



Perceptually Motivated Generative Model for Magnetic Resonance Image Denoising

Hazique Aetesam¹ · Suman Kumar Maji¹

Received: 21 March 2022 / Revised: 1 November 2022 / Accepted: 17 November 2022 / Published online: 6 December 2022
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2022

Abstract

Image denoising is an important preprocessing step in low-level vision problems involving biomedical images. Noise removal techniques can greatly benefit raw corrupted magnetic resonance images (MRI). It has been discovered that the MR data is corrupted by a mixture of Gaussian-impulse noise caused by detector flaws and transmission errors. This paper proposes a deep generative model (*GenMRIDenoiser*) for dealing with this mixed noise scenario. This work makes four contributions. To begin, Wasserstein generative adversarial network (WGAN) is used in model training to mitigate the problem of vanishing gradient, mode collapse, and convergence issues encountered while training a vanilla GAN. *Second*, a perceptually motivated loss function is used to guide the training process in order to preserve the low-level details in the form of high-frequency components in the image. *Third*, batch renormalization is used between the convolutional and activation layers to prevent performance degradation under the assumption of non-independent and identically distributed (non-iid) data. *Fourth*, global feature attention module (GFAM) is appended at the beginning and end of the parallel ensemble blocks to capture the long-range dependencies that are often lost due to the small receptive field of convolutional filters. The experimental results over synthetic data and MRI stack obtained from real MR scanners indicate the potential utility of the proposed technique across a wide range of degradation scenarios.

Keywords Magnetic resonance imaging · Image denoising · Gaussian-impulse noise · Perceptually motivated loss · Adversarial training · Attention block

Introduction

Magnetic resonance imaging (MRI) is a popular and effective medical imaging technique used to visualize anatomical details and analyse physiological processes in the human body's soft biological tissues. This aids in making clinical decisions at the appropriate time. MRI differs from CT scan in that it does not use ionizing radiation. It takes advantage of the magnetic properties of hydrogen protons, which are abundant in the human body. Strong primary static magnetic field and low variable secondary field in the presence of radio-frequency (RF) coils help in the spatial encoding of the received signal. An acquired MRI sequence is limited

by its low temporal resolution. Further, images acquired in temporary k-space are composed of real and imaginary parts in the frequency domain [1]. Image reconstruction using the inverse Fourier transform (IFT) yields magnitude and phase images [2]. However, the resulting image is found to have Gaussian noise due to multiple factors. These include thermal agitation caused by the human body in the receiver's coil elements, tissue inhomogeneity, static field intensity, and receiver element bandwidth [3]. Impulse noise is caused by factors such as analog-to-digital conversion, bit error transmission, and improper RF coil alignment. MR image restoration [4] is a longstanding problem in the scientific literature and different methods appear to tackle this problem. Clinically accurate restoration results enable successful downstream analysis such as disease detection, tissue localization, and quantitative MRI analysis [5].

Images in MRI are obtained in the complex domain using the method proposed by Kumar et al. [6] and Henkelman [1]. This complex domain representation includes both real and imaginary parts. Images obtained from the RF coils are

✉ Hazique Aetesam
hazique.pcs16@iitp.ac.in

Suman Kumar Maji
smaji@iitp.ac.in

¹ Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna 801106, India

in the frequency domain and are represented in a temporary space known as k -space. The resulting data contains both signal and noise components along both real and imaginary planes. Due to thermal noise in the patient [2], the noise is complex white Gaussian in nature. The signal must be converted to imaging domain before any useful information can be extracted from it. Inverse discrete Fourier transform (IDFT) is the most common tool available to accomplish this. The resultant image obtained from IDFT retains the Gaussian noise in the data due to the orthogonality of the Fourier transform [2]. We present an extensive literature review on recent developments in MR image denoising works in the following paragraphs. Filtering-based and model-driven approaches for noise removal are presented in the following first and second paragraphs respectively. Learning-based techniques are discussed in the third paragraph while the final paragraph explains MR image denoising schemes devoted exclusively to the removal of impulse and mixed Gaussian-impulse noise.

In an attempt to investigate previous works on MR image denoising, it was discovered that many of them used the non-local means (NLM) filter in various forms [7]. The authors of [8] used NLM for 3D denoising of MR stacks through block-wise and parallel implementation with automatic tuning of the smoothing parameter. In [9], authors propose a locally adaptive noise estimation method to accommodate parallel imaging environments such as Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA) [10] where noise is spatially variant. The authors of [11] used an adaptive soft wavelet coefficient mixing to perform adaptive MR denoising in a multi-resolution framework. In another work, the properties of self-similarity and sparsity in MR data are explored [12]. This is accomplished through hard thresholding of the discrete cosine transform (DCT) and a rotationally invariant version of the NLM filter. Other attempts at NLM-based denoising in natural images in general, and MR data in particular, include [13] and [14]. Authors in [7] provide a comprehensive overview of denoising in MRI using NLM. Here, NLM methods are classified into four categories: fast NLM, adaptive NLM, multi-resolution NLM, and statistical methods. However, because the weighted average of the neighbouring voxels must be estimated to find the target pixel, NLM-based methods are mostly computationally intensive.

Authors in [15] investigate a model-free approach for denoising diffusion-weighted (DW) MR data using principal component analysis (PCA). Similarly, authors in [16] attempted to denoise DW-MRI using random matrix theory and PCA of noise by exploiting the fact that eigenvalues of noise-only data are corrupted by the Marchenko-Pastur distribution. Prior to obtaining magnitude data, Wirestam et al. [17] using Wiener filtering in the wavelet domain, performed complex-valued DW-MRI denoising. A linear

minimum mean square error (LMMSE) approach is used to denoise 3D MR data by leveraging data redundancy and estimating local SNR [18]. In another work [19], authors employ an adaptive combination of fuzzy logic and bilateral filtering to denoise MR data. Total Variation (TV) [20] has also been used in several methods for removing noise from MR data. Contributions by Liu et al. [21] include a two-step wavelet filtering method to estimate noise maps and a TV-based hyper-Laplacian parameter to adaptively model spatially varying noise parameters. Similar approach is explored in [22]. Zhang et al. [23] discuss an alternating minimization approach that uses a Gaussian mixture model (GMM) to cluster non-local similar patches of MR data and a hyper-Laplacian prior to reduce ringing artefacts. Authors in [24] investigate a comparison of Gaussian denoising and wavelet-based approaches for denoising functional MRI (fMRI) data. However, TV-based priors are limited in their ability to characterize different levels of data and noise information [25], resulting in smoothed results in highly textured regions. In general, model-driven methods are restricted by their prior assumptions on noisy/clean data. These assumptions, however, are frequently violated in data acquired with diverse commercially available MRI scanners.

Since the last few years, there has been a surge of machine learning approaches in general and deep learning in particular for a wide class of computer vision tasks in medical imaging; including magnetic resonance imaging [26]. In [27], spatial-temporal denoising of dynamic contrast enhanced (DCE) MR data is performed using a deep learning framework to study the blood-brain barrier using pharmacokinetic parameters. Authors in [28] incorporate a 3D channel residual learning strategy for denoising MR images. The authors of [29] and [30] used a wide residual learning strategy to remove noise from MR data in order to increase the receptive field size and speed up training. A 3D MR image denoising is performed in [31] using a Wasserstein generative adversarial network (WGAN) in a residual setting. In [32], the authors propose a two-phase denoising model in which they estimate a signal-dependent and spatially variant noise level map using a maximum-likelihood based unbiased estimator. This estimated noise level map is appended as prior to the corrupted input data into the denoiser model. Bermudez et al. [33] and Moreno López et al. [34] are two other data-driven MRI denoising schemes. Most deep learning methods, however, necessitate a large amount of training data. In practice, generating labelled groundtruth of this size is not feasible. This lack of labelled groundtruth limits the model's ability to accommodate the intricate noise patterns observed in real MR data. This is further exacerbated by computationally demanding training of 3D MR patches.

It was discovered that the combination of Gaussian and impulse noise significantly contributes to the corruption of MR data. Most of the works in the previous paragraphs

emphasized the presence of Gaussian noise in complex frequency domain images obtained from RF coils. There is a wealth of literature available that denoises MR data under the assumption of impulse noise. The authors of [35] evaluated the denoising results for the removal of Gaussian and impulse from T2-weighted MR images using three different filters: adaptive, averaging, and median filters. A dual fidelity elastic net regularization is proposed in [36] to collectively handle the effect of Gaussian and impulse noise. In [37], a joint approach using linear and median filters for the elimination of impulse, fluctuation, and geometric noise from heart MRI was proposed. Two successful adaptive median filtering (AMF) strategies for the removal of impulse noise from MRI data are [38] and [39]. In [38], authors employed a switch-based AMF where a bell-shaped membership function instead of triangular membership function was used for obtaining better results. The same set of authors extended their work in [39] where neuro-fuzzy logic was used for both decision-making and filtering sub-operations in addition to a multilayer perceptron (MLP) architecture for detecting and removing impulse noise. Another AMF approach was proposed in [40] which judges a noisy pixel in a sub-image based on standard deviation of that sub-image and average value of the filtering window. In [41], a fixed weighted mean filter and an adaptive median filter were designed to remove high-intensity noise from MR images. More realistic random-valued impulse noise was considered in [42] to design a hardware device which emphasized on several different noisy scenarios in corrupted data. In [43], a local similarity with overlapped block partitioning approach was proposed for the detection of noisy pixels and their replacement with appropriate values. In [44], a quaternion median filtering approach was proposed, which selects a noisy pixel based on the sum of pixel differences with other pixels in the filtering window. Recently, authors in [45] developed an adaptive switching modified decision based unsymmetric trimmed median filter (ASMDBUTMF) for noise reduction in grayscale MRI data. However, all of these methods have two major flaws. To begin, most works cannot handle both Gaussian and impulse noise mixtures in MRI data at the same time. Second, in designing denoising techniques, mostly fixed valued impulse noise (FVIN) in the form of salt-and-pepper noise is considered, with little consideration given to more realistic random-valued impulse noise (RVIN) [46, 47].

This work attempts to overcome the limitations in present approaches. The paper's main contributions are summarized below:

- The Wasserstein generative adversarial network (WGAN) is used for training to reduce the effect of diminishing gradient, which is more conspicuous in adversarial training.
- To achieve visually appealing denoising results, a perceptually motivated loss function is used to guide the training process for data corrupted by mixed Gaussian-impulse noise.
- Batch renormalization facilitates the use of non-iid data for training under variable noise levels.
- 3D depth-wise separable convolution enables computationally intensive training processes to be performed on high-dimensional data.
- A global feature attention module (GFAM) is plugged at the beginning and end of the proposed architecture to capture the long-range dependencies that are oftenly lost in vanilla deep convolutional layers.

The remainder of the paper is organized as follows. We explain the image formation model under the Gaussian-impulse noise assumption in “[Image Formation Model and Objective](#)”. The proposed methodology is explained in the “[Adversarial Training](#)” to “[Training Details](#)”, while the experimental results and discussions are presented in “[Results](#)” and “[Discussion](#)” respectively. Finally, the paper concludes in “[Conclusion](#)”.

Methods

Image Formation Model and Objective

The image formation model for data corrupted by mixed Gaussian-impulse (G-I) noise is given by:

$$v = u + g + s \quad (1)$$

where $u \in R^{w \times h \times c}$ is the latent clean data. w , h and c are the data dimensions along horizontal, vertical and depth axes respectively. Clean data u is corrupted by the Gaussian noise component g and impulse noise component s . g is a random variable following Gaussian distribution with mean $\mu_g = 0$ and variance σ_g^2 ; denoted by $g \sim \mathcal{N}(\mu_g, \sigma_g^2)$. Similarly, the random variable s follows Laplace distribution [48, 49] with $\mu_s = 0$ as the location parameter and σ_s as the scaling parameter; denoted by $s \sim \mathcal{L}(\mu_s, \sigma_s)$. Under impulse noise, the pixel value u_i is replaced with a uniform random variable s_i in the range $[s_{\min}, s_{\max}]$. However, due to pixel value range clipping, the resultant distribution does not remain uniform or multimodal, but follows Laplacian distribution [49]. Hence, the resultant data is corrupted by mixed G-I noise to give the composite noisy signal v . The dimensions of other variables are the same as that of u . The noise level map for Gaussian as well as for impulse component can be spatially variant along the depth axis.

The objective of this work is to obtain clean estimate \hat{u} from the noisy observation v which is as close as possible to

the clean data ($\hat{u} \sim u$). This is made possible in an adversarial setting under the generative model [50]. More precisely, using a large pair of noisy-clean image pairs $\{v_i, u_i\}_{i=1}^N$, we intend to achieve the following objective:

$$\min_G \max_D \mathcal{L}(D, G) = E_{u \sim P_r} [\log D(u)] + E_{v \sim P_n} [1 - \log D(G(v))] \quad (2)$$

by training the generator G and discriminator D model in succession. Once the model is trained, the discriminator is discarded and the generator is trained to the extent so as to produce perceptually appealing visual results. In Eq. (2), P_r and P_n are the real and noisy data distribution respectively and $\mathcal{L}(D, G)$ is the loss function in the min-max game [50]. However, to generate visually realistic results, Eq. (2) needs to be modified.

Adversarial Training

The vanilla generative adversarial network (GAN) model proposed by Goodfellow et al. [50] has several flaws. To begin, GAN is known to suffer from the vanishing gradient problem, which occurs when a discriminator network becomes overly successful in discarding samples, as observed by [51]. Second, mode collapse prevents the generator network from producing diverse samples, according to [51]. Thirdly, non-convergence is also a prevalent issue [51]. Lastly, according to Eq. (2), the loss function used in GAN training does not correlate to the visual quality of the generated images.

To address the aforementioned limitations, several changes are being made to the vanilla GAN model. Wasserstein GAN (WGAN) is one example of such a modification. Rather than determining whether a sample is real or fake, the discriminator in WGAN (here referred to as the *critic* network) assigns a score based on the quality of the image generated. This requires a different objective in the minimization of the distance between true and generated data distributions [51] (of Eq. (2)). As a result, for model training, an alternative distance measure known as *Earth-mover distance* or Wasserstein distance is used [51]. This distance measure is known to be continuous and differentiable and allows smooth back-propagation. Furthermore, unlike discriminators, the critic gradient in WGAN does not saturate and continues to provide useful gradient to the generator network. This avoids the vanishing gradient issue. In the min-max game, the loss function of WGAN is modified as follows:

$$\min_G \max_D \mathcal{L}(D, G) = E_{u \sim P_r} [D(u)] - E_{v \sim P_n} [D(G(v))] \quad (3)$$

It is important to remember that the implementation of the critic network requires the former to be in the space of 1-Lipschitz functions. This is enforced by constraining the weight gradient to be in the range $[-c, c]$ [51]. This, however, leads to suboptimal critic performance because the

critic network is unable to capture the complexity of data under the limitations of 1-Lipschitz functions. The problem is mitigated by providing an alternative way of enforcing the Lipschitz constraint. It is known that a function is 1-Lipschitz continuous if it has gradient norm at most 1 everywhere in the manifold [52]. Hence, Eq. (3) can be modified to provide the loss of the discriminator as:

$$L_{\text{WGAN}}(D) = \min_G \max_D \mathcal{L}(D, G) = E_{u \sim P_r} [D(u)] - E_{v \sim P_n} [D(G(v))] + \underbrace{\lambda E_{\hat{v} \sim P_{\hat{v}}} [(\|\nabla_{\hat{v}} D(\hat{v})\|_2 - 1)^2]}_{\text{gradient penalty factor}} \quad (4)$$

Here, the gradient is calculated over the output of the critic network with respect to its input [52]. λ denotes the penalty coefficient and $P_{\hat{v}}$ is a probability distribution that is used to uniformly sample along straight lines between the real data distribution P_r and the generating data distribution P_g . Similarly, loss of the generator is given by:

$$L_{\text{WGAN}}(G) = -E_{v \sim P_n} [D(G(v))] \quad (5)$$

We conclude this section with the following exploitation of notations. D and G are the critic and generator models respectively. The aim of the generator is to generate a sample distribution P_g which is as close as possible to the real data distribution P_r . During the course of training, the generator excels in generating samples that mimic real (clean) data distribution.

Perceptually Motivated Loss Function

It is observed that generative models are not inherently suited for image restoration tasks. The generated samples lack low-level features (high-frequency components in the form of edges and textures) in the restored data. Therefore, in this section, we intend to augment Eq. (5) with a perceptually motivated content loss $L_{\text{MSSIM}}(u, G(v))$. The final loss of the generator is given as:

$$L = \underbrace{L_{\text{MSSIM}}(u, G(v))}_{\text{content loss}} + \underbrace{L_{\text{WGAN}}(G)}_{\text{adversarial loss}} \quad (6)$$

Perceptual Loss

From the next section, \hat{u} ($= G(v)$) is used to denote the clean estimated data obtained from the generator network. Content loss will be discussed shortly in this section.

In contrast to the work of residual encoder-decoder WGAN (RED-WGAN) [31], which uses a combination of extracted VGG features and mean squared error (MSE) loss to design generator loss, we provide an alternative proposition here. We use a perceptually motivated loss function because MSE does not synchronize with the image visual quality [53] and VGG network is intrinsically designed for

2D data [54, 55]. Structural similarity index (SSIM) [56] has been around for nearly 15 years; used as perceptually motivated image quality assessment metric according to Human Visual System (HVS). This encourages us to use this metric to modify the generator loss in Eq. (5). In a traditional setting, any gradient-based minimization technique requires the loss function to be differentiable in nature; this is also true during back-propagation steps in neural network training. As a result, before it can be used for network training, the derivative of SSIM must be calculated. SSIM is written as:

$$SSIM(u, \hat{u}) = l(u, \hat{u})^\alpha \cdot c(u, \hat{u})^\beta \cdot s(u, \hat{u})^\gamma \tag{7}$$

The three terms used in the product above have the same usual meaning as defined in [56].

$$l(u, \hat{u}) = \frac{2\mu_u\mu_{\hat{u}} + C_1}{\mu_u^2 + \mu_{\hat{u}}^2 + C_1}; c(u, \hat{u}) = \frac{2\sigma_u\sigma_{\hat{u}} + C_2}{\sigma_u^2 + \sigma_{\hat{u}}^2 + C_2}; s(u, \hat{u}) = \frac{2\sigma_{u\hat{u}}C_3}{\sigma_u\sigma_{\hat{u}} + C_3} \tag{8}$$

Here, l , c and s are the luminosity, contrast and structural information. C_1 , C_2 and C_3 are small constants used to avoid numerical instability where $C_3 = C_2/2$ and α , β and γ all are set to be 1. μ_u , σ_u and $\sigma_{u\hat{u}}$ are the mean, variance and covariance respectively calculated over entire pixels of the image. $\mu_{\hat{u}}$ and $\sigma_{\hat{u}}$ are calculated similarly for \hat{u} .

However, it is observed that SSIM calculated over entire image yields inaccurate results for l , c and s in Eq. (8) due to spatially variant values of these quantities. In order to circumvent this problem, we use image patch P_u which is centred at pixel p for finding patch-wise SSIM for $u(p)$ and $\hat{u}(p)$. SSIM for all the patches are averaged together to obtain mean SSIM (MSSIM). We make use of Gaussian filter G_σ with standard deviation σ to obtain patch results such that $\mu_u(p) = G_\sigma \star P_u$, $\sigma_u(p) = G_\sigma \star P_u^2 - \mu_u^2(p)$ and $\sigma_{u\hat{u}} = G_\sigma(P_u \cdot P_{\hat{u}}) - \mu_u(p) \cdot \mu_{\hat{u}}(p)$. $\mu_{\hat{u}}(p)$ and $\sigma_{\hat{u}}(p)$ are defined analogously. \star and \cdot represent convolution and point-wise multiplication operations respectively. For SSIM to be used as a loss function, we make use of three different properties of it: (i) boundedness (≤ 1), (ii) convexity and (iii) symmetry ($SSIM(u, \hat{u}) = SSIM(\hat{u}, u)$) [57]. As a result, MSSIM between $u(p)$ and $\hat{u}(p)$:

$$L_{MSSIM} = 1 - \frac{1}{M} \sum_p (SSIM(u(p), \hat{u}(p))) \tag{9a}$$

where M = total number of image patches. Upon manipulation of Eqs. (7), (8) and (9a), we obtain:

$$SSIM(u(p), \hat{u}(p)) = \underbrace{\left(\frac{2\mu_u\mu_{\hat{u}} + C_1}{\mu_u^2 + \mu_{\hat{u}}^2 + C_1} \right)}_{l(p)} \cdot \underbrace{\left(\frac{2\sigma_{u\hat{u}} + C_2}{\sigma_u^2 + \sigma_{\hat{u}}^2 + C_2} \right)}_{cs(p)} \tag{9b}$$

In order to apply back-propagation during model training, we need to find out the derivative of loss function (here, MSSIM).

$$\frac{\partial SSIM(u(p), \hat{u}(p))}{\partial u(q)} = cs(p) \frac{\partial}{\partial u(q)} \cdot l(p) + l(p) \frac{\partial}{\partial u(q)} \cdot cs(p) \tag{10}$$

$$\frac{\partial l(p)}{\partial u(q)} = \frac{\partial l(p)}{\partial \mu_u(p)} \cdot \frac{\partial \mu_u(p)}{\partial u(q)} = 2G_\sigma(p - q) \left[\frac{\mu_{\hat{u}} - \mu_u l(p)}{\mu_u^2 + \mu_{\hat{u}}^2 + C_1} \right] \tag{11a}$$

where $\partial \mu_u(p) / \partial u(q) = G_\sigma(p - q)$. Similarly, $\partial cs(p) / \partial u(q)$ can be evaluated as:

$$\frac{\partial cs(p)}{\partial u(q)} = \frac{\partial cs(p)}{\partial \mu_u(p)} \cdot \frac{\partial \mu_u(p)}{\partial u(q)} = 2G_\sigma(p - q) \left[(\hat{u}(q) - \mu_y) - cs(p)(u(q) - \mu_u) \right] \tag{11b}$$

Feeding the expressions of Eqs. (11a) and (11b) in Eq. (10), we obtain the final expression for MSSIM as:

$$\frac{\partial MSSIM(u, \hat{u})}{\partial u(q)} = \frac{2}{M} \sum_p \left[cs(p) \left(\frac{\mu_{\hat{u}} - \mu_u l(p)}{\mu_u^2 + \mu_{\hat{u}}^2 + C_1} \right) G_\sigma(p - q) + l(p) \left(\frac{G_\sigma(p - q)}{\sigma_u^2 + \sigma_{\hat{u}}^2 + C_2} \right) (\hat{u}(q) - \mu_y - cs(p)(u(q) - \mu_u)) \right] \tag{12}$$

Batch Renormalization and 3D Depth-wise Separable Convolution

In its pristine form, batch normalization (BN) is used to reduce internal covariate shift in data [58]. However, they are constrained by the assumption of independent and identically distributed (iid) data. In any image restoration problem, model needs to be trained over a wide range of noise levels in data, thus exhibiting non-iid property. Further, the small batch size makes the use of BN suboptimal. This is primarily due to differences in the BN operation during training and inference [59]. Moving average statistics are calculated during model training but only used during inference. By incorporating moving average statistics into the training process, batch renormalization (BRN) helps to mitigate the aforementioned issues [59].

Model training suffers from latency when using 3D data. To address this issue, we employ 3D depth-wise separable convolution [60]. This aids in reducing learnable parameters by a factor of 1/10 without compromising noisy to clean image mapping during inference.

Network Design

The generator model’s architecture is depicted in Fig. 1. Images are permuted between $y \leftrightarrow z$ and $x \leftrightarrow z$ dimensions

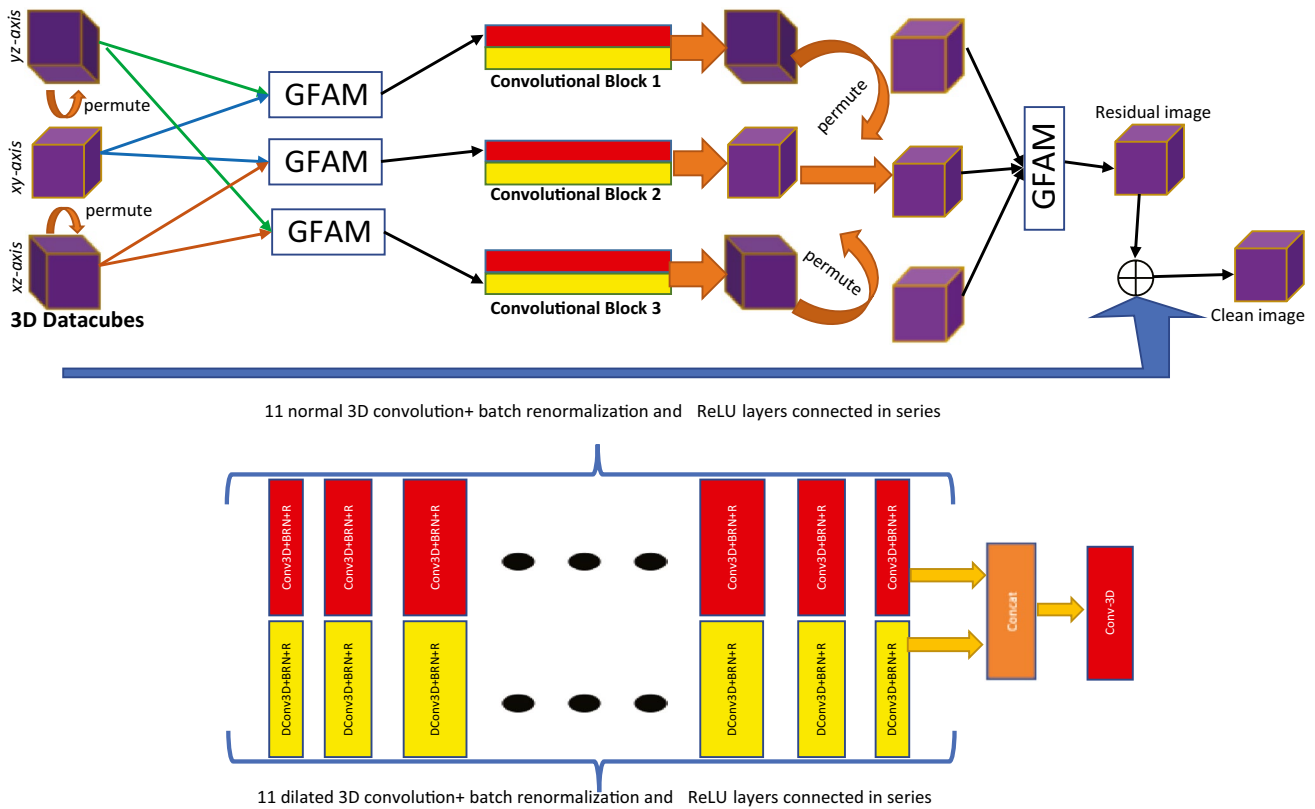


Fig. 1 The architecture of the proposed generator model. The 3D data patches are permuted along the three dimensions to obtain three data patches such that $xyz = \text{permute}(y, z)$, $yxz = \text{permute}(x, y)$ and $zyx = \text{permute}(x, z)$. The global feature attention module (GFAM) blocks can accept arbitrary number of inputs (see Fig. 2). The \rightarrow (green), \rightarrow (blue) and \rightarrow (brown) arrows coming out from the three

from their original xyz dimensions to exploit spatial information from different anatomical planes (coronal, sagittal, and transverse). The resultant dimensions are xyz , yxz and zyx . The results of these permutations are sent to global feature attention modules (GFAM, which will be discussed later). The GFAM output is fed into two sets of convolutional blocks. This is made up of two parallel blocks that use 3D normal and dilated convolutions. There are 11 blocks in sequence; each consisting of convolutional layer, BRN layer, and rectified linear units (CBR). The convolutional filter size is fixed at $3 \times 3 \times 3$, and the number of filters is increased from 4 to 128 from block 1 to 6 and symmetrically decreased to 1 from 7th to 11th block. Finally, the permutation layer is used once more to bring back images to the original xyz plane. The outputs are passed through GFAM once more to acquire global contextual features learned in the previous parallel ensemble blocks. The clean estimated image is obtained by subtracting the resultant residual image from the corrupted input data. The discriminator network architecture is the same as used in [54]. BN layers, on the other hand, are removed from the original implementation to

permutated 3D patches are fed as input to the three GFAM blocks. The top row shows the overall model while the bottom row shows the expanded views of the red and yellow convolutional blocks. The output features obtained from the convolutional blocks are again permuted to their original dimensions (xyz) and fused together into one 3D image using the final GFAM block

avoid correlations between samples in the same batch. This affects the gradient penalty term used in [52]. Furthermore, a linear activation function is used at the end of the critic model rather than a sigmoid function.

Adaptive Global Feature Attention Module

Conventional convolutional filters are limited in three different ways. *Firstly*, they are unable to map long-range dependencies existing in the 3D data blocks due to small receptive field size [61]. *Secondly*, all spatial locations in the image are treated similarly. This is more problematic in image restoration problems where low- and high-frequency details need to be properly separated into distinct regions [62]. A typical convolutional layer does not account for these low-level details. *Thirdly*, most of the traditional supervised learning solutions for image restoration like DnCNN [63] rely on the final feature maps obtained in the last layer. However, features obtained from initial and intermediate layers provide complementary information about the clean estimation. Therefore, we incorporate global feature attention

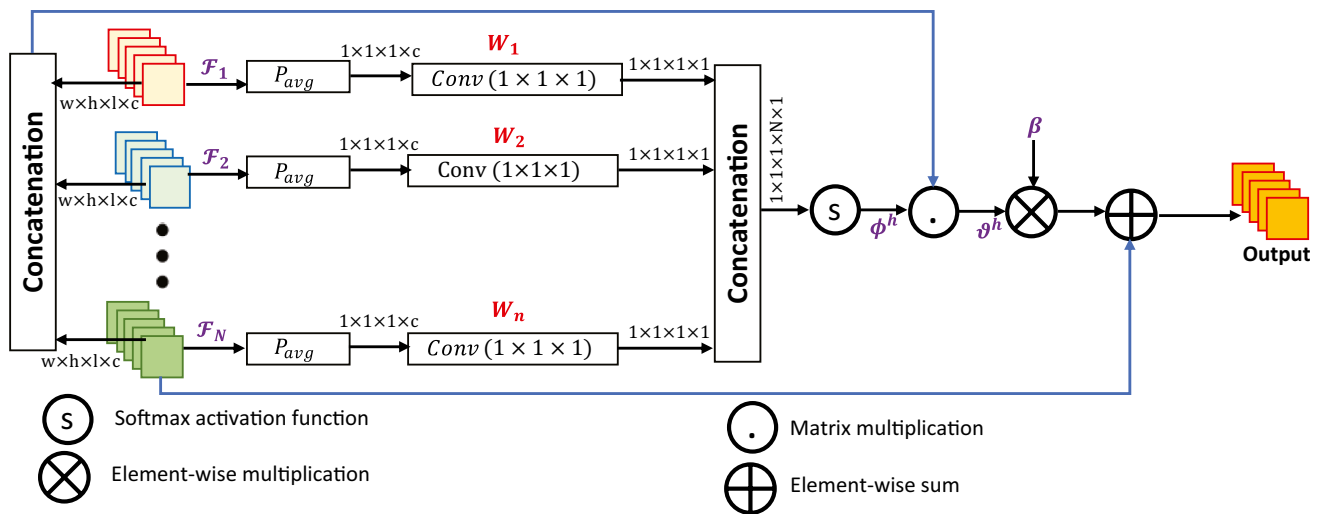


Fig. 2 Adaptive Global Feature Attention Module (GFAM)

module (GFAM) at the beginning and end of the set of parallel blocks. Since different features in the permuted input data and different feature maps along the parallel paths may provide hidden contextual information, we merge these features under a global setting. As shown in Fig. 2, the feature maps obtained from the previous layers are denoted as $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_N$ such that $\mathcal{F}_i \in R^{w \times h \times l \times c}$; where $\{w, h, l\}$ are the dimensions of the 3D data and c is the number of channels. Global features from each feature map \mathcal{F}_i are obtained by the application of average pooling P_{avg} followed by the convolutional filter W_i to obtain the features; denoted by $A_i = W_i P_{avg} \mathcal{F}_i$; $A_i \in R^{1 \times 1 \times 1 \times 1}$. All these features are then concatenated together to obtain $A \in R^{1 \times 1 \times 1 \times N \times 1}$. Softmax function is applied over the concatenated features to obtain $\omega \in R^{1 \times 1 \times 1 \times N \times 1}$.

$$\omega_j = \frac{\exp(A_j)}{\sum_{i=1}^N \exp(A_i^h)} \tag{13}$$

Further, the original \mathcal{F}_i 's are concatenated together to obtain $B \in R^{w \times h \times l \times N \times c}$. Global contextual modelling is performed by the matrix multiplication between ω and B .

$$v = \sum_{i=1}^N \omega_i B_i \tag{14}$$

where $v \in R^{w \times h \times l \times c}$. Further, an adaptive learnable weight β is point-wise multiplied with B . The final output of the GFAM is obtained by element-wise sum of previous output and the last feature input \mathcal{F}_N .

$$GFAM_{out} = \mathcal{F}_N + \beta \sum_{i=1}^N \omega_i B_i \tag{15}$$

Training Details

We have trained the model for MR denoising using simulated MR images obtained from Brainweb [64] database.¹ It consists of normal and multiple sclerosis (MS) datasets containing T1, T2 and PD-weighted images in five different slice thicknesses: 1 mm, 3 mm, 5 mm, 7mm and 9 mm. This gives rise to 181, 60, 36, 26 and 20 slices respectively. The spatial dimension of each image is 181×217 . Images of size $181 \times 217 \times 181$ from each modality are used for testing while the rest are used for model training. Image non-uniformity is set to 0% and initial noise levels are set to zero. The levels of noise used for training are: Gaussian noise in the signal to noise (SNR) range [2, 20] dB and impulse noise in the range [1%, 30%]. Ten noise levels are uniformly and randomly sampled from the range of each noise type. This gives a combination of 100 noise levels = (10 Gaussian noise levels \times 10 impulse noise levels) that can be added to the training data. We have used patch-based training in our model [65]. Non-overlapping patches of size $40 \times 40 \times 6$ are extracted from each image. Total number of patches generated are 108, 000.

Batch size is selected to 16. Adaptive moment estimation (ADAM) is used as the optimization technique as the de facto standard due to its superiority over other stochastic gradient descent methods. Its parameters are set as $\alpha = 0.01, \beta_1 = 0.9, \beta_2 = 0.99$ and $\epsilon = 1e - 6$. We could find that 200 epochs were sufficient for the convergence of the model.

¹ Obtained from: <https://brainweb.bic.mni.mcgill.ca/>

Results

In this section, we conduct experiments on synthetically corrupted and real MR data obtained from clinical level MRI scanners.

Experimental Setup

Image of size $181 \times 217 \times 181$ from each imaging modality (T1- and PD-weighted) was separated from the Brainweb [64] database during training time for evaluation of the trained model during test time. The selected 3D images correspond to the 1 mm slicing of simulated data. For comparison, seven different techniques are used. Four of them are filtering-based approaches² (optimized non-local means (ONLM) [8], adaptive optimized NLM (AONLM) [9], prefiltered rotationally invariant NLM (PRINLM) [12] and multi-resolution NLM (MRNLM) [11]) while the other three competing methods are learning-based (DnCNN) [63], Deep Image Prior (DIP) [66] and residual encoder-decoder-Wasserstein generative adversarial network (RED-WGAN) [31]). Under ONLM, a computationally efficient and parallelized version of NLM is proposed with features of automatic parameter tuning and relevant voxels selection. An adaptive version of NLM is used in AONLM where the strength of filter is decided based on the local noise level in the image. Authors in PRINLM exploit the sparsity and self-similarity properties by using moving window discrete cosine transform. Based on the spatial and frequency information in noisy data, a multi-resolution analysis is presented in MRNLM. For the NLM-based methods, optimal filtering parameters are chosen so as to generate best visual and metrics results [7, 67]; tabulated in Table 1. DnCNN is an implementation of neural network based denoiser where the network predicts the residual data rather than the clean estimation. Since DnCNN is inherently designed for grayscale/RGB data, we retrained the model using our own 3D MRI training data. DIP uses a novel proposition that assumes that an untrained or randomly trained neural network acts as a natural prior over the image to be estimated. RED-WGAN uses a generative model based on residual encoder-decoder architecture.

Two full-reference image quality metrics are used for the quantitative evaluation of restoration results: peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [56], which are defined as follows:

Table 1 Optimal filtering parameters for comparing techniques. S is the radius of search window and N is the radius of the patch, h is the smoothing parameter and σ is the standard deviation of Gaussian distribution

Comparing Techniques	Filtering Parameters
ONLM	$S = 5, h = \sigma, N = 2$
AONLM	$S = 3, N1 = 2, N2 = 2, h = \sqrt{2\beta\sigma^2 V }$ where β is a constant and $ V $ is the number of voxels
PRINLM	$S = 5, N = 2, h = 0.4\sigma$
MRNLM	$S = 11, h = \sigma, N = 2$
UKR	Radius of the second order regression = 4

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_u}{\text{MSE}^2} \right) \quad (16a)$$

$$\text{SSIM}(u, \hat{u}) = \frac{(2\mu_u\mu_{\hat{u}} + C_1)(2\sigma_{u\hat{u}} + C_2)}{(\mu_u^2 + \mu_{\hat{u}}^2 + C_1)(\sigma_u^2 + \sigma_{\hat{u}}^2 + C_2)} \quad (16b)$$

Here, u and \hat{u} are the patches of clean and recovered images from the local window respectively, μ_u , σ_u^2 and $\sigma_{u\hat{u}}$ are the mean, variance and covariance matrix, respectively.

Experiments on real data are conducted on IXI datasets³ obtained from Institute of Psychiatry (IoP). The full dataset consists of brain MRI obtained from five different imaging formats: T1, T2, PD-weighted, MRA and diffusion-weighted images from 600 healthy subjects. There are 74 images each from T1, T2 and PD modalities. Dimension of each image is $256 \times 256 \times 136$. A PD-weighted image was used for evaluation purpose.

Discussion

Simulated MR Data

Each simulated MR image (corresponding to T1 and PD-weighted images) is corrupted with six different levels of Gaussian-impulse noise: (15 dB, 3%), (12 dB, 6%), (10 dB, 10%), (6 dB, 15%), (3 dB, 20%) and (1 dB, 25%). PSNR and SSIM corresponding to these noise levels are shown in Tables 2 and 3 respectively. It can be observed that our proposed technique can maintain a considerable gap with comparing techniques. Visual inspection is necessary to confirm the validity of the metric results. For this, we have presented the visual results for T1 and PD-weighted images in Figs. 3 and 4 respectively. To present a fair comparison on 3D reconstruction ability

² Code available at: <https://sites.google.com/site/pierrickcoupe/software/denoising/mri-denoising>

³ Obtained from: <https://brain-development.org/ixi-dataset/>

Table 2 Peak Signal to Noise Ratio (PSNR) for three different imaging modes (T1 and PD) at different levels of noise for different methods

Methods/Noise (dB,%)	(15, 3)	(12, 6)	(10, 10)	(6, 15)	(3, 20)	(1, 25)	(15, 3)	(12, 6)	(10, 10)	(6, 15)	(3, 20)	(1, 25)
	T1-Weighted						PD-Weighted					
Noisy	38.47	28.22	22.24	18.90	17.25	16.53	38.21	28.74	22.60	20.27	18.22	16.92
ONLM	44.98	37.17	35.53	33.31	30.16	24.99	42.70	37.16	33.16	34.10	30.55	28.26
AONLM	43.23	35.54	30.13	25.85	23.12	20.06	41.79	35.94	30.10	27.64	23.09	20.37
PRINLM	45.88	39.86	35.99	33.35	29.60	23.88	45.83	38.86	35.86	30.49	31.66	31.08
MRNLM	44.85	37.32	34.95	33.11	30.36	25.04	42.10	36.69	32.62	34.00	29.46	26.84
DnCNN	41.02	33.38	29.90	27.19	25.00	22.66	35.70	33.27	28.40	27.33	25.25	23.71
DIP	45.24	37.59	32.23	28.45	25.98	23.55	45.19	37.77	30.35	29.14	26.52	24.55
RED-WGAN	44.60	38.70	33.68	29.19	25.48	23.56	38.92	37.74	34.66	30.07	26.97	23.87
GenMRIDenoiser	47.09	41.21	37.12	34.09	32.28	30.03	46.61	39.50	35.72	33.66	31.13	28.80

of different methods, we have presented the results on all three anatomical planes (the first image in each tabular block represents the transverse/axial section which when divided into cross-sectional planes gives coronal (shown by the red line) and sagittal (shown by the yellow line) plane results). For T1-weighted images, results are presented for the noise level (10 dB, 10%) and for PD-weighted images, results are displayed for the noise level (6 dB, 15%). It can be observed that results produced by DnCNN are not able to reduce noise considerably. Also, visuals produced in sagittal and coronal planes in DIP result in post-processing artefacts. However, images generated by the proposed neural network model can maintain fine details in recovered images. This is more prominent for PD-weighted images.

Results on running time are mandatory for testing the efficiency of different methods under the same hardware conditions. All experiments on conventional methods are conducted on MATLAB-2019 using 16GB RAM and an i7 processor. All deep learning models are trained and evaluated using Google Colab under Nvidia P100-PCIE-16GB GPU. Results on running time are tabulated in Table 4.

Real MR Data

Experiments on real data are performed on the IXI datasets. This database consists of real MR images from three different medical centres: Guys hospital (GH), Hammersmith hospital (HH) and Institute of Psychiatry (IoP). The results presented in Fig. 5 are obtained from IoP. Size of the image is $256 \times 256 \times 136$. Just like the synthetic data, we have visualized results on all three anatomical planes. The results produced by the proposed technique provide considerably better visuals with detail preserving capability. This is attributed to the perceptually motivated loss being used for model training. Results on running time (in seconds) are specified in the caption of the figure itself.

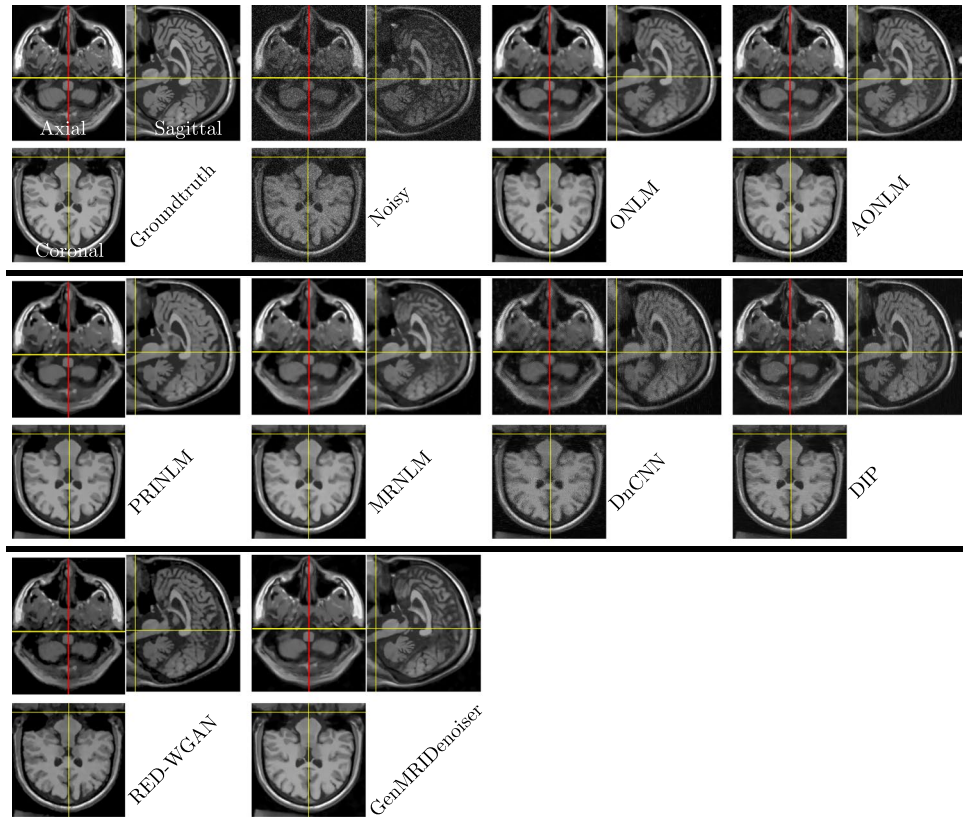
Ablation Study

As discussed in the proposed section, Batch renormalization (BRN) [59] is explored in the network training as it is a feasible alternative when data is non-iid and batch size is small. In Fig. 6 (left), we have validated this proposition by plotting validation loss for two different batch sizes: 40 and 200. We

Table 3 Structural Similarity (SSIM) for three different imaging modes (T1 and PD) at different levels of noise for different methods

Methods/Noise (dB,%)	(15, 3)	(12, 6)	(10, 10)	(6, 15)	(3, 20)	(1, 25)	(15, 3)	(12, 6)	(10, 10)	(6, 15)	(3, 20)	(1, 25)
	T1-Weighted						PD-Weighted					
Noisy	0.704	0.361	0.216	0.144	0.106	0.084	0.671	0.349	0.208	0.155	0.115	0.090
ONLM	0.966	0.941	0.899	0.847	0.796	0.704	0.958	0.921	0.890	0.841	0.773	0.681
AONLM	0.921	0.657	0.449	0.336	0.268	0.201	0.904	0.638	0.622	0.537	0.457	0.408
PRINLM	0.963	0.936	0.801	0.863	0.803	0.670	0.953	0.918	0.891	0.854	0.846	0.829
MRNLM	0.966	0.942	0.889	0.866	0.818	0.715	0.959	0.921	0.896	0.857	0.800	0.713
DnCNN	0.806	0.746	0.673	0.614	0.564	0.516	0.770	0.611	0.528	0.493	0.451	0.417
DIP	0.965	0.777	0.526	0.456	0.356	0.280	0.955	0.762	0.573	0.564	0.468	0.418
RED-WGAN	0.958	0.927	0.873	0.805	0.721	0.665	0.945	0.916	0.890	0.851	0.810	0.742
GenMRIDenoiser	0.986	0.946	0.897	0.878	0.832	0.742	0.988	0.935	0.901	0.875	0.860	0.831

Fig. 3 Results on simulated Brainweb-T1 dataset for slice thickness 1mm. The results of the comparing methods are shown for the noise level $(G, I) = (10 \text{ dB}, 10\%)$. Each figure in a group shows the axial (top left), sagittal (right) and coronal (down) planes. Upon careful visualization, it can be observed that the proposed methodology removes noise along with the preservation of details. This is more evident in the axial and coronal sections



can observe that models trained using both BN [58] and BRN for batch size 200 yields good results. However, we get more benefits over validation loss when batch size is chosen to be 40 and BRN is used as the normalization strategy. This is not true

for model trained using BN. However, training is more stable when large batch size is used or when BN is used.

Further, since training over 3D data is computationally costly; we used 3D depth-wise separable convolution to reduce

Fig. 4 Results on simulated Brainweb-PD dataset for slice thickness 1mm. The results of the comparing methods are shown for the noise level $(G, I) = (6 \text{ dB}, 15\%)$. Each figure in a group shows the axial (top left), sagittal (right) and coronal (down) planes. Upon careful visualization, it can be observed that the proposed methodology removes noise along with the preservation of details. This is more evident in the axial and coronal sections

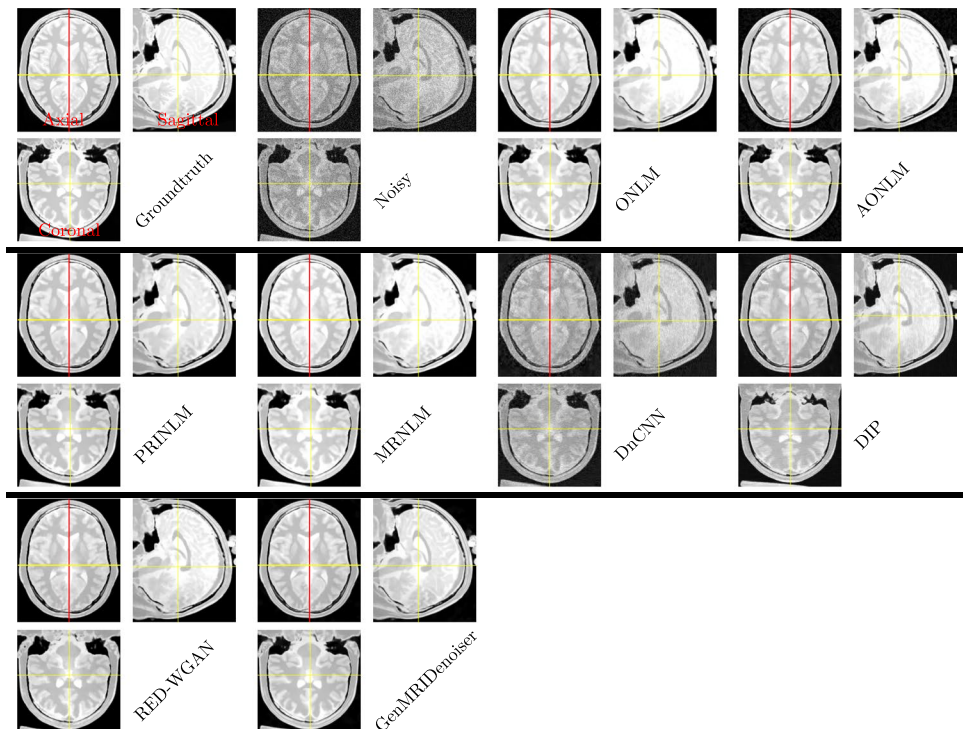


Table 4 Results on running time (in seconds) using different methods

Methods	ONLM	AONLM	MRNLM	PRINLM	DnCNN	DIP	RED-WGAN	GenMRIDenoiser
Running Time (in seconds)	73.21	513.49	315.03	130.59	18.82	990.75	34.66	28.14

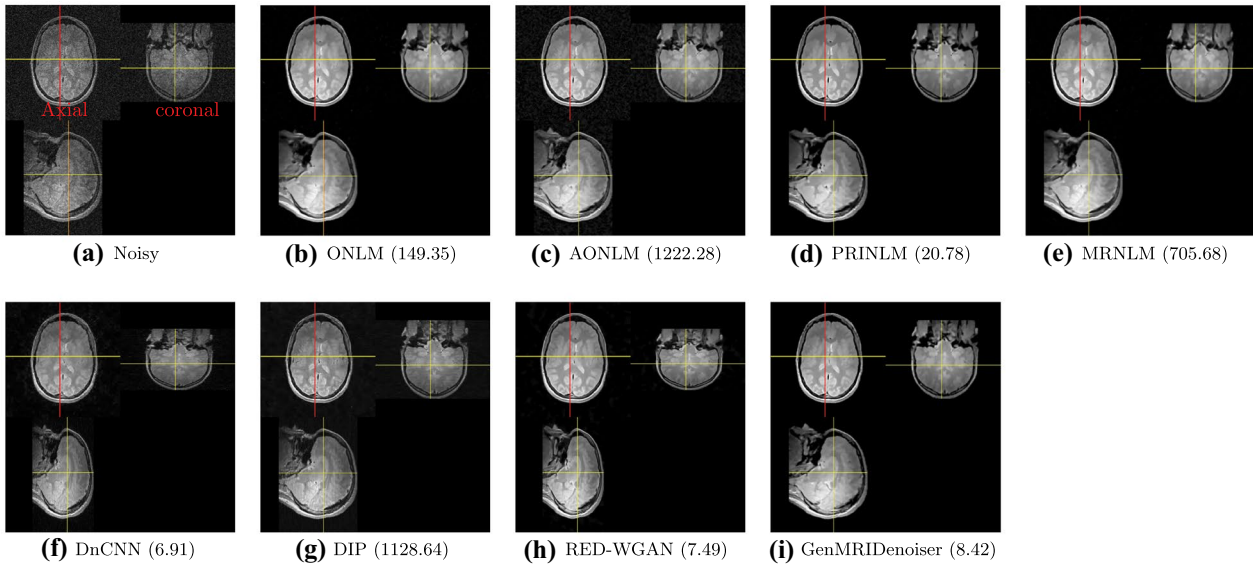


Fig. 5 Real PD-weighted data obtained from IXI database Institute of Psychiatry. Each figure in a group shows the axial (top left), coronal (right) and sagittal (right) planes

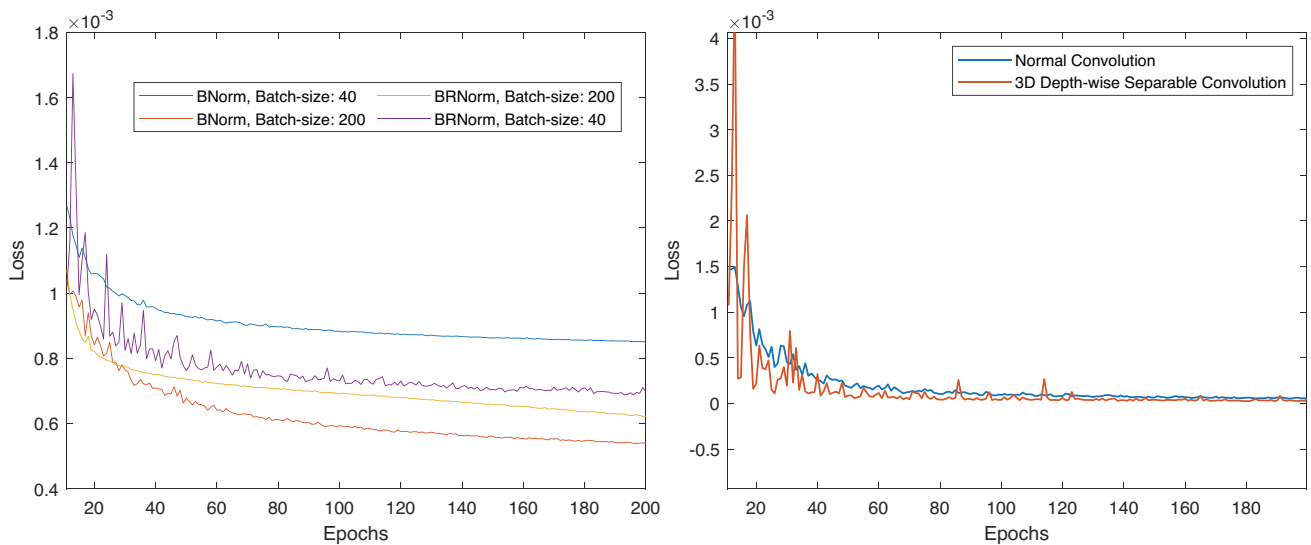


Fig. 6 **Left:** Validation loss for large (200) and small (40) batch size using BN and BRN. **Right:** Validation loss using normal and 3D depth-wise separable convolution

the number of model parameters without significant drop in model performance. To validate this result, we plotted validation loss for over 200 epochs for models trained using normal

and 3D depth-wise separable convolution. We can observe negligible difference in the two training paradigms. The plot is shown in Fig. 6 (right).

Conclusion

In this paper, we discussed the image denoising process for MR data under the assumption of Gaussian-impulse noise. A deep generative model was proposed with emphasis on perceptually motivated loss to guide the training process suited for low-level computer vision tasks. The problem of non-iid data during training process was handled using batch renormalization while the problem of computational burden during training was mitigated using 3d depth-wise separable convolution. The augmentation of global feature attention block further boosted the performance of the network by separately paying attention to the low- and high-frequency details in the image during training and inference. As an extension to this work, we intend to explore the effect of attention blocks in further improving the denoising accuracy of the network in the same adversarial setting. Furthermore, we intend to explore noise removal in the complex MR data and other imaging modalities like diffusion-weighted MRI (DW-MRI). Noise modelling in multi-coil acquisition systems and deep restoration frameworks under such settings can also be explored in subsequent works.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-022-00744-2>.

Declarations

Ethics Approval This is a numerical simulation study for which no ethical approval was required. The research did not involve any human/non-human subjects.

Conflict of Interest The authors declare no competing interests.

References

- Henkelman, R.M.: Measurement of signal intensities in the presence of noise in mr images. *Medical physics* **12**(2), 232–233 (1985)
- Nowak, R.D.: Wavelet-based rician noise removal for magnetic resonance imaging. *IEEE Transactions on Image Processing* **8**(10), 1408–1419 (1999)
- Luisier, F., Blu, T., Wolfe, P.J.: A cure for noisy magnetic resonance images: Chi-square unbiased risk estimation. *IEEE Transactions on Image Processing* **21**(8), 3454–3466 (2012)
- You, S., Lei, B., Wang, S., Chui, C.K., Cheung, A.C., Liu, Y., Gan, M., Wu, G., Shen, Y.: Fine perceptive gans for brain mr image super-resolution in wavelet domain. *IEEE transactions on neural networks and learning systems* (2022)
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., Shen, Y.: Bidirectional mapping generative adversarial networks for brain mr to pet synthesis. *IEEE Transactions on Medical Imaging* **41**(1), 145–157 (2021)
- Kumar, A., Welti, D., Ernst, R.R.: Nmr fourier zeugmatography. *Journal of Magnetic Resonance* (1969) **18**(1), 69–83 (1975)
- Bhujle, H.V., Vadavadagi, B.H.: Nlm based magnetic resonance image denoising—a review. *Biomedical Signal Processing and Control* **47**, 252–261 (2019)
- Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *IEEE transactions on medical imaging* **27**(4), 425–441 (2008)
- Manjón, J.V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M.: Adaptive non-local means denoising of mr images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging* **31**(1), 192–203 (2010)
- Breuer, F.A., Kellman, P., Griswold, M.A., Jakob, P.M.: Dynamic autocalibrated parallel imaging using temporal grappa (tgrappa). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **53**(4), 981–985 (2005)
- Coupé, P., Manjón, J.V., Robles, M., Collins, D.L.: Adaptive multiresolution non-local means filter for three-dimensional magnetic resonance image denoising. *IET image Processing* **6**(5), 558–568 (2012)
- Manjón, J.V., Coupé, P., Buades, A., Collins, D.L., Robles, M.: New methods for mri denoising based on sparseness and self-similarity. *Medical image analysis* **16**(1), 18–27 (2012)
- Tasdizen, T.: Principal neighborhood dictionaries for nonlocal means image denoising. *IEEE Transactions on Image Processing* **18**(12), 2649–2660 (2009)
- Lu, K., He, N., Li, L.: Nonlocal means-based denoising for medical images. *Computational and mathematical methods in medicine* **2012** (2012)
- Gurney-Champion, O.J., Collins, D.J., Wetscherek, A., Rata, M., Klaassen, R., van Laarhoven, H.W., Harrington, K.J., Oelfke, U., Orton, M.R.: Principal component analysis f0sr fast and model-free denoising of multi b-value diffusion-weighted mr images. *Physics in Medicine & Biology* **64**(10), 105015 (2019)
- Veraart, J., Novikov, D.S., Christiaens, D., Ades-Aron, B., Sijbers, J., Fieremans, E.: Denoising of diffusion mri using random matrix theory. *Neuroimage* **142**, 394–406 (2016)
- Wirestam, R., Bibic, A., Lätt, J., Brockstedt, S., Ståhlberg, F.: Denoising of complex mri data by wavelet-domain filtering: Application to high-b-value diffusion-weighted imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **56**(5), 1114–1120 (2006)
- Golshan, H.M., Hasanzadeh, R.P., Yousefzadeh, S.C.: An mri denoising method using image data redundancy and local snr estimation. *Magnetic resonance imaging* **31**(7), 1206–1217 (2013)
- Kala, R., Deepa, P.: Adaptive fuzzy hexagonal bilateral filter for brain mri denoising. *Multimedia Tools and Applications*, 1–18 (2019)
- Zhu, Y., Shen, W., Cheng, F., Jin, C., Cao, G.: Removal of high density gaussian noise in compressed sensing mri reconstruction through modified total variation image denoising method. *Heliyon* **6**(3), 03680 (2020)
- Liu, R.W., Shi, L., Huang, W., Xu, J., Yu, S.C.H., Wang, D.: Generalized total variation-based mri rician denoising model with spatially adaptive regularization parameters. *Magnetic resonance imaging* **32**(6), 702–720 (2014)
- Wang, Y., Zhou, H.: Total variation wavelet-based medical image denoising. *International Journal of Biomedical Imaging* **2006** (2006)
- Zhang, Y., Yang, Z., Hu, J., Zou, S., Fu, Y.: Mri denoising using low rank prior and sparse gradient prior. *IEEE Access* **7**, 45858–45865 (2019)
- Wink, A.M., Roerdink, J.B.: Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging* **23**(3), 374–387 (2004)
- Simi, V., Edla, D.R., Joseph, J., Kuppli, V.: Analysis of controversies in the formulation and evaluation of restoration algorithms for mr images. *Expert Systems with Applications* **135**, 39–59 (2019)

26. Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik* **29**(2), 102–127 (2019)
27. Benou, A., Veksler, R., Friedman, A., Raviv, T.R.: Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced mri sequences. *Medical image analysis* **42**, 145–159 (2017)
28. Jiang, D., Dou, W., Vosters, L., Xu, X., Sun, Y., Tan, T.: Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network. *Japanese journal of radiology* **36**(9), 566–574 (2018)
29. You, X., Cao, N., Lu, H., Mao, M., Wanga, W.: Denoising of mr images with rician noise using a wider neural network and noise range division. *Magnetic Resonance Imaging* **64**, 154–159 (2019)
30. Panda, A., Naskar, R., Rajbans, S., Pal, S.: A 3d wide residual network with perceptual loss for brain mri image denoising. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–7 (2019). IEEE
31. Ran, M., Hu, J., Chen, Y., Chen, H., Sun, H., Zhou, J., Zhang, Y.: Denoising of 3d magnetic resonance images using a residual encoder–decoder wasserstein generative adversarial network. *Medical image analysis* **55**, 165–180 (2019)
32. Aetesam, H., Maji, S.K.: Attention-based noise prior network for magnetic resonance image denoising. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–4 (2022). IEEE
33. Bermudez, C., Plassard, A.J., Davis, L.T., Newton, A.T., Resnick, S.M., Landman, B.A.: Learning implicit brain mri manifolds with deep learning. In: *Medical Imaging 2018: Image Processing*, vol. 10574, p. 105741 (2018). International Society for Optics and Photonics
34. Moreno López, M., Frederick, J.M., Ventura, J.: Evaluation of mri denoising methods using unsupervised learning. *Frontiers in Artificial Intelligence* **4**, 75 (2021)
35. Isa, I.S., Sulaiman, S.N., Mustapha, M., Darus, S.: Evaluating denoising performances of fundamental filters for t2-weighted mri images. *Procedia Computer Science* **60**, 760–768 (2015)
36. Aetesam, H., Maji, S.K.: L2- l1 fidelity based elastic net regularization for magnetic resonance image denoising. In: 2020 International Conference on Contemporary Computing and Applications (IC3A), pp. 137–142 (2020). IEEE
37. Shlykov, V., Kotovskyi, V., Višniakov, N., Šešok, A.: Model for elimination of mixed noise from mri heart images. *Applied Sciences* **10**(14), 4747 (2020)
38. Toprak, A., Güler, I.: Impulse noise reduction in medical images with the use of switch mode fuzzy adaptive median filter. *Digital signal processing* **17**(4), 711–723 (2007)
39. Toprak, A., Özerdem, M.S., Güler, İ.: Suppression of impulse noise in mr images using artificial intelligent based neuro-fuzzy adaptive median filter. *Digital signal processing* **18**(3), 391–405 (2008)
40. Lin, L., Meng, X., Liang, X.: Reduction of impulse noise in mri images using block-based adaptive median filter. In: 2013 IEEE International Conference on Medical Imaging Physics and Engineering, pp. 132–134 (2013). IEEE
41. Mafi, M., Martin, H., Adjouadi, M.: High impulse noise intensity removal in mri images. In: 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), pp. 1–6 (2017). IEEE
42. HosseinKhani, Z., Hajabdollahi, M., Karimi, N., Soroushmehr, S., Shirani, S., Samavi, S., Najarian, K.: Real-time impulse noise removal from mr images for radiosurgery applications. *arXiv preprint arXiv:1707.05975* (2017)
43. HosseinKhani, Z., Hajabdollahi, M., Karimi, N., Soroushmehr, R., Shirani, S., Najarian, K., Samavi, S.: Adaptive real-time removal of impulse noise in medical images. *Journal of medical systems* **42**(11), 216 (2018)
44. Chanu, P.R., Singh, K.M.: Impulse noise removal from medical images by two stage quaternion vector median filter. *Journal of medical systems* **42**(10), 197 (2018)
45. Sheela, C.J.J., Suganthi, G.: An efficient denoising of impulse noise from mri using adaptive switching modified decision based unsymmetric trimmed median filter. *Biomedical Signal Processing and Control* **55**, 101657 (2020)
46. HosseinKhani, Z., Karimi, N., Soroushmehr, S.M.R., Hajabdollahi, M., Samavi, S., Ward, K., Najarian, K.: Real-time removal of random value impulse noise in medical images. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 3916–3921 (2016). IEEE
47. HosseinKhani, Z., Hajabdollahi, M., Karimi, N., Najarian, K., Emami, A., Shirani, S., Samavi, S., Soroushmehr, S.M.R.: Real-time removal of impulse noise from mr images for radiosurgery applications. *International Journal of Circuit Theory and Applications* **47**(3), 406–426 (2019)
48. Jiang, J., Zhang, L., Yang, J.: Mixed noise removal by weighted encoding with sparse nonlocal regularization. *IEEE transactions on image processing* **23**(6), 2651–2662 (2014)
49. Huang, T., Dong, W., Xie, X., Shi, G., Bai, X.: Mixed noise removal via laplacian scale mixture modeling and nonlocal low-rank approximation. *IEEE Transactions on Image Processing* **26**(7), 3171–3186 (2017)
50. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. *arXiv preprint arXiv:1406.2661* (2014)
51. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017)
52. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: *Advances in Neural Information Processing Systems*, pp. 5767–5777 (2017)
53. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging* **3**(1), 47–57 (2016)
54. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., *et al*: Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690 (2017)
55. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*, pp. 694–711 (2016). Springer
56. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
57. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing* **21**(4), 1488–1499 (2011)
58. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, pp. 448–456 (2015). PMLR
59. Ioffe, S.: Batch renormalization: Towards reducing minibatch dependence in batch-normalized models. In: *Advances in Neural Information Processing Systems*, pp. 1945–1953 (2017)
60. Ye, R., Liu, F., Zhang, L.: 3d depthwise convolution: Reducing model parameters in 3d vision tasks. In: *Canadian Conference on Artificial Intelligence*, pp. 186–199 (2019). Springer
61. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803 (2018)
62. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082* (2019)
63. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* **26**(7), 3142–3155 (2017)
64. Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Pike, G.B., Evans, A.C.: Brainweb: Online interface to a 3d mri simulated brain database. In: *NeuroImage* (1997). Citeseer

65. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 International Conference on Computer Vision, pp. 479–486 (2011). IEEE
66. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
67. Aetesam, H., Maji, S.K.: Noise dependent training for deep parallel ensemble denoising in magnetic resonance images. *Biomedical Signal Processing and Control* **66**, 102405 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.