Chapter 6

Denoising of Magnetic Resonance Images Using Discriminative Learning-Based Deep Convolutional Neural Network

Highlights of the Chapter

- *The present chapter proposes a deep learning based method for denoising of MR images.*
- *Proposed method is based on depth wise separable convolution and LRN.*

Contribution of the chapter

The noise in MR images causes severe issues for medical diagnosis purposes. This chapter presents a discriminative learning based denoiser to denoise the MR image data, contaminated with noise. The proposed method incorporates the use of depthwise separable convolution along with local response normalization with modified hyperparameters and internal skip connections to denoise the contaminated MR images. Moreover, the addition of Parametric RELU instead of normal conventional RELU in our proposed architecture gives more stable and fine results. The denoised images were further segmented to test the appropriateness of the results. The network is trained on one dataset and tested on other dataset produces remarkably good results. Our proposed network was used to denoise the images of different noise levels, and it yields better performance as compared with various networks. The SSIM and PSNR showed an average improvement of (7.2 ± 0.002) % and (8.5± 0.25) % respectively when tested on different datasets without retaining the network. An improvement of 5% and 6% was achieved in the values of mean intersection over union (mIoU) and BF Score when the denoised images were segmented for testing the relevancy in biomedical imaging applications. The statistical test suggests that the obtained results are statistically significant as $p < 0.05$. The

denoised images obtained are more clinically suitable for medical image diagnosis purposes, as depicted by the evaluation parameters. Further, external clinical validation was performed by an experienced radiologist for testing the validation of the resulting images.

6.1 Introduction

The removal of noise present in MR images is a topic of utmost importance. Many methods have been proposed for the retrieval of noise-free images [46] [15]. Denoising methods based on discriminative model learning are attracting considerable attention [149]. In low level vision images, denoising is still an active topic as it aims to remove randomness in the form of noise. Noise appears as grains in the images and it is a random variation in the intensity of images. In quantitative image processing, the accuracy and reliability of diagnosis depends on the noise level; as the noise level increases, the diagnosis process is affected. Hence, denoising of images is an important pre-processing step in image analysis. A commonly encountered noise in MR images is Rician noise [49]. The spatial distribution of noise in MR images is usually modeled as Rician distribution because of the nature of the Gaussian noise in both real and imaginary parts of the K-space raw data. According to reported literatures, Rician noise induces problems in the clinical diagnosis of MR images and may lead to spurious diagnosis [150] [151]. In biomedical image processing, noise removal has been a challenging task, as it is of particular importance in computer based post-processing procedures such as registration, classification, segmentation, etc. The noisy images creates a lot of issues in accurate segmentation of the region of interests in biomedical images. Thus image denoising is quite necessary step prior to any image processing task. Over the past few years, many image denoising models have been explored to reduce/remove the level of noise present in any biomedical image. In the non-local means (NLM) method [152] [153], the block representation is exploited to restore the noisy slices from the desired slices. A filtering method based on linear minimum mean square error (LMMSE) [154] that deploys the property of self-similarity to observe noise data was proposed

by Golshan and Hasanzadeh. Markov random fields were exploited in Baselice et al. [155] for reduction in noise level without any supervision. To automate the denoising process, Chang and Chang used an artificial neural network with image texture feature analysis [156]. Martin-Fernandez proposed a denoising method with shrinkage of wavelet coefficients based on conditional probability theory [157]. Another denoising approach based on patches is block matching and 3D filter (BM3D) [158], which combines the idea of non-locality and domain transform. It first groups patches into an array and then transforms the array in the frequency domain using wavelet transform. Finally, it produces noise estimation at each location.

To enhance the quality of MR images, one can directly apply discriminant learning. Hence, we extended the Denoising Convolutional Neural Network (DnCNN) model, proposed by Zhang et al. to extract the noise free image from noisy observations [58]. Further, to reduce the level of Rician noise in MR images, we used skip connections with local response normalization . To reduce the noise percentage in order to improve clinical diagnostic capabilities, we introduced a pixel-by-pixel calculation method which is capable of reproducing images which are closer to their expected counterparts. The proposed network avoids the optimization of non-convex cost function as required in the discussed methods.

6.1.1 Nature of Rician noise

The main source of noise in MR images is thermal in nature, which originates from the stochastic motion of free electrons [159]. Thermal noise is white additive and follows Gaussian distribution having variance (σ) and zero mean (μ) [160]. The K-space data is first acquired and is delineated in the presence of Gaussian noise. Further, the Fourier transform of K-space data is used to get the MR images. For image computer vision analysis, the magnitude of the MR images is used. The magnitude reconstruction is the square root of the sum of two

independent Gaussian random variables, which is described as the magnitude image data by a Rician distribution.

Suppose 'R' and 'I' are the real and imaginary parts of the noisy MR data (corrupted with zero mean Gaussian, stationary noise with the standard deviation σ) with mean values μR and μI, respectively. Then the probability distribution function (PDF) of the magnitude data will be a Rician distribution, as described by equation 1:

$$P_{mag}(M) = \frac{M}{\sigma^2}\, e^{-\frac{M^2+A^2}{2\sigma^2}}\, I_0\left(\frac{AM}{\sigma^2}\right), \qquad M \geq 0 \tag{6.1}$$

Where, $$M = (R^2 + I^2)^{\frac{1}{2}} \tag{6.2}$$

$$A = (\mu R^2 + \mu I^2)^{\frac{1}{2}} \tag{6.3}$$

I_0 is the Bessel function of the first kind and order zero.

The Rician distribution tends to be a Rayleigh distribution when the SNR approaches zero, as described by equation 4.

$$P_{mag}(M) = \frac{M}{\sigma^2} e^{-\frac{M^2}{2\sigma^2}}, \mathrm{M} \geq 0 \tag{6.4}$$

The Rician distribution tends to be a Gaussian distribution when SNR is high, as described by equation 5:

$$P_{mag}(M) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(M-\sqrt{A^2+\sigma^2})^2}{2\sigma^2}}, \mathrm{M} < 0 \tag{6.5}$$

Thus in low SNR regions in MR images the Rician noise behaves as a Rayleigh noise, while in high SNR regions it follows Gaussian distribution.

6.2 Methodology

Zhang et al. developed DnCNN from a VGG network [66] for the purpose of image denoising. To apply this model for denoising of MR images with improved performance, we have introduced depth wise separable convolution, local response normalization and internal skip

connections. It has been observed that with the increase in the number of layers, the problem of vanishing/exploding gradient may prove critical [161] [72]. Moreover, the residual network is easier to optimize and can gain accuracy with the increased number of layers. Thus to set the trade-off between the number of layers with the problem of vanishing/exploding gradient, we proposed modifications in the existing DnCNN model with skip connections, local response normalization [97] [65] and parametric rectified linear unit (PRELU) [98] at $\alpha = 0.05$. The use of a max-pooling layer is avoided to obtain an output image that is the same size as the input image. In this approach, the denoiser attempts to formulate the mapping function for noise instead of image to be denoised. It predicts N(y) = n instead of N(y) = a, where N(y) is the mapping function, a is the image to be recovered and n is the noise. The depthwise separable is computationally cheaper as the number of multiplications is reduced to a more considerable extent. Moreover, it reduces overfitting as the number of parameters to adjust are less in the case of depthwise separable convolution (as discussed in chapter 5 and 6). Figure 6.1 shows the depthwise separable convolution block with LRN.

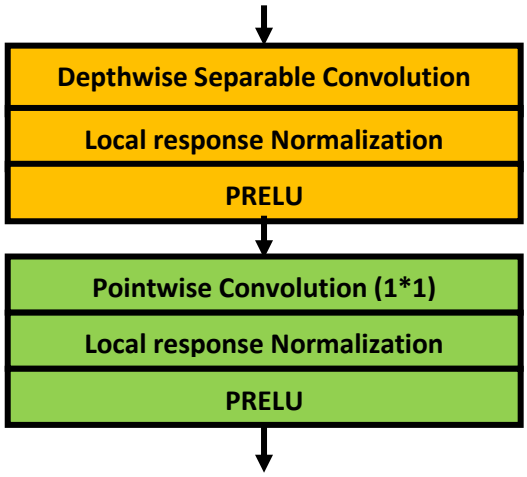

Figure 6.1: Depthwise Separable block with LRN

The feature maps produced by depthwise separable convolution possesses more information about boundary and object details. In depthwise separable convolution, the image is

transformed only once and then it is elongated across M channels, where M is the number of channels while for normal convolution the image is transformed M times. In this process, the computational power is saved to a large extent. Moreover, the number of multiplications gets reduced by a large extent which is notable merit of this approach. With each convolution, an extra multiplication is performed, and the boundary detail is lost during every such operation. The higher number of multiplications will lead to the blurred boundary details. Depthwise separable convolution eliminates the issue of blurring boundary by reducing the number of multiplications. The proposed changes yield better results than the existing denoising methods, as shown by the comparison in the results and discussion section. The network is designed to predict noise at each pixel of the original image. The hidden layers of the proposed model remove the noise from the noisy observations. In the proposed network, the estimation of noise level parameters is not needed. Hence, an increase in denoising accuracy is achieved without any dependence on estimation of noise level.

The training and testing images were contaminated programmatically with Rician noise. The Rician noise percentage ranged from 5% to 35%.The proposed network is shown in figure 6.2 which is composed of depthwise separable convolution blocks normalized with LRN. The size of the output image is the same as the input image. In the proposed network, 64 filters with filter size 3×3× c are used, where the value of c is 1 for 2D grey scale images. For optimizing the parameters of the proposed network, the mean square error between the desired image and the estimated image from the noisy input was calculated using the loss function of the network as given by equation 6:

$$l(\theta) = \frac{1}{2}\sum_{i=1}^{R}(t_i - y_i)^2 \tag{6.6}$$

Where, t_i is the target output and y_i is the network's prediction for response i. The MSE is minimized to achieve higher PSNR values which indicates minimization of noise i.e. Rician in case of MR images.

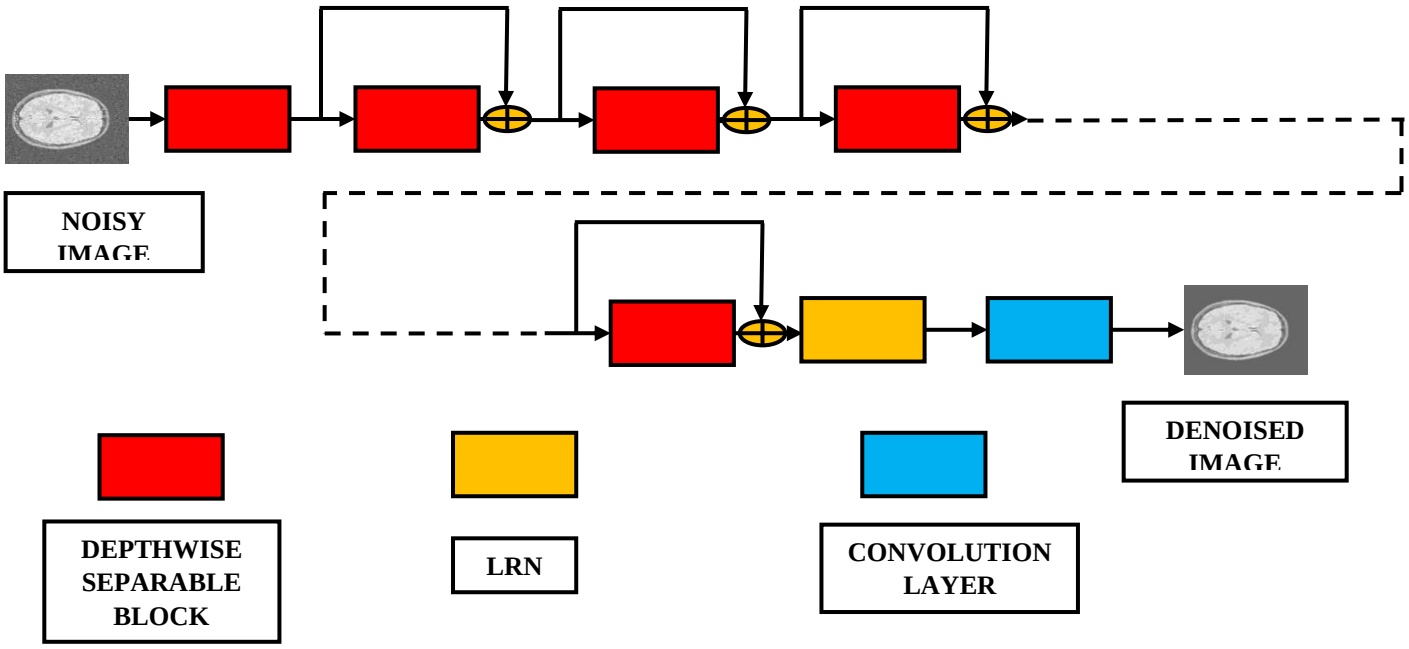

Figure 6.2: Architecture of the proposed denoising model.

In most existing methods it is desired to estimate the level of noise, and then carry out denoising based on the estimated noise level. Accordingly, the denoising results obtained are largely dependent on the correctness of the initial noise estimation, and good results are not achieved if there is variability of noise level. Therefore, we have proposed a method that automatically estimates and denoises the image in one go, in contrast to other existing methods, and which is capable of handling noise level variability. We trained our model using local response normalization (LRN). LRN creates competition for big activities among neuron output using different kernels which are inspired by the type found in real neurons. The LRN uses lateral inhibition, i.e. the capacity of a neuron to reduce the activity of its neighbors. The motive of the lateral inhibition is to carry out local contrast enhancement so that locally maximum pixel values are used as excitation for the next layers. The pixel values are square normalized in

feature maps within the local neighborhood. The LRN accurately detects the high-frequency features among the large inputs. It dampens the response which is uniformly large in the local neighborhood and performs normalization in multiple directions. LRN reduces the effect of uniformly distributed Rician noise in the MR images as it has the capability of dampening the common uniform high frequency effects. The common and uniform high frequency effects in the neighborhood (noise) once taken in to account are dampened with the use of LRN as noise is uniformly distributed over the range of pixels. In deep learning networks, there is a loss of information while reconstructing the images because of degradation issues. Skip connections are used to avoid the degradation issues caused in deep neural networks. These skip connections send the feature maps from the earlier layer directly to the next layers. Thus the information contained in images is well preserved when skip connections are used in the process of denoising. The hyperparameters of local response normalization were fixed as: k = 2, α = 0.1, β = 0.75. The constants k, α and β are hyper-parameters whose values are determined using a validation set. *K* is used to avoid any singularities (division by zero), *α* is used as a normalization constant, while β is a contrast constant. In local response normalization, $a_{x,y}^{i}$ denotes the activity of a neuron computed by applying kernel i at position (x, y) and then applying the RELU. The response-normalized activity $b_{x,y}^{i}$ is given by equation 7:

$$b_i = \frac{a_i}{\left(k + \alpha \sum_{j=\max\left(0,i-\frac{n}{2}\right)}^{\min\left(N-1,i+\frac{n}{2}\right)} \left(a_{x,y}^{j}\right)^2\right)^{\beta}} \tag{6.7}$$

where the sum runs over n "adjacent" kernel maps at the same spatial position, and N is the total number of kernels in the layer. In the proposed network, local response normalization is infused with the parametric RELU (α = 0.05). The use of PRELU instead of normal RELU prevents the saturation in RELU as they have negative slope. Thus they produce some output

values for negative inputs instead of zero as in the case of normal RELU. This infusion gives remarkable results compared to other methods.

6.2.1 Clinical external validation of the denoised data

For validating the effectiveness of the proposed network, the trained network was tested on the dataset acquired under different environment. Moreover, manual inspection was performed subsequent to the careful visual inspection of the images by a senior radiologist having experience of more than ten years in clinical reporting magnetic resonance imaging. This was done for the purpose of external validation from the clinical application point of view.

6.2.2 Testing the applicability of denoised data by segmentation

Even the minute details have greater importance in biomedical imaging as the incorrect analysis may lead to the blunder in the process of diagnosis. Thus the denoised images are further segmented for testing the applicability of the denoising method. Segmentation is done using local center of mass approach. Images denoised by our proposed method are well segmented in comparison to the other methods in the study as indicated by the segmentation evaluation metrics. The variable intensity components areas of necrosis cystic change and calcification were however included with in the region of segmentation. The segmented images of our denoising model were compared with the other methods in this study.

6.2.3 Comparison with other methods

The performance evaluation of the proposed network was done by comparing the proposed method with the other three well-established methods, which are discussed below.

6.2.3.1 Weighted Nuclear Norm Minimization (WNNM) Method [162]

This method is an extension of the nuclear norm minimization method. It is convex and the singular values are assigned to different weights. The property of image non-local self-

similarity is exploited in this method. It leads to visible PSNR improvement, preserves better image local structure, and generates many fewer visual artefacts.

6.2.3.2 Non-local Means (NLM) Denoising Method [159]

This method performs non-local averaging of all the pixels in the image. NLM compares the geometrical configuration in the whole neighborhood and the gray level in the single point. Thus, a more robust comparison is obtained compared to the other neighborhood algorithms.

6.2.3.3 Denoising Convolutional Neural Network (DnCNN) [58]

The DnCNN method proposed discriminative model learning for image denoising. The use of residual learning separates the noise from the noisy image. For speeding up the training process, the batch normalization and residual learning are integrated. Moreover, the denoising performance is also enhanced. Thus, this method proposes a different approach than the other methods, as it can handle Gaussian denoising with any noise level.

6.3 Dataset acquisition

The first dataset used in this paper is the IXI Guys dataset with images acquired at Guys Hospital (http://brain-development.org/ixidataset/) using the Philips 1.5 Tesla system at the following specifications:

Flip angle = 8°, reconstruction diameter = 240, number of phase encoding steps = 192, time = 9.813 msec, and echo time = 4.063 msec. The images were rotated at different angles of 30^0, 60^0,90^0,180^0. A total of 2000 images were used, along with rotations and original images.

The second dataset used is the simulated brain database (https://brainweb.bic.mni.mcgill.ca/brainweb/). This dataset contains a set of realistic MRI data produced by an MRI simulator, which is widely used in the performance evaluation of different

denoising methods. A total of 1500 images of brainweb dataset were used, along with rotations and original images.

For stochastic optimization, we used the Adam optimizer with a mini batch size of 25 and the value of learning rate was 0.01 . We trained the proposed model for 50 epochs and the learning rate was decayed exponentially. The experiment was performed using MATLAB 2019a on the system having 16GB RAM and 6GB Nvidia GPU. The denoised results were evaluated using PSNR [163] and SSIM [164] metrics and the segmented results of denoised images were evaluated using mIoU [140] and BF score [99].

6.4 Results

6.4.1 Results of Brain web dataset

The comparison of qualitative image statistics metrics (SSIM and PSNR) shows the superior performance of our proposed network as shown in Figure 6.3 when the network was trained on Brain web dataset and tested on 20% part of same dataset. It was found that the performance of the proposed model is better than all other methods. Figure 6.4 shows an example of denoised images by different methods at 15% noise level. It can be clearly seen from figure 6.4 that the information content and clinical visibility are enhanced.

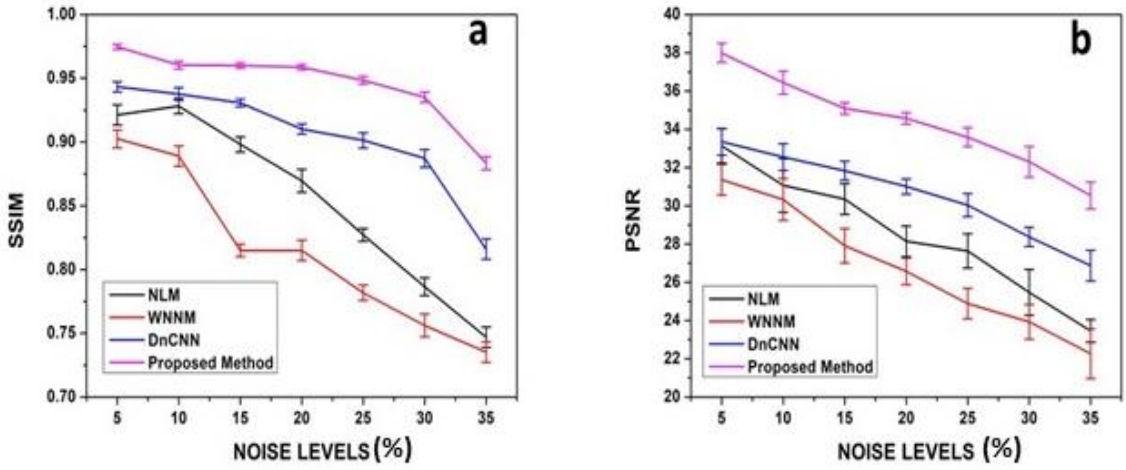

Figure 6.3: Denoising results of network trained on Brainweb dataset (a) SSIM and (b) PSNR.

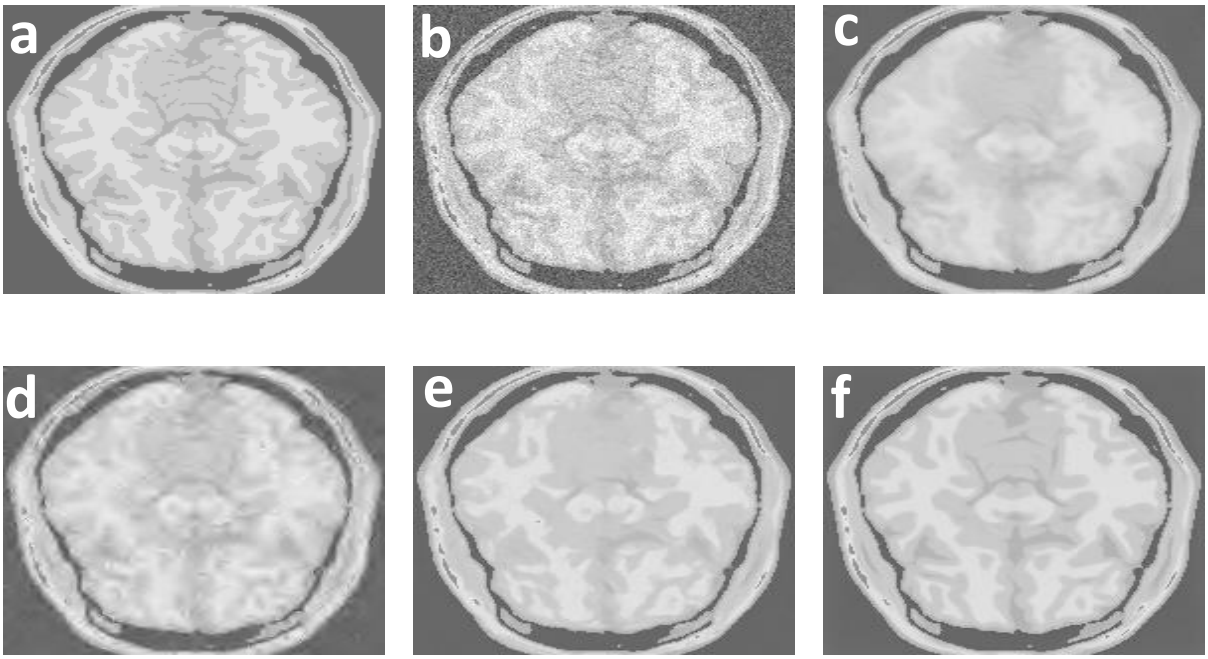

Figure 6.4: BrainWeb dataset denoising example: (a) original image, (b) with noise (c) NLM, (d) WNNM, (e) DnCNN, and (f) Proposed Model.

6.4.2 Results from IXI-Guys dataset

PSNR and SSIM values (as shown in Figure 6.5) obtained after training the network on the Brainweb dataset and tested on IXI Guys dataset suggests the better performance of our proposed network. Without retraining the network, we further applied the IXI Guys dataset directly to observe and evaluate the denoising performance. It is clearly evident from (Figure 6.5) that the values of SSIM and PSNR show remarkable increments. Figure 6.6 shows an example of denoised images by different methods at 15% noise level.

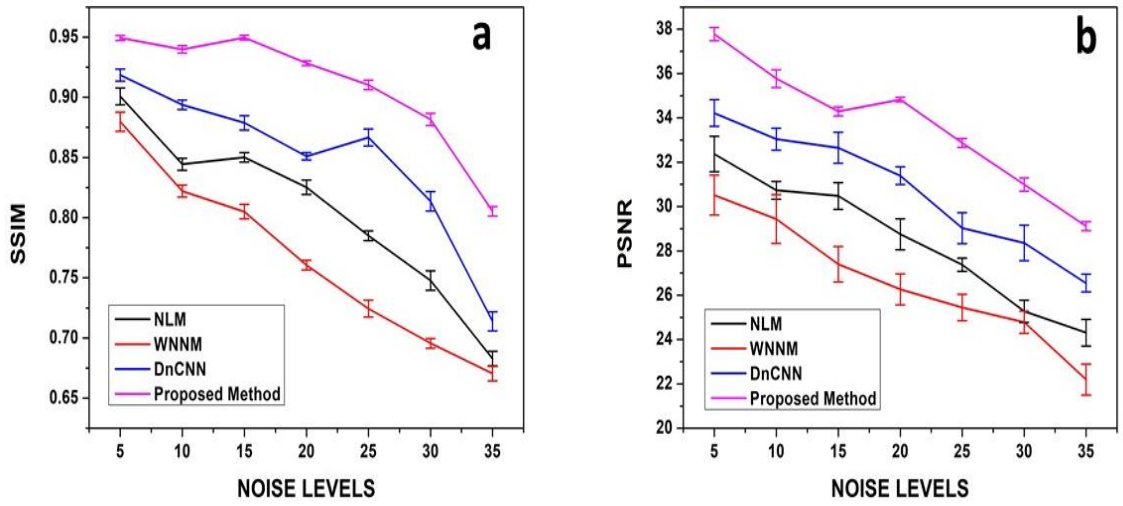

Figure 6.5: Denoising results of IXI-Guys dataset with network trained on IXI Guys dataset: (a) SSIM and (b) PSNR.

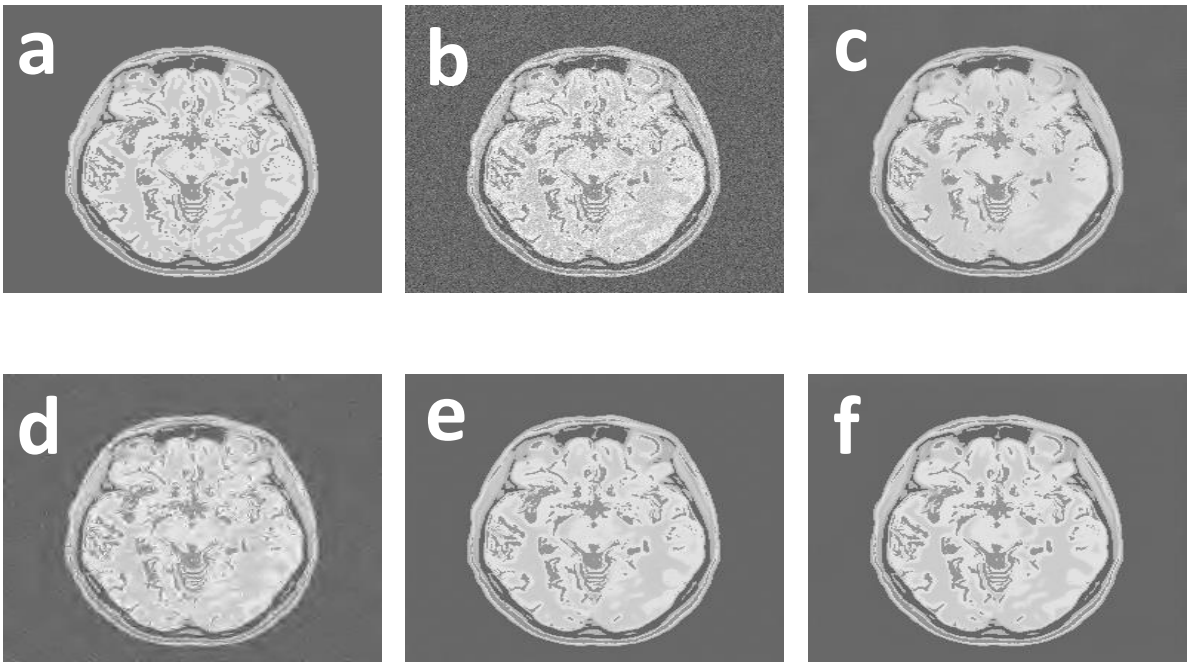

Figure 6.6: IXI Guys dataset denoising example: (a) original image, (b) image with noise, the corresponding denoised image from (c) NLM, (d) WNNM, (e) DnCNN, and (f) Proposed Model.

For statistical validation of the results, the Friedman test [165] and Wilcoxon test [166] were performed. The null hypothesis assumed that all the methods performed equally. The Friedman test ranks every method, and then it calculates the p-value. The lower p-value signifies that there is a significant statistical difference between the results obtained by various methods.

Hence, the null hypothesis is rejected. The results obtained by the Friedman test applied to SSIM values are shown in Table 6-1. The proposed method achieved the best rank. Further, the Wilcoxon test was performed in which the method with the best rank is selected as the controlling method. A comparison of the control method is performed with all other methods pairwise. The proposed method has achieved the best rank; thus, it was selected as the controlling method. The p-values obtained using the pairwise evaluation of methods are shown in Table 6-2. The comparative statistical analysis of the p-values indicates that the proposed method performed better in comparison to other methods in the study. The Table 6-2 shows that the p-values are less than $\alpha = 0.05$ [167]. Thus the results produced by the proposed method are statistically better as compared to other methods.

Table 6-10: Results of Friedman Test showing the mean ranks of the methods.

Methods	**NLM**	**WNNM**	**DnCNN**	**Proposed Method**
Mean of Rank	4.603	2.9839	1.9643	1.1098

Table 6-11: Statistical analysis results using Wilcoxon signed rank test with $\alpha = 0.05$.

Comparative hypothesis	**Proposed Method vs NLM**	**Proposed Method vs WNNM**	**Proposed Method vs DnCNN**
p-values	0.00764	0.003584	0.01629

6.4.3 Results of denoised segmented images:

The segmentation results of the images denoised for 15% noise level obtained by the networks in this study shows better performance of our proposed network as shown in figure 6.7. Table 6-3 shows the values of segmentation metrics mIou and BF score which depicts the better segmentation results. It can be observed that the segmented results of our proposed model are more in line with the human quality of visual perception.

Table 6-3: Comparison of segmentation evaluation metrics.

Networks	mIoU	Mean BF Score
NLM	0.6726	0.6028
WNNM	0.5258	0.4589
DnCNN	0.7732	0.7614
Proposed Model	0.8214	0.8098

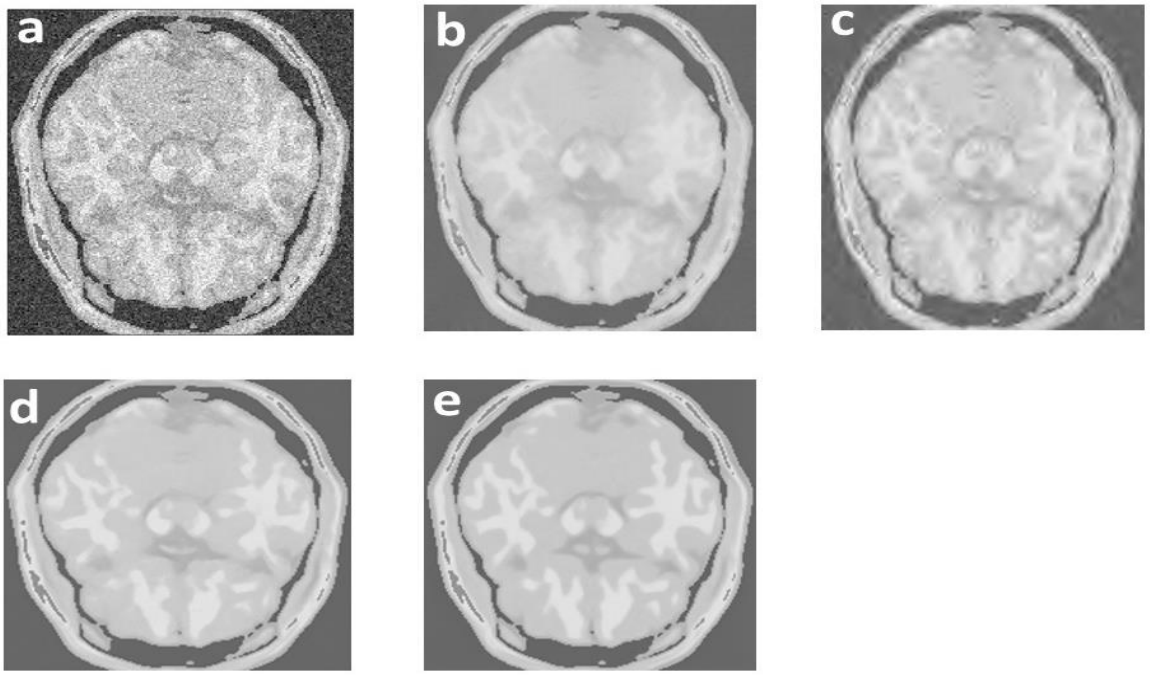

Figure 6.7: The segmentation results of denoised images (a) noisy image (15%), (b) NLM, (c) WNNM, (d) DnCNN, (e) Proposed Model.

6.5 Discussion

In this work, we propose an approach to denoising MR images corrupted with a modulating percentage of Rician noise (5% to 35%). The proposed model uses a depthwise separable convolution-based feed-forward neural network with Local Response Normalization(LRN). LRN effectively detects the high-frequency noise and dampens the effect of noise. The depth wise separable convolution reduces the blurring effects of normal convolution as it possesses the advantage of reduced computation complexity as the number of multiplications are reduced

to a more considerable extent. In normal convolution, the spatial resolution is lost at every convolution multiplication [168]. The reduction of multiplications in the case of depth wise separable convolution preserves the spatial information. Thus the detailed description of boundaries and contours is well preserved in the network which incorporates depth wise separable convolution. The incorporation of skip connections reduces the loss of information in the reconstructed images as the feature maps are directly transferred to the successive layers. Moreover, introducing skip connections in (any) network will enhance the flexibility of adding more layers to the network in order to avoid the problem of vanishing/exploding gradients which can occur in deeper networks [169] [170]. The center layers get updated with a decent amount when skip connections are used in the network. Whereas, the use of parametric RELU removes the problem of dying RELU, as it does not have a zero slope part for negative values as it is in normal RELU activation. Unlike RELU, parametric RELU is more balanced, so it learns faster compared to normal RELU, which makes the learning of the proposed network faster compared to other existing networks [124]. The proposed network can remove noise ranging to higher values in the MR images, and as the noise level increases, the proposed network outperforms all other methods. We trained the proposed network on the BrainWeb dataset and directly applied this trained network to the IXI Guys dataset without retraining the network on the second dataset. The performance achieved was remarkable in terms of measures of global structure similarity index (SSIM) and peak signal-to-noise ratio (PSNR) as specified in figure 6.3 and figure 6.5. At higher values of the noise, the improvement of approximately 8% was noted in the values of SSIM when the network was tested on 20% part of the dataset on which it was trained initially [171]. When the trained network was tested on a different dataset, the SSIM values showed an improvement of about 7%. After comparing the results obtained from the two datasets, we find that even after not being retrained on a different dataset, the proposed network gives an excellent performance. In addition, the proposed model can be

directly applied to the noisy datasets without prior estimation of noise, as noise estimation can further degrade the performance of denoising methods [172]. The BrainWeb dataset used in this work is simulated noise-free data and the results obtained using this dataset are expected to be the most promising in terms of denoising performance. This could be inferred as the general applicability of the proposed denoising model. The error bars obtain lower amplitude in the proposed network at nearly all noise levels (as shown in Figure 6.3 and Figure 6.5). This validates the reliability of the proposed denoising method as the error at the same noise level is less in the proposed method than in the other methods [173]. Higher SSIM and PSNR values indicate that the results are closer to the expected ground truth data. We obtained higher value of SSIM and PSNR, in the results of our proposed network, which indicates the better denoising results as compared with the other networks in this paper. The literature reports that less noisy MR images find more suitability inaccurate MR image diagnosis. So, the images denoised by the proposed model have shown more information content in the MR images. The use of max-pooling is avoided to keep the output size the same as the input size.

By visual inspection of figure 6.7, it can be seen that the boundaries of segments are well preserved in the output produced by our proposed network in comparison to other networks and the contours of the regions are clearly delineated. It can be noted that the values of the segmentation evaluation metrics are higher for our proposed model. The higher values of mIoU indicate that the area of overlap with the expected ground truth is more in case of our segmented denoised images. Moreover, the larger BF score depicts that the boundary details are well preserved in case of results produced by our network [174]. Higher values of both these metrics depict the better quality of segmentation [175]. The segmentation results obtained by the images of our proposed denoising model shows more than 6% and 5% improvement in the mIou and BF score in comparison to the second-best method. Thus the results obtained are having more area of overlap with the ground truth images as the mIou has the higher value in

case of our proposed model. The higher value of BF score indicates that the boundary details are well preserved in the denoised-segmented images produced by our model. Table 6-1 and Table 6-2 shows the statistical analysis using the Friedman test and Wilcoxon signed rank test. The proposed method attains the lowest value of mean rank among the methods under study. Thus validating the statistical difference between the methods and the lowest rank signifies the best ability to denoise the images. The Wilcoxon signed rank test gives the p-values less than 0.05 ($\alpha = 0.05$) which indicates the better performance of the proposed method [144]. It was hereby noted that the denoising achieved by the proposed network was clinically more pertinent in terms of accurate segmentation of brain regions and exclusion of unwanted noise.

6.6 Conclusion

The proposed model effectively removes the noise in MR images and outperforms all other methods, as indicated by statistical parameters. Our network model gives the cleanest results of all compared methods without retraining the network on different datasets. The results obtained were more clinically pertinent from the biomedical imaging point of view. The performance of the proposed network can be enhanced by training it on diverse data. We have achieved remarkable PSNR and SSIM levels compared with existing methods and improved the visual quality of the MR images to ensure the correct diagnosis. The better segmentation results of the proposed denoised images states the better applicability of images for the biomedical image analysis purposes. The images with reduced noise contents helps the medical practitioners to diagnose the disease more accurately.