to images due to the fact as stated in [15], Gaussian noise is additive in which shows contribution of Gaussian noise in each pixel. The images are used in the form of array with normalized values.

Number of epochs for which model is trained is 20 with batch size of 10.Early stopping is also used in the model for better performance.

C. Metrics Used

The results of the approach planned is matched with the current methods of denoising of medical dataset using the following metrics:

1) Peak Signal To Noise Ratio (PSNR): This evaluation metric is the ratio between the maximum power of the signal and the power of noise signal that affects reconstruced image quality. If the PSNR is high, reconstructed image would be better. The formula is given as follows:-

$$PSNR = 20 \log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \quad (2)$$

where MSE(Mean Squared Error) is calculated as follows:-

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} ||f(i,j) - g(i,j)||^2 \quad (3)$$

where,

- f represents the data matrix of noiseless image;
- g represents the data matrix of noisy image;
- m represents the total number of rows of pixels in the image;
- n represents the total number of columns of pixels in the image;
- MAX_f is the maximum signal value exists in the noiseless image.

2) Structural Similarity Index Measure (SSIM): It is one of the evaluation metrics which measures the quality of the reconstructed image with that of the reference image. Hence, higher would be the SSIM score, better would be the quality of the reconstructed image relating to structural similarity with reference image. The formula is given as follows:-

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

where:-

- μ_x is the mean of x;
- μ_y is the mean of y;
- σ_x^2 is the variance of x;

- σ_y^2 is the variance of y;
- σ_{xy} is the covariance of x and y;
- $c_1=(k_1L)^2$, $c_2=(k_2L)^2$ are to stabilize the division with weak denominator;
- L is the dynamic range of the pixel values;
- k_1 =0.01 and k_2 =0.03 (default values) are constants.

D. Results Analysis

Table I shows the SSIM scores and Table II depicts the PSNR scores. From the results, it can be interpreted that the proposed model outperforms the benchmarks on all the datasets by a significant margin.

TABLE I. SSIM SCORES

	CNN DAE [9]	CAE [10]	Proposed
Chest X-Ray Images Dataset	0.92	0.75	**0.93**
Dental X-Ray Images Dataset	0.86	0.66	**0.89**
Covid CT Dataset	0.74	0.70	**0.87**

TABLE II. PSNR SCORES

	CNN DAE [9]	CAE [10]	Proposed
Chest X-Ray Images Dataset	51.17	46.48	**52.51**
Dental X-Ray Images Dataset	49.35	45.61	**51.72**
Covid CT Dataset	43.44	42.55	**48.27**

For the Chest X-Ray dataset [11], Fig. 6, Fig. 7 and Fig. 8 show the results of CNN DAE model [9], CAE model [10] and the proposed model respectively. In each figure, the first row shows the noiseless image, the second row shows the noisy version of the corresponding first row image, and third row shows the output of the respective model on the corresponding first row image.

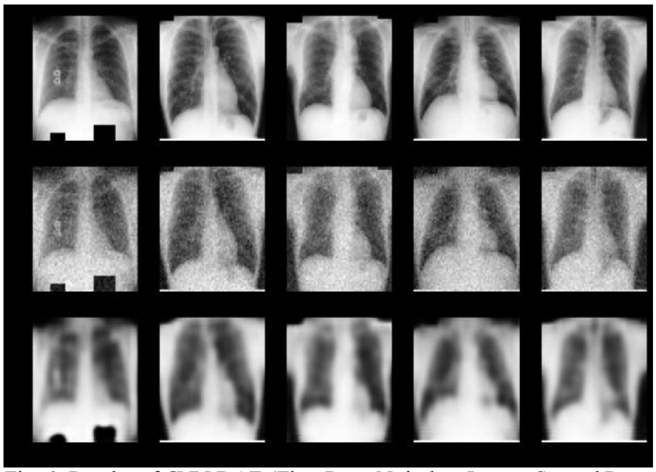

Fig. 6. Results of CNN DAE (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of CNN DAE)

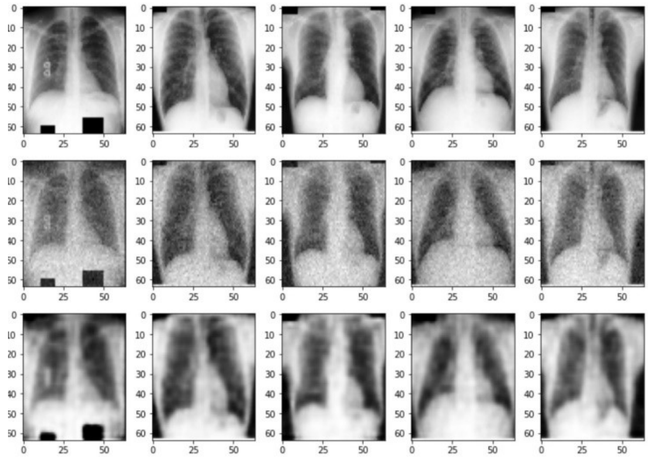

Fig. 7. Results of CAE (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of CAE)

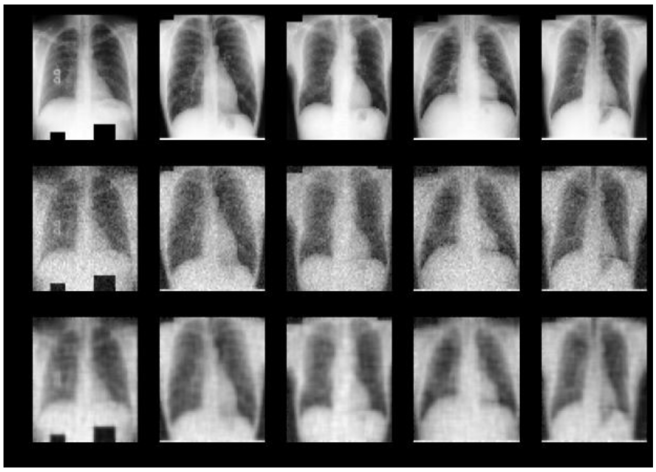

Fig. 8. Results of Proposed Model (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of Proposed model)

V. CONCLUSION AND FUTURE WORK

Medical image denoising is considered extremely crucial in medical image analysis. In this work, on a Convolutional Autoencoder based approach with Shortcut connections has been proposed for medical image denoising. The shortcut connections pass the features from one level to another by skipping the in-between layers. For the experiments, three medical images datasets have been used. The proposed approach is compared with several benchmark models of medical image denoising. The results demonstrated that the proposed approach outperforms the benchmarks by a significant margin on all the datasets.

Autoencoders have limitations of overfitting in case of large dataset due to their dual-network architecture of encoder and decoder. Since, the proposed model is based on Convolutional Autoencoder, it also faces the limitations of overfitting due to the high number of learning parameters of encoder as well as decoder. This limitation of dual-network architecture can be explored in future.

REFERENCES

[1] Ilesanmi, Ademola E., and Taiwo O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review." Complex & Intelligent Systems 7, no. 5 (2021): 2179-2198.

[2] Michael, Peter F., and Hong-Jun Yoon, "Survey of image denoising methods for medical image classification.", In Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, pp. 892-899. SPIE, 2020.

[3] Jain, Viren, and Sebastian Seung. "Natural image denoising with convolutional networks." Advances in neural information processing systems 21 (2008).

[4] Baldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." In Proceedings of ICML workshop on unsupervised and transfer learning, pp. 37-49. JMLR Workshop and Conference Proceedings, 2012.

[5] Vincent, Pascal, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103. 2008.

[6] Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." Journal of machine learning research 11, no. 12 (2010).

[7] L. Robinet. (2020), "Autoencoders and the denoising feature: From theory to practice," [Online]. Available: https://towardsdatascience.com/autoencoders-and-the-denoising-feature-from-theory-to-practice-db7f7ad8fc78

[8] Masci, Jonathan, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. "Stacked convolutional auto-encoders for hierarchical feature extraction." In International conference on artificial neural networks, pp. 52-59. Springer, Berlin, Heidelberg, 2011.

[9] Gondara, Lovedeep. "Medical image denoising using convolutional denoising autoencoders." In 2016 IEEE 16th international conference on data mining workshops (ICDMW), pp. 241-246. IEEE, 2016.

[10] Venkataraman, Prashanth. "Image Denoising Using Convolutional Autoencoder." arXiv preprint arXiv:2207.11771 (2022).

[11] Shiraishi, Junji, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules." American Journal of Roentgenology 174, no. 1 (2000): 71-74.

[12] Dental X-Ray Images Dataset: https://www.kaggle.com/datasets/parthplc/medical-image-dataset

[13] Yang, Xingyi, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, and Pengtao Xie. "COVID-CT-dataset: a CT scan dataset about COVID-19." arXiv preprint arXiv:2003.13865 (2020).

[14] Sharif, S. M. A., Rizwan Ali Naqvi, and Mithun Biswas. "Learning medical image denoising with deep dynamic residual attention network." Mathematics 8, no. 12 (2020): 2192.

[15] Verma, Rohit, and Jahid Ali. "A comparative study of various types of image noise and efficient noise removal techniques." International Journal of advanced research in computer science and software engineering 3, no. 10 (2013).