

Deep Learning-based MRI Reconstruction with Artificial Fourier Transform (AFT)-Net

Yanting Yang

Department of Biomedical Engineering
Columbia University
yy3189@columbia.edu

Jeffery Tian

Department of Computer Science
University of Maryland
jtian12@columbia.edu

Jia Guo

Department of Psychiatry
Columbia University
jg3400@columbia.edu

Abstract

The deep complex-valued neural network provides a powerful way to leverage complex number operations and representations, which has succeeded in several phase-based applications. However, most previously published networks have not fully accessed the impact of complex-valued networks in the frequency domain. Here, we introduced a unified complex-valued deep learning framework - artificial Fourier transform network (AFT-Net) - which can be readily used to solve the image inverse problems in domain-transform, especially for accelerated magnetic resonance imaging (MRI) reconstruction. During MRI acquisition, the data matrix is commonly obtained in what is known as k-space, an array of numbers representing spatial frequencies before any preprocessing and Fourier transform (FT) application. Conventional methods only accept magnitude images, while the proposed method takes raw k-space data in the frequency domain as inputs, allowing a mapping between the frequency domain and the image domain to be determined through cross-domain learning. We show that AFT-Net achieves superior accelerated MRI reconstruction and is comparable to existing approaches. Also, our approach can be applied to different tasks like denoised MRI reconstruction, different species with various contrasts, and magnetic field strengths. The AFT-Net presented here is a valuable preprocessing component for different preclinical studies which were acquired using customized MRI software without standard configuration.

1. Introduction

The domain shift from real coordinate space to complex coordinate space in the context of deep neural networks has uncovered the potential of utilizing the rich representational capacity of complex numbers and boosted the development of complex-valued neural architectures [4,6,21]. A similar but inverse domain shift is mirrored in the preprocessing of magnetic resonance imaging (MRI), where the raw data is acquired and stored in complex-valued k-space with each pixel representing the spatial frequency information in two or three dimensions of an object, and the image data is reconstructed by domain transforms (for example, Fourier, for fully-sampled Cartesian data [3]). Image reconstruction from k-space data to image data is an essential step in the preprocessing pipeline and lays the foundation for the overall image quality. It has been shown that k-space, as a low-dimensional feature space, can be leveraged in deep neural networks to determine the between-manifold mapping of domain transforms in low signal-to-noise settings [27].

Deep learning methods based on artificial neural networks with multiple layers between input and output can automate the analytical model-building process and outperform traditional data processing techniques and shallow machine learning models [11]. In recent years, many deep learning networks have been developed, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), transformers, generative adversarial networks (GANs), and diffusion models, which have revolutionized certain applications, including speech recognition [5], image classification [13] and medical image analysis [20].

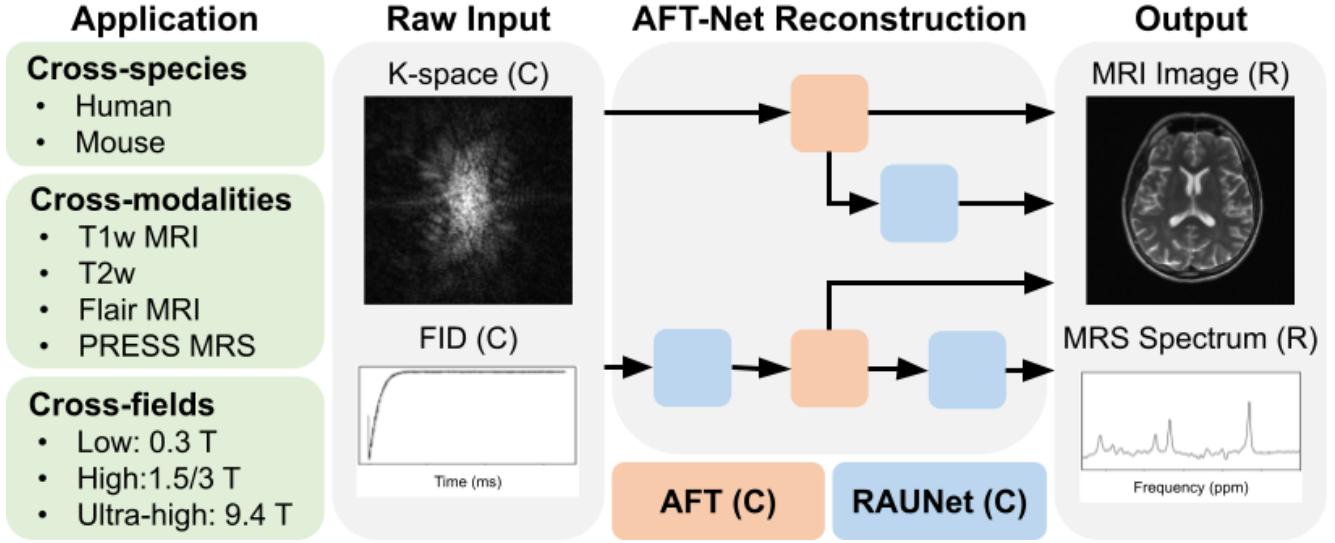


Figure 1. Schematics of general AFT-Net based deep-learning reconstruction. The application (1st column), acquired data (2nd column), reconstruction methods (3rd column), and reconstructed results (4th column) are shown for human T2w 1.5T MRI and human 3T PRESS MRS, respectively. C: complex-valued and R: real-valued.

With the accumulation of medical image data acquired from different sites, numerous open-access medical image datasets have become available for deep learning research, which allows the potential clinical uses of deep learning gradually come into practice. Medical images from various imaging systems contain rich structural and functional information, which is vital to consider for disease diagnosis and treatment. Once these medical images are obtained, analysis tasks, including lesion detection, classification, and diagnosis, are usually performed by radiologists [17]. However, due to the significant variation in pathology and the possible weariness of human experts [20], they can not guarantee that the diagnosis to be consistent and unbiased. The computer-aided diagnosis (CAD) has begun to benefit radiologists and save their lives from these time-consuming tasks, which usually require tedious interaction via software.

Conventionally, numerical and machine learning methods require human experts to select task-related features and establish models that highly represent the mapping between the input domains to the target domains [2]. With the improvement of computing hardware, it has become more feasible to use deep neural networks, which incorporate feature engineering extraction steps and simple preprocessing steps into the deep learning layers. With forward and backward propagation, the deep neural network can discover and determine the mapping of an input-target manifold through a self-taught learning [15].

Feature extraction engineering becomes problematic if the input or the target is not in the image domain, which means that the human experience is unreliable in such conditions. The domain conversion or data reconstruction from the sensor domain to the image domain is a critical task because it lays the foundation for all the following steps in the clinical workflow. How we reconstruct and process the raw data acquired from the imaging system sensors directly influences the overall image quality. The distortion caused by reconstruction can not be restored from conventional imaging processing methods due to information loss inevitably caused during the domain conversion.

To solve the problem mentioned above, we propose a unified complex-valued image reconstruction approach for MRI, which aims to get rid of any non-deep learning method in the workflow and incorporate data processing into deep learning frameworks. The framework we describe in this study is the artificial Fourier transform (AFT) which has the full functionality of a state-of-the-art Fourier transform (FT). It consists of fully adjustable parameters that can be fine-tuned through further training. We utilize AFT combined with deep complex networks [?] to design our AFT-Net for MRI reconstruction and preprocessing.

The main contributions of this work are shown in Figure 1 and can be summarized as follows:

- We introduce AFT-Net, a novel data-driven unified MRI reconstruction approach based on our AFT framework. Experimental results indicate that the AFT-Net achieves superior accelerated MRI reconstruction performance and comparable denoised MRI reconstruction.
- With extensive experiments, we demonstrate that AFT-Net can be applied to both human and mouse scans with different

modalities and different system fields.

- With simple modifications of the network structure, we demonstrate that AFT-Net can be applied to 1D magnetic resonance spectroscopy (MRS). Given free induction decay (FID) input with low signal-to-noise ratio (SNR), AFT-Net is able to determine the clean and high SNR spectrum.

1.1. Deep complex networks

Deep complex networks are architectures for deep learning based on complex-valued operations and representations. The complex-valued deep neural network leverages the richer representational capacity of complex numbers and facilitates noise-robust memory retrieval mechanisms [?]. Learning in the frequency domain also forces the neural network to implement learning-based frequency selection [25], which is not achievable with conventional building blocks and techniques. Complex-valued neural networks (CVNNs) [8, 9] have been proposed in many early works which focus on solving the basics of learning and have expanded widely in recent years. An extensive study on complex-valued CNNs has been conducted in recent years. Guberman [7] proposed a complex model as a generalization of the real-valued CNN model, which is significantly less vulnerable to overfitting. Oyallon *et al.* [18] introduces a deep scattering convolution network with complex wavelet filters over spatial and angular variables. Tygert *et al.* [22] presented a mathematical argument for complex-valued convolutional networks. Mathematical formation and practical implementation are discussed in [?, 7].

1.2. Medical image reconstruction

AUTOMAP [27] is a data-driven unified image reconstruction approach that learns the low-dimensional joint manifold of data represented in two domains. Their work is used to solve more generalized image reconstruction problems that find the best inverse function and sparse convolutional filters. Cole *et al.* [1] investigates complex-valued CNNs for accelerated MRI reconstruction and in several phase-based applications. They examine different complex-valued activation functions and compare complex-valued network with 2-channel real-valued network architectures. DeepcomplexMRI [23] is an end-to-end parallel imaging reconstruction framework based on complex residual CNN network that takes into account the correlation between the real and imaginary parts of MR images. CINENet [14] is a complex-valued 4D spatio-temporal CNN net with multi-coil data processing.

2. Background

2.1. Complex-valued neural network

The definition of the conventional real-valued neural network can be extended to the complex domain. We denote a complex operator as $W = W_{real} + iW_{imag}$, where W_{real} and W_{imag} are real-valued operators. The input complex vector can be represented as $x = x_{real} + ix_{imag}$. The output of complex operator W acting on x is derived by multiplication:

$$\begin{aligned} y &= W * x \\ &= (W_{real} * x_{real} - W_{imag} * x_{imag}) \\ &\quad + i(W_{imag} * x_{real} + W_{real} * x_{imag}). \end{aligned} \tag{1}$$

As the linear operator and convolution operator are distributive [?], we obtain:

$$\begin{aligned} \mathbb{C}\text{Linear}(z) &= (\text{Linear}_1(\Re(z)) - \text{Linear}_2(\Im(z))) \\ &\quad + i(\text{Linear}_2(\Re(z)) + \text{Linear}_1(\Im(z))) \end{aligned} \tag{2}$$

and

$$\begin{aligned} \mathbb{C}\text{Conv}(z) &= (\text{Conv}_1(\Re(z)) - \text{Conv}_2(\Im(z))) \\ &\quad + i(\text{Conv}_2(\Re(z)) + \text{Conv}_1(\Im(z))) \end{aligned} \tag{3}$$

where $z \in \mathbb{C}$ and we use subscripts 1 and 2 instead of *real* and *imag* to avoid misleading.

The complex version of the ReLU activation function we used in this study is simply applying separate ReLU on both the real and the imaginary part of the input as follows:

$$\mathbb{C}\text{ReLU}(z) = \text{ReLU}(\Re(z)) + i\text{ReLU}(\Im(z)) \tag{4}$$

which satisfies Cauchy–Riemann equations when both the real and the imaginary parts have the same sign or $\theta_z \in [0, \frac{1}{2}\pi] \cup [\pi, \frac{3}{2}\pi]$.

Normalization is a common technique widely used in deep learning to accelerate training and reduce statistical covariance shift [10]. This is also mirrored in the complex-valued neural network, where we want to ensure that both the real and the imaginary parts have equal variance. Extending the normalization equation to matrix notation we have:

$$\tilde{z} = V^{-\frac{1}{2}}(z - \mathbb{E}(z)) \quad (5)$$

where $x - \mathbb{E}(x)$ simply zero centers the real and the imaginary parts separately

$$z - \mathbb{E}(z) = \begin{bmatrix} \Re(z) - \text{Mean}(\Re(z)) \\ \Im(z) - \text{Mean}(\Im(z)) \end{bmatrix} \quad (6)$$

and V is the covariance matrix

$$\begin{aligned} V &= \begin{bmatrix} V_{rr} & V_{ri} \\ V_{ir} & V_{ii} \end{bmatrix} + \epsilon I \\ &= \begin{bmatrix} \text{Cov}(\Re(z), \Re(z)) & \text{Cov}(\Re(z), \Im(z)) \\ \text{Cov}(\Im(z), \Re(z)) & \text{Cov}(\Im(z), \Im(z)) \end{bmatrix} + \epsilon I. \end{aligned} \quad (7)$$

V is a 2×2 matrix, and the existence of the inverse square root is guaranteed by the addition of ϵI (Tikhonov regularization). Therefore, the solution of the inverse square root can be expressed analytically as

$$V = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \Rightarrow V^{-\frac{1}{2}} = \begin{bmatrix} (D+s)/d & -B/d \\ -C/d & (A+s)/d \end{bmatrix} \quad (8)$$

where $s = \sqrt{AD - BC}$, $t = \sqrt{A + D + 2s}$ and $d = st$.

The complex normalization is defined as

$$\text{Norm}(z) = \gamma \tilde{z} + \beta = \begin{bmatrix} \gamma_{rr} & \gamma_{ri} \\ \gamma_{ri} & \gamma_{ii} \end{bmatrix} \tilde{z} + \begin{bmatrix} \Re(\beta) \\ \Im(\beta) \end{bmatrix}. \quad (9)$$

Considering the limitation of GPU RAM and large memory consumption of complex-valued networks, we use the complex group normalization in our framework to avoid possible inaccurate batch statistics estimation caused by a small batch size.

3. Methods

3.1. Artificial Fourier Transform

Since 2D discrete Fourier transform (DFT) is a linear operation and can be represented by two successive 1D DFTs as

$$\begin{aligned} \mathcal{F}_{x,y}\{f(x,y)\} &= \mathcal{F}_x\{\mathcal{F}_y\{f(x,y)\}\} \\ &= \mathcal{F}_y\{\mathcal{F}_x\{f(x,y)\}\}, \end{aligned} \quad (10)$$

each dimension of 2D DFT can be modeled as a single-hidden-layer neural network with a linear activation function [16]. We further implement this idea with complex-valued neural networks and proposed AFT with two repeated blocks. Each block consists of two complex linear layers followed by a transpose operation as shown in Figure 2.

From the definition of the discrete Fourier transform of a sequence of N complex numbers which can be represented in the real and imaginary parts as

$$Z_k = \sum_{n=0}^{N-1} z_n \left[\cos\left(\frac{2\pi}{N}kn\right) - i \sin\left(\frac{2\pi}{N}kn\right) \right], \quad (11)$$

rewrite Equation (11) as

$$Z_k = W_{real}z + iW_{imag}z, \quad (12)$$

where W_{real} and W_{imag} are the real and the imaginary coefficients. Use matrix notation to represent real and imaginary parts of the DFT operation. We have:

$$\begin{bmatrix} \Re(Z_k) \\ \Im(Z_k) \end{bmatrix} = \begin{bmatrix} W_{real} & -W_{imag} \\ W_{imag} & W_{real} \end{bmatrix} \begin{bmatrix} \Re(z) \\ \Im(z) \end{bmatrix}. \quad (13)$$

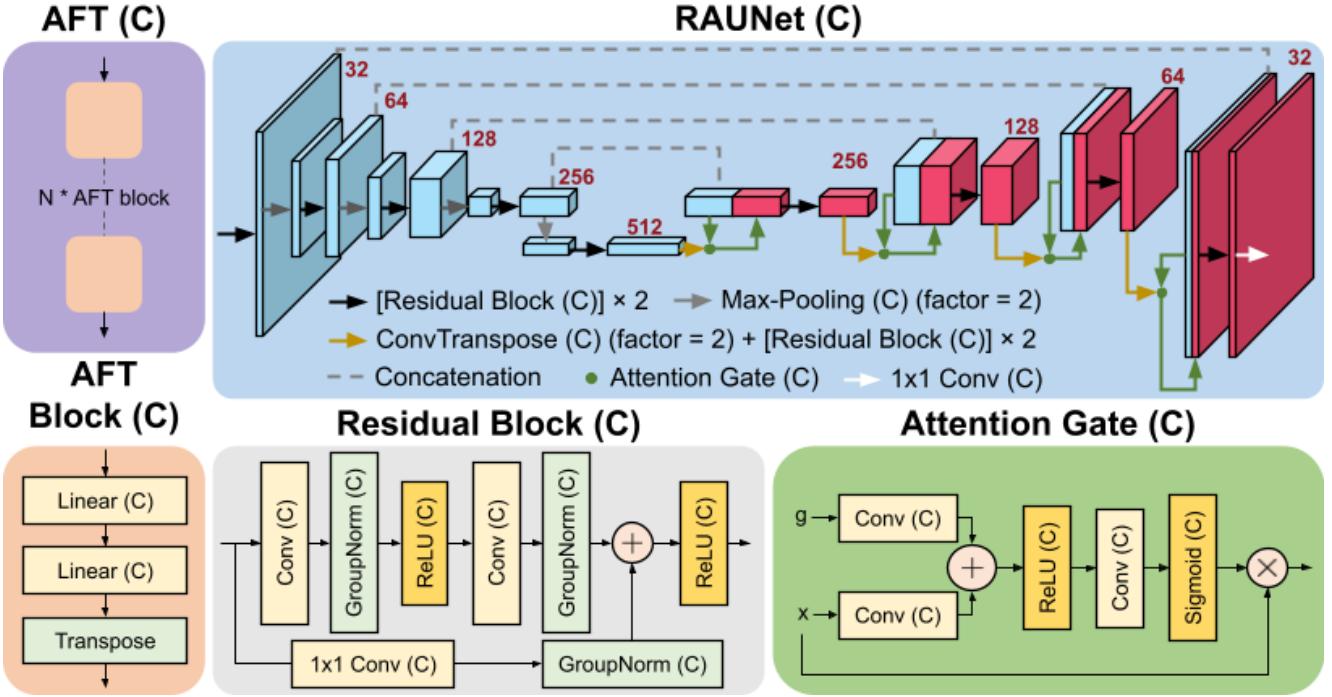


Figure 2. The structure of an N -dimensional AFT-Net (N can be 1, 2, or 3). Components include the complex-valued AFT block, the complex-valued residual attention UNet, the complex-valued residual block, and the complex-valued attention gate. All convolutional layers have a kernel size of 3, except those pointed out specifically. C: complex-valued. Red numbers indicate the number of channels produced by each layer.

Compare Equation (13) with Equation (1), a multi-layer complex-valued neural network with linear activation function can represent 1D DFT with appropriate weights. We use AFT_N to denote the complex-valued Fourier transform deep learning block on the input vector with N elements. The Fourier transform of the input data with dimension $H \times W$ can be represented as

$$Z = AFT_H(AFT_W(z^\top)^\top). \quad (14)$$

3.2. Network framework

The network structure and general workflow are shown in Figure 2 and Figure 3. We apply our AFT to the multi-coil k-space data acquired directly from the scanner for the reconstruction task. The target is derived from the inverse fast Fourier transform on the input data. The AFT does not compress the coil channel so that the input and output shapes/sizes are the same. The network performance is evaluated within magnitude images obtained by the Fourier application and coil compression. For the reconstruction plus denoising task, we combine our AFT with an entirely complex U-Net [24], which extracts higher features in the k-space and/or image domain and forces the network to represent sparsely in those domains. Multiple network architectures are evaluated to verify the effectiveness of both AFT and CUNet in different domains. We refer each of them to **AFT**, **AFT-Net (I)**, **AFT-Net (K)**, and **AFT-Net (KI)**, respectively as shown in Figure 3. We first train an AFT-only network to see if, without a non-linear activation function, the AFT can remove noise and enhance quality. Then a network with the AFT followed by the CUNet is trained to simulate a typical deep learning workflow where conventional numerical methods are used to preprocess the image, and CNNs are utilized to map the input domain to the target domain. We also evaluate the network with CUNet first implemented directly on the k-space domain. Given that each position in k-space contains the information of the whole image, CNNs implemented in k-space can leverage the complete information of all space, even if they have a fixed field of view. Finally, a CUNet-AFT-CUNet structure is evaluated with the first CUNet extracts k-space domain features and the second CUNet extracts image domain features.

The architecture of the CUNet presented in Figure 2 is generally based on the residual attention U-Net but with all the real-valued components replaced by complex-valued components as shown in Figure 2, including complex convolutional layers and complex ReLU layers introduced in Section 2.1. We further optimize the network for smaller batch sizes by replacing

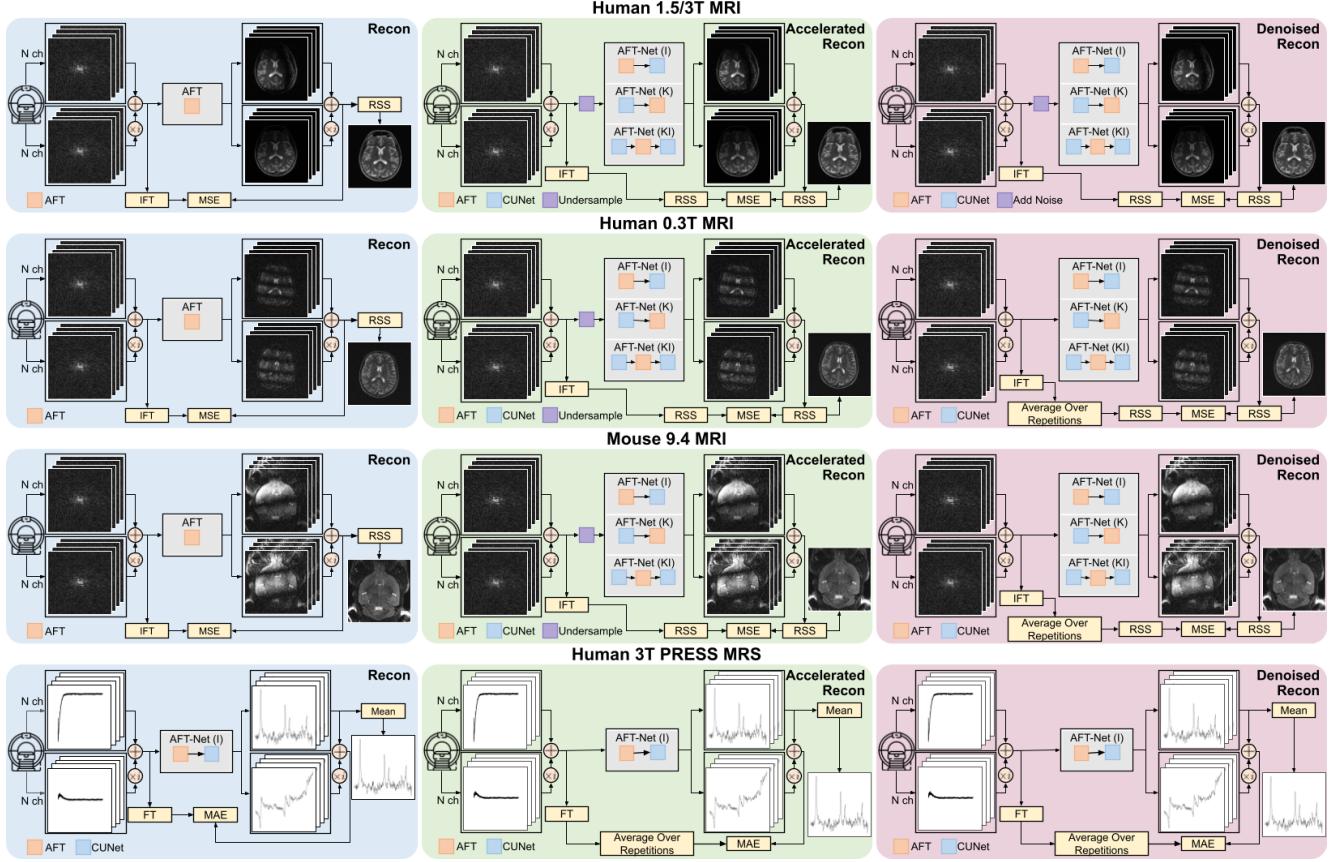


Figure 3. The workflows of experiments on each dataset.

batch normalization with group normalization. Other complex components are implemented in the same way. For example, the complex transposed convolution operator can be mirrored from Equation (3), complex sigmoid is applied like complex ReLU, and complex max pooling is almost the same as the real-valued version with indices inferred from absolute values.

3.3. Loss function

In the context of image reconstruction and processing, the impact of the loss function is vital if the final results are to be evaluated by human observers. One common and safe choice is ℓ_2 loss which works under the assumption of white Gaussian noise. For training AFT for MRI reconstruction, the loss value is determined in the frequency domain as

$$\mathcal{L}^{recon} = \mathcal{L}^{\ell_2}(\Re(x), \Re(y)) + \mathcal{L}^{\ell_2}(\Im(x), \Im(y)) \quad (15)$$

so that both real and imaginary outputs are optimized to match the conventional Fourier transformation. For training AFT-Net for accelerated MRI reconstruction, we only want to minimize the error of magnitude images. Therefore, the loss value for accelerated MRI reconstruction is determined in the image domain after coil combination. The root-sum-of-squares (RSS) approach [19] is applied to complex-valued output from the model to generate to optimal, unbiased estimate of magnitude image which is used for loss calculation.

3.4. Experimental data

The main experiments, including the training of the AFT-Net, were performed with the fastMRI brain dataset [26], comprising raw k-space data obtained on 3 and 1.5 Tesla magnets. We train the network with 4-channel 640×320 size axial T1 weighted and T2 weighted images. The inputs were normalized by the max intensity value of magnitude images accordingly.

Among 993 volumes of data, we randomly split them into training, validation and testing set, with 794, 99 and 100 volumes respectively. This results in 12622, 1582 and 1598 slices of training, validation and testing data.

3.5. Implementation details

We construct a batch size of 1 and optimize the network using the ADAM [12] optimizer. The initial learning rate is set to 10^{-3} , and we used a learning rate scheduler based on the SSIM in the validation set. When the metric has stopped improving, the learning rate is reduced by a factor of $\sqrt{10}$. We set patience to 2 and the lower bound of the learning rate to 10^{-6} . The training will stop early once learning stagnates and the learning rate reaches the lower bound. All experiments are done using PyTorch 1.11.0 and a Quadro RTX 6000 GPU.

3.6. Measurement of Reconstruction Quality

We adopt Pearson correlation coefficient (PCC), Spearman’s rank correlation coefficient (SCC), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) as evaluation metrics for quantitative comparison.

4. Results

4.1. Human normal-field MRI study

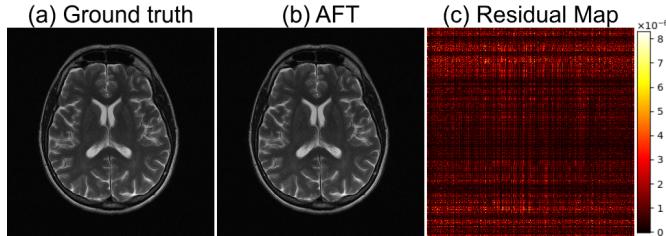


Figure 4. Human 1.5/3T MRI reconstruction results. (a) Ground truth, (b) proposed method, (c) difference magnitude of (a) and (b) (in Hot colormap).

First, we show the results of human 1.5/3T MRI reconstruction using raw fully-sampled fastMRI kspace data in Fig. 5. All the images shown here and in the following sections are cropped so that the anti-aliasing placed outside the field of view (FOV) in phase-encoding directions is removed. The ground truth image is derived by applying conventional Fourier transformation to the kspace data. It can be seen that the ground truth image obtained from FT is identical to the AFT prediction, which human observers can not distinguish. The results adhere to the mathematical description we discussed in Section 2.1. The residual map (pixel-wise difference between the ground truth image and the AFT prediction) shows that no brain structural information is presented. The grid-like remaining error is mainly caused by precision loss during floating-point calculation in matrix multiplication.

In Fig. 5, we show the results of human 1.5/3T accelerated reconstruction using under-sampled fastMRI kspace data. In the first row, we see the reconstructions from 1D 4x equal-spaced sampling, in which 8% of low-frequency columns are retained. Here, we compare different AFT-Net structures with zero filling method. AFT-Net (KI) performs outstanding reconstruction, where less structural difference can be seen from the residual map in the second row. The third row shows zoomed-in areas of both images and residual maps. AFT-Net (I) produces more blurry reconstruction which loses the structural details. Reconstruction through AFT-Net (K) induces foggy artifact, which is reflected in terms of SSIM. Fig. 6 shows the accelerated reconstruction results by comparing AFT-Net (I, K and KI) and zero filling in terms of SSIM across acceleration rates 2x, 4x, 8x and 16x. The performance of zero filling drops linearly as the acceleration rate increases while the AFT-Net methods are more robust to the acceleration scale. The t-test between each AFT-Net structure indicates that the AFT-Net (KI) overcomes all other AFT-Net structures significantly. The results of AFT-Net on different acquisition types and system field strength in Fig. 7 demonstrate that AFT-Net is robust to contrast difference and image quality.

Next, we illustrate the results of human 1.5/3T denoised reconstruction using fastMRI kspace data with added Gaussian noise in Fig. 8. Unlike the results of accelerated reconstruction, AFT-Net (I) performs the best across all three proposed AFT-Net structures, which can be proved from the t-test results in Fig. 9. It is also demonstrated in Fig. 9 that AFT-Net (KI) only shows comparable performance against AFT-Net (K) when the noise scale is 0.02 and underperforms other AFT-Net structures in noise scale 0.005 and 0.01, indicating that increasing the depth of AFT-Net does not necessarily increase the overall performance especially for denoised reconstruction task. The second row shows the pixel-wise difference between AFT-Net output and noiseless ground truth. It can be indicated that the noise in the background is attenuated significantly. Although the brain structure can be seen from the residual map, the zoomed-in version of the image shows that the AFT-Net reconstruction preserves the anatomy structure. The results of AFT-Net on different acquisition types and system field strength in Fig. 7 also demonstrate the generality of AFT-Net against different imaging modalities.

A comprehensive comparison of quantitative metrics on the test set is provided in Tab. 1, Tab. 2 and Tab. 3 for reconstruction, accelerated reconstruction and denoised reconstruction accordingly. AFT-Net (KI) significantly outperforms other AFT-Net structures on all the different acceleration rates. On all the different noise scales, AFT-Net (I) performs significantly better than other AFT-Net structures. More detailed quantitative metrics of human 1.5/3T MRI accelerated reconstruction and denoised reconstruction results are provided in the appendix, which is grouped by image contrast and system field strength.

It is worth mentioning that although AFT-Net (K) does not outperform other AFT-Net structures in both accelerated reconstruction and denoised reconstruction tasks, it demonstrates the ability to learn in a sparse frequency domain and its sparse representations with a complex-valued convolutional network.

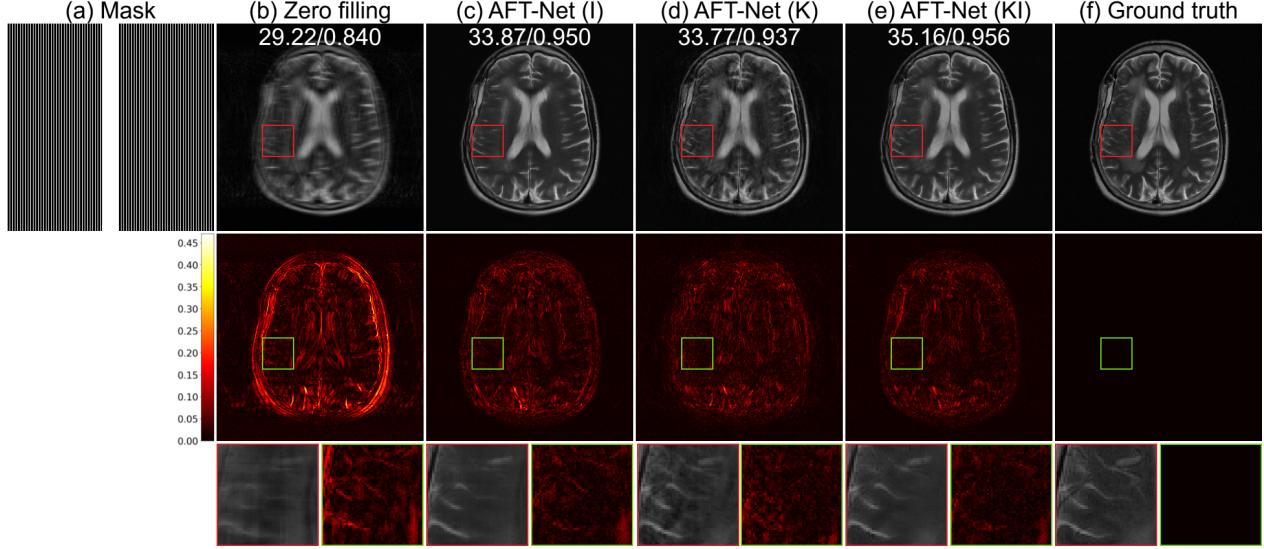


Figure 5. Human 1.5/3T MRI accelerated reconstruction results. (a) The sub-sampling mask used to generate the under-sampled image, (b) zero filling, (c)-(e) proposed methods, and (f) ground truth. 1st row: 1D 4x equal-spaced sampling (8% of low-frequency columns are retained), 2nd row: difference magnitude against (f) (in Hot colormap), 3rd row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

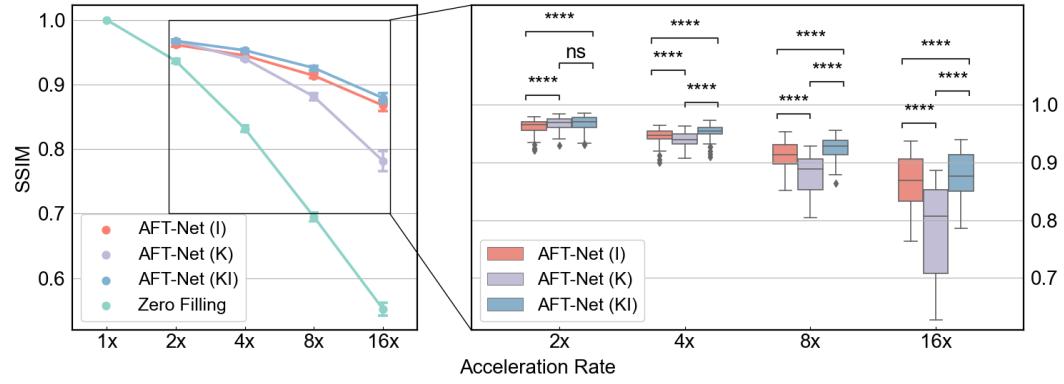


Figure 6. Results of human 1.5/3T MRI accelerated reconstruction by comparing AFT-Net (K Model, I Model, and KI Model) and Zero Filling in terms of SSIM for acceleration rates 2x, 4x, 8x, and 16x. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

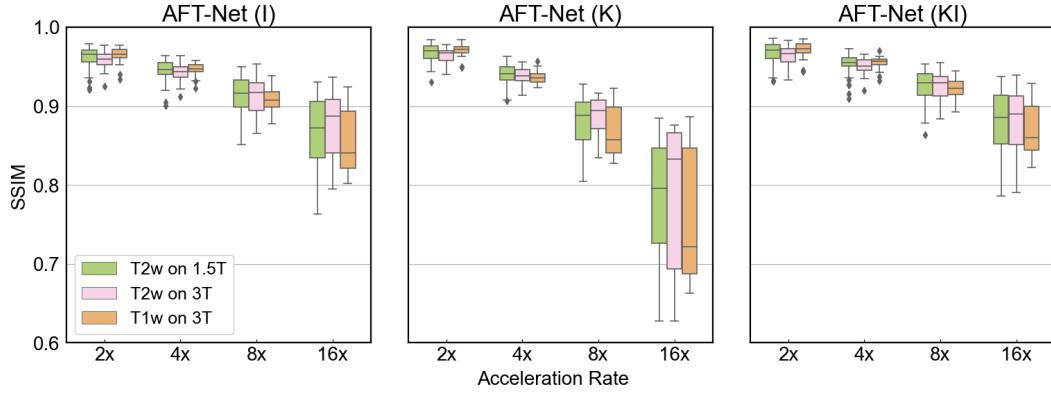


Figure 7. Results of human 1.5/3T MRI accelerated reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM on different acquisition types and system field strength for acceleration rates 2x, 4x, 8x, and 16x.

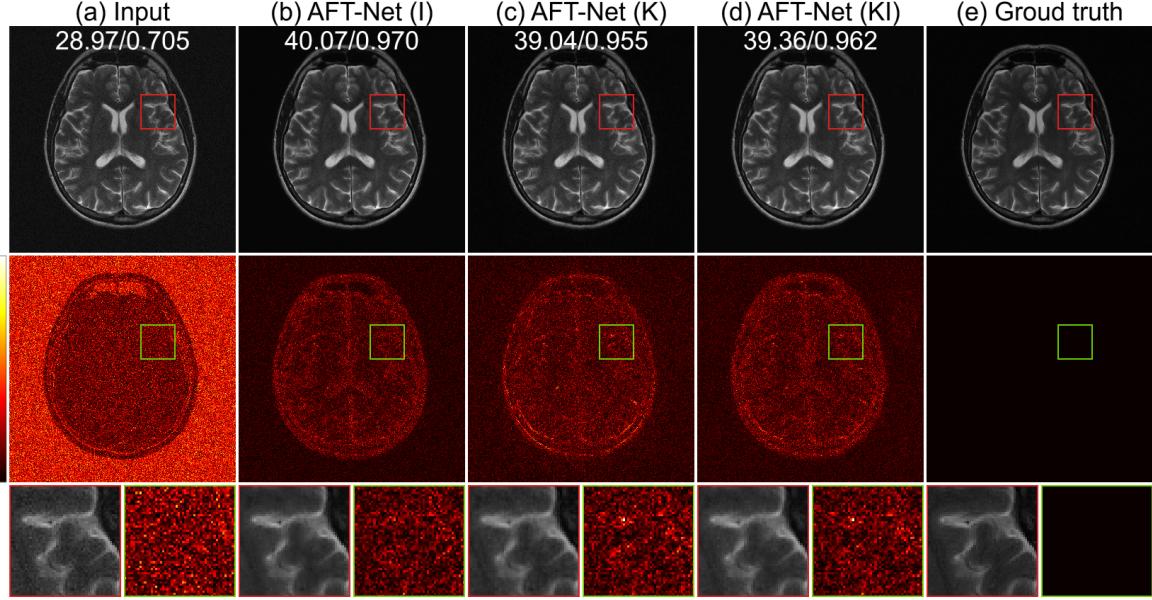


Figure 8. Human 1.5/3T MRI denoised reconstruction results. (a) Input, (b)-(d) proposed methods, and (e) ground truth. 1st row: Randomly added Gaussian noise in k-space (scale = 0.02), 2nd row: difference magnitude against (e) (in Hot colormap), 3rd row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

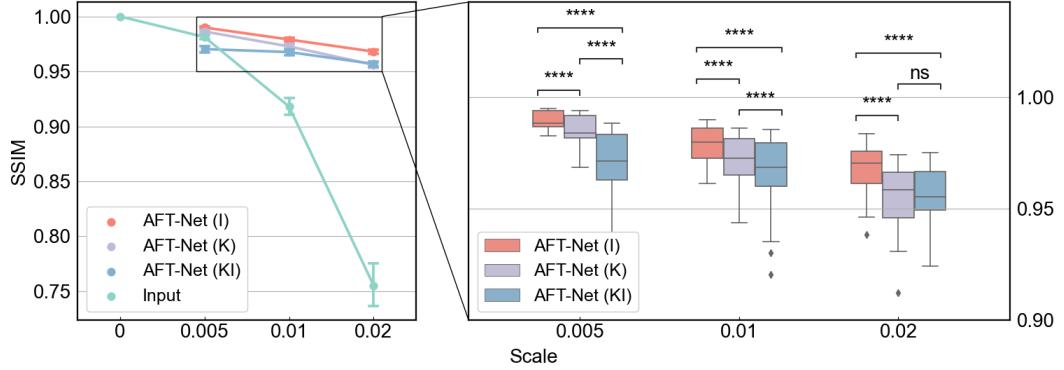


Figure 9. Results of human 1.5/3T MRI denoised reconstruction by comparing AFT-Net (K Model, I Model and KI Model) and input in terms of SSIM for noise scales 0.005, 0.01 and 0.02. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

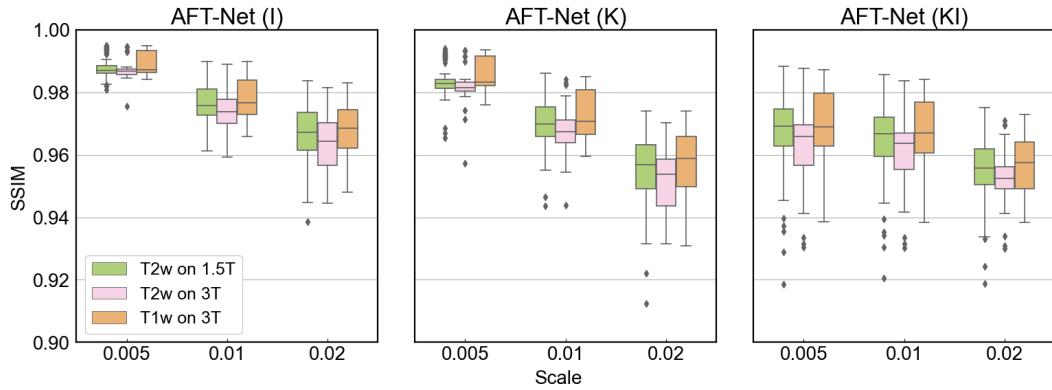


Figure 10. Results of human 1.5/3T MRI denoised reconstruction by comparing AFT-Net (I Model, K Model and KI Model) in terms of SSIM on different acquisition types and system field strength for noise scales 0.005, 0.01 and 0.02.

	SSIM	PSNR	NRMSE ($\times 10^{-6}$)
	1.000 ± 0.000	119.6 ± 1.4	6.519 ± 1.732

Table 1. Quantitative metrics of human 1.5/3T MRI reconstruction results. Numbers are presented as mean value \pm standard deviation.

Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.937 ± 0.017	0.961 ± 0.012	0.967 ± 0.011	0.968 ± 0.013
	PSNR (dB)	34.6 ± 1.9	37.2 ± 1.8	39.0 ± 1.7	39.2 ± 1.7
	NRMSE	0.113 ± 0.020	0.083 ± 0.010	0.068 ± 0.011	0.067 ± 0.011
4x (SR = 25%)	SSIM	0.832 ± 0.026	0.945 ± 0.012	0.940 ± 0.012	0.953 ± 0.012
	PSNR (dB)	28.7 ± 2.1	33.5 ± 2.2	33.7 ± 2.2	35.0 ± 2.1
	NRMSE	0.221 ± 0.031	0.126 ± 0.013	0.125 ± 0.020	0.107 ± 0.014
8x (SR = 12.5%)	SSIM	0.695 ± 0.035	0.914 ± 0.022	0.881 ± 0.030	0.926 ± 0.018
	PSNR (dB)	25.3 ± 2.3	30.2 ± 2.3	29.1 ± 2.3	31.2 ± 2.3
	NRMSE	0.326 ± 0.036	0.187 ± 0.021	0.212 ± 0.030	0.165 ± 0.021
16x (SR = 6.25%)	SSIM	0.552 ± 0.048	0.868 ± 0.043	0.782 ± 0.077	0.879 ± 0.039
	PSNR (dB)	23.2 ± 2.4	27.6 ± 2.5	25.4 ± 2.5	28.1 ± 2.5
	NRMSE	0.417 ± 0.038	0.251 ± 0.029	0.323 ± 0.028	0.237 ± 0.027

Table 2. Quantitative metrics of human 1.5/3T MRI accelerated reconstruction. Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
0.005	SSIM	0.981 ± 0.010	0.990 ± 0.004	0.986 ± 0.006	0.970 ± 0.015
	PSNR (dB)	44.9 ± 0.4	46.8 ± 0.4	45.6 ± 0.5	42.0 ± 1.1
	NRMSE	0.036 ± 0.011	0.028 ± 0.007	0.032 ± 0.008	0.049 ± 0.012
0.01	SSIM	0.918 ± 0.040	0.979 ± 0.007	0.973 ± 0.009	0.968 ± 0.014
	PSNR (dB)	37.4 ± 0.7	42.8 ± 0.9	41.8 ± 0.8	40.8 ± 1.1
	NRMSE	0.086 ± 0.030	0.044 ± 0.008	0.050 ± 0.010	0.056 ± 0.011
0.02	SSIM	0.755 ± 0.100	0.968 ± 0.010	0.956 ± 0.013	0.957 ± 0.011
	PSNR (dB)	29.6 ± 0.9	39.4 ± 1.3	38.3 ± 1.3	37.7 ± 1.4
	NRMSE	0.212 ± 0.077	0.065 ± 0.009	0.073 ± 0.011	0.078 ± 0.011

Table 3. Quantitative metrics of human 1.5/3T MRI denoised reconstruction. Numbers are presented as mean value \pm standard deviation. Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

4.2. Human low-field MRI study

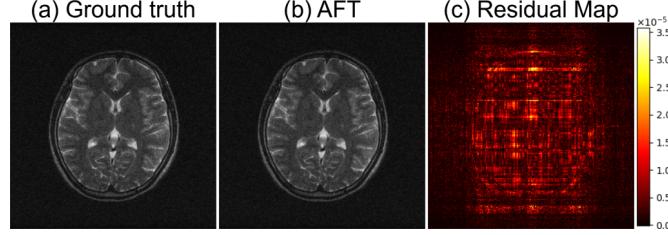


Figure 11. Quantitative metrics of human 0.3T MRI reconstruction. Numbers are presented as mean value \pm standard deviation.

First, we show above the results of human 0.3T MRI reconstruction using raw fully-sampled M4Raw kspace data in Fig. 11. All images were processed at the size of 256 x 256, with phase encoding in the X (LR) direction, and no cropping or reshaping was done due to it having been done already by the original M4Raw authors. The ground truth image was derived from the raw kspace data using a conventional Fourier transform method. From the image, we note that the image generated by AFT-Net is essentially identical to the Ground Truth. The results adhere to the mathematical description discussed in Section 2.1. The residual map shows minor brain structural information around the edges of the brain. Most of the remaining grid error is from floating point errors during matrix multiplication.

In Fig. 12, we show the results of human 0.3T accelerated reconstruction using under-sampled M4Raw kspace data. In the first row, we see the reconstructions from 1D 4x equal-spaced sampling, in which 8% of low-frequency columns are retained. We compare different AFT-Net structures against the Zero-Filling method. AFT-Net (KI) performs the best reconstruction, where the least structural difference can be seen from the residual map in the second row. The third row shows zoomed-in areas of both images and residual maps. AFT-Net (K) produces more blurry reconstruction which loses the structural details. Reconstruction through AFT-Net (I) produces somewhat similar results to AFT-Net (KI) but loses some structural detail. Fig. 13 shows the accelerated reconstruction results by comparing AFT-Net (I, K and KI) and zero filling in terms of SSIM across acceleration rates 2x, 4x, 8x and 16x. The performance of zero filling drops linearly as the acceleration rate increases while the AFT-Net methods are more robust to the acceleration scale. The t-test between each AFT-Net structure indicates that the AFT-Net (KI) clearly performs better than all other AFT-Net structures. The results of AFT-Net on different acquisition types and system field strength in Fig. 14 demonstrate that AFT-Net performs better on T1w images at 0.3T, but is robust in terms of image quality and retains excellent performance on other contrasts.

Next, we illustrate the results of human 0.3T denoised reconstruction using M4Raw kspace data with added Gaussian noise in Fig. 15 at a scale of 4.8. The noise scale was determined using the averaged maximum values of the dataset to stay in line with noise scales used for 1.5/3T tests. Unlike the results of accelerated reconstruction, AFT-Net (I) performs slightly better among three proposed AFT-Net structures, which can be seen in the t-test results in Fig. 16. Note that AFT-Net (I) does not hold a very significant advantage in SSIM, PSNR, or NRMSE compared to AFT-Net (KI) as demonstrated in Tab. 6. The second row shows the pixel-wise difference between AFT-Net output and noiseless ground truth. It can be indicated that the noise in the background is attenuated significantly. Although the brain structure can be seen from the residual map, the zoomed-in version of the image shows that the AFT-Net reconstruction preserves the anatomical structure. The results of AFT-Net on different acquisition types and system field strength in Fig. 14 also demonstrate that AFT-Net performs better on T1w images at 0.3T, but has good generality against different imaging modalities.

A comprehensive comparison of quantitative metrics on the test set is provided in Tab. 4, Tab. 5 and Tab. 6 for reconstruction, accelerated reconstruction and denoised reconstruction accordingly. AFT-Net (KI) significantly outperforms other AFT-Net structures on all the different acceleration rates. On denoised reconstruction, AFT-Net (I) performs slightly better than other AFT-Net structures. More detailed quantitative metrics of human 0.3T MRI accelerated reconstruction and denoised reconstruction results are provided in the appendix, which is grouped by image contrast and system field strength.

It is worth mentioning that although AFT-Net (K) does not outperform other AFT-Net structures in both accelerated reconstruction and denoised reconstruction tasks, it demonstrates the ability to learn in a sparse frequency domain and its sparse representations with a complex-valued convolutional network.

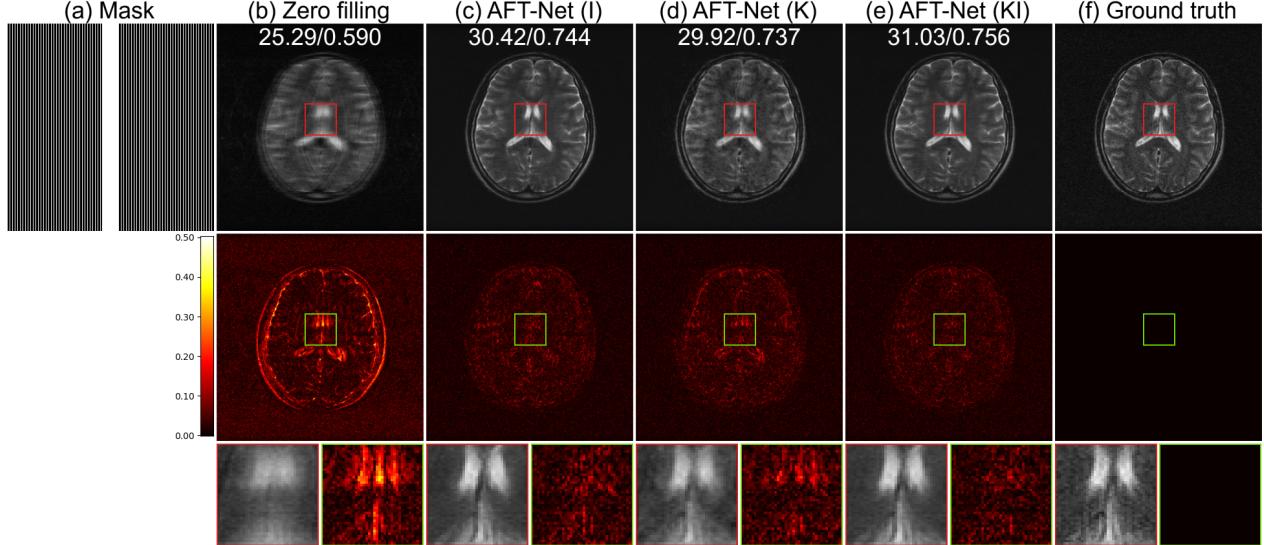


Figure 12. Human 0.3T MRI accelerated reconstruction results. (a) Sub-sampling mask used to generate under-sampled image, (b) zero filling, (c)-(e) proposed methods, and (f) ground truth. 1st row: 1D 4x equal-spaced sampling (8% of low-frequency columns are retained), 2nd row: difference magnitude against (f) (in Hot colormap), 3rd row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

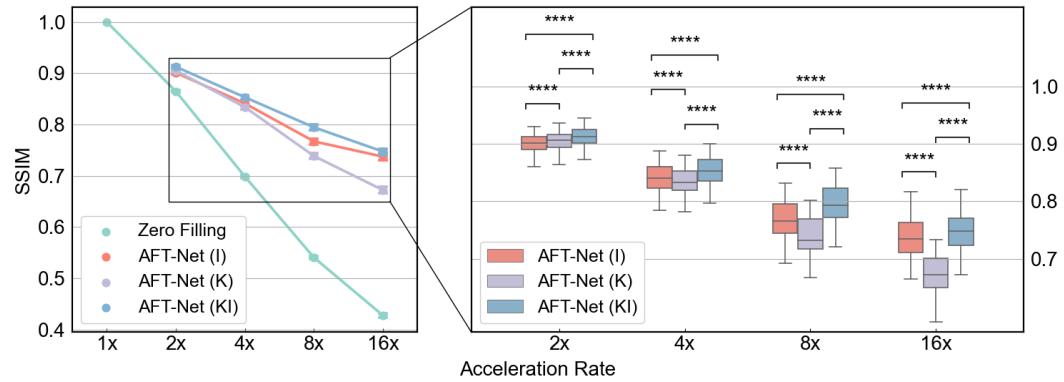


Figure 13. Results of human 0.3T MRI accelerated reconstruction by comparing AFT-Net (K Model, I Model and KI Model) and Zero Filling in terms of SSIM for acceleration rates 2x, 4x, 8x and 16x. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

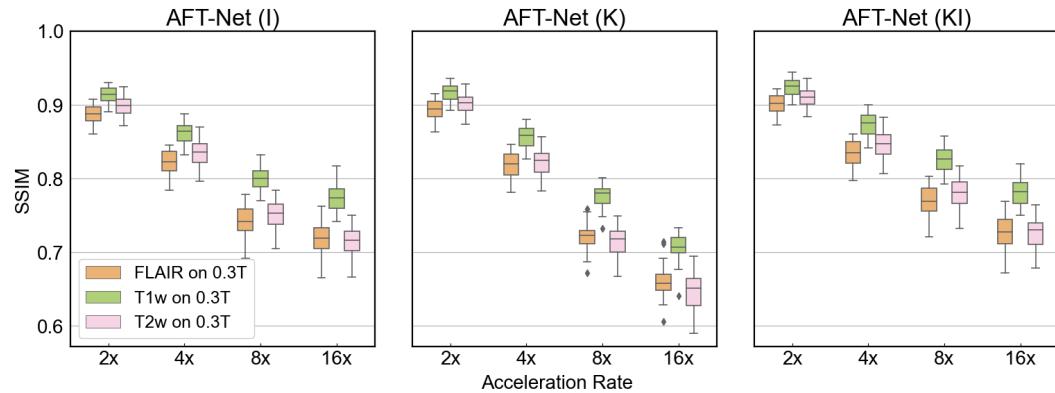


Figure 14. Results of human 0.3T MRI accelerated reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM on different acquisition types and system field strength for acceleration rates 2x, 4x, 8x and 16x.

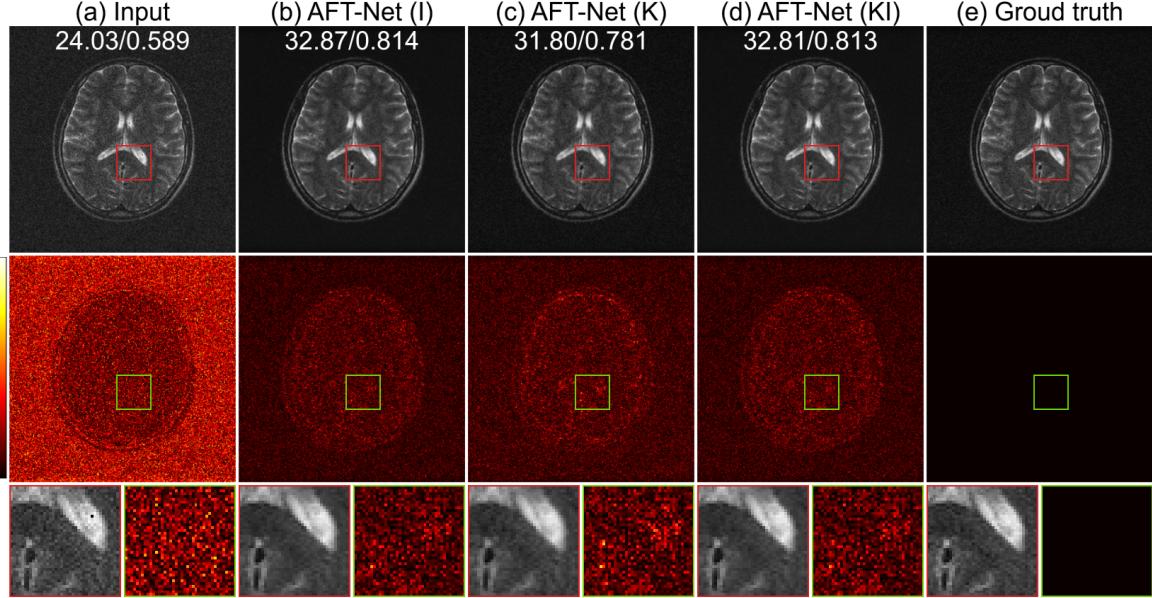


Figure 15. Human 0.3T MRI denoised reconstruction results. (a) Input, (b)-(d) proposed methods, and (e) ground truth. 1st row: Randomly added Gaussian noise in k-space (scale = 4.8), 2nd row: difference magnitude against (e) (in Hot colormap), 3rd row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

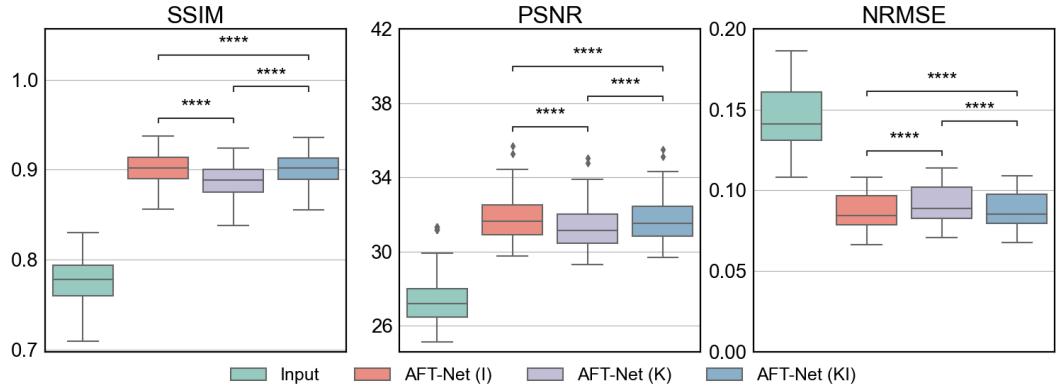


Figure 16. Results of human 0.3T MRI denoised reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM, PSNR and NRMSE. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

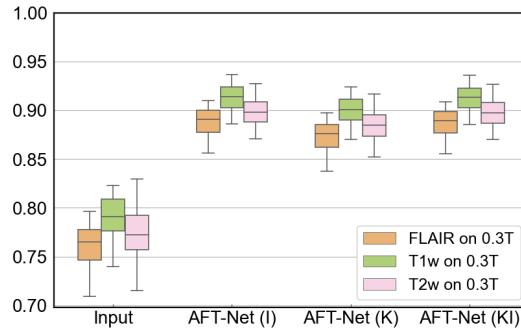


Figure 17. Results of human 0.3T MRI denoised reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM.

SSIM	PSNR	NRMSE (x10-5)
1.000 ± 0.000	106.9 ± 0.9	2.414 ± 0.166

Table 4. Quantitative metrics of human 0.3T MRI reconstruction. Numbers are presented as mean value \pm standard deviation.

Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.861 ± 0.012	0.898 ± 0.013	0.901 ± 0.013	0.910 ± 0.013
	PSNR (dB)	27.9 ± 0.5	30.8 ± 0.6	31.1 ± 0.6	31.6 ± 0.6
	NRMSE	0.143 ± 0.004	0.101 ± 0.004	0.099 ± 0.004	0.093 ± 0.004
4x (SR = 25%)	SSIM	0.688 ± 0.013	0.835 ± 0.018	0.823 ± 0.018	0.847 ± 0.019
	PSNR (dB)	23.1 ± 0.6	27.4 ± 0.6	26.9 ± 0.6	28.0 ± 0.6
	NRMSE	0.247 ± 0.009	0.151 ± 0.004	0.159 ± 0.005	0.141 ± 0.005
8x (SR = 12.5%)	SSIM	0.524 ± 0.013	0.751 ± 0.020	0.714 ± 0.021	0.781 ± 0.021
	PSNR (dB)	20.7 ± 0.6	24.5 ± 0.7	23.7 ± 0.7	25.4 ± 0.6
	NRMSE	0.326 ± 0.012	0.211 ± 0.008	0.231 ± 0.010	0.189 ± 0.006
16x (SR = 6.25%)	SSIM	0.394 ± 0.015	0.714 ± 0.021	0.646 ± 0.025	0.727 ± 0.023
	PSNR (dB)	19.0 ± 0.6	23.7 ± 0.7	22.5 ± 0.8	24.0 ± 0.7
	NRMSE	0.397 ± 0.013	0.230 ± 0.008	0.265 ± 0.013	0.222 ± 0.008

Table 5. Quantitative metrics of human 0.3T MRI accelerated reconstruction. Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
4.8	SSIM	0.777 ± 0.025	0.901 ± 0.017	0.888 ± 0.017	0.901 ± 0.016
	PSNR (dB)	27.4 ± 1.2	31.8 ± 1.2	31.4 ± 1.1	31.7 ± 1.1
	NRMSE	0.145 ± 0.019	0.086 ± 0.011	0.091 ± 0.012	0.087 ± 0.011

Table 6. Quantitative metrics of human 0.3T MRI denoised reconstruction (scale = 4.8). Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

4.3. Mouse high-field MRI study

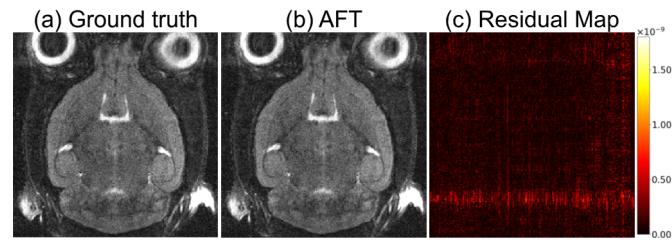


Figure 18. Mouse 9.4T MRI reconstruction results. (a) Ground truth, (b) proposed method, (c) difference magnitude of (a) and (b) (in Hot colormap).

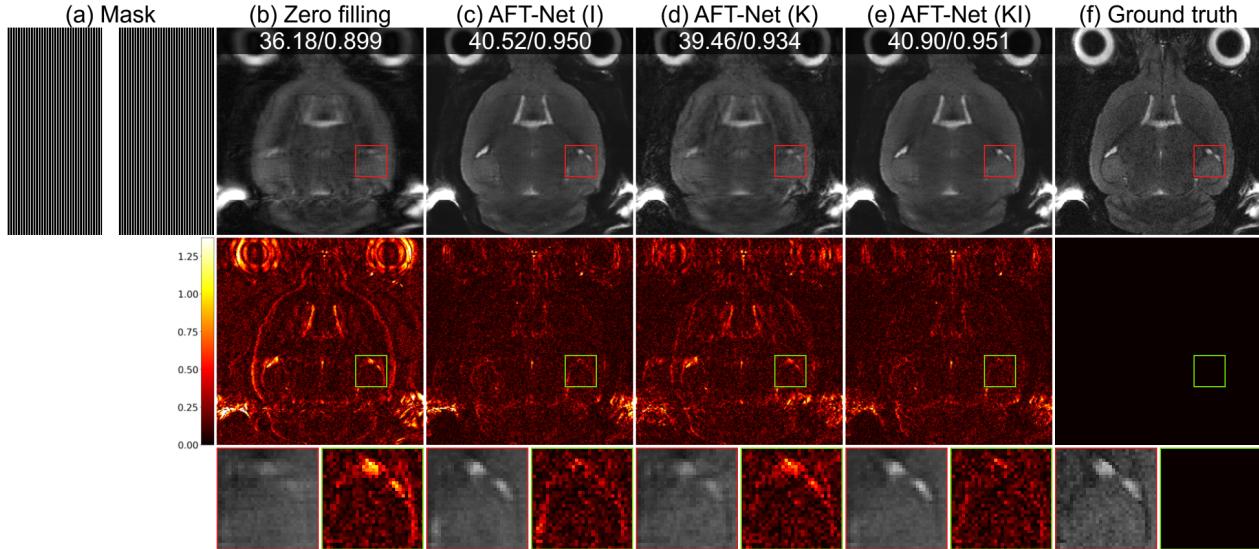


Figure 19. Mouse 9.4T MRI accelerated reconstruction results. (a) Sub-sampling mask used to generate under-sampled image, (b) zero filling, (c)-(e) proposed methods, and (f) ground truth. 1st row: 1D 4x equal-spaced sampling (8% of low-frequency columns are retained), 2nd row: difference magnitude against (f) (in Hot colormap), 3rd row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

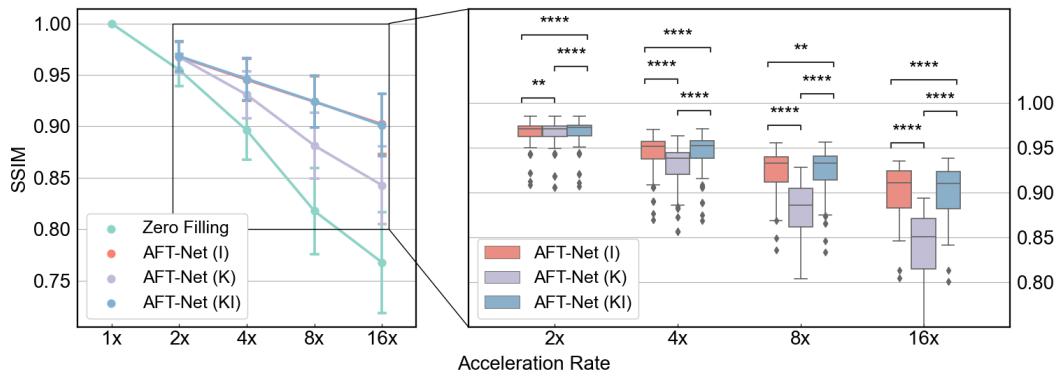


Figure 20. Results of mouse 9.4T MRI accelerated reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM, PSNR, and NRMSE. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

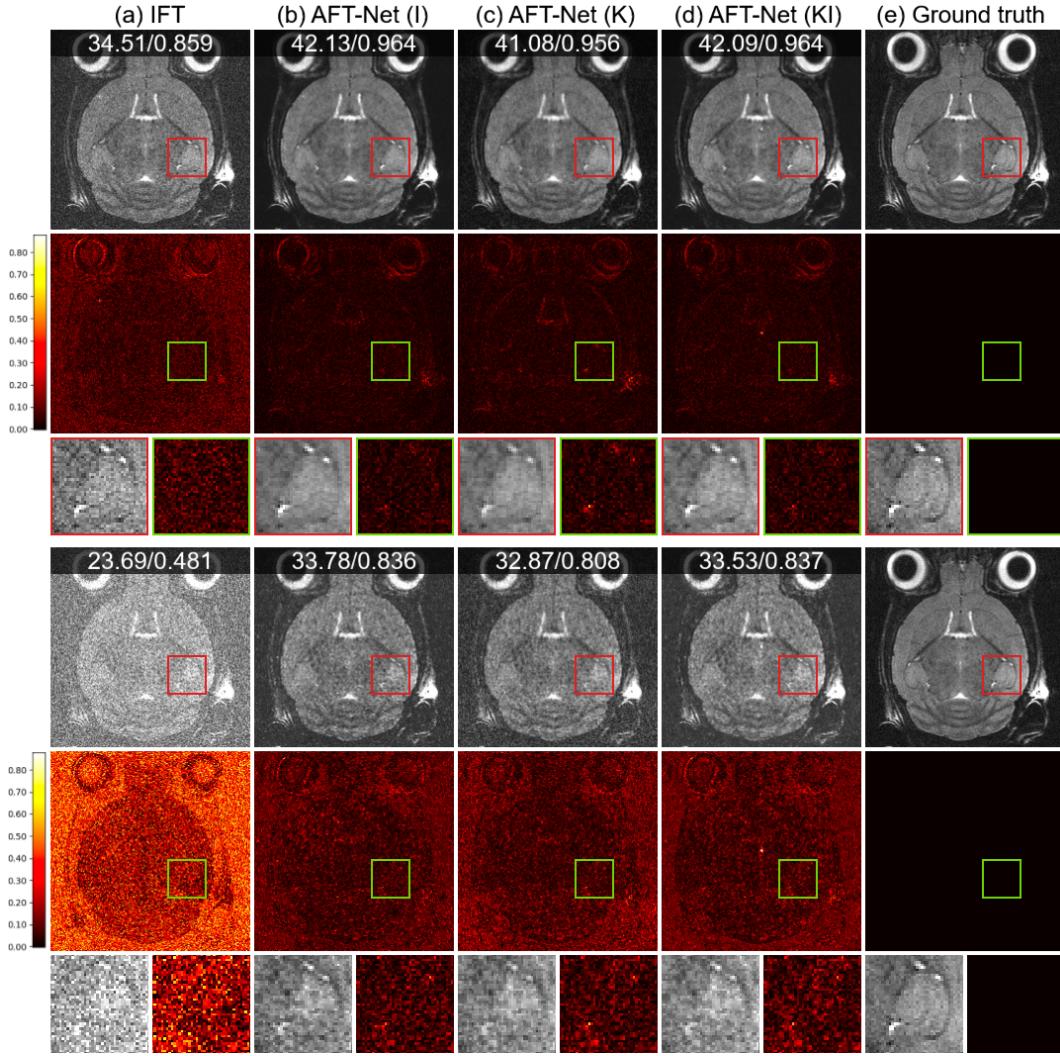


Figure 21. Mouse 9.4T MRI denoised reconstruction results. (a) Input, (b)-(d) proposed methods, and (e) ground truth. 1st row: Single repetition denoise, 4th row: Randomly added Gaussian noise in k-space (scale = 0.4), 2nd and 5th row: difference magnitude against (e) (in Hot colormap), 3rd and 6th row: zoomed-in version of the indicated box in 1st row and 2nd row. Yellow numbers in the upper center location indicate PSNR (db), and SSIM, respectively.

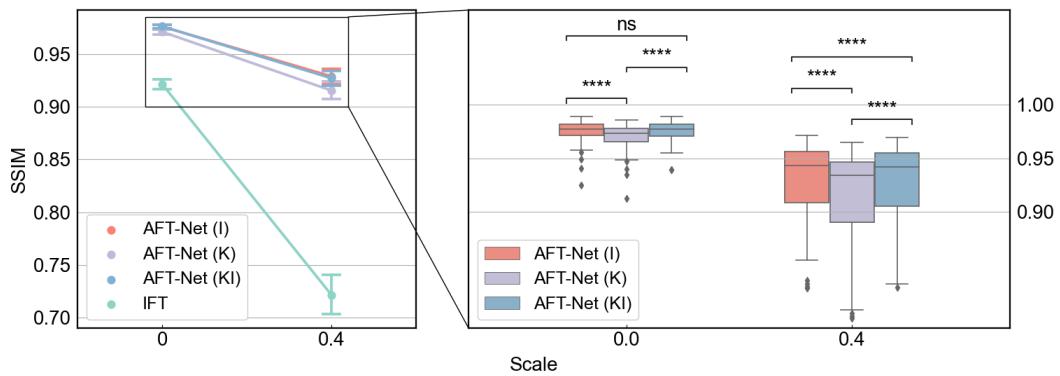


Figure 22. Results of mouse 9.4T MRI denoised reconstruction by comparing AFT-Net (I Model, K Model, and KI Model) in terms of SSIM, PSNR, and NRMSE. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.955 ± 0.016	0.968 ± 0.014	0.967 ± 0.015	0.968 ± 0.015
	PSNR (dB)	40.7 ± 2.0	43.1 ± 2.1	43.3 ± 2.1	43.3 ± 2.1
	NRMSE	0.162 ± 0.016	0.123 ± 0.012	0.122 ± 0.014	0.121 ± 0.013
4x (SR = 25%)	SSIM	0.896 ± 0.029	0.945 ± 0.020	0.931 ± 0.023	0.946 ± 0.020
	PSNR (dB)	36.1 ± 1.9	40.3 ± 1.9	39.3 ± 1.9	40.6 ± 1.9
	NRMSE	0.276 ± 0.015	0.170 ± 0.009	0.192 ± 0.011	0.165 ± 0.010
8x (SR = 12.5%)	SSIM	0.818 ± 0.042	0.924 ± 0.025	0.881 ± 0.032	0.924 ± 0.025
	PSNR (dB)	32.8 ± 1.8	38.3 ± 1.8	35.9 ± 1.8	38.5 ± 1.8
	NRMSE	0.403 ± 0.019	0.214 ± 0.017	0.284 ± 0.022	0.211 ± 0.016
16x (SR = 6.25%)	SSIM	0.767 ± 0.049	0.902 ± 0.029	0.843 ± 0.038	0.901 ± 0.030
	PSNR (dB)	31.1 ± 1.8	36.8 ± 1.8	34.1 ± 1.8	36.9 ± 1.8
	NRMSE	0.489 ± 0.028	0.257 ± 0.027	0.349 ± 0.032	0.253 ± 0.024

Table 7. Quantitative metrics of mouse 9.4T MRI accelerated reconstruction and evaluation on dataset with various acceleration rates. Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the lowest NRMSE and the highest PSNR/SSIM values

		AFT-Net			
		IFT	I Model	K Model	KI Model
T2w	SSIM	0.921 ± 0.024	0.976 ± 0.010	0.971 ± 0.011	0.976 ± 0.009
	PSNR (dB)	38.2 ± 2.0	43.8 ± 2.2	42.7 ± 2.0	43.7 ± 2.2
	NRMSE	0.211 ± 0.041	0.112 ± 0.025	0.126 ± 0.023	0.112 ± 0.024
T2w with added noise	SSIM	0.722 ± 0.097	0.929 ± 0.036	0.915 ± 0.042	0.927 ± 0.036
	PSNR (dB)	29.9 ± 3.0	39.2 ± 2.8	38.3 ± 2.8	38.9 ± 2.8
	NRMSE	0.564 ± 0.179	0.192 ± 0.053	0.214 ± 0.059	0.198 ± 0.055

Table 8. Quantitative metrics of mouse 9.4T MRI denoised reconstruction and evaluation on dataset with added Gaussian noise (scale = 0.4). Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate a significant difference against all other metrics, derived from paired two-sample t-tests under the null hypothesis.

4.4. Human normal-field MRS study

Magnetic resonance spectroscopy, namely MRS, is widely used for measuring human metabolism. While MRS has the potential to be highly valuable in clinical practice, it poses several challenges such as low signal-to-noise ratio, overlapping metabolite signals, experimental artifacts, and long acquisition times. Here, the AFT-Net is leveraged as a unified MRS reconstruction approach, which aims to reconstruct and process the FID in parallel, as shown in Fig. 3.

We trained our model on the MEGA-PRESS spectra from the Big GABA dataset for two reasons. First, as a proof of concept study, to guarantee the convergence of the supervised learning task, we need the dataset to be sufficient in the number of samples, good in data quality and publicly available. Thus the Big GABA dataset perfectly meets our requirements. Second, the smaller targeted signals are revealed by the subtraction of 2 spectra containing strong signals (OFF and ON), which provide a good way to verify the performance of the proposed method by measuring the subtraction artifacts. A total number of 101 subjects acquired by the Philips scanners were used in the training. For each subject, a standard GABA ON OFF edited MRS acquisition was run, where ON editing pulses were placed at 1.9 ppm and OFF editing pulses were placed at 7.46 ppm. The acquisition number is 320 (160 ON and 160 OFF transients) per subject. The AFT-Net was trained with an input size of 2048. The ground truth of the ON/OFF spectra is derived by taking the average over 160 acquisitions. We denote the ground truth as noiseless signals. For the training, we combined randomly sampled acquisitions of each subject to retrieve a noisy signal. By decreasing/increasing the number of sampled acquisitions, we can generate signals with higher/lower noise. We use the reduction rate (R) to denote the level of noise, which is defined as the ratio of the total acquisition number and the number of acquisitions sampled. This quantity is very handy to assess the power of denoising methods in practical terms. Retrieving accurate denoised signals at a high R has implications for the potential reduction of total experimental time.

The results of the AFT-Net approach and conventional numerical methods with Gaussian line broadening are illustrated in Fig. 23. The first row shows the reconstructed spectrum from the numerical methods and the proposed AFT-Net. The second row indicates the reconstructed spectrum overlaid with the ground truth. The third row plots the difference between the reconstructed spectrum and the ground truth. Under an acceleration rate of 80, where only 2 acquisitions were used over all 160 acquisitions, the AFT-Net shows excellent performance at high acceleration factors. The AFT-Net outperforms other methods for the DIFF spectra, indicating that the AFT-Net removes the noise in the FIDs while preserving the subject-level features. We used Goodness-of-Fit Coefficient (GFC) to measure the similarity between the reconstructed spectra and the ground truth, as shown in the Tab. 9. The metric value increases as the reduction rate decreases, but the absolute difference between high and low acceleration rates is tiny (0.9798 for OFF spectra under a reduction rate of 10 vs. 0.9688 for OFF spectra under a reduction rate of 160). In addition, AFT-Net outperforms the DFT+GLB (Gaussian Line Broadening) method across all metrics on the right table.

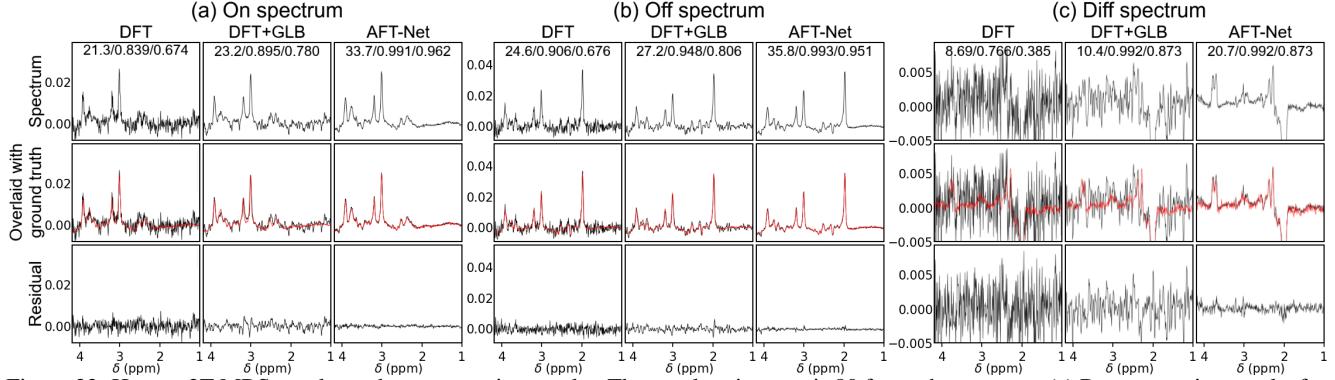


Figure 23. Human 3T MRS accelerated reconstruction results. The acceleration rate is 80 for each spectrum. (a) Reconstruction results for on spectrum, (b) reconstruction results for off spectrum, and (c) results for diff spectrum derived from (a) and (b). 1st row: reconstructed spectra, 2nd row: reconstructed spectra overlaid with ground truth (in red line), 3rd row: Difference of reconstructed spectra against ground truth. Yellow numbers in the upper center location indicate PSNR (db), PCC, and SCC, respectively.

R	Spectrum	DFT	DFT+GLB	AFT-Net
10	ON	0.9827 ± 0.0047	0.9686 ± 0.0086	0.9850 ± 0.0085
	OFF	0.9641 ± 0.0104	0.9617 ± 0.0108	0.9798 ± 0.0124
	DIFF	0.9403 ± 0.0164	0.9461 ± 0.0126	0.9868 ± 0.0037
20	ON	0.9660 ± 0.0090	0.9622 ± 0.0111	0.9843 ± 0.0092
	OFF	0.9314 ± 0.0192	0.9443 ± 0.0170	0.9794 ± 0.0127
	DIFF	0.8897 ± 0.0283	0.9208 ± 0.0206	0.9849 ± 0.0055
40	ON	0.9359 ± 0.0162	0.9496 ± 0.0155	0.9831 ± 0.0098
	OFF	0.8768 ± 0.0318	0.9152 ± 0.0281	0.9776 ± 0.0139
	DIFF	0.8129 ± 0.0418	0.8792 ± 0.0325	0.9815 ± 0.0078
80	ON	0.8826 ± 0.0280	0.9280 ± 0.0214	0.9803 ± 0.0120
	OFF	0.7890 ± 0.0486	0.8598 ± 0.0452	0.9748 ± 0.0160
	DIFF	0.7010 ± 0.0566	0.8077 ± 0.0480	0.9745 ± 0.0154
160	ON	0.7981 ± 0.0403	0.8854 ± 0.0325	0.9747 ± 0.0160
	OFF	0.6730 ± 0.0619	0.7759 ± 0.0638	0.9688 ± 0.0200
	DIFF	0.5710 ± 0.0654	0.7047 ± 0.0662	0.9616 ± 0.0245

Table 9. Quantitative metrics of human 3T MRS accelerated reconstruction. Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

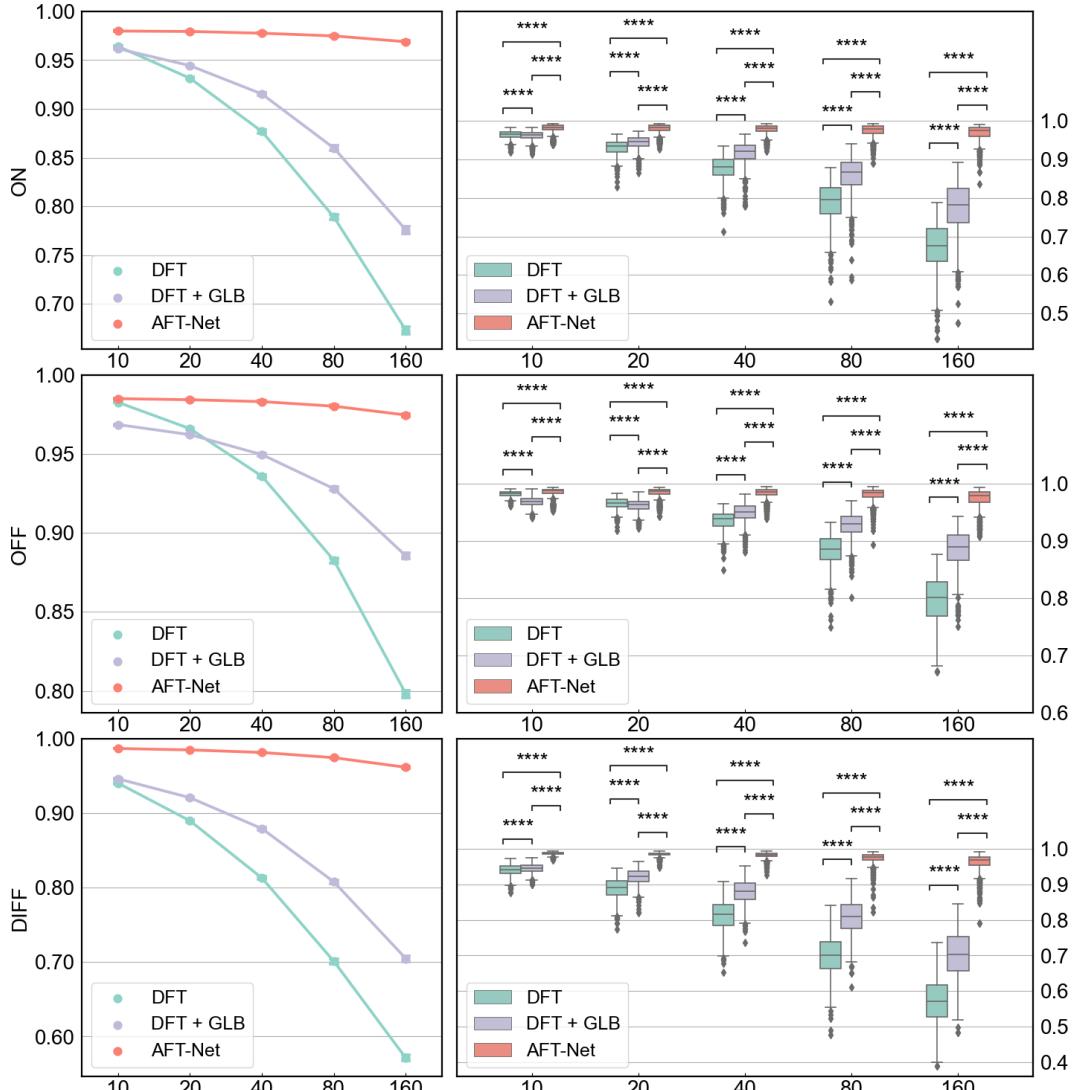


Figure 24. Results of human 3T MRS denoised reconstruction by comparing AFT-Net (I Model), DFT and DFT with Gaussian line broadening in terms of GFC. p-values indicate results from two-sided t-test for paired samples. (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$)

5. Conclusion

In conclusion, we propose AFT, a novel artificial Fourier transform framework that determines the mapping between k-space and image domain as conventional DFT while having the ability to be fine-tuned/optimized with further training. The flexibility of AFT allows it to be easily incorporated into any existing deep learning network as learnable or static blocks. We then utilized AFT to design our AFT-Net, which implements complex-valued U-Net to extract higher features in k-space and/or image domain. We aim to combine reconstruction and acceleration/denoising tasks into a unified network that simultaneously enhances the image quality by removing artifacts directly from the k-space and/or image domain. The proposed methods are evaluated on datasets with additional artifacts, different contrasts, and different modalities. Our AFT-Net achieves competitive results compared with other methods and proves to be more robust to noise and contrast differences. An extensive study on transfer learning demonstrates that our approach applies to other modalities.

However, one remaining limitation is that we only implement a complex-valued network with linear layers and CNNs, which are less effective than some advanced architectures. In our future work, we will aim to replace multi-layer perceptron and CNNs with transformer-based and diffusion-based models while extending the concept of AFT to more medical imaging tasks.

References

- [1] Elizabeth Cole, Joseph Cheng, John Pauly, and Shreyas Vasanawala. Analysis of deep complex-valued convolutional neural networks for mri reconstruction and phase-focused applications. *Magnetic resonance in medicine*, 86(2):1093–1109, 2021. 3
- [2] Marleen De Bruijne. Machine learning approaches in medical image analysis: From detection to diagnosis, 2016. 2
- [3] Jeffrey A Fessler. Model-based image reconstruction for mri. *IEEE signal processing magazine*, 27(4):81–89, 2010. 1
- [4] G.M. Georgiou and C. Koutsougeras. Complex domain backpropagation. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 39(5):330–334, 1992. 1
- [5] Alex Graves and Jürgen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in neural information processing systems*, 21, 2008. 1
- [6] Nitzan Guberman. On complex valued convolutional neural networks, 2016. 1
- [7] Nitzan Guberman. On complex valued convolutional neural networks. *CoRR*, abs/1602.09046, 2016. 3
- [8] Akira Hirose. *Complex-valued neural networks*, volume 400. Springer Science & Business Media, 2012. 3
- [9] Akira Hirose and Shotaro Yoshida. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and learning systems*, 23(4):541–551, 2012. 3
- [10] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 4
- [11] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, 2021. 1
- [12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7
- [13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1
- [14] Thomas Küstner, Niccolò Fuin, Kerstin Hammernik, Aurelien Bustin, Haikun Qi, Reza Hajosseiny, Pier Giorgio Masci, Radhouene Neji, Daniel Rueckert, René M Botnar, et al. Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions. *Scientific reports*, 10(1):1–13, 2020. 3
- [15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 2
- [16] Javier López-Randulfe, Tobias Duswald, Zhenshan Bing, and Alois Knoll. Spiking neural network for fourier transform and object detection for automotive radar. *Frontiers in Neurorobotics*, 15:69, 2021. 4
- [17] Emmanuel Montagnon, Milena Cerny, Alexandre Cadrian-Chênevert, Vincent Hamilton, Thomas Derennes, André Ilinca, Franck Vandenbroucke-Menu, Simon Turcotte, Samuel Kadoury, and An Tang. Deep learning workflow in radiology: a primer. *Insights into imaging*, 11(1):1–15, 2020. 2
- [18] Edouard Oyallon and Stéphane Mallat. Deep roto-translation scattering for object classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2865–2873, 2015. 3
- [19] Peter B Roemer, William A Edelstein, Cecil E Hayes, Steven P Souza, and Otward M Mueller. The nmr phased array. *Magnetic resonance in medicine*, 16(2):192–225, 1990. 6
- [20] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221, 2017. 1, 2
- [21] Chiheb Trabelsi, Oleksa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. In *International Conference on Learning Representations*, 2018. 1

- [22] Mark Tygert, Joan Bruna, Soumith Chintala, Yann LeCun, Serkan Piantino, and Arthur Szlam. A mathematical motivation for complex-valued convolutional networks. *Neural computation*, 28(5):815–825, 2016. 3
- [23] Shanshan Wang, Huitao Cheng, Leslie Ying, Taohui Xiao, Ziwen Ke, Hairong Zheng, and Dong Liang. Deepcomplexmri: Exploiting deep residual network for fast parallel mr imaging with complex convolution. *Magnetic Resonance Imaging*, 68:136–147, 2020. 3
- [24] EX Wu, D Sikka, N Igra, S Gjerwold-Sellec, C Gao, and J Guo. Cu-net: A completely complex u-net for mr k-space signal processing. In *Proceedings of International Society of Magnetic Resonance in Medicine*, 2021. 5
- [25] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1740–1749, 2020. 3
- [26] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzał, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastmri: An open dataset and benchmarks for accelerated mri, 2019. 6
- [27] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018. 1, 3

Human 1.5/3T MRI accelerated reconstruction

(a) T2w images on 1.5T						
Acceleration Rate	Metrics	Zero Filling	AFT-Net			
			I Model	K Model	KI Model	
2x (SR = 50%)	SSIM	0.937 ± 0.017	0.962 ± 0.012	0.968 ± 0.011	0.968 ± 0.013	
	PSNR (dB)	34.7 ± 1.9	37.3 ± 1.8	39.1 ± 1.7	39.3 ± 1.7	
	NRMSE	0.112 ± 0.020	0.082 ± 0.009	0.067 ± 0.011	0.066 ± 0.011	
4x (SR = 25%)	SSIM	0.833 ± 0.027	0.945 ± 0.013	0.941 ± 0.013	0.954 ± 0.012	
	PSNR (dB)	28.8 ± 2.0	33.6 ± 2.2	33.8 ± 2.2	35.1 ± 2.1	
	NRMSE	0.220 ± 0.031	0.125 ± 0.012	0.123 ± 0.020	0.106 ± 0.015	
8x (SR = 12.5%)	SSIM	0.699 ± 0.034	0.915 ± 0.022	0.883 ± 0.029	0.927 ± 0.018	
	PSNR (dB)	25.3 ± 2.2	30.3 ± 2.4	29.2 ± 2.3	31.4 ± 2.3	
	NRMSE	0.324 ± 0.035	0.185 ± 0.022	0.209 ± 0.031	0.163 ± 0.022	
16x (SR = 6.25%)	SSIM	0.557 ± 0.045	0.869 ± 0.044	0.785 ± 0.072	0.881 ± 0.039	
	PSNR (dB)	23.2 ± 2.3	27.7 ± 2.5	25.4 ± 2.4	28.2 ± 2.5	
	NRMSE	0.414 ± 0.035	0.249 ± 0.030	0.320 ± 0.027	0.233 ± 0.029	
(b) T2w images on 3T						
Acceleration Rate	Metrics	Zero Filling	AFT-Net			
			I Model	K Model	KI Model	
2x (SR = 50%)	SSIM	0.933 ± 0.016	0.958 ± 0.012	0.964 ± 0.010	0.964 ± 0.012	
	PSNR (dB)	35.0 ± 1.8	37.4 ± 1.9	39.2 ± 1.8	39.2 ± 1.8	
	NRMSE	0.113 ± 0.022	0.084 ± 0.010	0.069 ± 0.011	0.069 ± 0.011	
4x (SR = 25%)	SSIM	0.829 ± 0.027	0.943 ± 0.013	0.938 ± 0.010	0.950 ± 0.011	
	PSNR (dB)	29.2 ± 2.0	33.8 ± 2.1	34.0 ± 2.1	35.2 ± 2.0	
	NRMSE	0.219 ± 0.036	0.127 ± 0.014	0.126 ± 0.021	0.110 ± 0.015	
8x (SR = 12.5%)	SSIM	0.690 ± 0.035	0.914 ± 0.023	0.886 ± 0.027	0.925 ± 0.018	
	PSNR (dB)	25.7 ± 2.5	30.5 ± 2.2	29.6 ± 2.2	31.5 ± 2.2	
	NRMSE	0.325 ± 0.043	0.188 ± 0.021	0.210 ± 0.034	0.166 ± 0.021	
16x (SR = 6.25%)	SSIM	0.549 ± 0.053	0.874 ± 0.041	0.791 ± 0.087	0.881 ± 0.040	
	PSNR (dB)	23.5 ± 2.6	27.9 ± 2.3	25.8 ± 2.7	28.3 ± 2.3	
	NRMSE	0.417 ± 0.046	0.251 ± 0.024	0.322 ± 0.033	0.240 ± 0.020	
(c) T1w images on 3T						
Acceleration Rate	Metrics	Zero Filling	AFT-Net			
			I Model	K Model	KI Model	
2x (SR = 50%)	SSIM	0.939 ± 0.016	0.964 ± 0.013	0.969 ± 0.010	0.970 ± 0.012	
	PSNR (dB)	33.8 ± 1.9	36.5 ± 1.6	38.3 ± 1.5	38.5 ± 1.3	
	NRMSE	0.117 ± 0.014	0.085 ± 0.010	0.069 ± 0.010	0.068 ± 0.011	
4x (SR = 25%)	SSIM	0.829 ± 0.023	0.945 ± 0.010	0.938 ± 0.009	0.954 ± 0.010	
	PSNR (dB)	27.8 ± 2.2	32.7 ± 2.0	32.7 ± 2.1	34.0 ± 1.9	
	NRMSE	0.231 ± 0.020	0.132 ± 0.010	0.133 ± 0.010	0.113 ± 0.011	
8x (SR = 12.5%)	SSIM	0.684 ± 0.036	0.909 ± 0.018	0.868 ± 0.032	0.922 ± 0.014	
	PSNR (dB)	24.5 ± 2.3	29.2 ± 2.2	28.0 ± 2.3	30.2 ± 2.2	
	NRMSE	0.338 ± 0.031	0.196 ± 0.013	0.226 ± 0.015	0.175 ± 0.014	
16x (SR = 6.25%)	SSIM	0.535 ± 0.048	0.854 ± 0.040	0.756 ± 0.080	0.868 ± 0.034	
	PSNR (dB)	22.4 ± 2.2	26.6 ± 2.3	24.6 ± 2.3	27.2 ± 2.2	
	NRMSE	0.430 ± 0.035	0.265 ± 0.019	0.335 ± 0.022	0.249 ± 0.019	

Table 10. Quantitative metrics of human 1.5/3T MRI accelerated reconstruction. Numbers are presented as mean value ± standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Human 1.5/3T MRI denoised reconstruction

(a) T2w images on 1.5T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
0.005	SSIM	0.981 ± 0.009	0.990 ± 0.004	0.986 ± 0.005	0.971 ± 0.014
	PSNR (dB)	44.9 ± 0.4	46.9 ± 0.4	45.6 ± 0.5	42.0 ± 1.1
	NRMSE	0.036 ± 0.011	0.028 ± 0.007	0.032 ± 0.008	0.049 ± 0.011
0.01	SSIM	0.917 ± 0.039	0.979 ± 0.007	0.973 ± 0.009	0.968 ± 0.013
	PSNR (dB)	37.4 ± 0.7	42.8 ± 0.8	41.8 ± 0.8	40.8 ± 1.1
	NRMSE	0.086 ± 0.030	0.044 ± 0.008	0.050 ± 0.009	0.055 ± 0.010
0.02	SSIM	0.754 ± 0.098	0.968 ± 0.010	0.956 ± 0.013	0.957 ± 0.010
	PSNR (dB)	29.5 ± 0.9	39.4 ± 1.3	38.3 ± 1.3	37.7 ± 1.4
	NRMSE	0.212 ± 0.076	0.064 ± 0.009	0.073 ± 0.011	0.078 ± 0.010
(b) T2w images on 3T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
0.005	SSIM	0.979 ± 0.010	0.989 ± 0.003	0.985 ± 0.006	0.966 ± 0.014
	PSNR (dB)	44.9 ± 0.5	46.9 ± 0.4	45.6 ± 0.7	41.8 ± 1.4
	NRMSE	0.038 ± 0.013	0.029 ± 0.007	0.034 ± 0.009	0.052 ± 0.012
0.01	SSIM	0.910 ± 0.043	0.977 ± 0.007	0.970 ± 0.008	0.964 ± 0.013
	PSNR (dB)	37.3 ± 0.8	42.9 ± 0.9	41.8 ± 0.9	40.8 ± 1.3
	NRMSE	0.091 ± 0.034	0.046 ± 0.009	0.051 ± 0.010	0.058 ± 0.011
0.02	SSIM	0.735 ± 0.105	0.965 ± 0.009	0.953 ± 0.011	0.954 ± 0.010
	PSNR (dB)	29.5 ± 1.0	39.7 ± 1.4	38.6 ± 1.3	38.0 ± 1.4
	NRMSE	0.226 ± 0.086	0.065 ± 0.009	0.074 ± 0.011	0.079 ± 0.012
(c) T1w images on 3T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
0.005	SSIM	0.984 ± 0.009	0.991 ± 0.004	0.989 ± 0.005	0.974 ± 0.015
	PSNR (dB)	45.0 ± 0.4	46.8 ± 0.3	45.7 ± 0.3	41.9 ± 0.8
	NRMSE	0.033 ± 0.010	0.027 ± 0.006	0.030 ± 0.007	0.047 ± 0.014
0.01	SSIM	0.930 ± 0.036	0.981 ± 0.008	0.976 ± 0.009	0.971 ± 0.014
	PSNR (dB)	37.6 ± 0.6	42.5 ± 0.7	41.6 ± 0.6	40.5 ± 0.8
	NRMSE	0.079 ± 0.027	0.043 ± 0.009	0.048 ± 0.010	0.055 ± 0.012
0.02	SSIM	0.789 ± 0.095	0.971 ± 0.011	0.959 ± 0.013	0.961 ± 0.011
	PSNR (dB)	29.8 ± 0.8	38.9 ± 1.2	37.9 ± 1.2	37.3 ± 1.3
	NRMSE	0.195 ± 0.069	0.065 ± 0.009	0.073 ± 0.011	0.078 ± 0.010

Table 11. Quantitative metrics of human 1.5/3T MRI denoised reconstruction. Numbers are presented as mean value ± standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Human 0.3T MRI accelerated reconstruction

(a) FLAIR images on 0.3T					
Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.851 ± 0.014	0.887 ± 0.013	0.893 ± 0.013	0.900 ± 0.013
	PSNR (dB)	27.7 ± 0.7	31.0 ± 0.6	31.4 ± 0.7	31.8 ± 0.7
	NRMSE	0.124 ± 0.006	0.085 ± 0.003	0.081 ± 0.004	0.077 ± 0.004
4x (SR = 25%)	SSIM	0.681 ± 0.015	0.820 ± 0.018	0.817 ± 0.018	0.833 ± 0.019
	PSNR (dB)	22.6 ± 0.7	27.9 ± 0.6	27.7 ± 0.6	28.4 ± 0.6
	NRMSE	0.224 ± 0.015	0.121 ± 0.004	0.123 ± 0.004	0.114 ± 0.004
8x (SR = 12.5%)	SSIM	0.523 ± 0.015	0.742 ± 0.023	0.721 ± 0.020	0.769 ± 0.022
	PSNR (dB)	19.8 ± 0.9	25.1 ± 0.6	24.6 ± 0.6	26.0 ± 0.6
	NRMSE	0.309 ± 0.029	0.168 ± 0.007	0.176 ± 0.007	0.150 ± 0.006
16x (SR = 6.25%)	SSIM	0.423 ± 0.018	0.718 ± 0.026	0.661 ± 0.023	0.727 ± 0.025
	PSNR (dB)	18.5 ± 0.9	24.4 ± 0.7	23.2 ± 0.7	24.7 ± 0.6
	NRMSE	0.359 ± 0.032	0.182 ± 0.009	0.207 ± 0.011	0.175 ± 0.008
(b) T1w images on 0.3T					
Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.877 ± 0.011	0.913 ± 0.011	0.917 ± 0.011	0.923 ± 0.011
	PSNR (dB)	28.9 ± 0.9	32.7 ± 0.7	33.2 ± 0.8	33.6 ± 0.8
	NRMSE	0.123 ± 0.006	0.080 ± 0.005	0.076 ± 0.006	0.072 ± 0.006
4x (SR = 25%)	SSIM	0.721 ± 0.016	0.862 ± 0.014	0.856 ± 0.014	0.873 ± 0.015
	PSNR (dB)	23.5 ± 1.1	29.5 ± 0.7	29.3 ± 0.7	30.2 ± 0.7
	NRMSE	0.231 ± 0.014	0.115 ± 0.007	0.118 ± 0.008	0.106 ± 0.007
8x (SR = 12.5%)	SSIM	0.569 ± 0.018	0.800 ± 0.016	0.775 ± 0.015	0.826 ± 0.017
	PSNR (dB)	20.3 ± 1.2	26.7 ± 0.7	26.1 ± 0.7	27.8 ± 0.7
	NRMSE	0.332 ± 0.025	0.159 ± 0.009	0.171 ± 0.011	0.140 ± 0.009
16x (SR = 6.25%)	SSIM	0.464 ± 0.018	0.774 ± 0.019	0.706 ± 0.018	0.782 ± 0.018
	PSNR (dB)	18.8 ± 1.2	25.6 ± 0.6	24.0 ± 0.7	26.0 ± 0.6
	NRMSE	0.397 ± 0.032	0.181 ± 0.012	0.216 ± 0.014	0.172 ± 0.011
(c) T2w images on 0.3T					
Acceleration Rate	Metrics	Zero Filling	AFT-Net		
			I Model	K Model	KI Model
2x (SR = 50%)	SSIM	0.861 ± 0.012	0.898 ± 0.013	0.901 ± 0.013	0.910 ± 0.013
	PSNR (dB)	27.9 ± 0.6	30.8 ± 0.6	31.1 ± 0.6	31.6 ± 0.6
	NRMSE	0.143 ± 0.004	0.101 ± 0.004	0.099 ± 0.004	0.093 ± 0.004
4x (SR = 25%)	SSIM	0.688 ± 0.013	0.835 ± 0.018	0.823 ± 0.018	0.847 ± 0.019
	PSNR (dB)	23.1 ± 0.6	27.4 ± 0.6	26.9 ± 0.6	28.0 ± 0.6
	NRMSE	0.247 ± 0.009	0.151 ± 0.004	0.159 ± 0.005	0.141 ± 0.005
8x (SR = 12.5%)	SSIM	0.524 ± 0.013	0.751 ± 0.020	0.714 ± 0.021	0.781 ± 0.021
	PSNR (dB)	20.7 ± 0.6	24.5 ± 0.7	23.7 ± 0.7	25.4 ± 0.6
	NRMSE	0.326 ± 0.012	0.211 ± 0.008	0.231 ± 0.010	0.189 ± 0.006
16x (SR = 6.25%)	SSIM	0.394 ± 0.015	0.714 ± 0.021	0.646 ± 0.025	0.727 ± 0.023
	PSNR (dB)	19.0 ± 0.6	23.7 ± 0.7	22.5 ± 0.8	24.0 ± 0.7
	NRMSE	0.397 ± 0.013	0.230 ± 0.008	0.265 ± 0.013	0.222 ± 0.008

Table 12. Quantitative metrics of human 0.3T MRI accelerated reconstruction. Numbers are presented as mean value ± standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Human 0.3T MRI denoised reconstruction

(a) FLAIR images on 0.3T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
4.8	SSIM	0.763 ± 0.020	0.889 ± 0.014	0.875 ± 0.015	0.888 ± 0.013
	PSNR (dB)	26.8 ± 0.6	31.2 ± 0.6	30.9 ± 0.6	31.1 ± 0.6
	NRMSE	0.137 ± 0.007	0.082 ± 0.004	0.086 ± 0.004	0.083 ± 0.004
(b) T1w images on 0.3T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
4.8	SSIM	0.790 ± 0.022	0.913 ± 0.013	0.900 ± 0.014	0.912 ± 0.012
	PSNR (dB)	28.5 ± 0.9	33.0 ± 0.8	32.5 ± 0.8	32.9 ± 0.8
	NRMSE	0.129 ± 0.011	0.077 ± 0.006	0.081 ± 0.006	0.078 ± 0.006
(c) T2w images on 0.3T					
Noise Scale	Metrics	Input	AFT-Net		
			I Model	K Model	KI Model
4.8	SSIM	0.774 ± 0.024	0.898 ± 0.014	0.884 ± 0.015	0.898 ± 0.014
	PSNR (dB)	26.6 ± 0.7	31.1 ± 0.6	30.6 ± 0.6	31.0 ± 0.6
	NRMSE	0.166 ± 0.011	0.099 ± 0.005	0.105 ± 0.005	0.100 ± 0.005

Table 13. Quantitative metrics of human 0.3T MRI denoised reconstruction (scale = 4.8). Numbers are presented as mean value \pm standard deviation. Numbers in boldface indicate the best metric out of all the methods.

Appendix A. Evaluation of denoising model on mouse dataset with motions

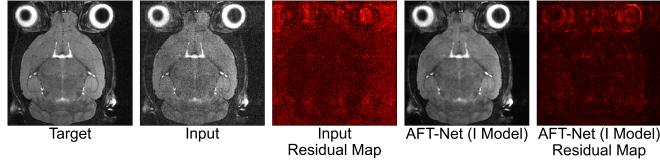


Figure A1. Evaluation of the AFT-Net (I Model), which was trained for denoising task, on the mouse dataset with motions. While denoising, the AFT-Net (I Model) also eliminated motions around eyes.

Appendix B. Accelerated MRI reconstruction on human dataset

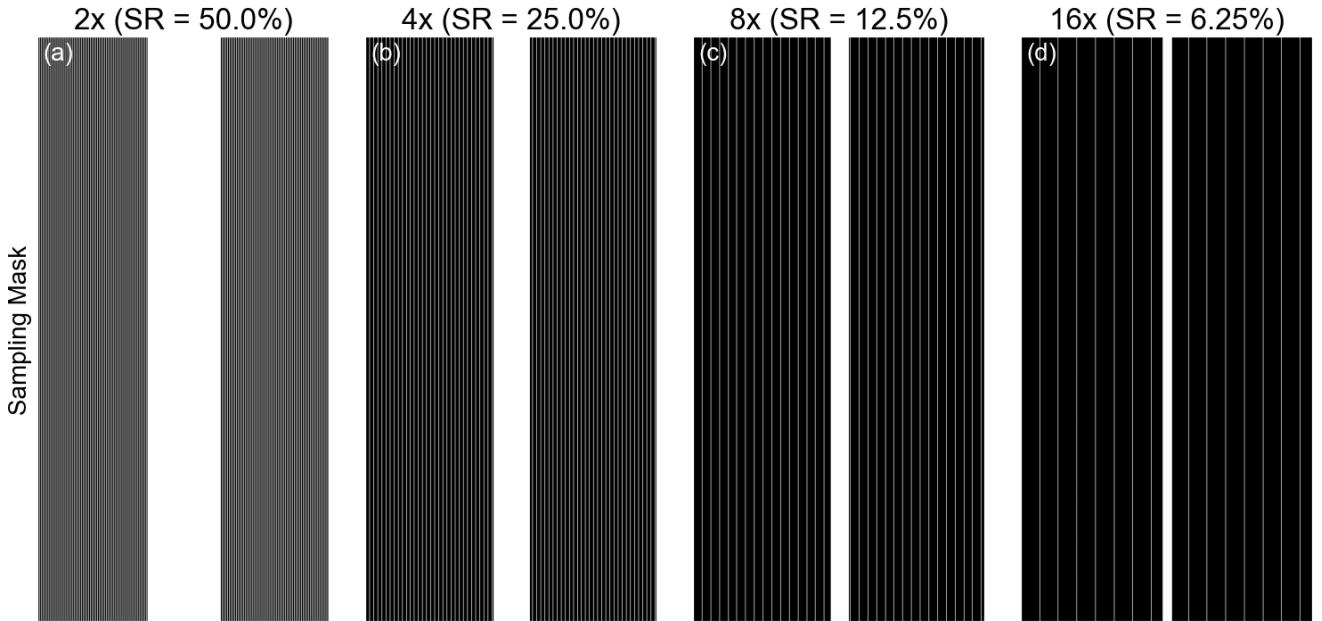


Figure B1. Binary undersampling masks used in this appendix (white = included, black = omitted). The frequency-encoding direction (y direction) is fully sampled. An equispaced mask with preserved high-frequency components and approximate acceleration matching in the phase-encoding direction (x direction) was used to generate undersampling masks with different acceleration rates or sampling ratios (SRs). The same undersampling masks are also used in Appendix C.

Acquisition Type	System Field Strength	
	1.5T	3T
T1	0	109
T2	692	147

Table B1. Number of scans for the different acquisition types and scanner field strengths. 948 fully sampled human brain MRIs from the fastMRI dataset were selected for training and evaluation. The dataset includes axial T1-weighted and T2-weighted images obtained on the 4-coiled scanner.