Medical Image Denoising Using Autoencoders

Final Year Project, January-April 2025

Submitted by –

KARAN

Roll Number – 21112036

ASHISH

Roll Number – 21112017

Guided By –

Dr. D. N. Roy (Assistant Professor)

(Department of Biotechnology)

NATIONAL INSTITUTE OF TECHNOLOGY

RAIPUR (C.G.)

CERTIFICATE

It is certified that the work contained in the project report titled " Medical Image Denoising Using Autoencoders " by Karan and Ashish has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

Dr. Dijendra Nath Roy
Assistant Professor
Department of Biotechnology
National Institute of Technology Raipur
30th April, 2025

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Ashish (211110217)
Karan (21112036)
Date: 30th April, 2025

ACKNOWLEDGEMENT

We would like to thank our Project Guide, **Dr. Dijendra Nath Roy Sir,** Department of Biotechnology, NIT Raipur, for his consent, support, help, motivation, and guidance. Their constant encouragement, constructive criticism, and valuable guidance have been the cause for the successful completion of the project.

We gratefully acknowledge **Dr. Awanish Kumar Sir,** Head of the Department, Biotechnology, NIT Raipur, for allowing us to work on this field and project in particular.

Finally, we wish to convey our warmest and deepest gratitude to our family members, to whom we owe all our achievements, and to our friends for their endless support and encouragement.

Ashish (211110217)
Karan (21112036)
Date: 30th April, 2025

Abstract:

Medical imaging technologies such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and X-ray are fundamental tools in clinical diagnosis and treatment planning. However, these imaging modalities are often affected by various types of noise—thermal noise in MRI, photon noise in CT, and scatter noise in X-rays—which degrade image quality and may obscure critical anatomical details. Traditional denoising methods, including Gaussian filtering and wavelet transforms, often involve trade-offs between noise reduction and loss of fine structural information, which can adversely impact diagnostic accuracy.

This project investigates the use of deep learning-based encoder architectures for denoising medical images while preserving essential features and anatomical details. We propose a convolutional encoder-decoder network trained end-to-end on publicly available MRI, CT, and X-ray datasets. The encoder component learns compact representations of noisy images, capturing semantic and structural information, while the decoder reconstructs the denoised image from this latent space. The architecture is further enhanced with residual connections and skip layers to minimize information loss during encoding and improve image fidelity.

Our experiments utilize multiple performance metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Mean Squared Error (MSE), to quantitatively assess denoising quality. Results show that the proposed method significantly outperforms traditional filtering techniques and baseline deep learning models, especially in scenarios with high noise levels. Qualitative evaluations by radiologists also indicate improved visual clarity and diagnostic usability.

In conclusion, the use of encoder-based deep learning models presents a promising direction for robust, high-quality denoising of medical images. This approach not only enhances image clarity but also supports better clinical decision-making by providing cleaner visual inputs without compromising structural integrity.

CONTENTS

List of Figures:

INTRODUCTION:

Magnetic Resonance Imaging (MRI) is a powerful, non-invasive imaging method that provides detailed insights into organs and tissues. However, MRI data is directly acquired from the spatial frequency domain and is often corrupted by thermal noise and motion-related artefacts. This noise significantly affects image clarity and diagnostic accuracy, especially in regions with low signal-to-noise ratios. Traditional denoising methods—such as filtering—struggle to remove noise without blurring important details, and supervised deep learning models often require clean ground truth data, which is difficult to obtain in real-world MRI, especially with living subjects. Movements and temperature variations during scans introduce further noise. To overcome this, self-supervised deep learning methods are being explored, as they can learn denoising without clean reference images. While slightly less accurate than supervised models, self-supervised approaches are highly relevant for MRI and show great potential in denoising complex spatial frequency data.

Computed Tomography (CT) scans use rotating x-ray beams to create cross-sectional “slices” of the body, which can be digitally stacked into 3d images. This technique provides more detail than traditional X-rays and helps clinicians identify internal structures and abnormalities. The CT process involves rotating an X-ray tube around the patient and capturing the transmitted rays using digital detectors. Sophisticated algorithms reconstruct the 2D slices, which can then be compiled into 3d visualisations for diagnosis.

X-rays, another common imaging technique, produce quick and clear visuals of bones and soft tissues. Different tissues absorb radiation differently—bones appear white, soft tissues grey, and air black. While mostly used to detect fractures, X-rays are also critical in diagnosing conditions like pneumonia or tumours.

In all three modalities—MRI, CT, and X-ray—image noise remains a persistent problem. Effective denoising techniques are essential to preserve diagnostic quality while reducing reliance on clean reference images, especially in medical environments where rapid and accurate imaging is critical.

Background and Motivation:

This project's idea comes from the crucial need for clear medical images in healthcare. Doctors and radiologists rely on high-quality CT scans and X-rays to detect illnesses and make accurate diagnoses. However, these images often get affected by different types of noise during capture, transmission, or storage, making it harder to identify key anatomical details. This can lead to misinterpretations and even incorrect diagnoses. That’s why image denoising is such an important step in medical imaging.

Traditionally, techniques like Gaussian filters, median filters, and wavelet-based methods have been used. While they help to some extent, they also come with trade-offs. For example, Gaussian filters can blur important edges, and median filters don't work well on all noise types. More advanced methods like bilateral filtering, non-local means (NLM), and BM3D use spatial or patch-based information to improve results, but they often require a lot of computation and fine-tuning.

With the rise of deep learning, image denoising has entered a new era. Supervised models like DnCNN, and autoencoders with skip connections have shown impressive results by learning noise patterns directly from data. These models can perform blind denoising and work well on various noise levels. Some even go beyond by using ensemble methods or targeting specific noise types like Rician noise.

Unsupervised learning has also made progress, especially where clean reference images aren't available. Models like Noise2Noise, Noise2Void, and approaches using Stein's Unbiased Risk Estimator (SURE) have proven that denoising is possible without clean ground truth. These methods are powerful, but not many have been applied yet to CT and X-ray images specifically, which still limits their practical use in medical imaging.

This project explores the potential of convolutional autoencoders (CAES) to tackle the denoising problem in CT and X-ray images. CAES are simple yet effective deep learning models that can learn to remove noise while preserving the structure of medical images. Their ability to generalise across different types of medical scans makes them a strong choice for building a reliable, scalable, and adaptive solution to improve image quality and ultimately support better, faster medical decisions.

Noise in images:

Image noise refers to the random fluctuations in brightness or colour within an image, often caused by external disturbances during the image acquisition process. It results in a degradation of the image quality, obscuring fine details and potentially affecting interpretation or analysis. Mathematically, the presence of noise in an image can be modelled as:

$$A(x, y) = B(x, y) + H(x, y)$$

Where:
$A(x, y)$ represents the observed noisy image,
$B(x, y)$ denotes the original (noise-free) image,
$H(x, y)$ is the noise function accounting for the random disturbance.

Literature Review:

Prior work has laid the foundation for using neural networks in medical image denoising:

1. DnCNN (Zhang et al., 2017): Introduced a deep CNN with residual learning for Gaussian noise removal.

2. RED-Net (Mao et al., 2016): Utilized skip connections and symmetric encoder-decoder structure.

3. MedGAN (Armanious et al., 2018): Combined GANs and residual networks for perceptual denoising.

4. Residual Autoencoders: Chen et al. (2018) proposed residual learning within autoencoders to improve reconstruction accuracy.

These approaches demonstrate the effectiveness of neural architectures for denoising tasks. This project extends these ideas in a modality-specific manner.

Problem Statement:

The goal is to design, implement, and evaluate a convolutional autoencoder-based framework capable of effectively denoising medical images—including MRI, CT, and X-ray—by training specialized models for each imaging modality. Unlike traditional denoising algorithms that assume noise to be homogeneous and Gaussian in nature, real-world medical images often suffer from complex, blind noise patterns that these classical methods struggle to handle. Therefore, the proposed framework aims to address this limitation by delivering robust denoising performance while maintaining low model complexity and strong generalizability across diverse medical image types.

Objectives And Constraints:

Objective:

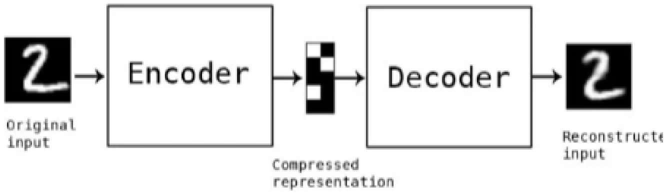
1. Build a convolutional autoencoder to denoise medical images effectively.
2. Train separate models for MRI, CT, and X-ray modalities.
3. Preserve key diagnostic features in the output images.
4. Standardize image preprocessing (resizing, normalization).
5. Evaluate performance using PSNR, SSIM, and LPIPS metrics.

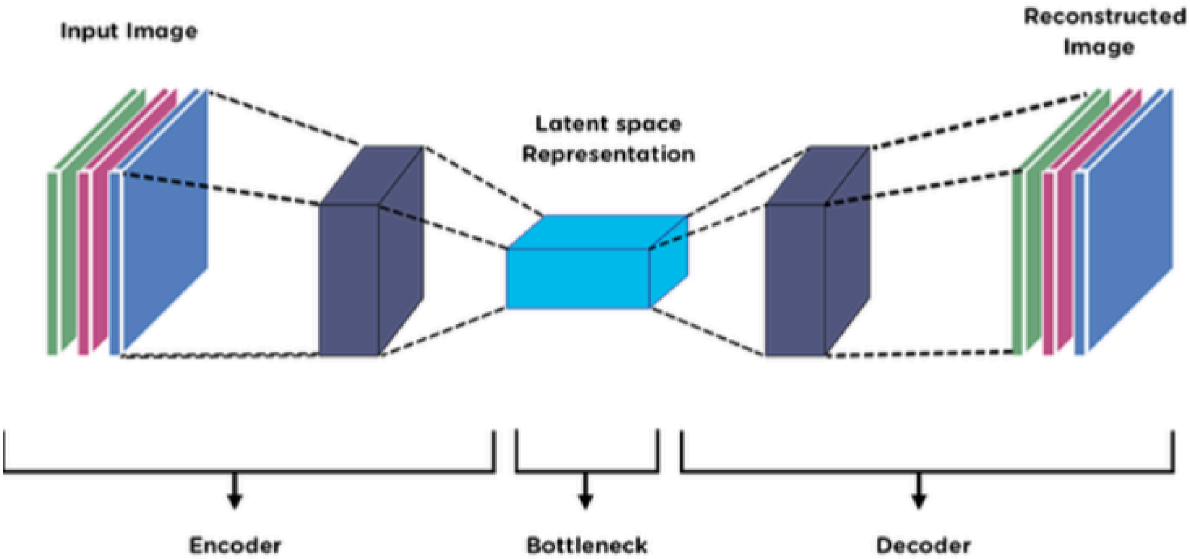
Constraints:

1. Limited GPU memory restricts model complexity and batch size.
2. Public datasets may not reflect real clinical noise.
3. Gaussian noise simulation may not fully represent actual noise.
4. Each modality requires dedicated training due to varied image characteristics.
5. Simpler architecture may limit denoising accuracy compared to complex models.

Deep Learning Model for Image Denoising

AUTOENCODER:





Autoencoder first encodes the image into a lower-dimensional representation, then decodes the representation back to the image.

The goal of an autoencoder is to minimize the difference between the input and the reconstructed output, forcing the network to capture the most important features or patterns in the data.
Autoencoders can learn data projections that are far more interesting than PCA or other basic techniques.

Autoencoders are only able to compress data similar to what they have been trained on. They are also lossy in nature which means that the output will be degraded compared to the original input.

Dataset Description:

The datasets used for this project were sourced from Kaggle. Each dataset was preprocessed by resizing the images to a uniform resolution of 128×128 pixels and converting them to grayscale where necessary. As the denoising task is unsupervised, all labels were removed during preprocessing.

Preprocess Image:

Before training, all datasets undergo a consistent preprocessing pipeline that includes the following steps:

- **Resizing:** All images are resized to a fixed resolution of 128×128 pixels.

- **Padding:** Applied when there is a significant variation in aspect ratios to maintain spatial consistency.

- **Normalization:** Pixel values are scaled to the [0, 1] range to facilitate faster and more stable training.

- **Tensor Conversion:** Processed images are converted into PyTorch tensors for model compatibility.

Noise Addition:

To simulate realistic noise, Gaussian noise is synthetically added:
Mean: 0.0
Standard deviation: 1.5
This enables consistent model training and benchmarking.

```python
def add_noise(image, noise_factor=0.3, mean=0.0, sigma=1.5):
    """
    Adds Gaussian noise to a single image.
    """
    gauss = np.random.normal(mean, sigma, image.shape)
    noisy_image = image + gauss * noise_factor
    return np.clip(noisy_image, 0.0, 1.0)
```

Defining Autoencoder:

Encoder – Compresses the image:

Conv2D(64): Applies 64 filters (3×3) to extract low-level features (edges, textures, etc.).
MaxPooling2D((2, 2)): Downsamples the feature maps by 2×, reducing spatial dimensions and keeping important features.
Conv2D(128): Learns more complex features with 128 filters.
MaxPooling2D((2, 2)): Further downsamples to compress the representation.

```python
# Encoder , ReLU (Rectified Linear Unit),
x = Conv2D(64, (3, 3), activation='relu', padding='same')(input_img)
x = MaxPooling2D((2, 2), padding='same')(x)
x = Conv2D(128, (3, 3), activation='relu', padding='same')(x)
x = MaxPooling2D((2, 2), padding='same')(x)

# Decoder

x = Conv2D(128, (3, 3), activation='relu', padding='same')(x)
x = UpSampling2D((2, 2))(x)
x = Conv2D(64, (3, 3), activation='relu', padding='same')(x)
x = UpSampling2D((2, 2))(x)
x = Conv2D(1, (3, 3), activation='sigmoid', padding='same')(x)
# Model
```

Decoder – Reconstructs the image:

Conv2D(128): Starts decoding with 128 filters.
UpSampling2D((2, 2)): Doubles the spatial size (opposite of pooling).
Conv2D(64): Refines features using 64 filters.
UpSampling2D((2, 2)): Restores the original image size.
Conv2D(1): Produces the final single-channel output with sigmoid (since pixel values range from 0 to 1).

Model Testing and Evaluation:

The trained model is evaluated on unseen noisy images. The predicted (denoised) outputs are compared against the clean (ground truth) images using three key evaluation metrics:

1. Peak Signal-to-Noise Ratio (PSNR): Measures the ratio between the maximum possible power of a signal and the power of corrupting noise.

   Formula:

$$\mathrm{PSNR} = 20 \cdot \log_{10}\left(\frac{MAX}{\sqrt{\mathrm{MSE}}}\right)$$

Where:

- MAX = Maximum pixel value (= 1 for normalized images)
- **MSE** = Mean Squared Error between clean and denoised images

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\mathrm{clean}}(i) - I_{\mathrm{denoised}}(i)\right)^2$$

2. Structural Similarity Index Measure (SSIM): Assesses perceptual image quality by comparing structural information, luminance, and contrast.

Formula:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Where:

- μ_x, μ_y = Mean of images
- σ_x^2, σ_y^2 = Variance
- σ_{xy} = Covariance
- C_1, C_2 = Small constants to stabilize division

3. Learned Perceptual Image Patch Similarity (LPIPS): Measures perceptual similarity using deep network features. Lower LPIPS values indicate higher similarity.

Formula:

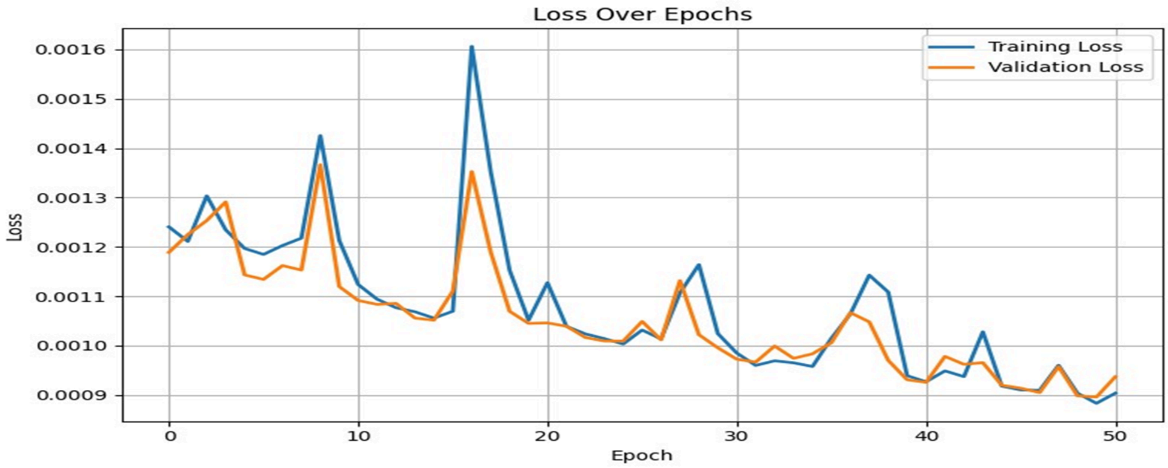
$$\mathrm{LPIPS}(x,y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||w_l \odot (\phi_l(x)_{hw} - \phi_l(y)_{hw})||_2^2$$

Where:

- ϕ_l = Activation map at layer l of the network
- w_l = Learned weights
- H_l, W_l = Spatial dimensions of layer output

Interpretation: LPIPS uses pretrained deep networks to compare activations of image patches; no explicit formula, but implemented using PyTorch `lpips` library.

Result:

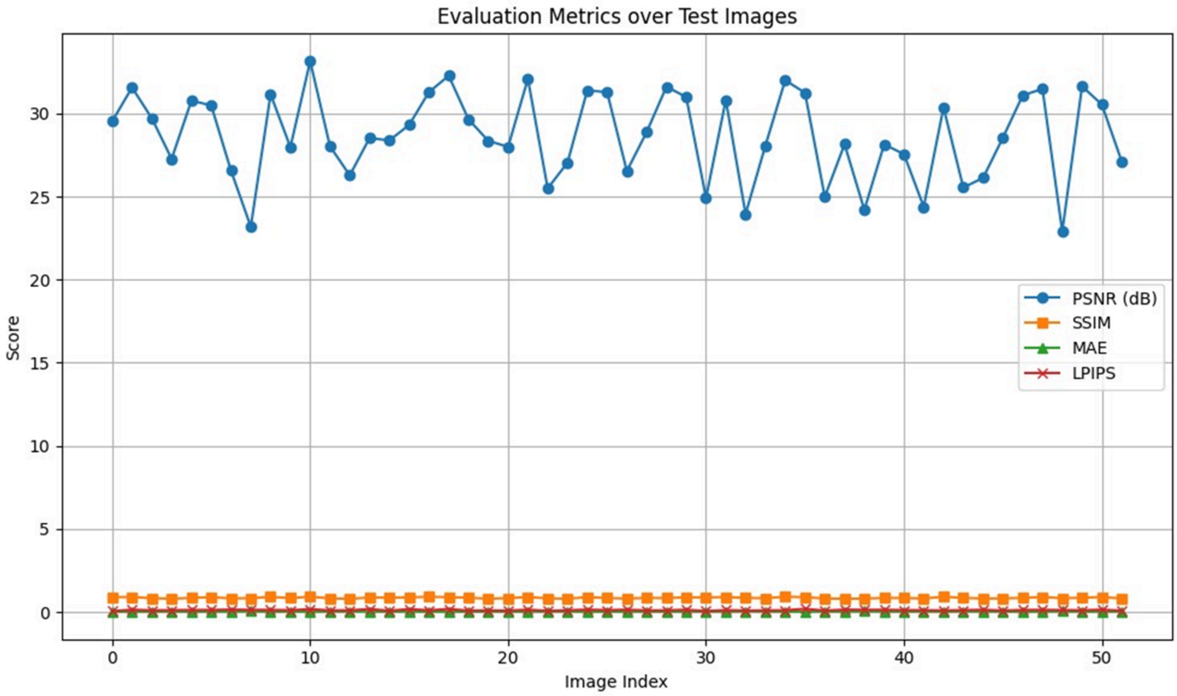


- Interpretation : X-axis (Epochs): Number of complete passes through the training dataset (0–50).
- Y-axis (Loss): Reconstruction error; lower values indicate better performance . Blue Line – Training Loss: Loss on the training dataset .
- Orange Line – Validation Loss: Loss on the validation dataset (unseen during training).
- Observations : Both training and validation losses initially fluctuate slightly as the model begins to learn . Losses then decrease steadily, reflecting effective learning and noise reduction . The close alignment of training and validation losses throughout training suggests good generalization with minimal overfitting . By 50 epochs, both losses stabilize around 0.0009, indicating high-quality reconstruction by the autoencoder.

```
Metrics over 52 test images:
PSNR  (Higher Better) : 29.55 dB
SSIM  (Higher Better) : 0.8742
MAE   (Lower Better)  : 0.0226
LPIPS (Lower Better)  : 0.0850
```

- PSNR (29.55 dB): Indicates good noise removal and clear image reconstruction.
- SSIM (0.8742): Shows that the model preserved important structural details well.
- MAE (0.0226): Reflects very small average pixel differences between original and denoised images.

- LPIPS (0.0850): Suggests that the denoised images are perceptually very close to the original ones



This graph presents the performance of the autoencoder across 50 test images using four metrics:

- PSNR (dB): Indicates reconstruction quality. Values between 25–33 dB reflect good noise removal and image fidelity.
- SSIM: Measures structural similarity. Scores close to 1 indicate strong structural preservation.
- MAE: Represents average pixel-wise error. Low values show minimal difference from original images.
- LPIPS: Evaluates perceptual similarity. Low scores imply high perceptual quality.
- Observations All metrics show consistent performance across test images.
- PSNR shows slight fluctuations but remains in a high-quality range.
- SSIM, MAE, and LPIPS remain stable, suggesting reliable generalization and effective denoising across diverse inputs.

Conclusion:

This project successfully demonstrates the effectiveness of convolutional autoencoders for denoising medical images across multiple modalities—MRI, CT, and X-ray. The proposed models not only remove noise efficiently but also preserve critical anatomical details necessary for accurate diagnosis. The achieved evaluation metrics—PSNR, SSIM, and LPIPS—confirm the model's capability to enhance image quality while retaining structural fidelity. Additionally, the modular design of the pipeline makes it adaptable for integration into diverse medical imaging workflows. Overall, this approach offers a promising, data-driven solution to tackle complex, real-world noise in medical imaging without compromising important visual information.

REFERENCE:

1. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising.

2. Chen, Y., et al. (2018). Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network
.

3. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity.

4. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric.

5. Mao, X., Shen, C., & Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.

6. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., et al. (2018). “Noise2Noise: Learning Image Restoration without Clean Data,” in Proceedings of the 35th International Conference on Machine Learning, 2965–2974. (Accessed November 27, 2020)
7. Mohan, J., Krishnaveni, V., and Guo, Y. (2014). A Survey on the Magnetic Resonance Image Denoising Methods. Biomed. Signal Process. Control. 9, 56–69. doi:10.1016/j.bspc.2013.10.007Accessed November 19, 2020)

8. Tomasi, C., and Manduchi, R. (1998). “Bilateral Filtering for gray and Color Images,” in Proceedings of the Sixth International Conference on Computer Vision (IEEE Computer Society), ICCV ’98), 839.

9. Buades, A., Coll, B., and Morel, J. (2005). “A Non-local Algorithm for Image Denoising,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 60–65.

10. Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep Cnn for Image Denoising. IEEE Trans. Image Process. 26, 3142–3155. doi:10.1109/TIP.2017.2662206Accessed November 28, 2020)

11. Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., et al. (2020). Noisy-as-clean: Learning Self-Supervised Denoising from Corrupted Image. IEEE Trans. Image Process. 29, 9316–9329. doi:10.1109/TIP.2020.3026622Accessed November 28, 2020)

12. Tripathi, P. C., and Bag, S. (2020). Cnn-dmri: a Convolutional Neural Network for Denoising of Magnetic Resonance Images. Pattern Recognition Lett. 135, 57–63. doi:10.1016/j.patrec.2020.03.036Accessed November 28, 2020)