



Denoising of volumetric magnetic resonance imaging using multi-channel three-dimensional convolutional neural network with applications on fast spin echo acquisitions

Shutian Zhao^{1,2,3,4^}, Fan Xiao^{3,4}, James F. Griffith^{3,4}, Ruokun Li^{1,2}, Weitian Chen^{3,4}

¹Department of Radiology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; ²College of Health Science and Technology, Shanghai Jiao Tong University School of Medicine, Shanghai, China; ³Department of Imaging and Interventional Radiology, the Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong SAR, China; ⁴CUHK Lab of AI in Radiology (CLAIR), Hong Kong SAR, China

Contributions: (I) Conception and design: S Zhao, W Chen; (II) Administrative support: JF Griffith, R Li, W Chen; (III) Provision of study materials or patients: S Zhao, F Xiao; (IV) Collection and assembly of data: S Zhao, F Xiao; (V) Data analysis and interpretation: S Zhao, W Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Weitian Chen, PhD. Department of Imaging and Interventional Radiology, the Chinese University of Hong Kong, Prince of Wales Hospital, Room 15, LG/F, Sir Yue Kong Pao Centre for Cancer, 30-32 Ngan Shing Street, Shatin, NT, Hong Kong SAR, China; CUHK Lab of AI in Radiology (CLAIR), Hong Kong SAR, China. Email: wtchen@cuhk.edu.hk.

Background: Three-dimensional (3D) magnetic resonance imaging (MRI) can be acquired with a high spatial resolution with flexibility being reformatted into arbitrary planes, but at the cost of reduced signal-to-noise ratio. Deep-learning methods are promising for denoising in MRI. However, the existing 3D denoising convolutional neural networks (CNNs) rely on either a multi-channel two-dimensional (2D) network or a single-channel 3D network with limited ability to extract high dimensional features. We aim to develop a deep learning approach based on multi-channel 3D convolution to utilize inherent noise information embedded in multiple number of excitation (NEX) acquisition for denoising 3D fast spin echo (FSE) MRI.

Methods: A multi-channel 3D CNN is developed for denoising multi-NEX 3D FSE magnetic resonance (MR) images based on the feature extraction of 3D noise distributions embedded in 2-NEX 3D MRI. The performance of the proposed approach was compared to several state-of-the-art MRI denoising methods on both synthetic and real knee data using 2D and 3D metrics of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

Results: The proposed method achieved improved denoising performance compared to the current state-of-the-art denoising methods in both slice-by-slice 2D and volumetric 3D metrics of PSNR and SSIM.

Conclusions: A multi-channel 3D CNN is developed for denoising of multi-NEX 3D FSE MR images. The superior performance of the proposed multi-channel 3D CNN in denoising multi-NEX 3D MRI demonstrates its potential in tasks that require the extraction of high-dimensional features.

Keywords: Three-dimensional fast spin echo (3D FSE); denoising; number of excitation (NEX); multi-channel; three-dimensional convolutional neural network (3D CNN)

Submitted Mar 27, 2024. Accepted for publication Jul 01, 2024. Published online Jul 29, 2024.

doi: 10.21037/qims-24-625

View this article at: <https://dx.doi.org/10.21037/qims-24-625>

[^] ORCID: 0000-0002-0956-2785.

Introduction

Magnetic resonance imaging (MRI) is one of the most widely used noninvasive diagnostic modalities, providing superior soft tissue contrast. Compared to two-dimensional (2D) MRI, three-dimensional (3D) MRI can provide higher through-plane spatial resolution and reduced partial volume effect, making it particularly suited for visualizing complex anatomical structures. However, increasing spatial resolution leads to diminished signal-to-noise ratio (SNR) in MRI, causing challenges to various MRI applications. Thus, denoising plays a major role in enhancing the clinical utilities of MRI.

Denoising of MRI is often performed through 2D operations (1-4). Although proven efficient, these methods do not fully exploit the through-plane signal correlations inherent in 3D MRI (5). In contrast, 3D denoising methods can naturally utilize the 3D features of volumes of MR images, providing a more comprehensive representation of the denoising problem and better utilization of inter-slice signal correlations. Therefore, it is desirable to employ 3D denoising algorithms for denoising tasks of 3D MR images.

Many methods have been proposed for 3D MR image denoising, most of which are extensions of traditional 2D denoising methods. These methods typically utilize image characteristics such as self-similarity in the image domain or sparse representation in the transform domain. Typical algorithms include spatial domain method nonlocal means (NLM) (6,7), transform domain method discrete cosine transform (DCT) (8), sparse representation method singular value decomposition (SVD) (9,10), and local principal component analysis (PCA) method (11,12). Among the traditional denoising methods, block matching with 4D filtering (BM4D) (13), an extension of block-matching 3D collaborative filtering (BM3D) (2), is widely regarded as a state-of-the-art approach. BM4D is capable of directly handling Rician noise and exhibits excellent performance in denoising MR images through the application of a variance stabilizing transformation prior to the denoising process.

Recently, deep learning-based models have emerged as highly effective approaches for image denoising. When it comes to processing 3D MR volumes, there are two groups of deep learning approaches. One group of approaches involves stacking multiple slices along the channel axis of a 2D network. For example, methods like McDnCNN (14) and DABN (15) leverage adjacent slices to denoise the central slice of a 3D MR volume. Compared to 3D models, these multi-channel 2D models offer advantages of

memory efficiency and pre-trained 2D models. However, they have limitations in fully capturing the through-plane information, as they primarily learn weighted features from neighboring slices in the initial layer.

An alternative group of approaches involves the utilization of 3D convolutional neural networks (CNNs) for processing volumetric data. It enables learning in three dimensions through the employment of 3D operations, such as 3D convolution (16). With the significant advancements in the graphics processing unit (GPU) power over the past decade, 3D CNNs have gained popularity, and their applications in MRI denoising have yielded promising results. Several studies have demonstrated its potential in denoising MR images (17-21). For example, Manjón *et al.* proposed a 9-layer 3D CNN PRI-PB-CNN (17) for denoising both Gaussian and Rician noise. By combining the multi-layer perceptron (MLP) and CNN, the 5-layer 3D-WRN-VGG (18) and the Residual-MLP-CNN-Mixer (19) have been proposed for Rician noise. Additionally, under the parallel CNN structure with normal and dilated convolutions, both Gaussian-impulse noise and Rician noise of MR images can be effectively suppressed (20,21). These advancements underscore the superiority of 3D CNN in MRI denoising.

The aforementioned 3D CNN MRI denoising approaches primarily rely on a single-channel 3D input/output structure for spatial feature extraction. However, MRI acquisitions are often conducted with additional dimensions beyond spatial ones, thus encapsulating a richer set of information than mere spatial features. For example, multiple number of excitation (NEX) acquisition is a widespread technique to enhance the image SNR through time integration (22) especially in low-field MRI (23). Studies have revealed that the inter-NEX information embedded in multi-NEX images could also be valuable for denoising tasks (4). Therefore, it becomes crucial to devise a network that can efficiently process both 3D spatial and inter-NEX information.

To address this need, we propose a novel multi-channel 3D denoising CNN. This network aims to harness the benefits of both 3D spatial and inter-NEX information within a single architecture. In this study, we investigate the effectiveness of our proposed network in reducing noise in 3D fast spin echo (FSE) images. We demonstrated the superiority of our proposed network over the state-of-the-art 3D denoising methods both in terms of 2D and 3D evaluation metrics, offering a promising solution for processing 3D spatial and inter-NEX correlated MRI data.

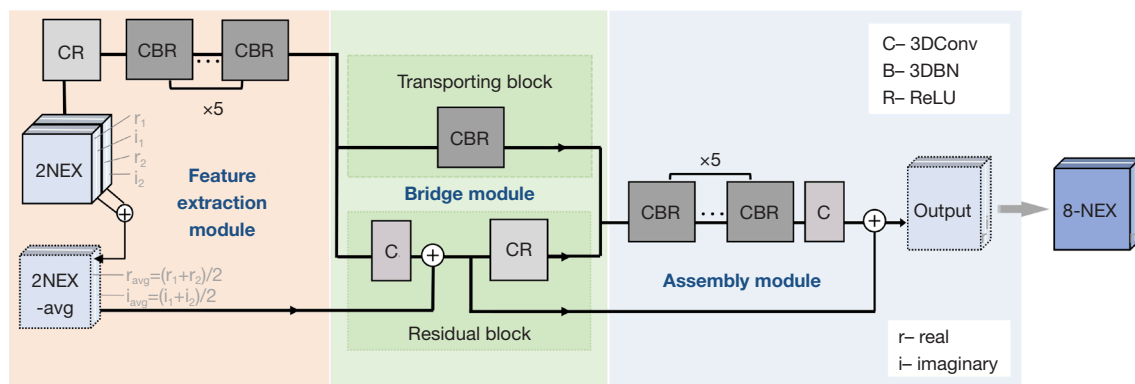


Figure 1 An illustration of the proposed multi-channel 3D denoising network. CR, 3DConv + ReLU; CBR, 3DConv + 3DBN + ReLU; NEX, number of excitation; avg, average; 3DConv, three-dimensional convolution; 3DBN, three-dimensional batch normalization; ReLU, rectified linear unit; 3D, three-dimensional.

The contributions of this work can be summarized as follows:

- (I) Given the spatially varying nature of noise, existing 3D denoising methods often fail to comprehensively address non-stationary noise distribution. This study fills this gap by exploring and demonstrating the capabilities of 3D CNNs in handling non-stationary 3D noise. Through experiments using both synthetic data and real data, we confidently demonstrate the adaptability and robustness of our model in enhancing the quality of MRI images in practical applications.
- (II) Our proposed method outperforms the existing state-of-the-art denoising methods in denoising 3D FSE MR images. We demonstrate this quantitatively using both 2D and 3D metrics.
- (III) We introduce a theoretical framework that establishes the alignment between the multi-channel 3D CNN and the multi-NEX 3D MR images. This alignment underscores the advantages of our approach, further validating its applicability in enhancing the quality of MRI scans.

Methods

Network and implementation details

Multi-NEX 3D MR images exhibit correlated features on two levels. Firstly, the 3D spatial information is directional in three dimensions, which is better extracted by 3D sliding convolution kernels than the 2D ones. Secondly, the inter-NEX images have no rank or directionality. For a neural

network, it is more appropriate to process this relationship along the channel dimension at the lowest computation cost. In this study, we employed a multi-channel 3D CNN to process both types of information in multi-NEX 3D MR images.

We extended our previously proposed denoising CNN (4) to 3D, with the filter number halved. The proposed 3D CNN has 14 layers, consisting of 3D convolution, 3D batch normalization (BN), and rectified linear unit (ReLU) (24). Each convolution has a filter size of $3 \times 3 \times 3$, stride 1, and padding 1. As shown in *Figure 1*, the model is composed of three modules: the feature extraction module, the bridge module, and the assembly module. To learn noise residuals, a two-step residual learning approach was employed over the parallel transporting and residual blocks. This structure enables the model to handle imbalanced input/output channels. In this experiment, we trained our network, denoted as 3D-Proposed, with an input channel of 4 to separately process the real and imaginary parts of each complex-valued NEX image. The whole network has about 1.2M trainable parameters. Our code is publicly available at <https://github.com/ShutianZ/Denoising3DFSE>.

In our experiments, we compared the proposed approach to BM4D (11), the 3D extension of DnCNN (3), 3D-EnsembleNet (20), and 3D-Parallel-RicianNet (21). BM4D supports blind-denoising of the Rician noise with two cascades of hard thresholding and Wiener filtering. DnCNN (3) is a thriving 2D denoiser proposed by Zhang *et al.*, with efficient and robust performance in MRI (25). In this work, we implemented 3D-DnCNN using 3D CNNs with a fourteen-layer architecture. 3D-EnsembleNet (20) is

proposed by Aetesam *et al.* which incorporates two parallel models using normal and dilated convolutions. 3D-Parallel-RicianNet (21) is a 19-layer 3D CNN with parallel residual learning architecture proposed by Wu *et al.* Among the above-mentioned 3D MRI denoising models, the input data format in 3D-Parallel-RicianNet (64×64×64 voxels)'s original settings is the closest to ours. Thus, it requires a minimum change in implementing this approach to process the data used in this study. Note that all the existing 3D MRI denoising CNNs, including the three implemented here, can only process the input and output with equal channels. Therefore, the input of these compared models is the real and imaginary parts of the averaged 2-NEX image in 2 separate channels, respectively.

We trained the networks with the Adam optimizer and the ReduceLROnPlateau monitor with an initial learning rate of 0.001, decaying by 0.2 when the loss stops decreasing for ten epochs. The batch size is eight. The model was optimized using the combination of L2 loss and 3D structural similarity index measure (SSIM) loss, as shown below:

$$\operatorname{argmin}_f \|f(I_{2\text{NEX}}) - I_{8\text{NEX}}\|_2^2 * (1 - 3\text{DSSIM}(f(I_{2\text{NEX}}), I_{8\text{NEX}})) \quad [1]$$

Here $I_{2\text{NEX}}$ is the input, and $I_{8\text{NEX}}$ is the high SNR images acquired using 8 NEX, which serves as the target in training. All experiments were implemented in Python 3.9.7 with Pytorch 1.10.0 on two NVIDIA (Santa Clara, CA, USA) RTX A6000 GPUs (48 GB).

Metrics

In our experiments, we implemented both slice-by-slice 2D and volumetric 3D metrics to evaluate the performance of 3D denoising methods. These metrics include the peak signal-to-noise ratio (PSNR), SSIM (26), and multiscale SSIM (MS-SSIM) (27). PSNR can be calculated in 2D and 3D. SSIM measures the perceived 2D image quality in luminance, contrast, and structure. MS-SSIM can be used to assess 3D volumetric data, and is calculated by:

$$\text{MS-SSIM}(f, g) = [l_M(f, g)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(f, g)]^{\beta_j} [s_j(f, g)]^{\gamma_j} \quad [2]$$

where the index M is the number of scales with a default setting of 5 in our experiment. Images are scaled ($M-1$) times by a downsampling factor 2 each time to incorporate image details at different resolutions. The luminance comparison is calculated only at scale M, denoted as $l_M(f, g)$. $c_j(f, g)$ and $s_j(f, g)$ refer to the contrast and structure comparison measures at scale j, respectively. α_M , β_j , and γ_j

are parameters to define the relative importance of three measures. The closer the SSIM or MS-SSIM values to 1, the more similar the two images are.

For clarity, we use the terms 2D PSNR and 2D SSIM to refer to the slice-by-slice 2D evaluations of PSNR and SSIM, respectively. Similarly, we use the terms 3D PSNR and 3D SSIM to refer to the volumetric measurements of PSNR and MS-SSIM, respectively. We used a paired sample *t*-test to compare the difference, with a P value of less than 0.001 indicating a significant difference after Bonferroni correction for multiple comparisons.

Dataset

In this work, we acquired multi-NEX 3D MR images with isotropic resolution for training and testing. These multi-NEX 3D images can be considered as 4D data with both 3D spatial features and 4th dimension along the NEX direction carrying inter-NEX information. Our datasets include 68 proton density-weighted 3D FSE knee MRIs. The datasets were collected using a 3D proton density-weighted VISTATM pulse sequence on a Philips Achieva TX 3.0T MRI (Philips Healthcare, Best, Netherlands) with an eight-channel receiver knee coil (Invivo, Gainesville, FL, USA). All MRI examinations were conducted under the approval of the Institutional Review Board. The MRI parameters were as follows: repetition time/echo time 900/33.6 ms; excitation flip angle 90; FOV 160×160×120 mm³; 150 slices with a 3D isotropic acquisition resolution 0.8×0.8×0.8 mm³; echo train length 42; sensitivity encoding (SENSE) acceleration factor 2×2 (AP × RL); and spectral attenuated inversion recovery (SPAIR) for fat suppression. Complex images were reconstructed for each NEX acquisition using the standard post-processing pipeline including the SENSE reconstruction provided by the vendor. These complex images were subsequently used for the denoising process in our study. Totally 8-NEX images were acquired for each subject with a total scan duration of 23.3 minutes. As it is assumed that the multi-NEX data can provide intact information about the inherent noise distribution, we used the first 2-NEX acquisitions in the 8-NEX acquisition as the low SNR 2-NEX input, and the high-SNR 8-NEX as the target to train the network. All voxels are interpolated to the same resolution of 0.714 mm³ with 168 slices of each dataset. We used the cubic voxel of 64×64×64 with a sliding stride of 32×32×32 for input. Outputs are 3D volumes instead of only the central slice. The output volume is generated with the same shape as the input for

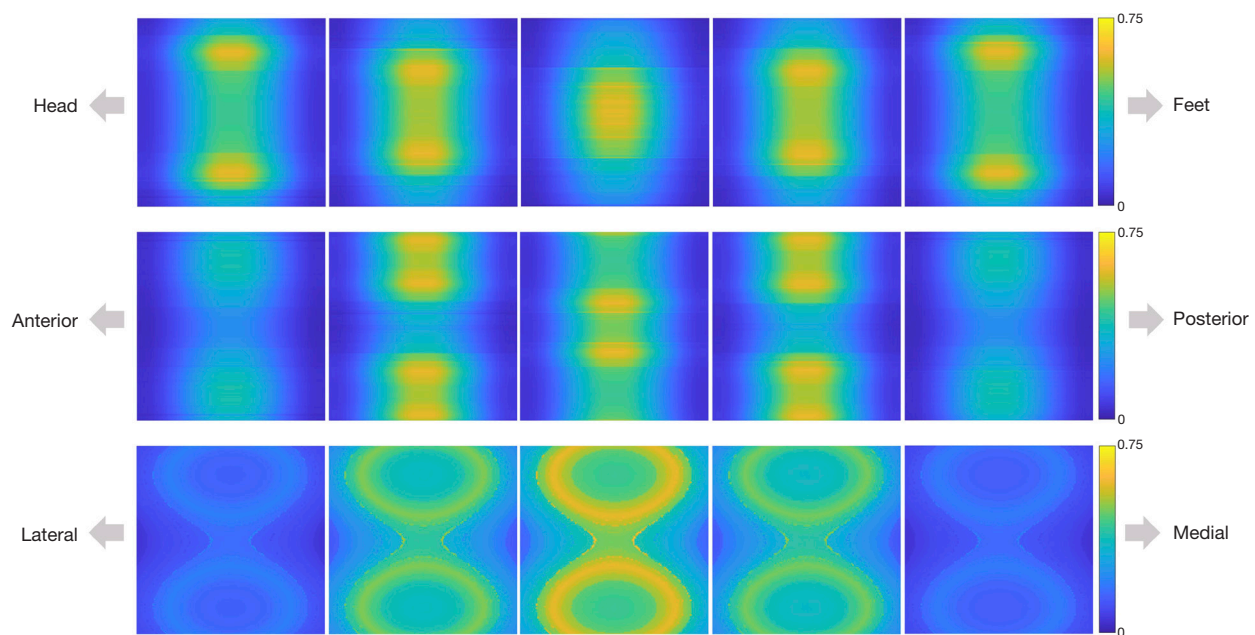


Figure 2 The synthetic 3D spatial-variant noise map considered in our experiments. From top to bottom row: the axial, coronal, and sagittal planes. Typical slices from the 3D noise map were arranged in the directions marked in the figure. 3D, three-dimensional.

computational efficiency. In total, 7,200 patches from 50 datasets were used for training. The other 18 3D datasets were cropped to a size of $168 \times 168 \times 168$ used for testing.

We also conducted experiments with synthetic noise. For the experiments with synthetic noise, we had ground truth to compare the performance of denoising methods. As the noise distribution in real MR images is spatially variant, we generated the 3D datasets incorporating Gaussian white noise with a non-stationary noise pattern to both the real and imaginary parts of the image. To approximate the real situation as much as possible, the noise level of the Gaussian noise is scaled by a factor between 0.8 and 1.2 among NEXs. The noise pattern we used was shown in three planes in *Figure 2*, generated using MATLAB R2021b (Mathworks, Natick, MA, USA). Note the noise distribution varies in 3D space.

Results

Results on synthetic data

Tables 1,2 present the performance of 3D denoising methods on synthetic data with non-stationary noise. The performance is measured using slice-by-slice 2D and volumetric 3D metrics, respectively. Compared to the noisy input, all approaches significantly improved 2D and 3D

PSNR and SSIM ($P < 0.001$). The deep learning methods outperformed the traditional denoising algorithm BM4D ($P < 0.001$). Among these methods, our proposed multi-channel model achieved the best performance in both metrics.

Figures 3-5 display typical slices of the denoising results and the corresponding differences to the ground truth image in the axial, coronal, and sagittal planes, respectively. Our proposed model achieved the best performance in all three planes. Compared to 3D-Parallel-RicianNet, 3D-EnsembleNet, and 3D-DnCNN, this model generated a better denoising result with an even reduced computational load.

Results on real data

Tables 3,4 present the models' performance on real data using both slice-by-slice 2D and volumetric 3D metrics. The four deep learning approaches achieved better denoising performance compared to both the noisy input ($P < 0.001$) and the conventional method BM4D ($P < 0.001$). Compared to 3D-Parallel-RicianNet, 3D-EnsembleNet, and 3D-DnCNN, the proposed 3D network shows improved performance in all metrics measured in 2D and 3D.

Figures 6-8 show typical slices in axial, coronal, and

Table 1 The performance of 3D denoising methods measured using 2D metrics on synthetic data with non-stationary noise

Plane	Metric	Input	BM4D	3D-Parallel-RicianNet	3D-DnCNN	3D-EnsembleNet	3D-Proposed
Axial	PSNR	28.44±0.23	33.19±1.53	38.52±1.14	38.52±1.15	38.52±0.94	38.74±1.03
	SSIM	0.5459±0.0404	0.7578±0.0443	0.9405±0.0061	0.9398±0.0064	0.9418±0.0060	0.9427±0.0060
Coronal	PSNR	28.54±0.23	33.37±1.60	39.43±1.47	39.47±1.52	39.51±1.35	39.74±1.47
	SSIM	0.5285±0.0394	0.7425±0.0444	0.9394±0.0064	0.9387±0.0067	0.9407±0.0063	0.9415±0.0062
Sagittal	PSNR	30.20±0.19	34.37±1.37	40.95±1.36	41.02±1.38	41.03±1.18	41.22±1.27
	SSIM	0.5491±0.0391	0.7570±0.0424	0.9402±0.0062	0.9395±0.0064	0.9415±0.0061	0.9422±0.0060

The PSNR and SSIM refer to the mean ± standard deviation of the PSNR and SSIM of all 2D slices. 3D, three-dimensional; 2D, two-dimensional; BM4D, block matching with 4D filtering; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

Table 2 The performance of 3D denoising methods measured using 3D metrics on synthetic data with non-stationary noise

Metric	Input	BM4D	3D-Parallel-RicianNet	3D-DnCNN	3D-EnsembleNet	3D-Proposed
3D PSNR	28.35±0.24	33.05±1.49	38.46±1.13	38.46±1.14	38.45±0.94	38.68±1.02
3D SSIM	0.9456±0.0075	0.9773±0.0041	0.9926±0.0008	0.9925±0.0009	0.9926±0.0007	0.9929±0.0007

The 3D PSNR and 3D SSIM are calculated volumetrically and presented as mean ± standard deviation. 3D, three-dimensional; BM4D, block matching with 4D filtering; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

sagittal planes, with annotated 2D PSNR and 2D SSIM values. Among these models, our approach 3D-Proposed achieved the best denoising performance in both metrics, which is consistent with the results obtained from the synthetic data. In regions such as the cartilage, 3D-Proposed not only effectively suppressed the noise but also preserved the intricate details.

Discussion

Clinical FSE sequences are often collected in 2D slices repeated in different planes for an overall visualization of anatomical structures, particularly in musculoskeletal applications. In contrast, 3D FSE with isotropic resolution can be acquired in a single acquisition and reformatted to arbitrary planes, providing the possibility of reducing the total scan time and flexibility of visualization. Additionally, 3D FSE can generate thinner slices, thereby reducing partial volume effects along slice direction, potentially improving sensitivity in detecting lesions. These advantages highlight the potential benefits of 3D FSE in clinical utilities. However, reduced SNR is a common drawback of 3D FSE compared to 2D FSE due to its high spatial resolution. To address this issue, we proposed a multi-channel 3D CNN and demonstrated its superior denoising performance in 3D

FSE. We believe that this multi-channel multi-dimensional network not only helps with denoising of 3D images, but also serves as a hint for processing MRI data in higher dimensions, such as quantitative MRI based on FSE (28,29). As there are few systematic studies of denoising of 3D MRI using multi-channel 3D CNN, below we provide further discussions of the underlying mechanism.

3D MRI naturally possesses redundant anatomy information along adjacent areas along both in-plane and through-plane directions. 3D MRI provides features that can be extracted in 3D volume, which fits well with 3D convolution. By applying 3D convolution, the image quality of 3D MRI can be boosted in both in-plane and through-plane directions by utilizing such redundancy. Therefore, 3D CNN can offer improved feature extraction to 3D MRI compared to 2D CNN.

In addition to 3D CNN, another key point in this experiment is the multi-channel design. In a standard multi-channel convolution, multi-channel kernels are employed to convolve with the multi-channel input or feature maps. The filtered outputs are then summed over the channels to yield a new feature map. This process is known as cross-correlation. It corresponds to the channel-wise summation of the convolution output where each input channel is convolved with an independent kernel. If these kernels

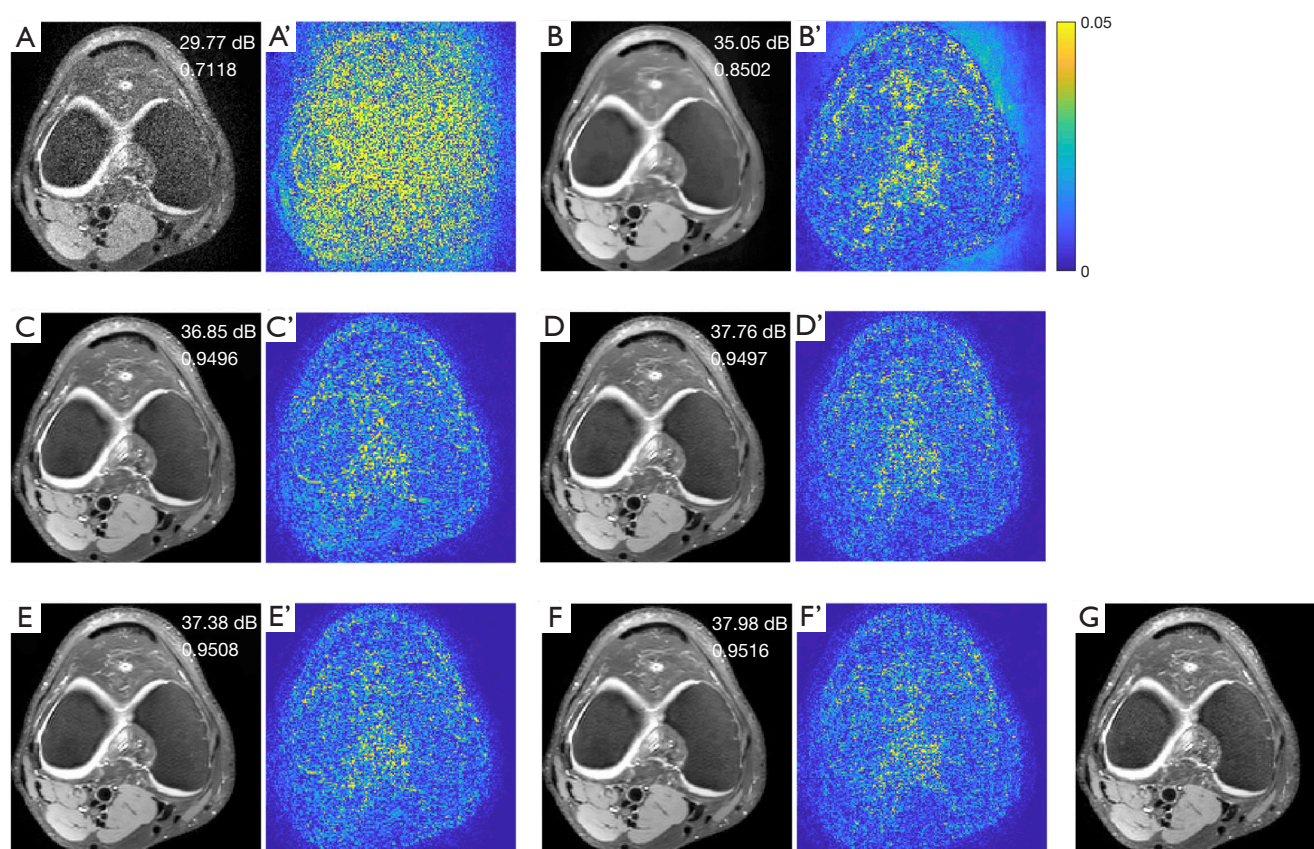


Figure 3 A typical axial slice of denoised results on synthetic 3D FSE data using different 3D denoising methods. Image (A) is the input with the non-stationary noise. Images (B-F) are the corresponding denoised images of (A), including (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (A'-F') refer to their difference to the ground truth (G). The 2D PSNR and SSIM values are displayed on the plots. 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

employ the same weights, the channel-wise summation can be considered as occurring prior to convolution, corresponding to directly filtering on a single-channel input that is channel-wise summed. However, in the proposed approach, the kernels in a multi-channel cross-correlation do not have the same weights and are independently updated towards a canonical expression of the fused feature map, tending to have different focuses. Therefore, a multi-channel input could support a more representative feature map than its channel-wise summated single-channel counterparts.

It is important to note that the number of model parameters increases with input channels, as well as the model complexity. Due to the complexity of processing 3D data, 3D networks naturally require more computing resources compared to 2D networks. Such increased

computational burden is typically noteworthy not only in the training phase but also in the testing phase, which may be unaffordable using the computer configurations in clinical environment. Therefore, it is important to develop a lightweight architecture for the 3D network. In this experiment, we introduced a 3D structure designed for multi-channel inputs with an asymmetric parallel bridge module that could efficiently extract comprehensive noise features with minimum computational memory usage. Table 5 shows a comparison of computer consumption of 3D models, including the forward/backward pass size, the number of trainable parameters, and the floating point of operations (FLOPs). Forward/backward pass size refers to the amount of data processed during the forward and backward passes in the training process. It is typically measured in terms of the amount of memory required to

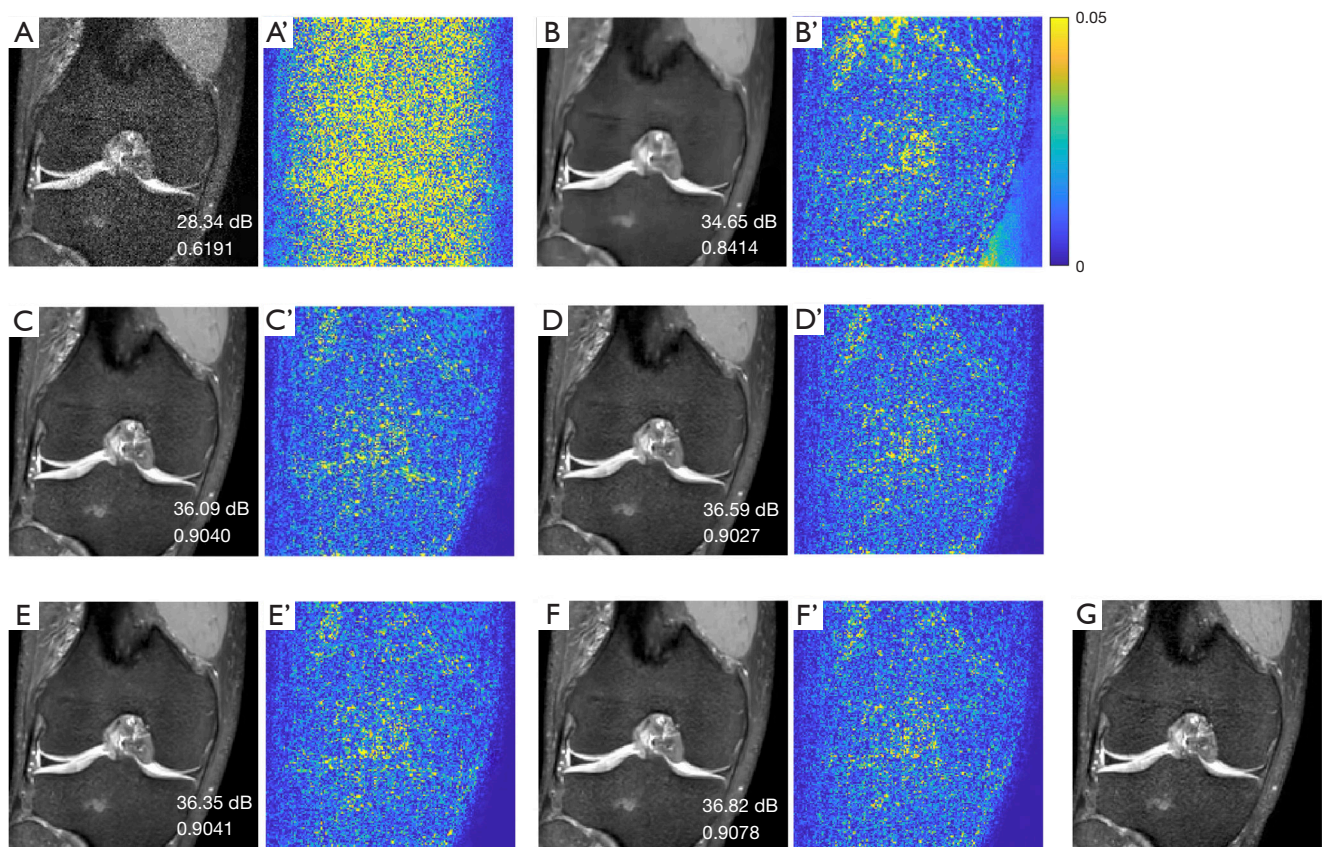


Figure 4 A typical coronal slice of denoised results on synthetic 3D FSE data using different 3D denoising methods. Image (A) is the input with the non-stationary noise. Images (B-F) are the corresponding denoised images of (A), including (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (A'-F') refer to their difference to the ground truth (G). The 2D PSNR and SSIM values are displayed on the plots. 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

store the intermediate results and gradients during the computation. Trainable parameters refer to the adjustable weights and biases in the model that are updated during the training process. FLOPs refer to the total number of floating-point operations required to run the model. A higher number of trainable parameters and FLOPs are generally associated with improved model performance, but they also signify the model's computational complexity. Among all the 3D models compared, our model stands out as the one with the smallest forward/backward pass size and the second-lowest FLOPs and trainable parameters. Our method is relatively efficient in terms of both computing and storage resources. With such lightweight architecture, our proposed network structure works well with the 2-NEX input data through its multi-channel design, providing efficient 3D MRI denoising.

Due to the challenge to collect high SNR *in vivo* data to train the neuro network, the existing 3D MRI denoising CNNs were trained using data with synthetic Gaussian or Rician noise (15-21). When applying these methods on real MRI data (17,20,21), their performance was measured either qualitatively (17,20) or compared to the pseudo-labels generated from pre-trained models (21) due to the lack of high SNR *in vivo* images serving as the target. Such evaluation criteria may be subjective. Therefore, it is important to explore the generalizability of 3D CNN with real noise and measure its performance quantitatively. In this work, the performance of the proposed multi-channel 3D CNN and other methods in denoising 3D MR images is quantitatively evaluated on high SNR real 3D FSE datasets acquired using 8 NEXs. Experiments show that our proposed model outperforms the state-of-the-art 3D MRI

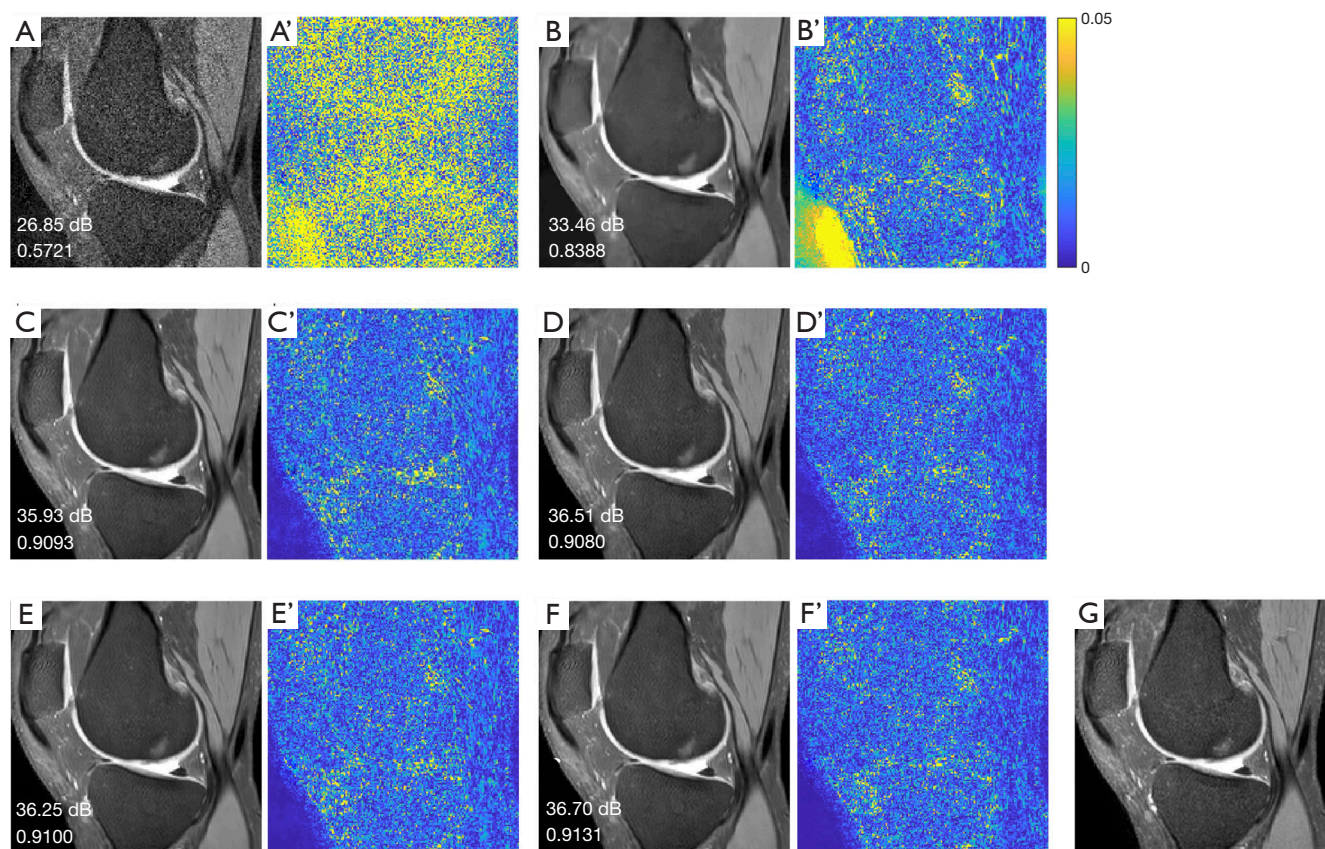


Figure 5 A typical sagittal slice of denoised results on synthetic 3D FSE data using different 3D denoising methods. Image (A) is the input with the non-stationary noise. Images (B-F) are the corresponding denoised images of (A), including (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (A'-F') refer to their difference to the ground truth (G). The 2D PSNR and SSIM values are displayed on the plots. 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

Table 3 The performance of 3D networks measured using 2D metrics on real data							
Plane	Metric	Input	BM4D	3D-Parallel-RicianNet	3D-DnCNN	3D-EnsembleNet	3D-Proposed
Axial	PSNR	34.89±2.10	35.75±1.84	37.78±1.92	38.07±2.12	38.14±1.97	38.31±2.43
	SSIM	0.9103±0.0258	0.9279±0.0165	0.9501±0.0133	0.9505±0.0146	0.9518±0.0133	0.9532±0.0144
Coronal	PSNR	36.01±2.53	36.64±2.18	38.82±2.43	39.17±2.70	39.20±2.50	39.44±2.98
	SSIM	0.9099±0.0261	0.9267±0.0169	0.9500±0.0135	0.9504±0.0147	0.9516±0.0136	0.9531±0.0147
Sagittal	PSNR	37.55±2.35	38.04±1.99	40.08±2.10	40.50±2.44	40.46±2.23	40.68±2.72
	SSIM	0.9101±0.0261	0.9266±0.0170	0.9501±0.0133	0.9506±0.0146	0.9517±0.0135	0.9532±0.0145

The PSNR and SSIM refer to the mean ± standard deviation of the PSNR and SSIM of all 2D slices. 3D, three-dimensional; 2D, two-dimensional; BM4D, block matching with 4D filtering; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 4 The performance of 3D networks measured using 3D metrics on real data

Metric	Input	BM4D	3D-Parallel-RicianNet	3D-DnCNN	3D-EnsembleNet	3D-Proposed
3D PSNR	34.77±2.12	35.58±1.88	37.63±1.96	37.92±2.15	37.98±2.00	38.16±2.45
3D SSIM	0.9852±0.0054	0.9882±0.0038	0.9916±0.0027	0.9918±0.0030	0.9921±0.0028	0.9924±0.0030

The 3D PSNR and 3D SSIM are calculated volumetrically and presented as mean ± standard deviation. 3D, three-dimensional; BM4D, block matching with 4D filtering; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

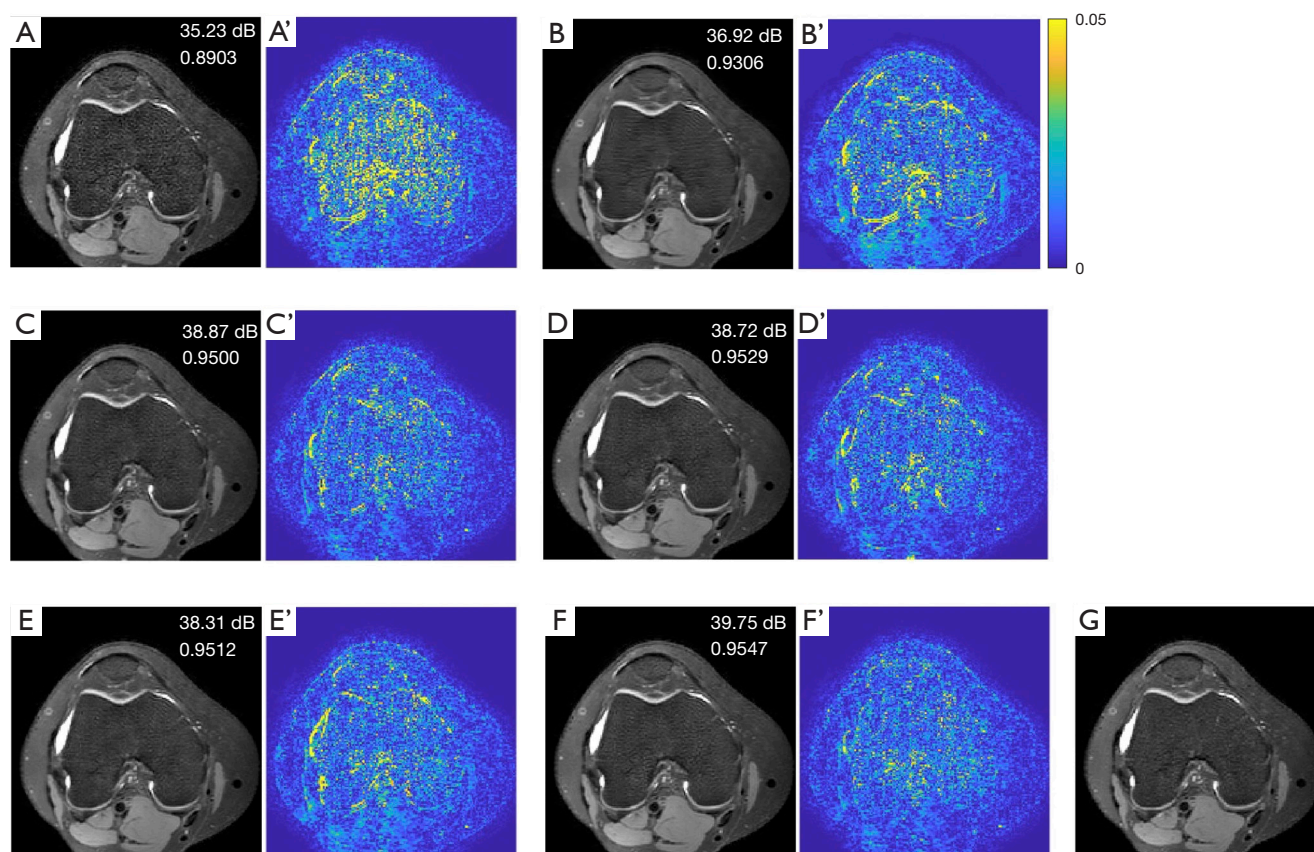


Figure 6 Representative denoised results and the corresponding differences to the target high-SNR image in an axial plane reformatted from a 3D FSE knee volume acquired in sagittal plane. (A) 2-NEX input, and denoised results obtained using (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (G) refers to the 8-NEX target high-SNR image. (A'-F') refers to the corresponding difference to the target image. The 2D PSNR and SSIM values are displayed on the plots. SNR, signal-to-noise ratio; 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

denoising methods in both 2D and 3D metrics.

Although the proposed method achieved satisfactory performance, there are several limitations of this work. First, the types of synthetic noise we employed were limited. We trained the network with only one nonstationary noise pattern. More noise patterns can be incorporated in future

work for comprehensive studies. Second, our proposed model is a supervised network that requires a training target. Note more training data is needed to train a 3D model compared to a 2D model to avoid overfitting. In this work, the 8-NEX 3D FSE volumes served as the target high SNR images, which takes over twenty minutes to

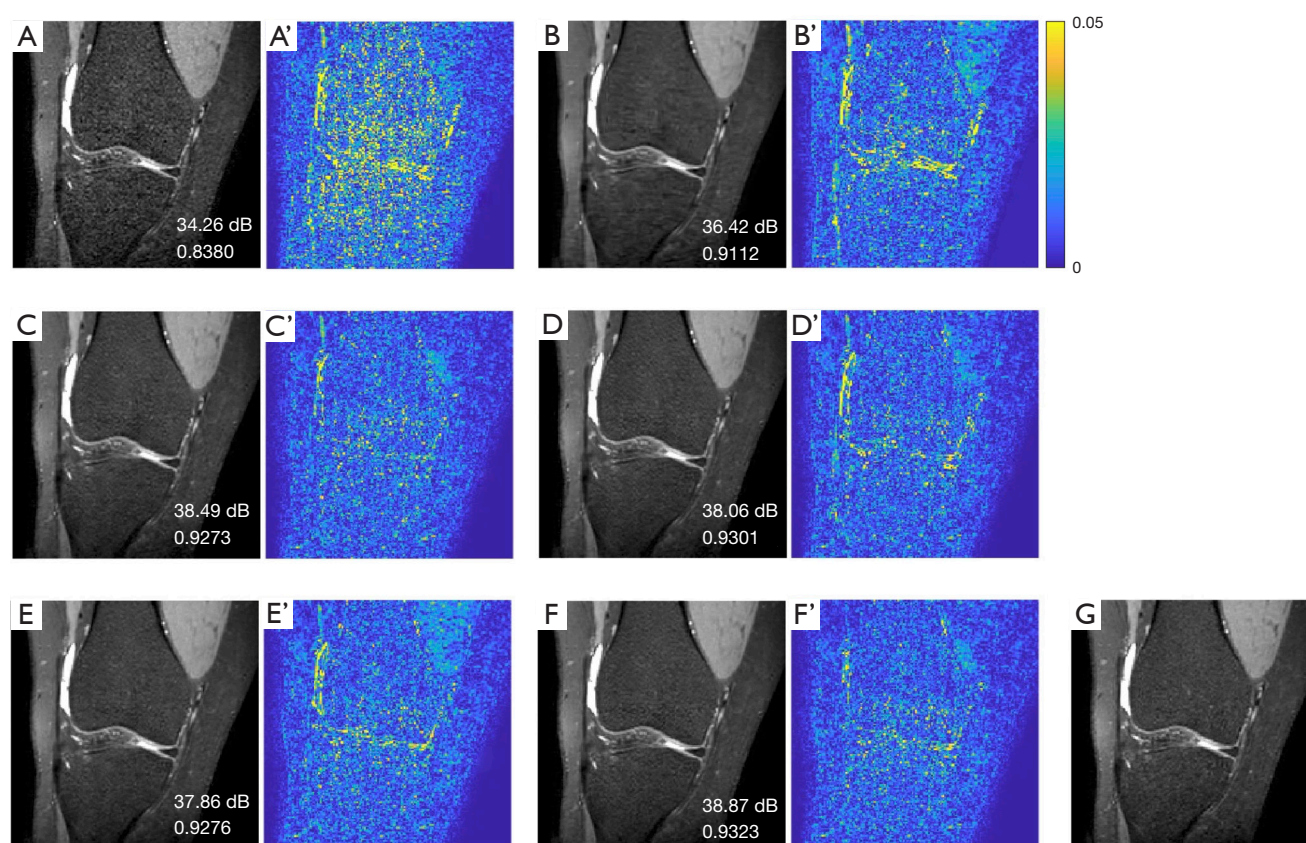


Figure 7 Representative denoised results and the corresponding differences to the target high-SNR image in a coronal plane reformatted from a 3D FSE knee volume acquired in sagittal plane. (A) 2-NEX input, and denoised results obtained using (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (G) refers to the 8-NEX target high-SNR image. (A'-F') refers to the corresponding difference to the target image. The 2D PSNR and SSIM values are displayed on the plots. SNR, signal-to-noise ratio; 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

acquire one dataset. Further work is needed to investigate self-supervised or unsupervised (30) approaches using the proposed multi-channel 3D CNN for denoising MRI. Third, the application of 3D CNN together with 3D datasets significantly increases the computational burden compared to 2D denoising. To build resource-efficient 3D CNNs, group convolutions (31) and depthwise separable convolutions (32) could be considered. If the computation burden is relieved, better performance can be expected by utilizing deeper models with more training data. Fourth, we demonstrated the denoising performance of the proposed method on 3D FSE knee MRI. Further work is needed to validate its performance on other anatomies or using other 3D MRI pulse sequences. Additionally, the generalization

capability of the network needs to be verified on datasets with varying acquisition parameters.

Conclusions

In this work, we proposed a multi-channel 3D CNN for denoising multi-NEX 3D FSE MR images and quantitatively measured its performance. Experiments on synthetic and real knee MRI data showed that our proposed multi-channel 3D CNN outperformed the state-of-the-art methods in denoising 3D FSE images. Our work demonstrated the potential of the proposed method in knee imaging and provided valuable guidance for extending its application in other anatomies.

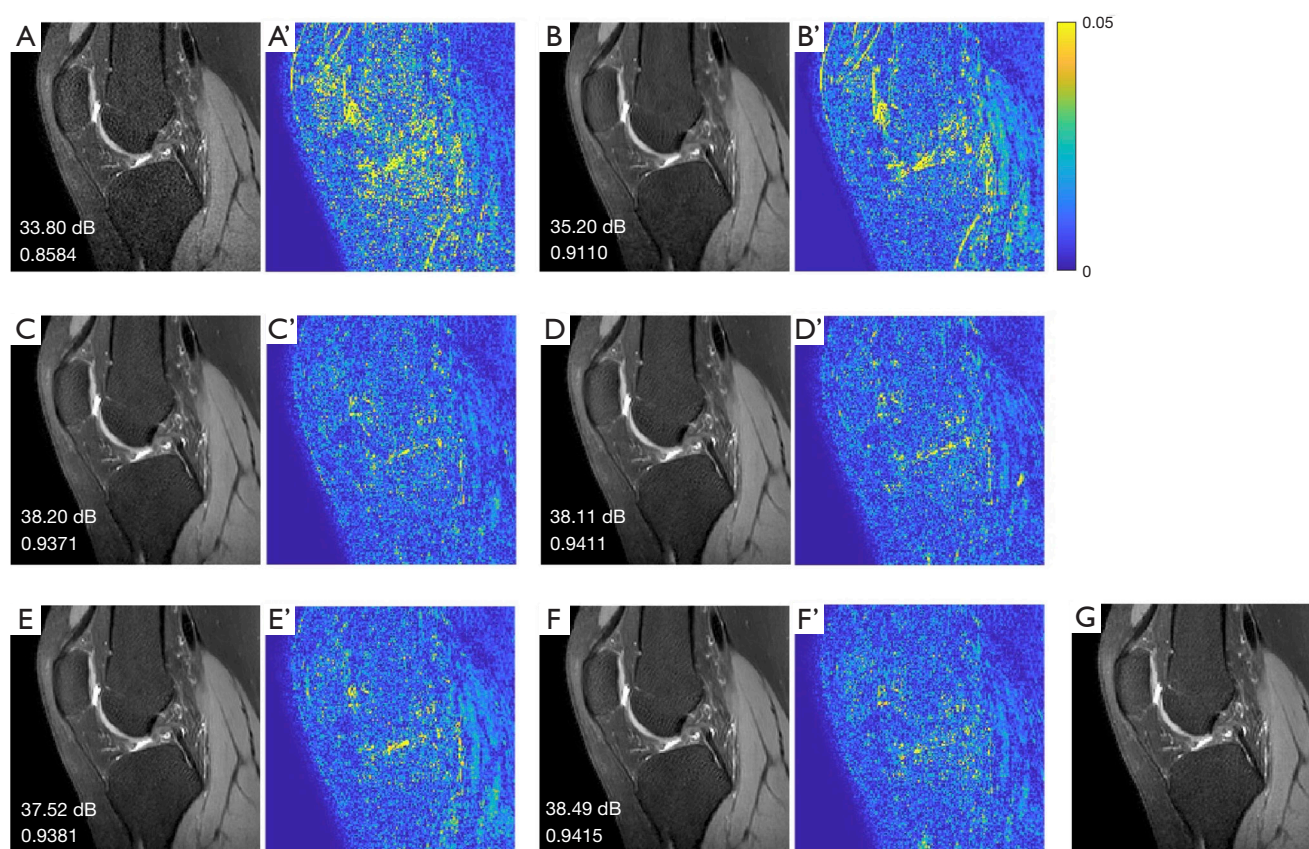


Figure 8 Representative denoised results and the corresponding differences to the target high-SNR image in a sagittal plane from a 3D FSE knee volume acquired in sagittal plane. (A) 2-NEX input, and denoised results obtained using (B) BM4D, (C) 3D-Parallel-RicianNet, (D) 3D-DnCNN, (E) 3D-EnsembleNet, and (F) our proposed 3D model. (G) refers to the 8-NEX target high-SNR image. (A'-F') refers to the corresponding difference to the target image. The 2D PSNR and SSIM values are displayed on the plots. SNR, signal-to-noise ratio; 3D, three-dimensional; FSE, fast spin echo; BM4D, block matching with 4D filtering; 2D, two-dimensional; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

Table 5 The comparison of the computer consumption of 3D deep learning models

Resource requirements	3D-Parallel-RicianNet	3D-DnCNN	3D-EnsembleNet	3D-Proposed
Forward/backward pass size (MB)	7,797.21	3,359.64	3,175.09	3,028.29
Trainable Parameters (M)	0.276	1.336	1.182	1.179
FLOPs (G)	73	351	310	309

The input size was set to (1, 2, 64, 64, 64). 3D, three-dimensional; FLOPs, floating points of operations.

Acknowledgments

Funding: This study was supported by a grant from the Innovation and Technology Commission of the Hong Kong SAR (Project No. MRP/001/18X).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-625/coif>). J.F.G.

serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. All authors report that this work is supported by a grant from the Innovation and Technology Commission of the Hong Kong SAR (Project No. MRP/001/18X). W.C. is a co-founder and a shareholder of Illuminatio Medical Technology Limited. S.Z. and W.C. are the inventors of a US patent SYSTEM AND METHOD FOR DENOISING IN MAGNETIC RESONANCE IMAGING (application number 18/128,193). The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Zhao Y, Yi Z, Xiao L, Lau V, Liu Y, Zhang Z, Guo H, Leong AT, Wu EX. Joint denoising of diffusion-weighted images via structured low-rank patch matrix approximation. *Magn Reson Med* 2022;88:2461-74.
2. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans Image Process* 2007;16:2080-95.
3. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans Image Process* 2017;26:3142-55.
4. Zhao S, Cahill DG, Li S, Xiao F, Blu T, Griffith JF, Chen W. Denoising of three-dimensional fast spin echo magnetic resonance images of knee joints using spatial-variant noise-relevant residual learning of convolution neural network. *Comput Biol Med* 2022;151:106295.
5. Zhou Y, Wang H, Liu C, Liao B, Li Y, Zhu Y, Hu Z, Liao J, Liang D. Recent advances in highly accelerated 3D MRI. *Phys Med Biol* 2023. doi: 10.1088/1361-6560/acc0cd.
6. Sahu S, Anand A, Singh AK, Agrawal AK, Singh MP. MRI de-noising using improved unbiased NLM filter. *J Ambient Intell Human Comput* 2023;14:10077-88.
7. Li S, Wang F, Gao S. New non-local mean methods for MRI denoising based on global self-similarity between values. *Comput Biol Med* 2024;174:108450.
8. Miri A, Sharifian S, Rashidi S, Ghods M. Medical image denoising based on 2D discrete cosine transform via ant colony optimization. *Optik* 2018;156:938-48.
9. Lv H, Wang R. Denoising 3D magnetic resonance images based on low-rank tensor approximation with adaptive multirank estimation. *IEEE Access* 2019;7:85995-6003.
10. Kang S, An DG, Ha H, Yang DH, Jang I, Song S. Denoising four-dimensional flow magnetic resonance imaging data using a split-and-overlap approach via singular value decomposition. *Phys Fluids* 2024;36:011906.
11. Fernandes FF, Olesen JL, Jespersen SN, Shemesh N. MP-PCA denoising of fMRI time-series data can lead to artificial activation "spreading". *Neuroimage* 2023;273:120118.
12. Olesen JL, Ianus A, Østergaard L, Shemesh N, Jespersen SN. Tensor denoising of multidimensional MRI data. *Magn Reson Med* 2023;89:1160-72.
13. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans Image Process* 2013;22:119-33.
14. 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. *Jpn J Radiol* 2018;36:566-74.
15. Xu Y, Han K, Zhou Y, Wu J, Xie X, Xiang W. Deep Adaptive Blending Network for 3D Magnetic Resonance Image Denoising. *IEEE J Biomed Health Inform* 2021;25:3321-31.
16. Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell* 2013;35:221-31.
17. Manjón JV, Coupe P. MRI denoising using deep learning. In: Bai W, Sanroma G, Wu G, Munsell B, Zhan Y, Coupé P. editors. *Patch-Based Techniques in Medical Imaging*. Patch-MI 2018. *Lecture Notes in Computer Science()*, Springer, 2018;11075:12-19.
18. Panda A, Naskar R, Rajbans S, Pal S. A 3D wide residual network with perceptual loss for brain MRI image denoising. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019:1-7.
19. Yang H, Zhang S, Han X, Zhao B, Ren Y, Sheng Y, Zhang XY. Denoising of 3D MR images using a voxel-

- wise hybrid residual MLP-CNN model to improve small lesion diagnostic confidence. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, Springer, 2022;13433:292-302.
20. Aetesam H, Maji SK. Noise dependent training for deep parallel ensemble denoising in magnetic resonance images. *Biomed Signal Proces* 2021;66:102405.
 21. Wu L, Hu S, Liu C. Denoising of 3D Brain MR Images with Parallel Residual Learning of Convolutional Neural Network Using Global and Local Feature Extraction. *Comput Intell Neurosci* 2021;2021:5577956.
 22. Tsuboyama T, Takei O, Okada A, Honda T, Kuriyama K. Comparison of HASTE with multiple signal averaging versus conventional turbo spin echo sequence: a new option for T2-weighted MRI of the female pelvis. *Eur Radiol* 2020;30:3245-53.
 23. Qiu Y, Dai K, Zhong S, Chen S, Wang C, Chen H, Frydman L, Zhang Z. Spatiotemporal encoding MRI in a portable low-field system. *Magn Reson Med* 2024;92:1011-21.
 24. Baker RR, Muthurangu V, Rega M, Walsh SB, Steeden JA. Rapid 2D (23)Na MRI of the calf using a denoising convolutional neural network. *Magn Reson Imaging* 2024;110:184-94.
 25. Sliusarenko D, Netreba A, Radchenko S. 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 2023:968-71.
 26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600-12.
 27. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, Pacific Grove, CA, USA, 2003;2:1398-402.
 28. Huang C, Qian Y, Yu SC, Hou J, Jiang B, Chan Q, Wong VW, Chu WC, Chen W. Uncertainty-aware self-supervised neural network for liverT(1ρ)mapping with relaxation constraint. *Phys Med Biol* 2022. doi: 10.1088/1361-6560/ac9e3e.
 29. Jordan CD, McWalter EJ, Monu UD, Watkins RD, Chen W, Bangerter NK, Hargreaves BA, Gold GE. Variability of CubeQuant T1ρ, quantitative DESS T2, and cones sodium MRI in knee cartilage. *Osteoarthritis Cartilage* 2014;22:1559-67.
 30. Li S, Zhao S, Zhang Y, Hong J, Chen W. Source-free unsupervised adaptive segmentation for knee joint MRI. *Biomed Signal Proces* 2024;92:106028.
 31. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.
 32. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017:1251-8.

Cite this article as: Zhao S, Xiao F, Griffith JF, Li R, Chen W. Denoising of volumetric magnetic resonance imaging using multi-channel three-dimensional convolutional neural network with applications on fast spin echo acquisitions. *Quant Imaging Med Surg* 2024;14(9):6517-6530. doi: 10.21037/qims-24-625