Auto-N2N: Self-supervised MRI Denoising using Synthetic Pair Generation and Mixture of Experts

Sergio Morell-Ortega[1] sermoor1@teleco.upv.es
[1] *Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*
Alexey Mengliev[1] almen3@inf.upv.es
Ángela González-Cebrián[1] angonce1@upv.edu.es
Jaume Ivars Grimalt [1] jivagri@inf.upv.es
Pierrick Coupe[2] pierrick.coupe@u-bordeaux.fr
[2] *CNRS, Univ. Bordeaux, Bordeaux INP, LABRI, UMR5800, in2brain, F-33400 Talence, France*

Jose V. Manjón[1] jmanjon@fis.upv.es

Editors: Under Review for MIDL 2026

Abstract

Magnetic Resonance Imaging (MRI) is the gold standard for neuroimaging, yet its acquisition process inherently introduces noise that degrades diagnostic quality and quantitative analysis. Deep learning (DL) methods have achieved state-of-the-art performance, but suffer from a critical limitation: they predominantly rely on supervised learning with clean ground-truth data, which is clinically usually unavailable or unsupervised approaches, such as Noise2Noise (N2N), requiring spatially aligned noisy image pairs. This paper presents a new self-supervised method that eliminates the need for double scanning. We introduce a synthetic data generation mechanism based on Non-Local Means (NLM) principles to create training pairs from single volumes. Furthermore, we propose a Mixture of Experts (MoE) framework in which each expert employs a 3D Residual CNN architecture, enabling the system to handle the heterogeneous noise levels typically encountered in clinical settings. Experiments on the Brainweb phantom and OASIS dataset demonstrate that our approach significantly outperforms state-of-the-art methods, particularly in high-noise regimes, while reducing inference time from minutes to seconds.

Keywords: MRI Denoising, Unsupervised Learning, Noise2Noise, Deep Learning, Mixture of Experts, Non-Local Means.

1. Introduction

Medical imaging plays a central role in modern healthcare, with Magnetic Resonance Imaging (MRI) being one of the most versatile and used image modality. However, the MRI acquisition pipeline is susceptible to various noise sources, including thermal fluctuations in the receiver coils and physiological motion. This noise, often modeled as Rician in magnitude images or Gaussian in high Signal-to-Noise Ratio (SNR) regions, obscures anatomical details and adversely affects downstream automated tasks such as segmentation, tumor detection, and cortical thickness measurement. Denoising is therefore a mandatory preprocessing step. Early approaches relied on linear filtering, such as Gaussian, which reduce noise by averaging neighboring pixels according to its distance. However, these methods

© 2026 CC-BY 4.0, S. Morell-Ortega, A. Mengliev, González-Cebrián, J.I. Grimalt, P. Coupe & J.V. Manjón.

act as low-pass filters, inevitably blurring edges and fine details. Non-linear methods, such as the Bilateral Filter (Tomasi and Manduchi, 1998), attempted to preserve edges by considering intensity differences, but still struggled with texture preservation.

The Non-Local Means (NLM) algorithm (Buades et al., 2005) represented a major leap, exploiting the self-similarity of images. Instead of local neighborhoods, NLM searches the entire image (or large search windows) for similar patches to average. In MRI, extensions like Rotationally Invariant NLM (RI-NLM) and Non-Local Principal Component Analysis (NL-PCA) have set the standard for quality (Coupe et al., 2008; Manjón et al., 2012, 2015). However, these methods are computationally expensive, often taking several minutes to process a single 3D volume, and they struggle to deal with high levels of noise.

The advent of Deep Learning (DL) has shifted the paradigm. Convolutional Neural Networks (CNNs) have shown a remarkable ability to map noisy inputs to clean outputs (Manjon and Coupe, 2019). However, standard supervised DL requires paired noisy/clean data. In MRI, a ”clean” ground truth is physically impossible to acquire, as the signal is corrupted at the moment of acquisition. Synthetic data generation is a common workaround, but it often fails to capture the complex, non-stationary noise distributions and anatomical variability of real scans.

On the other hand, self-supervised learning, specifically the Noise2Noise (N2N) framework (Lehtinen et al., 2018), offers a theoretical breakthrough: a network can learn to recover the clean signal by training on pairs of noisy observations (x_i, x_j) of the same underlying scene, provided the noise is zero-mean and independent. While powerful, N2N faces a practical barrier in clinical MRI, acquiring two perfectly aligned scans for every patient is cost-prohibitive, time-consuming, and uncomfortable for patients.

Subsequent works attempted to relax the requirement for paired data. Noise2Void (N2V) (Krull et al., 2018) applies a ”blind-spot” strategy, predicting a pixel's value from its surroundings. While this eliminates the need for pairs, it inherently limits performance: by masking the central pixel, the network is deprived of direct high-frequency information, often resulting in over-smoothed textures and a loss of fine anatomical detail compared to full-target methods like N2N or NLM.(Krull et al., 2018). Noisier2Noise (Moran et al., 2019) generates training pairs by additively corrupting the already noisy input. However, in this method, the generated image pairs share the same noise instance in input and output which dificults the estimation of the clean image (specialy for low noise levels).

In this work, we propose a new framework designed to solve the ”missing pair” problem. Our contributions are threefold, 1) Synthetic Pair Generation: We utilize a fast, NLM-based algorithm to generate a second noisy realization from a single input volume using the natural structural redundancy inherent in MRI data, enabling N2N training without repeated scans, 2) Mixture of Experts (MoE): We implement a curriculum learning strategy where distinct models are fine-tuned for specific noise levels (1% to 9%), preventing the over-generalization observed in single-model approaches and 3) Hybrid Inference Pipeline: We combine the speed of 3D CNNs with the robustness of classical Rotationally Invariant NLM (RI-NLM) in a post-processing stage, achieving superior restoration quality and computational efficiency compared to current state-of-the-art methods.

2. Material and Methods

2.1. N2N Pair Generation

The core innovation of our method is the generation of a synthetic ”second scan” from a single noisy input x, eliminating the dependency on acquiring two physically distinct noisy volumes. We leverage the principle of NLM not for direct denoising, but for searching for the most similar voxels.

For every voxel x_i in the input volume, the algorithm defines a search window Ω. It identifies the most similar 3D patch N_j within Ω based on Euclidean distance. Crucially, instead of averaging multiple patches, we simply substitute the value at x_i with the central value x_j of the most similar patch N_j to create synthetic image y.

Since MRI data possesses high spatial redundancy, the structural content of x and y is nearly identical, but the noise realization at the substituted pixel is nearly-independent. This satisfies the N2N condition.

This process, implemented in optimized C code, takes only a few seconds for a standard MRI volume. Figure 1 shows an example of the results of this approach.

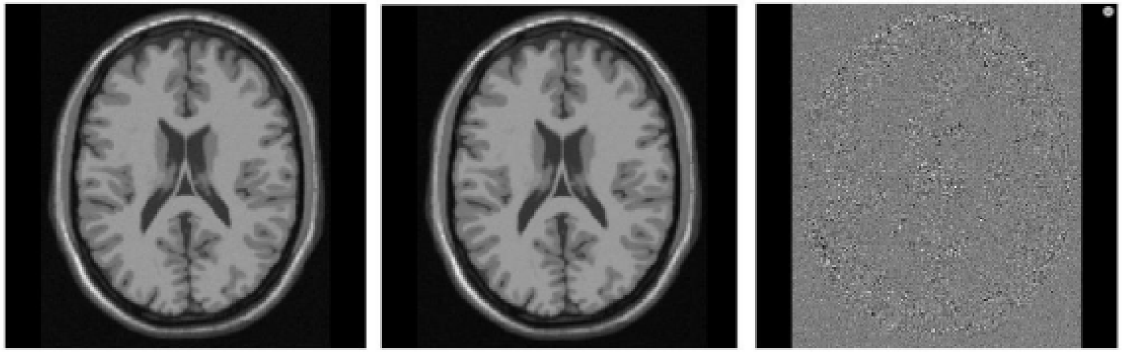

Figure 1: Left: 1% noisy Brainweb image, Center: Synthetic generated pair and Right: difference image showing just noise.

2.2. 3D Network Architecture

We propose a 3D Deep Convolutional Neural Network (DCNN) with residual learning (see Figure 2) to enhance the flow of gradients during training. The architecture details are as follows:

1. Fixed High-Pass Filter: The first layer is a non-trainable convolutional layer that subtracts a locally averaged version (3x3x3 mean kernel) from the input. This acts as a high-pass filter, removing low-frequency anatomical information and highlighting the noise map and high-frequency details (edges), thereby simplifying the task for the subsequent layers.

2. Residual Encoder: The backbone consists of 5 residual blocks. Each block contains two convolutional layers (3x3x3 with symmetric padding) with 16 filters, using ReLU activation. A skip connection adds the input block to the output.

3. Feature Compression and Reconstruction: A final 1x1x1 convolution compresses the 16 feature maps back to a single channel. This residual noise map is added to the original input via a global skip connection to produce the denoised volume.

The network has a total of 67,521 parameters, making it lightweight and fast, with an inference time of 1-2 seconds.

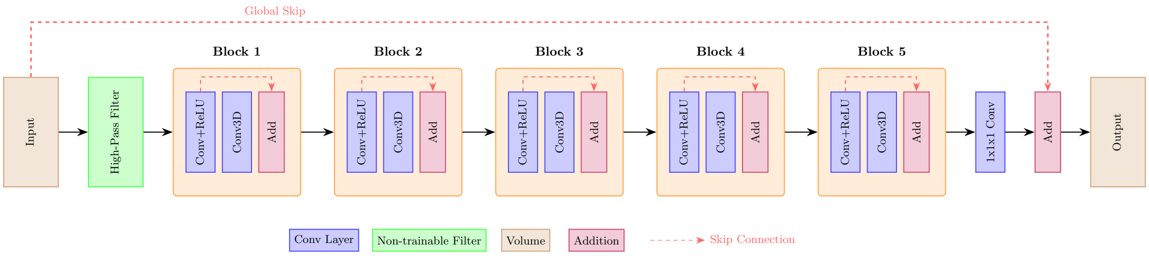

Figure 2: Architecture of DCNN network showing horizontal data flow through five residual blocks with local and global skip connections.

2.3. Loss function

The core purpose of our loss function is to enforce statistical consistency and counteract the tendency of deep learning denoisers to over-smooth. We designed a custom loss function to balance noise suppression, structural fidelity, and statistical consistency. The loss is defined as [1]:

$$\mathcal{L} = \frac{L_{\text{bias}} \cdot L_{\text{fidelity}}}{L_{\text{var}} + \epsilon} \tag{1}$$

where $\epsilon = 0.1$ for numerical stability, $R = x - f(x)$ is the residual image, and $f(x)$ is the network output. The components are defined as:

- **Bias Suppression (L_{bias}):** Defined as the mean of the local means of the residuals ($L_{\text{bias}} = |\mathbb{E}[|\mu_{\text{local}}(R)|]|$). This term explicitly enforces the Noise2Noise zero-mean noise assumption , preventing the network from introducing "photometric shifts" or low-frequency intensity artifacts.

- **Structural Fidelity (L_{fidelity}):** We employ the Mean Absolute Error (L1 norm), $L_{\text{fidelity}} = ||y - f(x)||_1$, for the primary fidelity term over L2 (MSE) as it is more robust to outlier noise and ensures superior preservation of sharp anatomical edges.

- **Texture Preservation (L_{var}):** Defined as the local standard deviation of the residuals ($L_{\text{var}} = \mathbb{E}[\sigma_{\text{local}}(R)]$). By placing this term in the denominator, it acts as a regularizer and forces to maximize noise extraction.

[1] For all the local losses the patch size is 3x3x3

2.4. Mixture of Experts (MoE) and Fine-Tuning

Initial experiments showed that a single model trained on a broad range of noise (1-9%) tends to overfit to high-noise cases while neglecting subtle noise in cleaner images. To address this, we employed a Mixture of Experts strategy. We trained five distinct ”expert” models, specialized for normalized noise variance levels corresponding to 1%, 3%, 5%, 7%, and 9%. We used a Fine-Tuning (or curriculum learning) approach. First, Train the 1% noise model from scratch and later use the weights of the 1% model to initialize the 3% model and so on. This strategy ensures faster convergence and better stability than training each model from random initialization. In inference, we estimate the input noise level using an automated NLM-based estimator to select the appropriate ”expert” model (see Section 3.1).

2.5. Postprocessing

At inference stage, we used Test-Time Data Augmentation (TTDA). We perform inference on the input and its flipped versions (axial, coronal, sagittal), averaging the results to improve robustness. In a similar way than the filter PRI-NLPCA filter (Manjón et al., 2015) the filtered image with the selected expert was used as a guide for the Rotationally Invariant NLM (RI-NLM) filter. The final image is an average of the CNN output and the RI-NLM output, leveraging the strengths of both DL and non-local self-similarity.

2.6. Datasets

We used 3 different data sources to conduct our experiments:

1. Training: We utilized the IXI Dataset (IXI), comprising 580 T1-weighted volumes collected from three hospitals (1.5T and 3T scanners). Since these are real clinical images, they contain intrinsic noise (estimated at approximately 1%), which serves as the baseline for our N2N pair generation.

2. Validation: We used the Brainweb synthetic phantom (Cocosco et al., 1997). This provides a noise-free ground truth, allowing us to add controlled Gaussian noise (1-9%) and accurately calculate PSNR metrics.

3. Qualitative Testing: A subset of 42 volumes from the OASIS dataset (Marcus et al., 2007) was used to evaluate performance on real-world data distinct from the training set.

3. Experiments and results

The framework was implemented in Python using pytorch 2.4 (Paszke et al., 2019) and trained on a L4 GPU (24GB VRAM) with the training set. We used data Augmentation consisting in random 3D flipping (axial, coronal, sagittal) and adds Gaussian noise on-the-fly to simulate different noise levels during training. The Adam optimizer was used with an adaptive learning rate of 0.001. Training was stopped early if no improvement was observed for a set patience period (200 epochs). We named the proposed method Auto-N2N.

3.1. Noise Estimation and Correction

To select the correct ”Expert” model during inference, we need to estimate the input noise level.

We used our pair generation method: if we subtract the generated image y from the input x, the standard deviation of the difference should theoretically be $\sigma \cdot \sqrt{2}$ being σ the noise standard deviation. Unfortunately, the most similar patch selection also affects the most similar noise realization giving a lower standard deviation than expected. However, we found a strong linear relationship between this estimate and the true added noise. Therefore, we estimated and applied a linear correction factor derived from regression analysis to achieve near-perfect noise estimation ($\sigma_{corrected} = 3.33\sigma - 1.89$, $(R^2 = 1.0, RMSE = 0.0462)$).

3.2. Quantitative Comparison

We compared our method against two benchmarks, PRI-NLPCA: A state-of-the-art classical method using non-local Principal Component Analysis and the PRI-PBCNN: A prior supervised deep learning approach. Table 1 summarizes the Peak Signal-to-Noise Ratio (PSNR) results on the BrainWeb phantom.

Table 1: PSNR comparison (dB) on BrainWeb phantom across methods and noise levels. Best results in bold.

Method	**1%**	**3%**	**5%**	**7%**	**9%**	**Mean**
Noisy Input	40.67	32.37	29.33	27.40	26.10	31.17
PRI-NLPCA	**45.38**	39.33	36.63	34.90	33.58	37.96
PRI-PBCNN	45.09	39.34	36.70	35.00	33.67	37.96
Auto-N2N (1 model)	43.93	38.80	37.43	36.47	35.79	38.48
Auto-N2N (experts)	44.39	39.67	38.02	37.19	36.78	39.21
Auto-N2N + RI-NLM	45.30	**40.62**	**39.03**	**38.25**	**37.83**	**40.20**

As can be noticed, at low noise (1%), PRI-NLPCA performs slightly better (+0.08 dB), likely due to the high self-similarity in clean images which PCA exploits well. At medium to high noise (3% - 9%), Auto-N2N significantly outperforms classical methods. At 9% noise, our method achieves 37.83 dB versus 33.58 dB for PRI-NLPCA, a massive improvement of + 4.25 dB. The combined approach (Auto-N2N + RI-NLM) yields the best overall performance, demonstrating that deep learning and classical priors are complementary. Figure 3 shows an example of the results.

3.3. Inference Speed

Efficiency is a major advantage of our framework. While classical PRI-NLPCA takes approximately 5 minutes per volume, and PRI-PBCNN takes 50 seconds, Auto-N2N processes a full 3D volume in 2-3 seconds. This real-time capability is transformative for high-throughput clinical workflows.

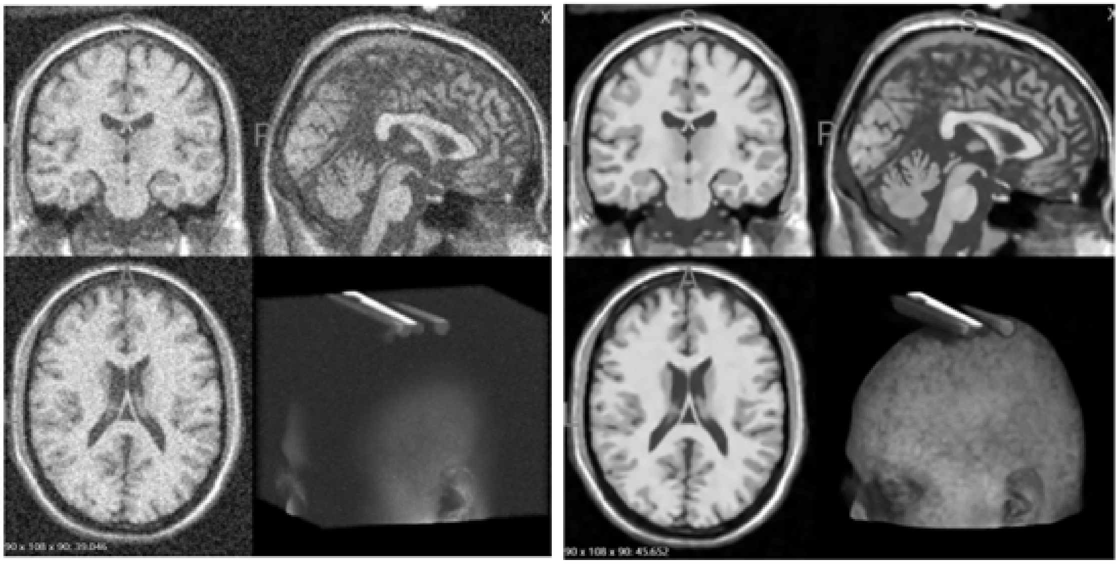

Figure 3: Visual comparison on BrainWeb phantom corrupted with 9% noise (left) and its denoised version (right).

3.4. Qualitative Results

We applied the method to the OASIS dataset (real clinical data with around 2.1% noise). Figure 4 illustrates the results. The filter successfully removes the granular noise in the white matter and cerebrospinal fluid while preserving the sharp contrast of the grey matter and anatomical boundaries. Unlike simple Gaussian filtering, there is no visible blurring of the cortex edges.

4. Discussion

The results validate the hypothesis that structural redundancy in MRI can be exploited to synthesize the ground truth alternative required by N2N. Unlike "blind-spot" approaches (e.g., Noise2Void) , which mask the central voxel and inevitably degrade high-frequency spatial resolution, our Auto-N2N approach preserves full spatial context by retrieving a homologous patch from the search window. By generating training targets that are structurally identical yet statistically independent, we unlock the power of N2N for clinical data. This approach allows the CNN to learn full texture distributions rather than merely interpolating from neighbors, resulting in the superior edge preservation seen in Figure (**?**).

The Mixture of Experts approach proved essential. A single model struggled to balance the subtle corrections needed for 1% noise with the aggressive filtering needed for 9% noise. Specialized experts allow for optimal performance across the dynamic range of scanner qualities. The Post-Processing stage, combining Test-Time Data Augmentation (TTDA) and fusion with RI-NLM, provides a final boost in quality.

Currently, the noise estimation relies on the NLM-based generation, which requires correction factors for high noise. This underestimation is intrinsic to the NLM selection process: by minimizing Euclidean distance to find a partner patch, the algorithm naturally biases selection towards noise realizations that reduce the total distance, thereby artificially dampening the measured variance. While the linear correction proved robust for this study, end-to-end learning where the network self-estimates noise variance would be a future solution.

Additionally, we aim to adapt the proposed method to deal with spatially varying Rician noise (basically using a local intensity/noise dependent bias).

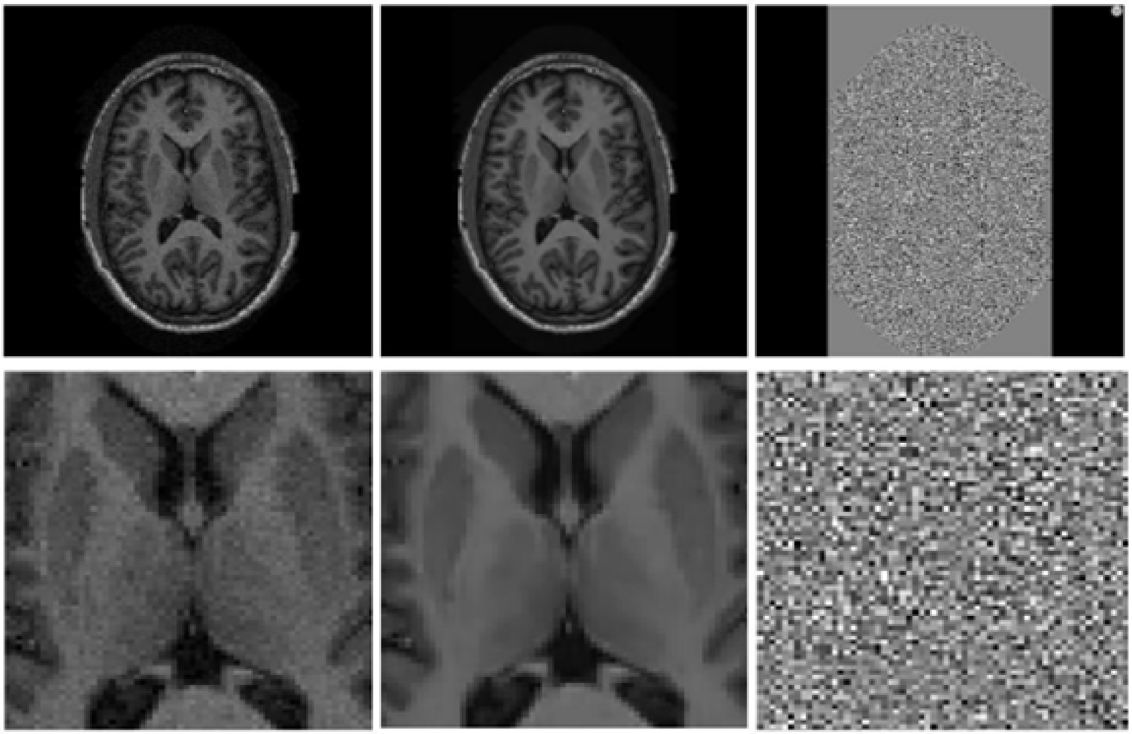

Figure 4: Denoising results on real OASIS data. From Left to right: noisy image, filtered image and the corresponding residuals. The zoom (bottom row) shows preservation of fine subcortical structures.

5. Conclusion

We have presented Auto-N2N, a robust, self-supervised, and highly efficient 3D MRI denoising framework. By successfully synthesizing training pairs from single volumes and employing a noise-specific Mixture of Experts, we achieve state-of-the-art denoising quality without requiring clean ground truth or repeated patient scans. The method reduces processing time by orders of magnitude compared to classical methods, making it an ideal candidate for integration into clinical MRI scanners and PACS systems.

Acknowledgments

This work has been developed thanks to the projects PID2020-118608RB-I00 and PID2023-152127OB-I00 of the Ministerio de Ciencia e Innovacion de España. This work also benefited from the support of the projects DeepvolBrain, HoliBrain and FOLDDICO of the French National Research Agency (ANR-18-CE45-0013, ANR-23-CE45-0020-01 and ANR-20-CHIA-0027-01). Finally, this study received financial support from the French government in the framework of the University of Bordeaux's France 2030 program / RRI "IMPACT, the PEPR StratifyAging and the IHU VBHI (ANR-23-IAHU-0001). We thank the support of ITACA (Institute of Information and Communication Technologies) at UPV (Universitat Politècnica de València).

References

Ixi dataset – brain development. URL https://brain-development.org/ixi-dataset/.

Antoni Buades, Bartomeu Coll, and Jean Michel Morel. A non-local algorithm for image denoising. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, II:60–65, 2005. doi: 10.1109/CVPR.2005.38. URL https://ieeexplore.ieee.org/document/1467423.

C. Cocosco, Vasken Kollokian, R. K. Kwan, and Alan C. Evans. Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage*, 1997.

Pierrick Coupe, Pierre Yger, Sylvain Prima, Pierre Hellier, Charles Kervrann, and Christian Barillot. An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *IEEE transactions on medical imaging*, 27:425–441, 4 2008. ISSN 1558-254X. doi: 10.1109/TMI.2007.906087. URL https://pubmed.ncbi.nlm.nih.gov/18390341/.

Alexander Krull, Tim Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:2124–2132, 11 2018. ISSN 10636919. doi: 10.1109/CVPR.2019.00223. URL https://arxiv.org/pdf/1811.10980.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *35th International Conference on Machine Learning, ICML 2018*, 7:4620–4631, 3 2018. ISSN 26403498. URL https://arxiv.org/pdf/1803.04189.

Jose V. Manjon and Pierrick Coupe. Mri denoising using deep learning and non-local averaging. 11 2019. URL https://arxiv.org/pdf/1911.04798.

José V. Manjón, Pierrick Coupé, Antonio Buades, D. Louis Collins, and Montserrat Robles. New methods for mri denoising based on sparseness and self-similarity. *Medical Image Analysis*, 16:18–27, 1 2012. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2011.04.003. URL https://www.sciencedirect.com/science/article/pii/S1361841511000491.

José V. Manjón, Pierrick Coupé, and Antonio Buades. Mri noise estimation and denoising using non-local pca. *Medical Image Analysis*, 22:35–47, 5 2015. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2015.01.004. URL https://www.sciencedirect.com/science/article/pii/S1361841515000171.

Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19:1498–1507, 9 2007. ISSN 0898-929X. doi: 10.1162/JOCN.2007.19.9.1498. URL https://pubmed.ncbi.nlm.nih.gov/17714011/.

Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 12061–12069, 10 2019. ISSN 10636919. doi: 10.1109/CVPR42600.2020.01208. URL https://arxiv.org/pdf/1910.11908.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 12 2019. ISSN 10495258. URL https://arxiv.org/pdf/1912.01703.

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. *Proceedings of the IEEE International Conference on Computer Vision*, pages 839–846, 1998. doi: 10.1109/ICCV.1998.710815. URL https://ieeexplore.ieee.org/document/710815.