

# Deep Learning-based MRI Reconstruction with Artificial Fourier Transform Network (AFTNet)

Yanting Yang<sup>a</sup>, Yiren Zhang<sup>a</sup>, Zongyu Li<sup>a</sup>, Jeffery Siyuan Tian<sup>b</sup>, Matthieu Dagommer<sup>a</sup>, Jia Guo<sup>c</sup>

<sup>a</sup>*Department of Biomedical Engineering, Columbia University, 500 W. 120th Street #351, New York, 10027, NY, United States*

<sup>b</sup>*Department of Computer Science, University of Maryland, 8125 Paint Branch Drive, College Park, 20742, MD, United States*

<sup>c</sup>*Department of Psychiatry, Columbia University, 1051 Riverside Drive, New York, 10032, NY, United States*

---

## Abstract

Deep complex-valued neural networks (CVNNs) provide a powerful way to leverage complex number operations and representations and have succeeded in several phase-based applications. However, previous networks have not fully explored the impact of complex-valued networks in the frequency domain. Here, we introduce a unified complex-valued deep learning framework—Artificial Fourier Transform Network (AFTNet)—which combines domain-manifold learning and CVNNs. AFTNet can be readily used to solve image inverse problems in domain transformation, especially for accelerated magnetic resonance imaging (MRI) reconstruction and other applications. While conventional methods typically utilize magnitude images or treat the real and imaginary components of k-space data as separate channels, our approach directly processes raw k-space data in the frequency domain, utilizing complex-valued operations. This allows for a mapping between the frequency (k-space) and image domain to be determined through cross-domain learning. We show that AFTNet achieves superior accelerated MRI reconstruction compared to existing approaches. Furthermore, our approach can be applied to various tasks, such as denoised magnetic resonance spectroscopy (MRS) reconstruction and datasets with various contrasts. The AFTNet presented here is a valuable preprocessing component for different preclinical studies

---

*Email address:* jg3400@columbia.edu (Jia Guo)

and provides an innovative alternative for solving inverse problems in imaging and spectroscopy. The code is available at: <https://github.com/yanting-yang/AFT-Net>.

*Keywords:* MRI, Deep Learning, Reconstruction

---

## 1. Introduction

The shift from real to complex coordinate space in deep neural networks has unveiled the potential of complex numbers' rich representational capacity, thereby spurring the development of complex-valued neural architectures [1] [2] [3]. A corresponding, though inverse, domain shift occurs in the preprocessing of magnetic resonance imaging (MRI), where raw data are acquired and stored in complex-valued k-space, with each pixel encoding spatial frequency information across two or three dimensions. Following acquisition, the raw k-space data are transformed into images via a reconstruction process that enables interpretation by MR operators, physicians, radiologists, or data scientists. This crucial reconstruction step underpins overall image quality. Proper image reconstruction methods can increase the signal-to-noise ratio (SNR) by suppressing thermal noise [4] [5], improve spatial inhomogeneities affected by point spread functions (PSFs), and correct unexpected signal artifacts [6].

Theoretically, image reconstruction is performed using domain transforms, for example Fourier transforms for fully sampled Cartesian data [7]. However, in clinical settings where signal non-idealities are prevalent, numerical methods and machine learning approaches become indispensable. Traditionally, human experts select task-related features and develop models to capture the mapping between k-space and image domains [8]. However, due to significant pathological variations and the risk of human oversight [9], a consistent and unbiased diagnosis cannot be guaranteed. In recent years, k-space, as a low-dimensional feature space, has been leveraged in deep neural networks to learn the manifold mapping of domain transforms in low signal-to-noise settings [10]. This image reconstruction process can be reformulated as a data-driven supervised learning task that determines the mapping between the k-space and the image domains, demonstrating superior immunity to noise and reconstruction artifacts. The conventional discrete Fourier transform algorithm, which is mathematically derived and is not based on learning, can be substituted by neural networks [11], as neural networks can

learn complex mappings and priors from data, enabling them to recover missing information and reduce artifacts more effectively. This is particularly important in highly undersampled scenarios, where traditional DFT-based methods struggle to achieve high-quality reconstructions. Thus, fundamental neural networks have been presented that avoid the difficulties of finding the network structure and optimizing the algorithm. A similar approach is adopted in domain transform manifold learning along the phase-encoding direction [12], where the front-end convolutional layers, an intermediate global transform, and the back-end convolutional layers are combined to perform data restoration in the k-space and image domains. Modern score-based diffusion models provide a powerful way to sample data from a conditional distribution given the measurements in the k-space domain, which can be used to solve inverse problems in imaging [13].

Although most deep learning-based MR image reconstruction algorithms have applied the concept of domain-transform learning that directly learns a low-dimensional joint manifold between the k-space and image domains, including end-to-end [14] and the unrolled [15] network, few works have fully leveraged complex-valued neural networks (CVNNs), which allow neural networks to implement learning-based frequency selection [16]. Some early works proposed CVNNs but focused mainly on solving the basics of learning [17] [18]. In recent years, extensive studies have been conducted on complex-valued CNNs. Generalizations of real-valued CNN models have been shown to be significantly less vulnerable to overfitting [2]. Mathematical arguments and implementations have also been discussed [19] [3], enabling the practical application of CVNNs. A combination of CVNNs and vision deep learning models (e.g., UNet [20]) has recently emerged and is being exploited for MR image reconstruction [21], [22], demonstrating superior and accelerated reconstruction compared to real-valued neural networks. However, the main drawback of previous CVNN works is that the potential of leveraging CVNNs in domain manifold learning has not been fully investigated.

To fully exploit the rich information contained in complex-valued MR data and overcome the limitations of traditional reconstruction methods, we propose (1) Artificial Fourier Transform (AFT) block which replaces conventional DFT with a learnable module and can adapt to noise and signal non-idealities, allowing the network to learn complex mappings between the k-space and image domains; (2) Artificial Fourier Transform Network (AFTNet), shown in Figure 1, which combines domain-manifold learning with CVNNs by integrating AFT with complex-valued convolutional encoder-

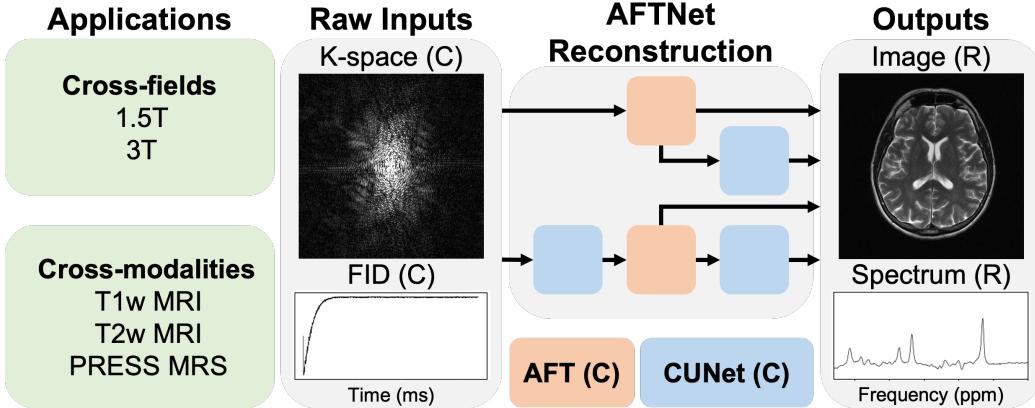


Figure 1: Schematics of general deep-learning MR imaging/spectroscopy reconstruction based on AFTNet. The inputs of AFTNet can be 2D MRI k-space data or 1D MRS FID data. The outputs are reconstructed MR images or spectra. Different structures of AFTNet are developed by appending front-end and/or back-end convolutional networks to the AFT block. Here we show the T2w 1.5T MRI and 3T PRESS MRS reconstruction results, respectively. C: complex-valued and R: real-valued.

decoder networks to facilitate MR reconstruction. AFTNet directly process the full complex information (both magnitude and phase) inherent in k-space, avoiding the common pitfall of separating or discarding phase information. By operating directly on raw k-space data, AFTNet minimizes information loss and learns subtle frequency domain features critical for accurate image reconstruction. The CVNNs in AFTNet process global features in the frequency (k-space) domain and local details in the image domain, leading to improved artifact removal and reconstruction quality, particularly in highly undersampled scenarios. Through its modular design, AFTNet can be extended to any dimension (e.g., 1D MR spectroscopy data). We conduct extensive experiments in various modalities, system field strengths, acceleration ratios, and noise levels on the MRI reconstruction, accelerated reconstruction and MRS denoised reconstruction, achieving superior performance in all scenarios.

## 2. Background

### 2.1. Complex-valued neural networks

The definition of a conventional real-valued neural network can be extended to the complex domain. We denote a complex operator as  $W(\cdot) =$

$\mathbf{W}_{real}(\cdot) + i\mathbf{W}_{imag}(\cdot)$ , where  $\mathbf{W}_{real}$  and  $\mathbf{W}_{imag}$  are real-valued operators. The input complex vector can be represented as  $z = \Re(z) + i\Im(z)$ . The output of complex operator  $\mathbf{W}$  acting on  $z$  is derived following the vector dot product:

$$\begin{aligned} y &= \mathbf{W} \cdot z \\ &= [\mathbf{W}_{real}(\Re(z)) - \mathbf{W}_{imag}(\Im(z))] \\ &\quad + i [\mathbf{W}_{imag}(\Re(z)) + \mathbf{W}_{real}(\Im(z))]. \end{aligned} \tag{1}$$

As the linear operator and convolution operator are distributive [3], we can directly replace  $\mathbf{W}(\cdot)$  with  $\text{Linear}(\cdot)$  or  $\text{Conv}(\cdot)$  to obtain the complex-valued version:

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

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 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 simply applies 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 [3] 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 the statistical covariance shift [23] [24] [25]. This is also mirrored in the CNNs, 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:

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

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

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

---

**Algorithm 1** Complex group normalization

---

**Input:**  $z, \gamma, \beta, G$  (number of groups),  $\epsilon$   
 Shape of  $z$  is  $(B, C, H, W)$   
 Reshape  $z$  to  $(B, G, C//G, H, W)$   
**for** each group **do**  
 $\tilde{z} \leftarrow V^{-\frac{1}{2}}(z - \mathbb{E}(z))$  based on Equations (6) to (8).  
**end for**  
 Reshape  $\tilde{z}$  to  $(B, C, H, W)$   
**return**  $\gamma\tilde{z} + \beta$

---

and  $V$  is the covariance matrix of real and imaginary parts of  $z$

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

$V$  is a  $2 \times 2$  matrix, and the existence of the inverse square root is guaranteed by adding  $\epsilon I$  (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

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

where  $\gamma$  and  $\beta$  are learnable parameters.

Considering the limitation of GPU RAM and the large memory consumption of complex-valued networks, we use group normalization [24] in our framework to avoid the possible inaccurate estimation of batch statistics caused by a small batch size. The group normalization in the complex domain can be represented as Algorithm 1

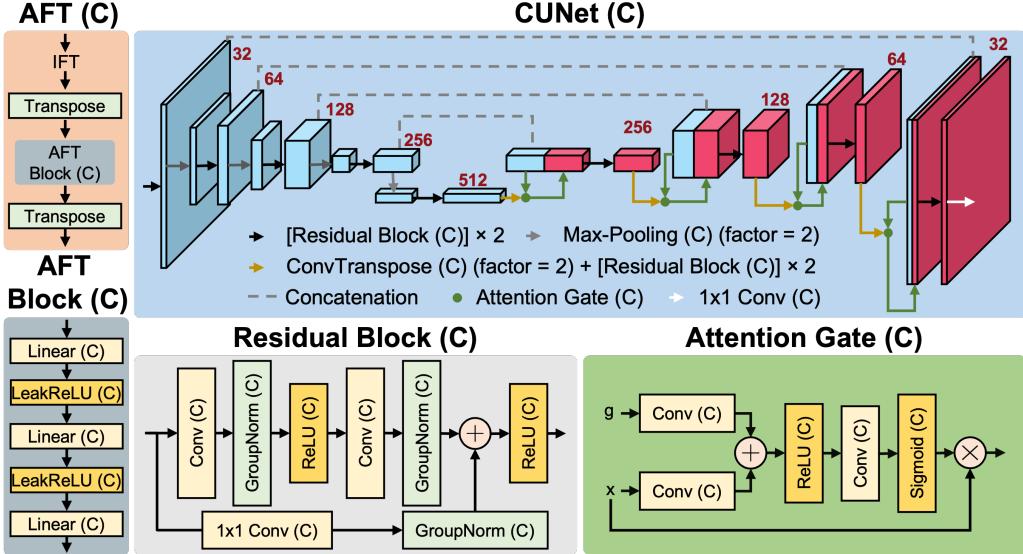


Figure 2: Structure of 2-dimensional AFTNet. Components include the complex-valued AFT block, the complex-valued residual attention UNet, the complex-valued residual block, and the complex-valued attention gate. All convolutional layers have a kernel size of 3, except those pointed out specifically. C: complex-valued. Red numbers indicate the number of channels produced by each layer.

### 3. Methods

#### 3.1. Artificial Fourier Transform

Take cartesian sampling which is widely used in conventional MRI acquisitions as an example, since 2D discrete Fourier transform (DFT) is a linear operation and can be represented by two successive 1D DFTs as

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

where the  $x$  and  $y$  are two separable independent variables. Each dimension of 2D DFT can be modeled as a trainable neural network [11]. We further proposed that this idea could be naturally implemented with CVNNs 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

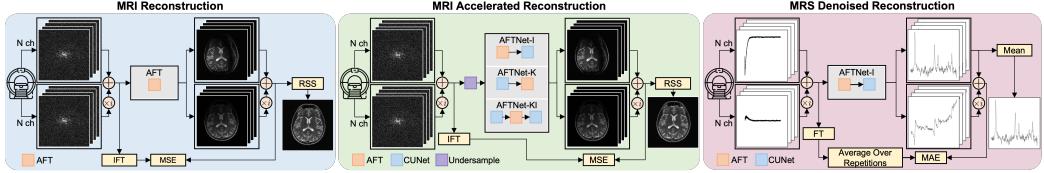


Figure 3: The workflows of experiments on each dataset.

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 the real and imaginary parts of the DFT operation. We have the following:

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

Compared Equation (13) with Equation (1), a multilayer CVNN 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. The AFT block is composed of an MLP with three complex linear layers linked by two complex LeakyReLU activation layers. As for conventional accelerated MRI acquisition, the undersampling is applied to the phase-encoding direction, and we also only apply the AFT block to that direction. For the accelerated reconstruction task, we combine our AFT with an entirely complex-valued UNet (CUNet) [21] with residual attention, which

extracts local and global features in the k-space and/or image domain. Multiple network architectures are evaluated to verify the effectiveness of both AFT and CUNet in different domains. We refer each of them to **AFT**, **AFTNet-I**, **AFTNet-K**, and **AFTNet-KI**, respectively as shown in Figure 3.

The architecture of the CUNet presented in Figure 2 is generally based on the residual attention UNet [20] but with all the components of real value replaced by complex value components as shown in Figure 2, including complex-valued convolutional layers and complex-valued ReLU layers introduced in Section 2.1. We further optimize the network for smaller batch sizes by replacing batch normalization with group normalization as illustrated in Algorithm 1. Other complex components are implemented in the same way. For example, the complex transposed convolution operator can be mirrored from Equation (3), the complex sigmoid is applied like the complex ReLU, and the complex max pooling is almost the same as the real-valued version except that the indices are inferred from absolute values.

### 3.3. Reconstruction Loss

In the context of image reconstruction and processing, the impact of the loss function is vital if human observers are to evaluate the final results. One common and safe choice is  $\ell_2$  loss which works under the assumption of Gaussian white 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.  $x$  and  $y$  are predicted and targeted complex-valued images. For training AFTNet for accelerated MRI reconstruction, we also want to minimize the error of magnitude images. Therefore, the loss value for accelerated MRI reconstruction is

$$\mathcal{L}^{acc} = \mathcal{L}^{recon} + \mathcal{L}^{\ell_2}(\text{RSS}(x), \text{RSS}(y)) \quad (16)$$

where the root-sum-of-squares (RSS) approach [26] is applied to complex-valued output from the model to generate the optimal, unbiased estimate of magnitude image which is used for loss calculation. For denoised MRS reconstruction, we use  $\ell_1$  loss following the previous practice [27] so the loss is

$$\mathcal{L}^{denoise} = \mathcal{L}^{\ell_1}(\Re(x), \Re(y)) + \mathcal{L}^{\ell_1}(\Im(x), \Im(y)) \quad (17)$$

Data consistency is applied after each RACUNet and AFT block so that the known positions of the output data are replaced by the original data samples to obtain better fidelity [28].

## 4. Experiments

### 4.1. Overview

In the MRI reconstruction experiment, we apply our AFT to the multicoil k-space data acquired directly from the scanner for the reconstruction task. The target is derived from the fast inverse 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. Network performance is evaluated within magnitude images obtained by the Fourier application and coil compression, which are crucial for radiologists in clinical applications. We also recognize some clinical settings that rely on the phase image, so we also provide evaluation on the phase images.

For the MRI accelerated reconstruction experiment, we first trained AFT, CUNet (DFTNet-I) and DFTNet-K models. Then, we replace the DFT with AFT in the DFTNet to train the AFTNet (I and K) models. At last, we concatenate the pretrained K model with I model to train the KI model. To be more specific, we train an AFT-only network to see if, without convolution layers, the AFT can remove artifacts 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 in 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 spaces, even if they have a fixed field of view. Finally, a CUNet-AFT-CUNet structure is evaluated with the first CUNet that extracts k-space domain features and the second CUNet extracting image domain features.

We also extended AFTNet to 1D scenario for spectral reconstruction to evaluate its generality. In such case, we use the AFTNet-I model and train it from scratch.

#### 4.2. Experimental data

In this study, two datasets were used: fastMRI dataset [29] and Big GABA dataset [30]. The proposed methods were trained on these datasets separately.

The fastMRI dataset contains fully sampled brain MRIs obtained on 3 and 1.5 Tesla magnets. We selected 4-channel axial T1 and T2-weighted scans from the raw fastMRI dataset. We applied a data filtering strategy by counting the number of images with similar shapes and field strength, and selecting the groups that have more than 100 samples. This process resulted in 671 T2-weighted images at 1.5T, 157 T2-weighted images at 3T, and 115 T1-weighted images at 3T. All selected images are acquired on Siemens scanner. A total of 943 scans were used with 794, 99, and 50 each for the training, validation, and test set. The detailed scan parameters are listed in Table A.1. Data preprocessing includes normalization to the maximum intensity value of one in the image domain and cropping to  $640 \times 320$ . For experiment on accelerated reconstruction, the fully sampled k-space data was undersampled by applying a mask in the phase-encoding direction. We use the acceleration rate (or acceleration factor) to denote the level of scan time reduced for the undersampled k-space data, which is defined as the ratio of the amount of k-space data required for a fully sampled image to the amount collected in an undersampled k-space data [31]. The sampling ratio, SR, is also used to denote the information retained in the undersampled k-space data, which is defined as the inverse of the acceleration rate. An equispaced mask with approximate acceleration matching is used to undersample the k-space data. The fraction of low-frequency columns to be retained for acceleration rates 2x, 4x, and 8x is 16%, 8%, and 4%, respectively.

The *in vivo* MRS dataset used in this study is from the BIG GABA [30] repository and contains GABA-edited MEGA-PRESS data acquired on 3T Philips scanners from multiple sites. The voxel was placed in the medial parietal lobe of healthy subjects aged 18 to 36 years. The data consist of 101 pairs of edit-ON/-OFF spectra, each with 160 ON averages and OFF averages. The AFTNet was trained with an input size of 2048. The ground truth of the ON/OFF spectra is derived by taking the average over 160 transients. We denote the ground truth as noiseless signals. For training, we randomly sampled transients from each subject to reconstruct the averaged signal. By varying the number of transients in computing the average sampled acquisitions, we can generate signals with different levels of signal-to-noise ratio (SNR). We use the term reduction rate (R) to refer to the ratio of the total

number of transients and the number of transients used in calculating the average. Being able to reconstruct accurate denoised signals at high R has implications for reducing acquisition time. From a total number of 101 scans in the dataset, 80, 10, and 11 scans were assigned for the training, validation, and test set, respectively. All spectra were normalized to the maximum value of its magnitude. In denoised MRS reconstruction, we use the reduction rate, R, to denote the noise level, which is defined as the ratio of the number of total repetitions (160 for this study) to the number of repetitions retained for a noisy input. We generate noisy free induction decay (FIDs) with 5 reduction rates of 10, 20, 40, 80, and 160.

#### 4.3. Measurement of Reconstruction Quality

Three metrics were adopted for the quantitative evaluation of image quality compared to ground truth: structural similarity (SSIM) [32], peak signal-to-noise ratio (PSNR), and normalized root mean squared error (NRMSE). For the quality measurement of the 1D spectra, three other metrics were used: Pearson’s correlation coefficient (PCC), Spearman’s rank correlation coefficient (SCC) and goodness fitting coefficient (GFC) [33]. The GFC is introduced to evaluate the goodness of the mathematical reconstruction with a value ranging from 0 to 1, where 1 indicates a perfect reconstruction. If  $\hat{y}_i$  is the predicted value of the  $i$ -th sample and  $y_i$  is the corresponding true value, then the GFC estimated over  $n_{\text{samples}}$  is defined as

$$\text{GFC}(y, \hat{y}) = \frac{|\sum_{i=0}^{n_{\text{samples}}-1} y_i \hat{y}_i|}{\sqrt{\sum_{i=0}^{n_{\text{samples}}-1} y_i^2} \sqrt{\sum_{i=0}^{n_{\text{samples}}-1} \hat{y}_i^2}}. \quad (18)$$

#### 4.4. Implementation details

We construct a batch size of 1 and optimize the network using the ADAM [34] optimizer. We use a cosine annealing learning rate scheduler [35]. The initial learning rate is  $10^{-3}$  and the final learning rate is  $10^{-5}$ . We set the negative slope of LeakyReLU to 0.1. All methods are trained for 50 epochs. All experiments were performed with PyTorch 2.2.2 and a Quadro RTX 6000 GPU.

For experiment on the MRI accelerated reconstruction, we compared AFTNet with zero-filling (DFT), numerical method GRAPPA [36], real-valued model KIKI-net [37] and diffusion model score-MRI [13]. For experiment on the MRS denoised reconstruction, we compared AFTNet with Gaussian line broadening (GLB).

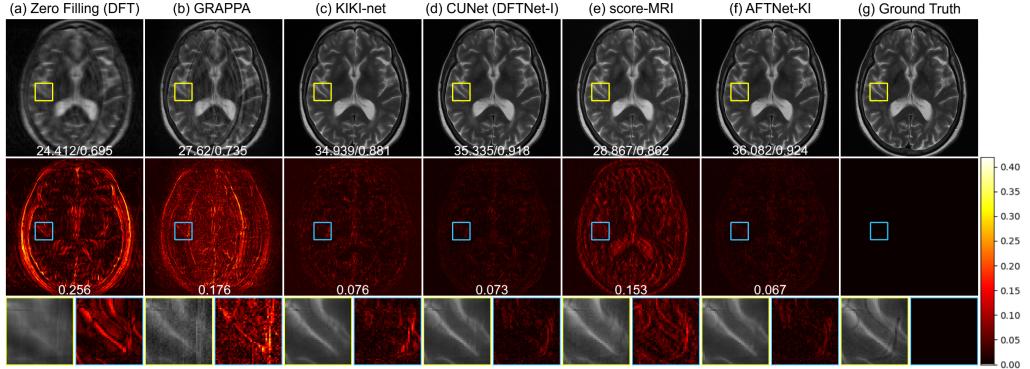


Figure 4: Comparison between baseline methods and AFTNet. Here we present the qualitative results for a T2w image on 3T

under 4x acceleration rate, where a 1D 4x equal-spaced sampling mask with 8% of low-frequency columns retained is implemented to undersample the input k-space data. Residual map shows the difference magnitude against ground truth (in Hot colormap). Yellow and blue boxes show the zoomed-in version of the indicated area. White numbers in the magnitude images indicate PSNR (dB) and SSIM. White numbers in the residual maps indicate NRMSE.

## 5. Results

### 5.1. Comparison study

The effectiveness of AFTNet was evaluated on the test set of the human normal-field MRI dataset through a comparative study as shown in Figure 4, where the up row shows the visualization of the baseline models, including conventional Fourier transform-based networks and score-MRI. The bottom row shows the proposed networks and ground truth derived from the fully sampled k-space data. Both the residual maps against the ground truth and the zoomed-in areas are presented. The residual maps are normalized to the maximum value of 0.5 such that they are comparable across all models. We first compare the AFT-only network with the results after DFT. It clearly shows that the light-trainable AFT block not only approximates the DFT but also facilitates artifact removal with the presence of signal non-ideality. Comparing the results of the AFT-only network and score-MRI, which solves the image reconstruction inverse problem based on score-based generative models, we demonstrate that even with such a simple structure, AFT can achieve similar performance in terms of PSNR and NRMSE. The score-MRI

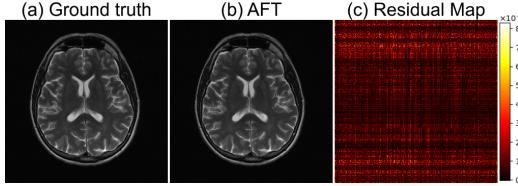


Figure 5: Fully sampled MRI reconstruction results. Here we present the qualitative results of a T2w image on 1.5T. (a) Ground truth, (b) proposed method, (c) difference magnitude of (a) and (b) (in Hot colormap).

was originally trained on the knee dataset and we fine-tuned it on our dataset. Qualitatively, AFTNet outperformed score-MRI, as can be seen from the difference magnitude map and the zoomed-in area in Figure 4 by comparing the results of AFTNet and score-MRI. Quantitative metrics demonstrate the superior performance of AFTNet over score-MRI and KIKI-net, as shown in Table I including SSIM, PSNR, and NRMSE in all acceleration rates. The statistical t-test between the score-MRI and AFTNet metrics also shows the superior accelerated reconstruction of the proposed method with a p-value below 0.0001. Furthermore, the AFT block can serve as a better replacement for DFT by comparing the results of DFTNet and AFTNet, where all AFTNet shows better performance than DFTNet, as can be seen from the residual maps. It should be pointed out that AFTNet-KI significantly outperforms DFTNet-KI, indicating that AFT can be implemented to connect two networks while preserving the capability of both networks.

### 5.2. Ablation study

Different AFTNet structures, as mentioned in Section 3.2, were compared in datasets with different field strengths and different modalities to verify the stability and generality of the AFTNet. In addition, the effectiveness of the front-end/back-end convolutional networks is also evaluated in this section. To validate the robustness of AFTNet to k-space artifacts, these proposed AFTNet structures were compared on image reconstruction and accelerated reconstruction as described in Section 4.2. Furthermore, the extended AFTNet was compared with numerical methods using 1-dimensional MRS FID data on the denoised reconstruction.

First, we show the results of human 1.5/3T MRI reconstruction using fully sampled fastMRI k-space data in Figure 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

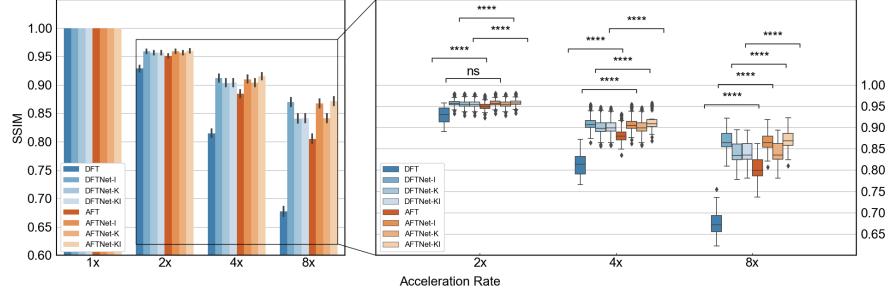


Figure 6: Results of human 1.5/3T MRI accelerated reconstruction by comparing AFTNet (K Model, I Model, and KI Model) and with their DFT counterparts in terms of SSIM for acceleration rates 1x, 2x, 4x, and 8x. p-values indicate results from two-sided t-tests for paired samples between the DFT-based model and AFT-based model. (ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ )

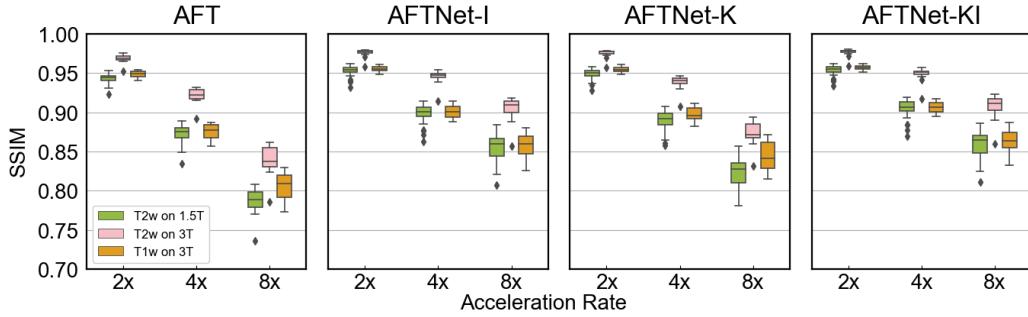


Figure 7: Results of human 1.5/3T MRI accelerated reconstruction by comparing AFTNet (I Model, K Model, and KI Model) in terms of SSIM on different acquisition types and system field strength for acceleration rates 2x, 4x, and 8x.

ground truth image is derived by applying the conventional Fourier transformation to the k-space 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 Figure 4, we show the results of MRI accelerated reconstruction using under-sampled fastMRI k-space data. In the first row, we see the reconstructions from the 1D 4x equal space sampling, in which 8% of low-frequency

columns are retained (see Figure A.1). Here, we compare different AFTNet structures with the zero-filling method. AFTNet-KI performs outstanding reconstruction, where less structural difference can be seen from the residual map in the second row. The third row shows the zoomed-in areas of both the images and residual maps. AFTNet-I produces a more blurry reconstruction which loses the structural details. Reconstruction through AFTNet-K induces foggy artifacts, which are reflected in terms of SSIM. Figure 6 shows the results of the accelerated reconstruction comparing AFT and AFTNet (I, K, and KI) with their counterparts in terms of SSIM in acceleration rates of 2x, 4x and 8x. The performance of DFT drops linearly as the acceleration rate increases, while the AFTNet methods are more robust to the acceleration scale. The t-test between AFT-based and DFT-based structures indicates that AFTNet significantly outperforms all DFTNet structures except the I model in the 2x acceleration rate. The results of AFTNet on different acquisition types and system field strength are shown in Figure 7. The difference of performance is mainly caused by the image quality between the 3T and 1.5T images and the inrisint contrast difference between T2w and T1w images.

A comprehensive comparison of quantitative metrics on the test set is provided in Table 1 and Table 3 for accelerated reconstruction. In addition to the metrics evaluated on the magnitude images, we provide the evaluation on the phase images in Table 2. AFTNet-KI outperforms other AFTNet structures on all the different acceleration rates, and both AFTNet-KI and AFTNet-I perform significantly better than other AFTNet-K structures. Table 3 shows detailed quantitative metrics of the human 1.5/3T MRI accelerated reconstruction results grouped by image contrast and system field strength. It is worth mentioning that although AFTNet-K does not outperform other AFTNet structures in the accelerated reconstruction task, it demonstrates the ability to learn in a sparse frequency domain and its sparse representations with a complex-valued convolutional network.

### 5.3. Generality of AFTNet

The results of the proposed AFTNet approach and traditional DFT method with and without GLB are illustrated in Figure 8. The first row shows the reconstructed spectra from the methods being compared. The second row shows the reconstructed spectra overlaid with the ground truth. The third row shows the difference between the reconstructed spectra and the ground truth. The reconstructed spectra illustrated here represent a reduction rate

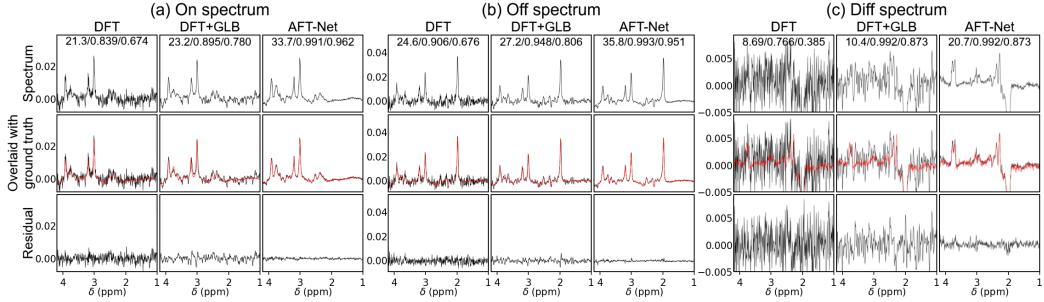


Figure 8: Human 3T MRS denoised reconstruction results. The acceleration rate is 80 for each spectrum. (a) Reconstruction results for the ON spectrum, (b) reconstruction results for the OFF spectrum, and (c) results for the DIFF spectrum derived from (a) and (b). 1st row: reconstructed spectra, 2nd row: reconstructed spectra overlaid with ground truth (in red line), 3rd row: Difference of reconstructed spectra against ground truth. Black numbers in the upper center location indicate PSNR (db), PCC, and SCC, respectively.

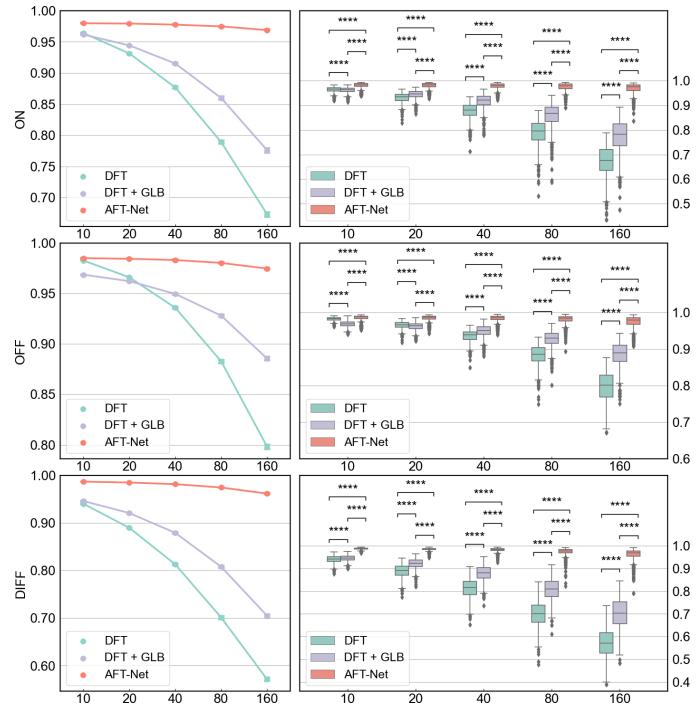


Figure 9: Results of human 3T MRS denoised reconstruction by comparing AFTNet (I Model), DFT, and DFT with Gaussian line broadening in terms of GFC. p-values indicate results from two-sided t-tests for paired samples. (ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ )

of 80, where only two out of a total of 160 transients were used to construct the spectrum. The AFTNet approach outperforms the other methods in all reduction rates listed in the table. We used the GFC to measure the similarity between the reconstructed spectra and the ground truth, as shown in Table 4. Although the GFC decreases as the reduction rate increases, the AFTNet method is able to maintain a high performance even at very high reduction rates. For example, the difference in GFC between the OFF spectra under a reduction rate of 10 and 160 is very small, around 0.9798 and 0.9688, respectively.

## 6. Discussion

In this study, a unified MR image reconstruction framework is proposed, which is composed of two main components: artificial Fourier transform block and complex-valued residual attention UNet. The AFT block is used to approximate conventional DFT, which is demonstrated in Table 1 and shown in Section 3.1. The front-end/back-end convolutional layers are used to extract features at different levels in the k-space/image domains and play different roles in various tasks. As shown in Table 1, both front- and back-end convolutional layers show superior accelerated reconstruction performance under all sampling ratios compared to single front-end/back-end convolutional layers. This is potentially because the undersampling is performed in k-space where the artifacts are separated from the non-artifact. While in the image domain, it is converted to an aliasing overlapped over the whole image. The artifact removal task can be recast as an image inpainting problem in the k-space domain which can be done more easily by the front-end convolutional layers. However, all structures with front-end-only convolutional layers show lower performance, indicating that the sparse representation of k-space data makes it harder for a convolutional network to extract noise information in the low-frequency areas and back-end convolutional layers are necessary to achieve the optimal performance.

Domain-transform manifold learning has been introduced for years, and several deep learning frameworks were developed based on this idea. The first model, AUTOMAP [10], proposed the simple FC-Conv structure which can only be applied to images with a small matrix size due to its redundant FC layers. DOTA-MRI [12] extended AUTOMAP to Conv-FC-Conv structure. However, it did not extend the model structure and implement CVNNs. The AFTNet we proposed in this study solves the problem mentioned above

through a modular-designed AFT block. We also demonstrated that the extended AFTNet can also be applied to 1D data in Section 5.3. In addition, previous works define the loss in the magnitude image, while we calculate the loss in the complex-valued image domain, which preserves the relations between the real and imaginary parts. The phase is then derived from the output of AFTNet, which is essential for several phase-based applications, such as flow quantification and fat–water separation.

CVNNs, especially complex-valued convolutional networks [22] [21], have been studied for MRI reconstruction, but have mainly focused on simple tasks or only applied them to the image domain. We investigate the different impacts of complex-valued convolutional networks on the k-space and image domain and extend the application to accelerated reconstruction and denoised reconstruction, which are more clinically important. We also incorporate domain-manifold learning by adding domain transform blocks that determine the mapping between the k-space and the image domain instead of the conventional discrete Fourier transform. It is more robust to noise and signal non-ideality due to imperfect acquisition. We also extend the application of complex-valued convolutional networks to 1D MRS denoised reconstruction, which has not been studied in previous work. Our results indicate that AFTNet is able to effectively decrease the contribution of the noise in the FIDs while preserving the quantifiable signals in the reconstructed spectra.

One remaining methodological limitation is that the FC layers used by AFTNet narrow the application to datasets with various image matrix sizes. Although the convolutional layers are not sensitive to the image matrix sizes and cropping/padding can be applied to match the desired sizes, the features of FC layers need to be selected carefully which requires further investigation. Another parameter that needs to be taken into account is the coil number. In this study, we selected especially 4-channel MRI data for convenience of data preprocessing, while deep learning-based coil combinations could be incorporated into the framework in future work. Furthermore, diffusion models are shown to be a powerful tool for image reconstruction across body regions and coil numbers [13]. However, the score-MRI we compared in this study does not demonstrate superior performance compared with AFTNet and the inference stage time is long. This is potential because the backbone of the score-MRI is still a real-valued UNet structure and the relation between the real and imaginary part is not considered during the calculation of the score function. For future works, the AFTNet could be further extended by leveraging diffusion-based models with complex-valued convolutional networks as

the backbone and careful optimization to reduce the inference time.

## 7. 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 AFTNet, which implements complex-valued UNet to extract features in the 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 AFTNet achieves competitive results compared with other methods and is found to be robust to noise differences. An extensive study on various system fields, various modalities, and various tasks demonstrates the effectiveness and generality of AFTNet. Our current claims are supported by quantitative and objective visual analyses. In future works, neuroradiologist evaluations could be incorporated to further strengthen the qualitative assessment.

## References

- [1] G. Georgiou, C. Koutsougeras, Complex domain backpropagation, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* (1992).
- [2] N. Guberman, On complex valued convolutional neural networks, arXiv preprint arXiv:1602.09046 (2016).
- [3] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, C. J. Pal, Deep complex networks, *International Conference on Learning Representations* (2018).
- [4] J. B. Johnson, Thermal agitation of electricity in conductors, *Physical review* (1928).

- [5] H. Nyquist, Thermal agitation of electric charge in conductors, *Physical review* (1928).
- [6] M. S. Hansen, P. Kellman, Image reconstruction: an overview for clinicians, *Journal of Magnetic Resonance Imaging* (2015).
- [7] J. A. Fessler, Model-based image reconstruction for mri, *IEEE signal processing magazine* (2010).
- [8] M. De Bruijne, Machine learning approaches in medical image analysis: From detection to diagnosis (2016).
- [9] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, *Annual review of biomedical engineering* (2017).
- [10] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, M. S. Rosen, Image reconstruction by domain-transform manifold learning, *Nature* (2018).
- [11] J. López-Randulfe, T. Duswald, Z. Bing, A. Knoll, Spiking neural network for fourier transform and object detection for automotive radar, *Frontiers in Neurorobotics* (2021).
- [12] T. Eo, H. Shin, Y. Jun, T. Kim, D. Hwang, Accelerating cartesian mri by domain-transform manifold learning in phase-encoding direction, *Medical Image Analysis* (2020).
- [13] H. Chung, J. C. Ye, Score-based diffusion models for accelerated mri, *Medical image analysis* (2022).
- [14] X. Zhao, T. Yang, B. Li, X. Zhang, Swigan: A dual-domain swin transformer-based generative adversarial network for mri reconstruction, *Computers in Biology and Medicine* (2023).
- [15] Y. Yan, T. Yang, X. Zhao, C. Jiao, A. Yang, J. Miao, Dc-siamnet: Deep contrastive siamese network for self-supervised mri reconstruction, *Computers in Biology and Medicine* (2023).
- [16] K. Xu, M. Qin, F. Sun, Y. Wang, Y.-K. Chen, F. Ren, Learning in the frequency domain, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020).
- [17] A. Hirose, Complex-valued neural networks, Wiley Online Library, 2012.

- [18] A. Hirose, S. Yoshida, Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence, *IEEE Transactions on Neural Networks and learning systems* (2012).
- [19] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, A. Szlam, A mathematical motivation for complex-valued convolutional networks, *Neural computation* (2016).
- [20] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18 (2015).
- [21] D. Sikka, N. Igra, S. Gjerwold-Sellec, C. Gao, E. Wu, J. Guo, Cu-net: A completely complex u-net for mr k-space signal processing, *ISMRM (International Society of Magnetic Resonance Imaging) Virtual Conference & Exhibition, 2021* (2021).
- [22] E. Cole, J. Cheng, J. Pauly, S. Vasanawala, Analysis of deep complex-valued convolutional neural networks for mri reconstruction and phase-focused applications, *Magnetic resonance in medicine* (2021).
- [23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International conference on machine learning* (2015).
- [24] Y. Wu, K. He, Group normalization, *Proceedings of the European conference on computer vision (ECCV)* (2018).
- [25] J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization, *arXiv preprint arXiv:1607.06450* (2016).
- [26] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, O. M. Mueller, The nmr phased array, *Magnetic resonance in medicine* (1990).
- [27] D. J. Ma, Y. Yang, N. Harguindeguy, Y. Tian, S. A. Small, F. Liu, D. L. Rothman, J. Guo, Magnetic resonance spectroscopy spectral registration using deep learning, *Journal of Magnetic Resonance Imaging* (2024).

- [28] Z. Wu, W. Liao, C. Yan, M. Zhao, G. Liu, N. Ma, X. Li, Deep learning based mri reconstruction with transformer, Computer Methods and Programs in Biomedicine (2023).
- [29] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, et al., fastmri: An open dataset and benchmarks for accelerated mri, arXiv preprint arXiv:1811.08839 (2018).
- [30] M. Mikkelsen, P. B. Barker, P. K. Bhattacharyya, M. K. Brix, P. F. Buur, K. M. Cecil, K. L. Chan, D. Y.-T. Chen, A. R. Craven, K. Cuypers, et al., Big gaba: Edited mr spectroscopy at 24 research sites, Neuroimage (2017).
- [31] A. Deshmane, V. Gulani, M. A. Griswold, N. Seiberlich, Parallel mr imaging, Journal of Magnetic Resonance Imaging (2012).
- [32] Z. Wang, E. P. Simoncelli, A. C. Bovik, Multiscale structural similarity for image quality assessment, The Thirity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003 (2003).
- [33] J. Romero, A. García-Beltrán, J. Hernández-Andrés, Linear bases for representation of natural and artificial illuminants, JOSA A (1997).
- [34] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [35] I. Loshchilov, F. Hutter, SGDR: stochastic gradient descent with warm restarts, 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings (2017).
- [36] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, A. Haase, Generalized autocalibrating partially parallel acquisitions (grappa), Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine (2002).
- [37] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, D. Hwang, Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images, Magnetic resonance in medicine (2018).

## Tables

Acc.	Metrics	Baseline					Ours			
		DFT	Grappa	KIKI-net	score-MRI	DFTNet-I	AFT	AFTNet-K	AFTNet-I	AFTNet-KI
1x	SSIM	-	1.000	0.998	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
	PSNR	-	<b>161.08</b>	46.8	63.9	153.3	153.3	153.3	153.3	153.3
	NRMSE	-	0.000	0.024	0.003	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
2x	SSIM	0.929	0.928	0.948	0.939	0.959	0.951	0.957	0.959	<b>0.960***</b>
	PSNR	33.6	35.27	39.2	38.5	39.5	37.0	38.9	39.6	<b>39.8***</b>
	NRMSE	0.105	0.088	0.06	0.052	0.053	0.070	0.057	0.052	<b>0.051***</b>
4x	SSIM	0.815	0.835	0.870	0.874	0.912	0.885	0.904	0.912	<b>0.915***</b>
	PSNR	27.3	29.16	34.7	32.8	35.2	32.0	33.7	35.3	<b>35.7***</b>
	NRMSE	0.214	0.173	0.092	0.099	0.086	0.125	0.103	0.085	<b>0.082***</b>
8x	SSIM	0.677	0.744	0.783	0.787	0.869	0.805	0.841	0.869	<b>0.872***</b>
	PSNR	23.8	25.29	31.4	27.5	31.3	27.6	29.4	31.3	<b>31.5***</b>
	NRMSE	0.32	0.271	0.134	0.189	0.135	0.208	0.171	0.137	<b>0.132***</b>

Table 1: Quantitative metrics of human 1.5/3T MRI accelerated reconstruction for magnitude images. Numbers are presented as mean value  $\pm$  standard deviation. Numbers in boldface indicate the best metric out of all the methods. Acc.: Acceleration Rate. p-values indicate results from two-sided t-tests for paired samples between best and second-best models. (ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ )

Acc.	Metrics	Baseline				Ours			
		DFT	DFTNet-K	DFTNet-I	DFTNet-KI	AFT	AFTNet-K	AFTNet-I	AFTNet-KI
1x	SSIM	-	<b>1.000</b>						
	PSNR	-	<b>126.36</b>						
	NRMSE	-	<b>0.000</b>						
2x	SSIM	0.687	0.730	0.746	0.749	0.717	0.730	0.747	<b>0.752****</b>
	PSNR	10.50	11.00	11.24	11.28	10.81	11.00	11.26	<b>11.32****</b>
	NRMSE	0.671	0.634	0.617	0.614	0.647	0.635	0.615	<b>0.611****</b>
4x	SSIM	0.585	0.610	0.666	0.669	0.633	0.610	0.669	<b>0.672****</b>
	PSNR	9.31	9.59	10.03	10.06	9.65	9.59	10.06	<b>10.10****</b>
	NRMSE	0.766	0.743	0.708	0.705	0.738	0.743	0.705	<b>0.702****</b>
8x	SSIM	0.561	0.561	0.618	0.621	0.590	0.559	0.620	<b>0.623****</b>
	PSNR	8.84	8.87	9.33	9.36	9.07	8.87	9.34	<b>9.40***</b>
	NRMSE	0.807	0.805	0.766	0.763	0.787	0.805	0.765	<b>0.760***</b>

Table 2: Quantitative metrics of human 1.5/3T MRI accelerated reconstruction for phase images. Numbers are presented as mean value  $\pm$  standard deviation. Numbers in boldface indicate the best metric out of all the methods. Acc.: Acceleration Rate. p-values indicate results from two-sided t-tests for paired samples between best and second-best models. (ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ )

(a) T2w images on 1.5T

Acc.	Metrics	DFT	Ours			
			AFT	AFTNet-K	AFTNet-I	AFTNet-KI
1x	SSIM	-	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>
	PSNR	-	<b>154.0 ± 1.3</b>	<b>154.0 ± 1.3</b>	<b>154.0 ± 1.3</b>	<b>154.0 ± 1.3</b>
	NRMSE	-	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>
2x	SSIM	0.911 ± 0.011	0.943 ± 0.008	0.948 ± 0.008	0.952 ± 0.008	<b>0.953 ± 0.007</b>
	PSNR	32.5 ± 1.1	36.5 ± 1.2	38.0 ± 1.3	39.1 ± 1.2	<b>39.2 ± 1.2</b>
	NRMSE	0.127 ± 0.008	0.080 ± 0.004	0.068 ± 0.005	0.060 ± 0.004	<b>0.059 ± 0.005</b>
4x	SSIM	0.792 ± 0.014	0.871 ± 0.013	0.890 ± 0.014	0.897 ± 0.014	<b>0.903 ± 0.013</b>
	PSNR	26.7 ± 1.0	31.7 ± 1.4	32.8 ± 1.3	35.0 ± 1.2	<b>35.3 ± 1.2</b>
	NRMSE	0.248 ± 0.015	0.141 ± 0.006	0.124 ± 0.007	0.095 ± 0.007	<b>0.093 ± 0.007</b>
8x	SSIM	0.662 ± 0.023	0.788 ± 0.016	0.823 ± 0.020	0.855 ± 0.019	<b>0.859 ± 0.018</b>
	PSNR	23.6 ± 1.1	27.0 ± 1.4	28.4 ± 1.5	<b>30.9 ± 1.3</b>	<b>30.9 ± 1.3</b>
	NRMSE	0.358 ± 0.022	0.240 ± 0.011	0.204 ± 0.010	<b>0.153 ± 0.008</b>	<b>0.153 ± 0.009</b>

(b) T2w images on 3T

Acc.	Metrics	DFT	Ours			
			AFT	AFTNet-K	AFTNet-I	AFTNet-KI
1x	SSIM	-	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>
	PSNR	-	<b>152.3 ± 1.2</b>	<b>152.3 ± 1.2</b>	<b>152.3 ± 1.2</b>	<b>152.3 ± 1.2</b>
	NRMSE	-	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>
2x	SSIM	0.948 ± 0.007	0.968 ± 0.006	0.975 ± 0.006	<b>0.976 ± 0.006</b>	<b>0.976 ± 0.006</b>
	PSNR	33.1 ± 1.1	36.6 ± 1.3	39.4 ± 1.5	40.2 ± 1.5	<b>40.4 ± 1.6</b>
	NRMSE	0.093 ± 0.009	0.063 ± 0.006	0.046 ± 0.006	0.042 ± 0.006	<b>0.041 ± 0.006</b>
4x	SSIM	0.841 ± 0.021	0.921 ± 0.011	0.937 ± 0.011	0.945 ± 0.010	<b>0.948 ± 0.010</b>
	PSNR	26.7 ± 1.2	31.1 ± 1.3	33.3 ± 1.4	35.2 ± 1.4	<b>35.5 ± 1.5</b>
	NRMSE	0.195 ± 0.017	0.117 ± 0.007	0.091 ± 0.006	0.074 ± 0.008	<b>0.071 ± 0.008</b>
8x	SSIM	0.694 ± 0.031	0.838 ± 0.021	0.872 ± 0.017	0.903 ± 0.017	<b>0.906 ± 0.017</b>
	PSNR	23.1 ± 1.3	26.6 ± 1.3	28.2 ± 1.3	30.5 ± 1.2	<b>30.7 ± 1.3</b>
	NRMSE	0.295 ± 0.022	0.198 ± 0.011	0.164 ± 0.009	0.126 ± 0.010	<b>0.124 ± 0.010</b>

(c) T1w images on 3T

Acc.	Metrics	DFT	Ours			
			AFT	AFTNet-K	AFTNet-I	AFTNet-KI
1x	SSIM	-	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>	<b>1.000 ± 0.000</b>
	PSNR	-	<b>153.0 ± 0.8</b>	<b>153.0 ± 0.8</b>	<b>153.0 ± 0.8</b>	<b>153.0 ± 0.8</b>
	NRMSE	-	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>	<b>0.000 ± 0.000</b>
2x	SSIM	0.941 ± 0.006	0.949 ± 0.004	0.955 ± 0.004	0.956 ± 0.004	<b>0.957 ± 0.003</b>
	PSNR	35.8 ± 0.7	38.3 ± 0.7	40.0 ± 0.8	40.1 ± 0.9	<b>40.4 ± 0.8</b>
	NRMSE	0.079 ± 0.008	0.059 ± 0.004	0.048 ± 0.003	0.048 ± 0.004	<b>0.046 ± 0.003</b>
4x	SSIM	0.830 ± 0.014	0.875 ± 0.010	0.898 ± 0.009	0.901 ± 0.008	<b>0.906 ± 0.008</b>
	PSNR	28.9 ± 0.7	33.2 ± 0.8	35.7 ± 0.9	35.9 ± 1.1	<b>36.4 ± 1.0</b>
	NRMSE	0.174 ± 0.020	0.106 ± 0.008	0.080 ± 0.006	0.077 ± 0.009	<b>0.073 ± 0.007</b>
8x	SSIM	0.687 ± 0.024	0.805 ± 0.019	0.844 ± 0.019	0.858 ± 0.016	<b>0.863 ± 0.015</b>
	PSNR	24.7 ± 0.9	29.5 ± 0.8	32.1 ± 0.9	33.0 ± 1.3	<b>33.3 ± 1.2</b>
	NRMSE	0.281 ± 0.027	0.162 ± 0.013	0.121 ± 0.009	0.109 ± 0.015	<b>0.104 ± 0.013</b>

Table 3: 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.

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

Table 4: GFC metric of human 3T MRS denoised reconstruction. Numbers are presented as mean value  $\pm$  standard deviation. Numbers in boldface indicate the best metric out of all the methods.

## Appendix A.

System Model	Field Strength (T)	Institution Name	Protocol Name	Encoded Matrix Size	Field of View (mm)	TR (ms)	TE (ms)	Sequence Type	Count
Avanto	1.494	NYU	AX	640 x 272 x 1	440 x 186.9725 x 7.5	5120	103	TurboSpinEcho	1
Avanto	1.494	NYU	AX	640 x 320 x 1	440 x 220 x 7.5	4000	107	TurboSpinEcho	1
Avanto	1.494	NYU	AX	640 x 320 x 1	440 x 220 x 7.5	5120	103	TurboSpinEcho	373
Avanto	1.494	NYU	AX	640 x 320 x 1	440 x 220 x 7.5	5120	107	TurboSpinEcho	295
Avanto	1.494	NYU Medical Center	AX T2.FBB	640 x 320 x 1	440 x 220 x 7.5	5120	103	TurboSpinEcho	1
TrioTim	2.8936	NYU	AX	768 x 308 x 1	440 x 176.5392 x 7.5	6000	107	TurboSpinEcho	1
TrioTim	2.8936	NYU	AX	768 x 350 x 1	440 x 200.4446 x 7.5	6000	107	TurboSpinEcho	1
TrioTim	2.8936	NYU	AX	768 x 392 x 1	440 x 224.62 x 7.5	6000	107	TurboSpinEcho	141
TrioTim	2.8936	NYU	AX	768 x 392 x 1	460 x 234.83 x 7.5	6000	107	TurboSpinEcho	14
TrioTim	2.8936	NYU Clinical Cancer Ctr	AX T1 PRE_FBB	640 x 260 x 1	440 x 178.75 x 5	264	2.88	Flash	1
TrioTim	2.8936	NYU Clinical Cancer Ctr	AX T1 PRE_FBB	640 x 320 x 1	440 x 220 x 5	250	2.88	Flash	1
TrioTim	2.8936	NYU Clinical Cancer Ctr	AX T1 PRE_FBB	640 x 320 x 1	440 x 220 x 5	264	2.88	Flash	106
TrioTim	2.8936	NYU Clinical Cancer Ctr	AX T1 PRE_FBB	640 x 320 x 1	460 x 230 x 5	264	2.88	Flash	5
TrioTim	2.8936	NYU Clinical Cancer Ctr	AX T1 PRE_FBB	640 x 320 x 1	480 x 240 x 5	264	2.88	Flash	2

Table A.1: Imaging parameters. The name AX indicates T2w imaging protocol.

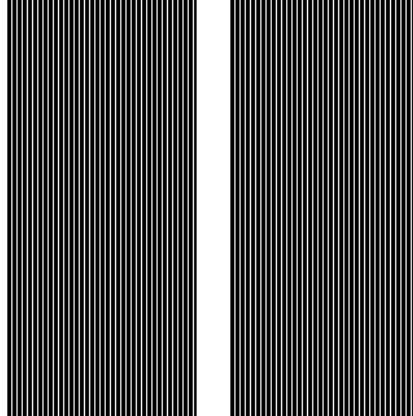


Figure A.1: 1D 4x equal-spaced sampling mask with 8% of low-frequency columns retained. White space indicates the signal retained and black space indicates the signal masked out.