Denoising 3D Magnetic Resonance Images: Replication and Testing of Residual Encoder–Decoder Wasserstein Generative Adversarial Network

John Kenneth Goebel
Joint Department of Biomedical Engineering
University of North Carolina at Chapel Hill
Chapel Hill, United States
jgoebel@unc.edu

Jane Qiu Luan Corah
Joint Department of Biomedical Engineering
University of North Carolina at Chapel Hill
Chapel Hill, United States
jcorah@unc.edu

Anh Duc Nguyen
Joint Department of Biomedical Engineering
University of North Carolina at Chapel Hill
Chapel Hill, United States
anguyen4@unc.edu

Regina Kay Smith
Joint Department of Biomedical Engineering
University of North Carolina at Chapel Hill
Chapel Hill, United States
reginaka@unc.live.edu

***Abstract*—Magnetic Resonance Imaging (MRI) is currently the gold standard for imaging clinical diagnostics. Techniques like image segmentation, identification, and deblurring can enhance diagnostic accuracy in MRI. This paper implements and evaluates the robustness of the Wasserstein-Generative Adversarial Network (WGAN) method as proposed in previous research [1]. The deep learning model specifically applies imaging denoising while examining the structural similarity of neighboring volumetric slices. The model uses residual autoencoders with convolutional and deconvolution operations to denoise three dimensional (3D) MRI scans with a WGAN network. To prevent oversmoothing, a custom linear combination loss function was developed, using Mean Squared Error (MSE), VGG-19 perceptual similarity, and MSE adversarial loss. The model is compared to the state-of-the-art BM3D method on PD-weighted images. Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics were calculated to evaluate the differences between these models. While BM3D outperformed the PD-weighted WGAN models in PSNR across all noise levels, the WGAN models achieved higher SSIM values at higher noise levels, demonstrating better preservation of structural details and mitigating BM3D's tendency to oversmooth the data.**

***Index Terms*—Magnetic Resonance Image, Encoder, Decoder, Generator, Discriminator, Wasserstein, Deep Learning**

I. Introduction

MRI is a critical tool in clinical diagnostics and research due to its ability to non-invasively generate high-resolution 3D images of internal tissues and organs. However, a drawback of MRI image acquisition is that they are often affected by noise. Depending on the specific MRI acquisition device, MRIs are compromised by noise of varying magnitudes in real-world scenarios, particularly when high speed or resolution is required. Noise degrades image quality and negatively affects the accuracy of clinical diagnostic registration, segmentation, and detection [2]–[4]. Among common noise types, such as Gaussian and Raleigh, Rician noise is most present in MRIs and is challenging to computationally approximate due to its non-Gaussian distribution. The complex composition of Rician noise necessitates robust algorithms to approximate the generation of this real-world noise and subsequent denoising algorithms [5].

Several approaches have been proposed for MRI denoising, which can be broadly categorized into traditional, statistical, and deep learning-based methods. Traditional techniques, such as anisotropic diffusion filters [6], non-local means (NLM) filters [7], [8], and their extensions [9], [10], operate in the spatial domain to suppress noise while preserving structural details. Although these methods have been widely used, they often suffer from computational inefficiency and oversmoothing effects. Statistical methods often attempt to model the noise distribution, such as the Rician noise model, and derive optimal denoising solutions based on these assumptions [11], [12]. Despite their mathematical rigor, statistical methods often require complex parameter tuning and lack adaptability to diverse clinical data.

In recent years, the rapid advancement of deep learning has revolutionized image processing tasks, including MRI denoising. Deep learning approaches leverage convolutional neural networks (CNNs), autoencoders, and generative adversarial networks (GANs) to achieve state-of-the-art performance in tasks such as denoising, super-resolution, and deblurring [13]–[15]. For example, CNN-based methods have demonstrated superior performance over traditional algorithms, including BM3D, NLM, and sparse representation techniques [14].

GANs, in particular, have been used to model complex data distributions and generate high-quality denoised images while addressing the oversmoothing effects of of CNN-based approaches [16], [17].

To further improve MRI denoising, Ran *et al.* [1] proposed the Residual Encoder–Decoder Wasserstein Generative Adversarial Network (RED-WGAN), which integrates several innovative components. First, the generator network employs a residual encoder-decoder structure with skip connections to preserve fine structural details and edges critical for medical analysis. Second, the WGAN framework stabilizes training and ensures the generator learns the true data distribution, effectively mitigating the challenges of traditional GANs [18]. Finally, by incorporating a perceptual loss term computed using a pre-trained VGG-19 network, the model enhances structural preservation and avoids oversmoothing [19].

In the proposed paper, the RED-WGAN outperformed state-of-the-art methods such as BM4D [10], PRI-NLM3D [9], and CNN3D in both simulated and clinical datasets. Its ability to suppress noise while maintaining structural integrity was validated across a range of noise levels and imaging modalities [1].

In this project, we replicate and evaluate the RED-WGAN framework to verify its efficacy in MRI denoising. This involves implementing the proposed architecture, training on a clinical dataset, and comparing its performance against traditional and deep learning methods. This work aims to contribute to the development and validation of robust MRI denoising approaches, improving diagnostic accuracy and reliability.

II. Methods

A. Noise Reduction Model

MRI images are commonly affected by Rician noise, which is constructed of real and imaginary components. This noise complicates clinical diagnoses and impacts tasks such as segmentation and registration [11], [20]. The goal of MRI denoising is to recover a high-quality image $y \in \mathbb{R}^{m\times n}$ from a noisy input $x \in \mathbb{R}^{m\times n}$, where:

$$x = \sigma(y), \tag{1}$$

and σ represents the noise function. Deep learning approaches learn an inverse mapping function f such that:

$$\hat{y} = f(x), \quad \hat{y} \approx y. \tag{2}$$

The task is treated as an optimization problem:

$$\arg\min_f \|\hat{y} - y\|^2, \tag{3}$$

where the model minimizes the difference between the reconstructed image $\hat{y}$ and the ground truth y.

B. Wasserstein GAN

GAN: GANs consist of two networks: a generator G and a discriminator D. The generator G maps noisy inputs x to denoised outputs $G(x)$, while the discriminator D distinguishes between real images y and generated images $G(x)$. Traditional GANs use a binary classification loss for the discriminator, which can lead to mode collapse and unstable training [16].

WGAN: The WGAN framework replaces the binary classification loss with a continuous "critic" score from 0 to 1 based on the Wasserstein distance, which improves training stability [18]. The WGAN loss for the discriminator is:

$$\begin{aligned} L_{\text{WGAN}}(D) = &-\mathbb{E}_{y\sim P_r}[D(y)] + \mathbb{E}_{x\sim P_n}[D(G(x))] \\ &+\lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \end{aligned} \tag{4}$$

where P_r and P_n represent the real and noisy distributions, respectively, and λ is a penalty coefficient enforcing the Lipschitz constraint [21]. The generator's objective is:

$$L_{\text{WGAN}}(G) = -\mathbb{E}_{x\sim P_n}[D(G(x))]. \tag{5}$$

C. Residual Encoder–Decoder Architecture

The generator in the RED-WGAN framework uses a residual encoder-decoder structure. The encoder is composed of four 3D convolutional layers with $3\times3\times3$ kernels, batch normalization, and LeakyReLU activation extract features from noisy inputs. The batch normalization is applied to stabilize and accelerate the training while the activation functions introduce non-linearity and help prevent a vanishing gradient. Additionally, the decoder is symmetrically made with four 3D deconvolutional layers that reconstruct the clean image, aided by skip connections linking encoder and decoder layers. With residual learning, skip connections enable the model to focus on reconstructing residual noise, rather than the entire image, preserving structural details in the process [22].

On the other hand, the discriminator is a 3D convolutional network with three layers, followed by batch normalization and LeakyReLU activation. The purpose for including batch normalization in the discriminator is that it stabilizes the adversarial training while the activation function, like before, ensures efficient gradient flow for learning the difference in real and generated images. The discriminator outputs a Wasserstein distance score that indicates the closeness of generated images to the real distribution.

D. Combined Loss Function

The generator is trained using a combined loss function that includes pixel-level accuracy, perceptual consistency, and adversarial feedback. The MSE penalizes pixel-wise differences between $G(x)$ and y:

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \|G(x_i) - y_i\|^2. \tag{6}$$

The perceptual loss measures feature-level differences using a pre-trained VGG-19 network [19]:

$$L_{\text{Perceptual}} = \frac{1}{N}\sum_{i=1}^{N} \|\phi(G(x_i)) - \phi(y_i)\|^2, \tag{7}$$

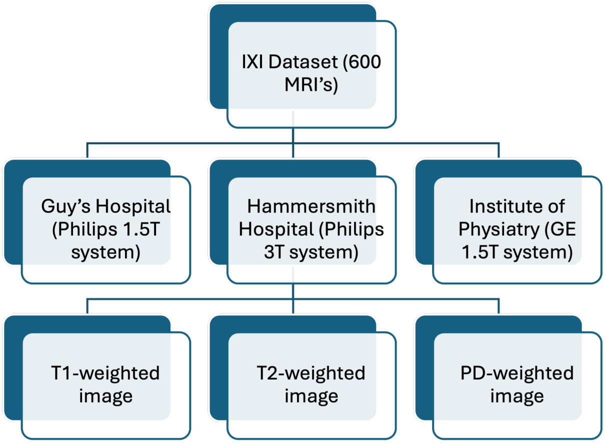

Fig. 1. Overview of the IXI Dataset: 600 MRIs collected including T1-, T2-, and PD-weighted images.

where ϕ is the feature extractor. The WGAN Loss Encourages the generator to align its output with the real data distribution. Overall, the total generator loss is:

$$L_{\text{GEN}} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{Perceptual}} + \lambda_3 L_{\text{WGAN}}, \tag{8}$$

where $\lambda_1, \lambda_2, \lambda_3$ are empirically determined.

III. Experiment

A. Dataset

We randomly selected 60 MRI scans from the IXI dataset, including 20 of each T1-weighted, T2-weighted, and PD-weighted images. The volumetric scans were used for training, validation, and testing of the proposed model. These images were acquired from the Hammersmith Hospital with a Phillips 3T system. This specific dataset was chosen for its high-contrast, size (600+ MRI volumes), and relevance to real-world clinical research. The organization of the selected dataset is shown in Figure 1

B. Preprocessing

A randomized position along the z-axis for each MRI was designated as the origin to generate transverse slices with a depth of 6. Additionally, we alternatively trained models using transverse slices with a depth of 6 extracted particularly from the middle position of the z-axis for each MRI. The original MRIs had dimensions of 256x256 with varying depths, which were reduced to 6 layers. Rician noise was then added to the data by combining real and imaginary components base on the phase and magnitude of the signal from the image. Rician noise, caused by thermal agitation of electrons during MRI acquisition, can be modeled by combining the orthogonal signal projection of the real and imaginary components. Rician noise follows a non-symmetric distribution and is essentially a shifted Raleigh distribution. Signal-to-noise ratios (SNRs) of 1%, 7%, and 13% with $\mu = 0.0$, $\sigma = 1$ were added to the training, validation, and testing sets.

Overlapping voxels were created as this approach has been shown to not only increase the size of our data set, but also improve the models ability to detect perceptual differences in patches. Additionally, deep learning methods require a large number of samples and thus a sliding kernel of size 32x32x6 with a stride of 16 was used across each MRI for a total of approximately 88,000 training patches.

C. Training Details

To validate the performance of the replicated network architecture, we trained the RED-WGAN model on T1-, T2-, and PD-weighted images. Two data sets were created, one using only PD-w images and one with T1-,T2- and PD-w images, to assess the effect of data diversity. Simulated Rician noise at varying SNRs (1%, 7%, 13%) was added and shuffled to generate noisy inputs for training. The parameters λ_1, λ_2, and λ_3 in the loss function were experimentally set to 1, 0.1, and 1×10^{-3}, respectively, based on the recommendations in [17] and [23]. Following [16], the penalty coefficient λ in the Wasserstein loss was set to 10.

The Adam algorithm [24] was employed to optimize the loss function. The optimizer parameters were set as: $\alpha = 5 \times 10^{-5}, \quad \beta_1 = 0.5, \quad \beta_2 = 0.9$

The implementation was carried out using PyTorch, and all training was conducted on the University of North Carolina (UNC) Longleaf Computing Cluster, which enabled efficient training for the 3D convolutional layers and adversarial components.

The learning rate employed by the original paper was a step-wise decay which halved the learning rate after every 4 epochs. This learning rate was initially set to 5×10^{-6}. To test the effect of learning rate, an exponential learning was also applied with an exponential decay rate of 0.97.

D. Evaluation Methods

To validate the performance of RED-WGAN, its results were compared against BM3D, a state-of-the-art denoising algorithm. To evaluate the performance of these methods, two quantitative metrics were employed, PSNR and SSIM. PSNR considers the root mean square error (RMSE) between the ground truth and the denoised images. Higher PSNR values indicate better fidelity to the original clean image. SSIM, introduced by Wang et al. [25], measures the structural similarity between the ground truth and the denoised images. It evaluates image quality by considering luminance, contrast, and structural information, with values closer to 1 indicating higher similarity.

IV. Results

A. Clinical WGAN Results

The initial data set contained only PD-w images with 88305 training and 12615 validation patches. The secondary data set contained T1-,T2-, and PD-w images with 28350 training and 4050 validation patches.

After running a baseline test for both data sets, where all the hyper parameters (excluding number of epochs) matched the original paper, a series of experiments was completed. First we enabled batch normalization in the discriminator, then

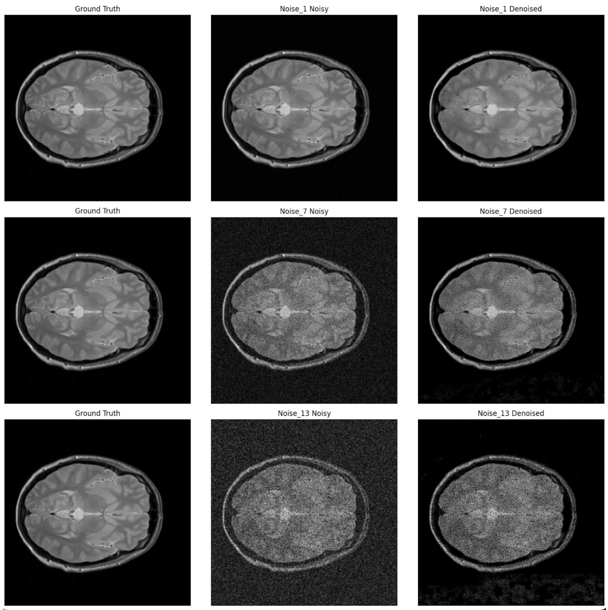

Fig. 2. Ground Truth, Noisy and Denoised PD-w images at noise levels 1,7,13. Images denoised using Baseline PD-w only model.

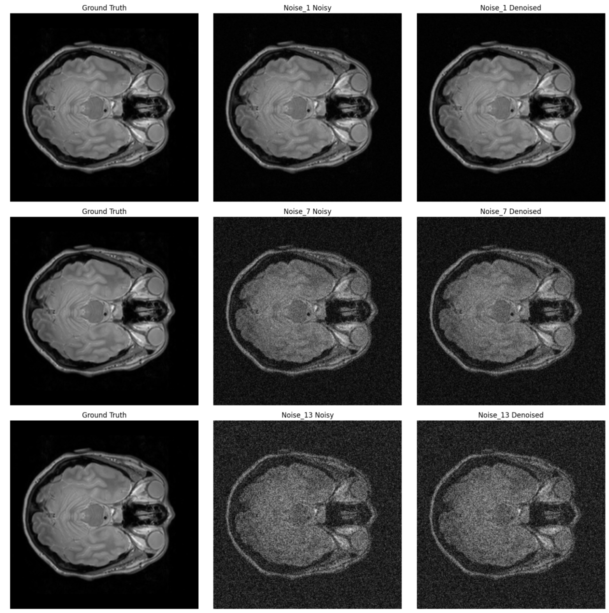

Fig. 3. Ground Truth, Noisy and Denoised PD-w images at noise levels 1,7,13. Images denoised using Baseline T1-,T2-, and PD-w model.

applied an exponential learning rate, and finally tested the ratio of discriminator to generator trainings (1 and 3 vs. the nominal 5). For each of these tests, only one aspect of training was changed compared to the baseline model. Although some models were trained for a longer number of batches, limited cluster resources kept some training to only 30 epochs. As such, for comparison of models, image metrics from epoch 30 will be listed in the table below.

Model	**Level 1 Noise**		**Level 7 Noise**		**Level 13 Noise**	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baseline	**29.28**	0.7430	24.98	.6547	23.99	0.5043
BN in D	29.22	0.7423	24.95	0.6534	23.96	**0.5083**
Exp.	28.49	0.7437	24.22	**0.6627**	**24.10**	0.4940
Iter 1	29.19	0.7429	25.00	0.6557	24.00	0.4887
Iter 3	**29.28**	**0.7439**	**25.02**	0.6583	24.04	0.5061

TABLE I
IMAGE METRICS FROM DENOISED PD-W IMAGES OUTPUTTED BY MODELS TRAINED ON THE PD-W ONLY DATA SET.

Model	**Level 1 Noise**		**Level 7 Noise**		**Level 13 Noise**	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baseline	48.47	0.9518	30.90	0.4415	25.35	0.2664
BN in D	49.07	0.9657	30.94	0.4502	25.36	0.2656
Exp.	**50.27**	**0.9788**	**31.38**	**0.4635**	**25.57**	**0.2665**
Iter 1	47.53	0.9522	30.87	0.4493	25.34	**0.2665**
Iter 3	48.37	0.9526	30.90	0.4497	25.35	0.2661

TABLE II
IMAGE METRICS FROM DENOISED PD-W IMAGES OUTPUTTED BY MODELS TRAINED ON THE T1-, T2-, AND PD-W DATA SET.

B. BM3D

In the original paper, BM4D was utilized as one of the state-of-the-art methods. For our purposes, BM3D was deemed sufficient, although it is less effective at processing 3D images, such as MRIs, due to its inability to track similarities across layers. In this study, BM3D was applied independently to each of the six layers within each patch, after which the image layers were stacked back together. The PSNR and SSIM values for the BM3D-processed PD images under noise levels 1, 7, and 13 are presented in Table III.

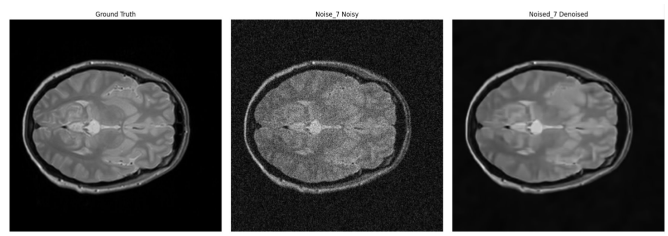

Fig. 4. Ground Truth, Noisy and Denoised PD-w images at noise level 7. Images denoised using BM3D. There is significant oversmoothing shown in the Denoised image

Metrics	Level 1 Noise	Level 7 Noise	Level 13 Noise
PSNR	40.37	25.97	24.72
SSIM	0.8211	0.5070	0.4595

TABLE III
TABLE WITH PSNR AND SSIM MEASURES FOR BM3D PROCESSED IMAGES OF NOISE LEVELS 1, 7, AND 13

While these results are inferior to the BM4D outcomes reported in the original paper, BM3D adequately denoised the images but performed significantly worse than BM4D in maintaining structural integrity. For instance, the SSIM value for BM4D at noise level 13 was 0.8124, whereas the SSIM for BM3D at the same noise level was only 0.4595. [1] Both BM3D and BM4D cause higher amounts of oversmoothing, which can be seen in Figure 4. Since BM3D is not a deep learning-based method, we did not expect the performance of BM3D to improve with a broader dataset. The PD-w BM3D results can be used as a comparison for both models trained.

V. Discussion

A. Interpretation of Results

Table I contains the PSNR and SSIM results for our first dataset, with the best image metrics for each noise level highlighted in bold. The model trained with a Discriminator-Generator training ratio of 3:1 consistently performed best at lower noise levels, while the model with an exponential learning rate performed well at higher noise levels. This suggests that reducing the ratio from 5:1 to 3:1 and adopting exponential learning rates could improve training.

Looking at Figure 2, it can be seen that for noise levels 7 and 13, the PD-w-only model significantly denoised the MRI image. However, for noise level 1, the denoised image actually loses some of its structural features and the image metrics worsen as training continues. This is possibly due to a bias toward higher noise levels.

The image metrics for the T1-, T2-, PD-w trained models are shown in Table II, with the exponential learning rate model outperforming the other models every time. This warrants further investigation comparing the baseline step decay learning rate to an exponential learning rate. Although the T1-, T2-, PD-w trained models seem superior to the PD-w-only models based on PSNR image metrics, Figure 3 demonstrates that the baseline model did not perform well at denoising noise levels 7 and 13. This suggests that the SSIM image metric, in which the PD-w-only models performed better, is more important when evaluating the performance of RED-WGAN models. The deceased performance using the second data set may have been caused by the smaller number of training and validation patches. Unfortunately, this does not allow any conclusions to be made regarding the effect of image diversity on model performance. Another potential method to evaluate model performance is comparing the initial and final image metrics to assess how well a model improves, rather than just focusing on the final value. It should be noted that the original paper was only outperformed by the T1-, T2-, PD-w trained models in terms of PSNR for noise level 1 [1]. The original paper performed better for every other image metric and noise level combination.

Finally, while BM3D outperformed the PD-w weighted models in PSNR for all noise levels, the PD-w weighted models achieved higher SSIM values for noise levels 7 and 13 (Table III). This may be due to BM3D's tendency to oversmooth data, which is exemplified in Figure 4.

B. Limitations

A significant limitation of the first dataset was that it consisted exclusively of PD images. Additionally, during preprocessing, MRI slices were taken only from the same region of the brain, restricting the diversity of the training data. These issues were mitigated in the second dataset, which included T1, T2, and PD images and randomized slice acquisition along the Z-axis.

Another limitation pertains to the noise generation process. The datasets simulated noise at discrete levels rather than a random continuous spectrum, which would have better represented real-world scenarios. The greatest constraint of this study was computational power. We were unable to utilize multiple GPUs simultaneously, and time constraints prevented us from training for the suggested 100 epochs, potentially compromising the model's robustness. High memory demands due to the volume of training images also posed challenges. Unlike the target paper, which validated models using simulated data from the BrainWeb Database, our study relied on different validation methods.

C. Future Improvements

To address the limitations, future work could include generating noise at random SNR levels between two discrete values, preventing the model from learning overly specific noise patterns. Leveraging multiple GPUs for parallel processing would reduce epoch times and enable longer training durations. Hyperparameter tuning, such as optimizing the learning rate, could further enhance the model's performance. The paper mainly used a step-wise decay learning rate, while a possible improvement would be to go back and use the step-wise decay learning rate with our more robust dataset.

Expanding the dataset to include MRI scans from multiple hospital systems, with different resolutions, could improve model generalization. Additionally, using BM4D for evaluation would provide a more relevant comparison metric, as BM4D is specifically designed for 3D image denoising. Metrics such as the Information Fidelity Criterion (IFC) could also be introduced to better assess the model's ability to preserve critical anatomical features.

D. Conclusion and Practical Implementation

In this study, we attempted to replicate the RED-WGAN framework. We trained two models one with just PD images and one with a combined dataset of T1-, T2-, and PD-w images. There was simulated Rician noise at 1Our model was still able to denoise the images, particularly at higher noise levels, and outperformed the BM3D algorithm in terms of SSIM. Reducing the discriminator-to-generator training ratio from 5:1 to 3:1 and employing an exponential learning rate decay, showed improvement in model performance. However, the study was limited by the dataset and computational strength In conclusion, our model was able to denoise images and with further refinements and better datasets, our model would be better suited at replicating the results from the target paper. The residual autoencoder structure ensures high levels of structural preservation, making it valuable for applications like neurosurgical planning, where high-quality images are essential. Furthermore, the model enables faster MRI sampling times by effectively denoising lower-quality images acquired in shorter durations.

ABBREVIATIONS

3D – Three-Dimensional; BM3D - Block Matching and 3D Filtering; BM4D - Block Matching and 4D Filtering; CNNs – Convolutional Neural Networks; GAN – Generative Adversarial Network; IFC - Information Fidelity Criterion; MSE – Mean Squared Error; MRI – Magnetic Resonance Imaging; NLM – Non-Local Means; PSNR – Peak Signal-to-Noise Ratio; RED-WGAN – Residual Encoder–Decoder Wasserstein Generative Adversarial Network; RMSE - Root Mean Square Error; SNR – Signal-to-Noise Ratio; SSIM – Structural Similarity Index Measure; UNC - University of North Carolina; WGAN – Wasserstein-Generative Adversarial Network.

DATA AND CODE AVAILABILITY

The code for this project is available at GitHub Repository, and the MRI dataset was obtained from the IXI dataset available at IXI Dataset.

ACKNOWLEDGMENT

We would like to thank Dr. Arian Azarang, professor of BMME 575: Practical Machine Learning, for his guidance and support on this project.

AI USAGE STATEMENT

AI was used for text refinement and coding assistance, with all content remaining original to the authors and the referenced sources.

REFERENCES

[1] M. Ran, J. Hu, Y. Chen, H. Chen, H. Sun, J. Zhou, and Y. Zhang, "Denoising of 3d magnetic resonance images using a residual encoder-decoder wasserstein generative adversarial network," *Medical Image Analysis*, vol. 55, pp. 165–180, 2019.

[2] J. V. Manjón, J. Carbonell-Caballero, J. J. García-Martí, L. Martí-Bonmatí, and M. Robles, "Mri denoising using non-local means," *Medical Image Analysis*, vol. 12, no. 4, pp. 514–523, 2008.

[3] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomedical Signal Processing and Control*, vol. 9, pp. 56–69, 2014.

[4] X. Zhang, Z. Xu, N. Jia, W. Yang, Q. Feng, W. Chen, and Y. Feng, "Denoising of 3d magnetic resonance images by using higher-order singular value decomposition," *Medical Image Analysis*, vol. 19, no. 1, pp. 75–86, 2015.

[5] A. H. Andersen, "On the rician distribution of noisy mri data," *Magnetic Resonance Imaging*, vol. 36, pp. 331–333, 1996.

[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, 1990.

[7] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," *Proceedings of IEEE Computer Vision and Pattern Recognition*, vol. 2, pp. 60–65, 2005.

[8] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[9] J. V. Manjón, P. Coupé, A. Buades, C. D. Louis, and M. Robles, "New methods for mri denoising based on sparseness and self-similarity," *Medical Image Analysis*, vol. 16, no. 1, pp. 18–27, 2012.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[11] S. P. Awate and R. T. Whitaker, "Feature-preserving mri denoising: A nonparametric empirical bayes approach," *IEEE Transactions on Medical Imaging*, vol. 26, no. 9, pp. 1242–1255, 2007.

[12] H. M. Golshan, R. P. Hasanzadeh, and S. C. Yousefzadeh, "An mri denoising method using image data redundancy and local snr estimation," *Magnetic Resonance Imaging*, vol. 31, no. 7, pp. 1206–1217, 2013.

[13] C. Dong, C. L. Chen, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *Proceedings of the European Conference on Computer Vision*, pp. 184–199, 2014.

[14] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.

[16] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.

[17] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, 2017.

[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] M. Bouhrara, J. M. Bonny, B. Ashinsky, M. Maring, and R. Spencer, "Noise estimation and reduction in multispectral magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 181–193, 2016.

[21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," pp. 5767–5777, 2017.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, 2016.

[23] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2017.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.