# Accurate MRI Reconstruction via Multi-Domain Recurrent Networks

**Jinbao Wei**[1,2] , **Zhijie Wang**[1,2] , **Kongqiao Wang**[2*] ,
**Li Guo**[2] , **Xueyang Fu**[1] , **Ji Liu**[1] and **Xun Chen**[1]
[1]University of Science and Technology of China
[2]Zepp Health
{wei1998, wzj18895376976}@mail.ustc.edu.cn, {kongqiao.wang, guoli}@zepp.com,
{xyfu, lj1257, xunchen}@ustc.edu.cn

## Abstract

In recent years, deep convolutional neural networks (CNNs) have become dominant in MRI reconstruction from undersampled k-space. However, most existing CNNs methods reconstruct the undersampled images either in the spatial domain or in the frequency domain, and neglecting the correlation between these two domains. This hinders the further reconstruction performance improvement. To tackle this issue, in this work, we propose a new multi-domain recurrent network (MDR-Net) with multi-domain learning (MDL) blocks as its basic units to reconstruct the undersampled MR image progressively. Specifically, the MDL block interactively processes the local spatial features and the global frequency information to facilitate complementary learning, leading to fine-grained features generation. Furthermore, we introduce an effective frequency-based loss to narrow the frequency spectrum gap, compensating for over-smoothness caused by the widely used spatial reconstruction loss. Extensive experiments on public fastMRI datasets demonstrate that our MDR-Net consistently outperforms other competitive methods and is able to provide more details.

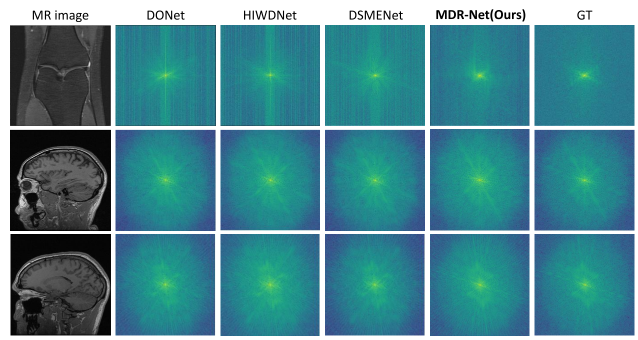


Figure 1: Frequency domain gaps between the real and reconstructed images by existing related models in the MR image reconstruction. Our MDR-Net is able to effectively synthesize the high-frequency components while preserving the low-frequency information, making it superior to other methods.

## 1 Introduction

Magnetic Resonance Imaging (MRI) is one of the most powerful imaging modalities for diagnosis, and can provide superior soft tissue contrast. However, MRI requires a long acquisition time that may occur patient discomfort, leading to significant artifacts in the reconstructed image caused by patient or physiological motions during acquisitions. Furthermore, the availability of MR scanners is also limited by the long acquisition process [Wang *et al.*, 2020a].

A common way to speed up the MRI acquisition process is to sample fewer k-space data instead [Aggarwal *et al.*, 2018]. However, the undersampled k-space data obtained using low-frequency results in aliasing artifacts in the reconstructed images. Therefore, various efforts have focused on developing advanced algorithms to reconstruct an artifact-free MR image from undersampled k-space data. The most widely used technology is compressed sensing (CS), which utilizes the sparsity of an MR image in a specific transform domain, to reconstruct the full image from undersampled k-space data [Lustig *et al.*, 2007]. However, the sparsity regularized CS-MRI methods are time-consuming due to the iterative nature of optimization solutions, which makes it challenging to deploy in real-time MRI scenarios, i.e., Cardiac-MRI.

Recently, deep learning-based methods have demonstrated superior performance over CS for MRI reconstruction [Sandino *et al.*, 2020; Liang *et al.*, 2020]. DL-based methods can be roughly categorized into two fashions, i.e., single-domain methods and dual-domain methods. On the one hand, the single-domain methods [Yang *et al.*, 2017; Mardani *et al.*, 2018; Quan *et al.*, 2018; Wang *et al.*, 2020a; Han *et al.*, 2019] reconstruct MR images from undersampled data solely in the spatial or frequency domain. On the other hand, the dual-domain methods restore the undersampled image in dual-domain [Zhu *et al.*, 2018; Eo *et al.*, 2018; Souza *et al.*, 2019; Zhou and Zhou, 2020]. In general, due to the limited domain knowledge, single-domain methods usually underperform the dual-domain techniques. Dual-domain methods process reconstruction in both domains to broaden domain information, where the features from different domains are serially delivered while ignoring the correlation be-

*Corresponding Author

tween these two representations.

Besides, most DL-based MRI reconstruction methods mainly focus on designing the spatial-domain loss as it can directly provide perceptual improvement. However, single spatial-domain optimization tends to guide the model to generate over-smoothed images due to the lack of frequency information. Many studies [Ronen *et al.*, 2019; Rahaman *et al.*, 2019] show that models tend to fit low-frequency components first that are easy to synthesize while losing high-frequency parts. We visualize the frequency spectra of reconstructed MR images in Figure 1, and previous methods based on spatial-domain optimization show an obvious frequency domain discrepancy between the reconstructed MR images and ground truth. Both HIWDNet and DSMENet fail to restore the high-frequency information. The focal frequency loss [Jiang *et al.*, 2021] is proven effective in synthesizing fine frequency components, but its potential to narrow the frequency domain gap in MRI reconstruction remains under-explored. We find that each frequency in the spectra is the statistical sum across all pixels in the MRI so that frequency-level supervision can offer a new solution for global guidance.

In this paper, we propose a multi-domain recurrent network (MDR-Net) to restore the undersampled MR image. The motivation comes from the spectral convolutional theorem in [Katznelson, 2004] and the dual-domain learning strategy in [Huang *et al.*, 2022]. Therefore, we designed a multi-domain learning (MDL) block in our MDR-Net to interactively learn the local spatial features and global frequency information to obtain complementary representations. Moreover, we introduce an effective frequency loss to make the model concentrate high frequencies that are difficult to synthesize. As a result, frequency spectra reconstructed by MDR-Net are closest to the ground truth in Figure 1, which shows the superiority of our method in narrowing the frequency difference. Furthermore, the feature-level learning in MDL and frequency-level refinement mutually promote common prosperity and ameliorate image quality. Lastly, our recurrent learning helps to avoid overfitting in directly optimizing networks in dual domains. Experiments show that our MDR-Net is able to accurately reconstruct MR images with sharp details.

The main contributions of this work are as follows:

- We propose a novel learning strategy, i.e., a multi-domain learning strategy, which allows us to explore and exploit the properties of undersampled images across different domains.
- We design a new MDL block that incorporates both spatial and frequency information to effectively merge local and global representations, which can provide complementary information.
- To compensate for the excessive smoothing caused by the spatial-domain loss, we design an effective frequency loss to narrow the frequency domain discrepancy by forcing the model to restore high frequencies adaptively.

# 2 Related Work

## 2.1 MRI Reconstruction

Generally, there are two kinds of MRI reconstruction methods: model-driven and data-driven. The former aims to design different optimization algorithms for reconstructing MR images from undersampled k-space data. For example, [Lustig *et al.*, 2007] proposed compressed sensing magnetic resonance imaging (CS-MRI) to accelerate the imaging speed from sparse MR signals. In [Ravishankar and Bresler, 2010], MR images are reconstructed from highly undersampled k-space data using dictionary learning. In addition, a new sparse reconstruction model with multi-class dictionaries is introduced in [Zhan *et al.*, 2015] to accelerate the learning process. However, efficiency issues and poor adaptability limit the effectiveness of these methods.

Inspired by the great success of computer vision, [Wang *et al.*, 2016] first proposed a deep learning model to reconstruct MR images in the spatial domain. With the development of deep learning, various networks [Fan *et al.*, 2018; Zhao *et al.*, 2019; Liu, 2021; Zhang *et al.*, 2021] have been applied to MRI reconstruction. For example, [Schlemper *et al.*, 2017] designed a data consistency layer in a deep cascade of CNNs to ensure the consistency between the constructed image and ground truth. [Wu *et al.*, 2019] introduced the self-attention mechanism with deep residual CNNs, called (SAT-Net). [Zheng *et al.*, 2019] proposed a cascaded dilated dense network with two-step data consistency to remove aliasing artifacts in the reconstructed MR image. However, the investigation of single-domain MRI reconstruction methods soon arrived at a serious bottleneck due to the limited domain knowledge. Therefore, [Eo *et al.*, 2018] proposed a KIKI-Net to reconstruct the image and k-spaces sequentially, which can improve the reconstruction quality progressively. [Zhou and Zhou, 2020] designed a Dual-Domain Recurrent Network (DudoRNet) with deep T1 prior for reconstructing the k-space and image information. [Ran *et al.*, 2020] proposed MR-Recon-Net employs parallel architecture to process the relationships between the spatial domain and frequency domain. However, MD-Recon-Net still does not consider the differences between the spatial domain and frequency domain as it adapts the same CNNs for restoring these two domains. [Feng *et al.*, 2021] proposed double-frequency convolution to learn multi-scale spatial frequency features for parallel MRI. To explore the characteristics of time-frequency features in the wavelet domain, [Tong *et al.*, 2022] propose a hybrid image-wavelet domain reconstruction network (HIWDNet) for accurate MRI reconstruction. In [Wang *et al.*, 2022], detail and structure mutually enhancing network (DSMENet) is proposed to effectively learn the mapping from undersampled input to truth images. In contrast to the above-mentioned data-driven methods, our approach leverages the correlations across different domains to improve the reconstruction performance.

## 2.2 Frequency Spectrum Analysis

The key point of frequency spectrum analysis is signal frequency characteristics. [Xu *et al.*, 2019] proved that DL-based networks attach more importance to low frequencies

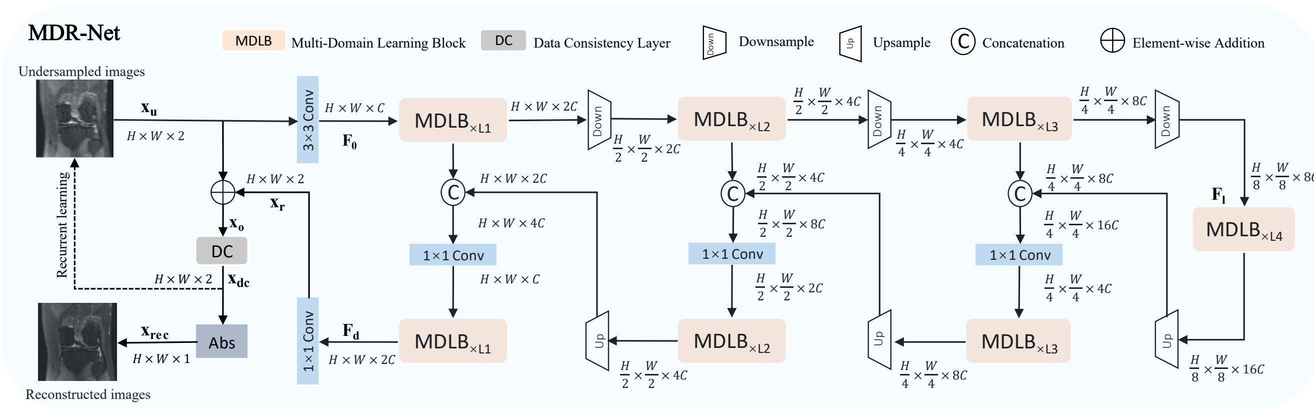


Figure 2: Overview of our multi-domain recurrent network (MDR-Net). It adopts an encoder-decoder architecture with efficient multi-domain learning (MDL) blocks to facilitate dual-domain complementary learning. The data consistency layer is used to ensure the consistency of the reconstructed image in k-space. Furthermore, recurrent learning facilitates the model to restore the refined details progressively.

to fit the objective, which inevitably leads to the frequency domain gap. Many studies [Wang *et al.*, 2020b; Zhang *et al.*, 2019] proved that the periodic patterns shown in the frequency domain are consistent with the artifacts in the spatial domain. Therefore, recent methods tend to improve the visual difference in the spatial domain by narrowing the gap between reconstructed images and ground truth. [Fritsche *et al.*, 2019] proposed the frequency separation method to treat low-frequency and high-frequency images differently. To solve the domain deviation in super-resolution, [Wei *et al.*, 2021] used domain-gap aware training and domain-distance weighted supervision. [Jiang *et al.*, 2021] indicated that focusing on hard frequencies can improve reconstruction performance. In MRI reconstruction, the over-fitting at low frequencies brings smooth textures and blurry structures. Therefore, exploring adaptive constraints at specific frequencies is essential for an accurate reconstruction.

# 3 Method

## 3.1 Overall Architecture

The overall network architecture and internal module of our MDR-Net are illustrated in Figure 2. We chose U-Net as the backbone to build our MDR-Net, which helps to show the superiority of the designed components.

In MR image reconstruction, our purpose is to reconstruct a desired image from measured k-space data. The binary masks $M$ is used for simulating the fast acquisition process of MR signals, projecting the fully k-space data $k \in R^{H\times W\times 2}$ into $k_u \in R^{H\times W\times 2}$. Then we formulate the transformation process as follows:

$$x_u = \mathcal{F}_{2D}^{-1}(k_u) = \mathcal{F}_{2D}^{-1}(M \odot k) = \mathcal{F}_{2D}^{-1}(M \odot \mathcal{F}_{2D}(x)), \quad (1)$$

where $x_u$ is the undersampled input image of MDR-Net; $F_{2D}$ and $F_{2D}^{-1}$ are 2D Fourier transform (FT), and inverse Fourier transform (IFT); $\odot$ is element-wise multiplication.

Given an undersampled image $x_u \in R^{H\times W\times 2}$, MDR-Net first applies a $3 \times 3$ convolution layer to extract shallow features $F_0 \in R^{H\times W\times C}$; where $H \times W$ denotes the spatial dimension and C is the number of channels. Next, these low-level features $F_0$ pass through a 4-level symmetric encoder-decoder and then transform into deep features $F_d \in R^{H\times W\times 2C}$. To prove the effectiveness and efficiency of the multi-domain learning block, each level of encoder-decoder only contains an MDL block, where the original U-Net has instead a sequence of two $3 \times 3$ convolution operations to learn features-maps. Starting from the high-resolution image, the encoder hierarchically reduces spatial size while expanding channel capacity. The decoder takes low-resolution latent features $F_l \in R^{\frac{H}{8}\times\frac{W}{8}\times 8C}$ as input and recovers the high-resolution representations progressively. To assist the recovery process, the encoder features with the same spatial dimension are concatenated with the decoder features via skip connections. Then the concatenation operation is followed by a $1 \times 1$ convolution to reduce channels at all levels except the top one. Finally, a $1 \times 1$ convolution layer is applied to the deep features $F_d$ to generate a residual image $x_r \in R^{H\times W\times 2}$ to which input image is added to obtain the restored image: $x_o = x_u + x_r$.

Then the restored image $x_o$ passes through the data consistency (DC) layer to ensure the consistency between the reconstructed k-space and sampled k-space ($k_u$) since the k-space is changed after inference through MDR-Net. Therefore, the purpose of the DC layer is to maintain the k-space fidelity at sampled locations $z$ of $M$. Denoting the FT output of $x_o$ as $k_o$, the corresponding output from DC layer [Sriram *et al.*, 2020] can be thus formulated as:

$$k_{dc} = \begin{cases} \frac{\lambda k_o(z)+k_u(z)}{\lambda+1} & \text{if } M(z)=1, \\ k_o(z) & \text{if } M(z)=0, \end{cases} \quad (2)$$

where $k_{dc}$ denotes the corrected k-space altered by the DC layer, and $\lambda$ is a trainable hyper-parameter that controls the level of linear combination between sampled k-space values and predicted values. When $\lambda = 0$, the sampled k-space directly substitutes the prediction at z in k-space. The corrected k-space is transformed into the spatial domain as $x_{dc}$ by IFT. The process mentioned above can be seen as the first iteration of our MDR-Net. The corrected restored image $x_{dc}$ will be

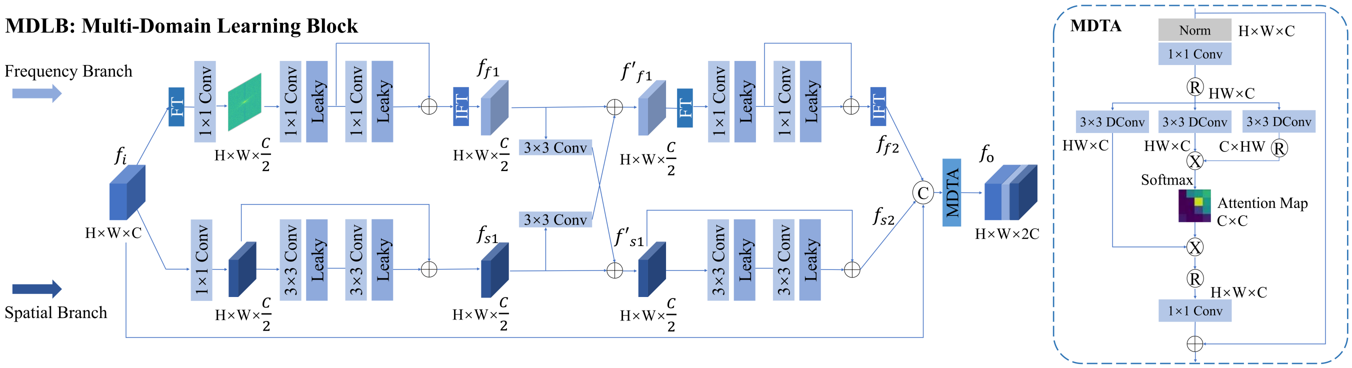


Figure 3: Overall architecture of multi-domain learning (MDL) block. The MDL block includes a frequency branch, a spatial branch and a multi-Dconv head transposed attention (MDTA). The frequency branch uses $1 \times 1$ convolution filters to process global information, while the spatial branch utilizes a residual block with $3 \times 3$ convolutions. There are interactions between the representations of these two branches to provide complementary information. Lastly, MDTA is applied to perform feature interaction across channels for the concatenated results.

adopted as the following input of MDR-Net according to the number of recurrent blocks ($N_{rec}$). Next, we will present the details of the MDL block.

### 3.2 Multi-domain Learning Block

To further explore the difference and relationship between the spatial domain and frequency domain, we propose the MDL block in our model to learn feature-maps from different domains with customized convolution blocks. According to Fourier theory, processing information in Fourier space is capable of capturing the global frequency representation in the frequency domain. In contrast, normal convolution focuses on learning local representations in the spatial domain. In this way, we propose the interactive block to combine these two representations, which is beneficial for learning more representative features. As shown in Figure 3, the MDL block comprises a spatial branch and a frequency branch for processing spatial and frequency representations. Denoting $f_i$ as the input features of the MDL block, the spatial branch first applies a $1 \times 1$ convolution operation to reduce channels (by half) and then adopts a residual block with $3 \times 3$ convolution layers to process information in the spatial domain and obtain $f_{s1}$. While the frequency branch uses Fourier transform to convert the feature-maps to the Fourier space and then adopts a $1 \times 1$ convolution to halve the frequency features. To process frequency-domain representation, we adopt a residual block with $1 \times 1$ convolution operation and then apply Inverse Fourier transform to convert it back to the image space that obtains $f_{f1}$. Thus,$f_{f1}$ is the processed result of the frequency-domain representation. Next, we interact the features from spatial branch $f_{s1}$ and frequency branch $f_{f1}$ as:

$$\begin{aligned} f_{f1}^{'} &= f_{f1} + \mathcal{P}_1(f_{s1}), \\ f_{s1}^{'} &= f_{s1} + \mathcal{P}_2(f_{f1}), \end{aligned} \tag{3}$$

where $\mathcal{P}_1$ and $\mathcal{P}_2$ denotes the $3 \times 3$ convolution operation, $f_{s1}^{'}$ and $f_{f1}^{'}$ are the output of the interacted spatial and frequency branch. As illustrated in Figure 4, both $f_{s1}^{'}$ and $f_{f1}^{'}$ get the complementary representation, which is beneficial for these two branches to obtain more representational features. The following spatial and frequency branches are formulated in the same way as above and output the results $f_{s2}$ and $f_{f2}$, respectively. To efficiently aggregate local and non-local pixel interactions across different domains, we introduce a multi-Dconv head transposed attention(MDTA) [Zamir *et al.*, 2022] to process the final concatenated results from $f_i$, $f_{s2}$ and $f_{f2}$. Specifically, the MDTA implicitly models global context by applying self-attention across channels rather than the spatial dimension, which allows MDL block to prioritize the feature-maps learning from different domains. These design yield quality improvements were shown in the experiment section.

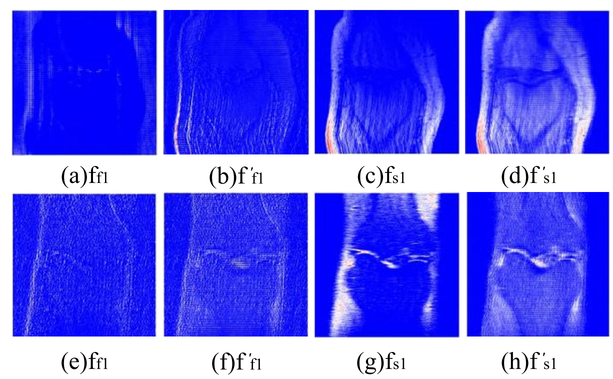


Figure 4: Feature visualizations in the MDL block reveals distinct representations across different branches. Notably, the interaction between features enables them to complement each other, resulting in notable variations between features that interact and those that do not, both in frequency and spatial domains. Specifically, $f_{f1}$ exhibits more spatial invariance, while $f_{s1}$ retains more spatial information. Furthermore, $f_{f1}^{'}$ captures both spatial and detailed information, whereas $f_{s1}^{'}$ highlights details.

### 3.3 Loss Functions

The inherent bias of CNNs makes it challenging to synthesize high-frequency features in the MR image reconstruction, which leads to the frequency domain discrepancy in other methods in Figure 1. One way is to introduce some con-

| | fastMRI 1D random mask | | | | | | CC-359 2D random mask | | | | | | CC-359 2D radial mask | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $RF = 4$ | | | $RF = 8$ | | | $RF = 5$ | | | $RF = 10$ | | | $RF = 5$ | | | $RF = 10$ | | |
| Methods | SSIM | PSNR | NMSE | SSIM | PSNR | NMSE | SSIM | PSNR | NMSE | SSIM | PSNR | NMSE | SSIM | PSNR | NMSE | SSIM | PSNR | NMSE |
| Zero-Filled | 0.6870 | 30.5 | 0.0438 | 0.5428 | 26.6 | 0.0839 | 0.6074 | 19.32 | 0.1951 | 0.5455 | 17.79 | 0.2778 | 0.6912 | 21.16 | 0.1289 | 0.5428 | 17.79 | 0.2778 |
| U-Net [Ronneberger *et al.*, 2015] | 0.7587 | 32.4 | 0.0318 | 0.6006 | 26.8 | 0.0798 | 0.8500 | 27.78 | 0.0279 | 0.8133 | 26.75 | 0.0349 | 0.8859 | 29.47 | 0.0191 | 0.8154 | 26.23 | 0.0396 |
| DCCNN [Schlemper *et al.*, 2017] | 0.7453 | 32.6 | 0.0315 | 0.6131 | 27.6 | 0.0677 | 0.8924 | 30.85 | 0.0140 | 0.8340 | 27.91 | 0.0272 | 0.8999 | 30.83 | 0.0138 | 0.8127 | 26.41 | 0.0383 |
| KIKI [Eo *et al.*, 2018] | 0.7662 | 33.2 | 0.0286 | 0.6133 | 27.6 | 0.0684 | 0.9054 | 29.96 | 0.0180 | 0.8717 | 28.12 | 0.0269 | 0.9155 | 30.50 | 0.0158 | 0.8637 | 27.89 | 0.0278 |
| CDDN [Zheng *et al.*, 2019] | 0.7578 | 33.0 | 0.0295 | 0.6251 | 27.8 | 0.0654 | 0.9267 | 32.96 | 0.0090 | 0.8967 | 30.26 | 0.0168 | 0.9321 | 32.92 | 0.0090 | 0.8798 | 29.32 | 0.0202 |
| MD-Recon-Net [Ran *et al.*, 2020] | 0.7590 | 33.3 | 0.0272 | 0.6260 | 28.5 | 0.0567 | 0.8954 | 30.93 | 0.0173 | 0.8476 | 28.02 | 0.0288 | 0.9021 | 30.93 | 0.0145 | 0.8233 | 26.78 | 0.0389 |
| DudoRNet [Zhou and Zhou, 2020] | 0.7717 | 33.3 | 0.0285 | 0.6256 | 27.9 | 0.0641 | 0.9315 | 33.26 | 0.0084 | 0.8983 | 30.27 | 0.0162 | 0.9354 | 33.21 | 0.0085 | 0.8857 | 29.34 | 0.0199 |
| DONet [Feng *et al.*, 2021] | 0.7645 | 33.0 | 0.0294 | 0.6193 | 27.6 | 0.0677 | 0.9313 | 33.35 | 0.0083 | 0.8949 | 30.41 | 0.0157 | 0.9087 | 30.68 | 0.0143 | 0.8863 | 29.71 | 0.0183 |
| HIWDNet [Tong *et al.*, 2022] | 0.7762 | 33.5 | 0.0278 | 0.6344 | 28.0 | 0.0629 | 0.9345 | 33.40 | 0.0083 | 0.9031 | 30.83 | 0.0144 | 0.9368 | 33.33 | 0.0084 | 0.8953 | 29.83 | 0.0180 |
| DSMENet [Wang *et al.*, 2022] | 0.7770 | 33.6 | 0.0271 | 0.6304 | 28.3 | 0.0605 | 0.9366 | 33.68 | 0.0079 | 0.9073 | 31.05 | 0.0140 | 0.9390 | 33.54 | 0.0081 | 0.8988 | 30.21 | 0.0167 |
| MDR-Net (Ours) | **0.7807** | **33.7** | **0.0271** | **0.6688** | **29.6** | **0.0474** | **0.9378** | **33.73** | **0.0067** | **0.9082** | **31.25** | **0.0132** | **0.9412** | **34.33** | **0.0065** | **0.8993** | **30.68** | **0.0153** |

Table 1: Comparisons with state-of-the-art methods on the fastMRI and CC-359 datasets with different sampling patterns and reduction factors (RF). The best results are **boldfaced**. $RF$ denotes the reduction factor.

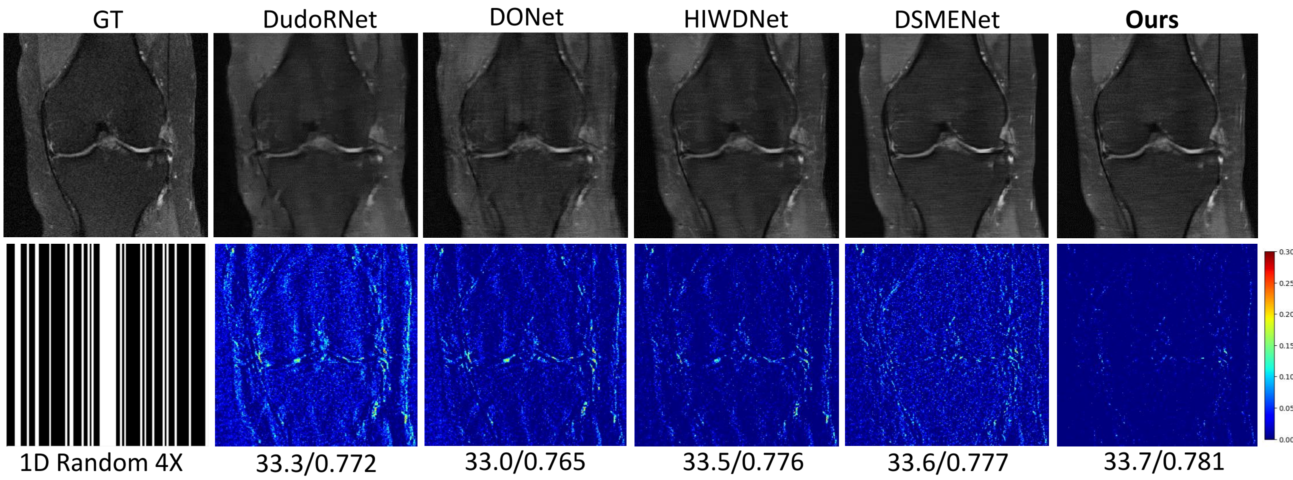


Figure 5: Visual comparisons on the fastMRI dataset with 1D random undersampling patterns and a reduction factor RF=4. The corresponding PSNR and SSIM values are also shown below the error maps.

straints to alleviate this. In this paper, we introduce a joint spatial and frequency loss to regularize the optimization of our model, which can improve the reconstruction results from both the spatial and frequency domains. The loss in the spatial domain is indispensable for the image reconstruction task. So, first of all, we use $SSIM$ loss in the spatial domain. It can be formulated as follows:

$$L_{SSIM} = 1 - SSIM(x_{rec}, x_{gt}), \tag{4}$$

where $x_{rec}$, $x_{gt}$ are the reconstruction and ground truth images, respectively. Although the MDL block in our model includes frequency domain learning, the model is still not optimized for the frequency domain. Therefore, the reconstructed images provided by MDR-Net have an undesirable spectrum, as shown in Figure. 1, so we introduce the focal frequency loss [Jiang *et al.*, 2021] to construct the frequency fidelity term. We define the reconstructed MR image and ground truth as $x_{rec}$ and $x_{gt}$ with the dimensions of $R^{H\times W\times 1}$. Specifically,the focal frequency loss $L_{FFL}$ between the ground truth $x$ and the reconstructed MR image $x_{rec}$ is formulated as:

$$\mathcal{L}_{FFL} = \frac{1}{HW}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} w(u,v)|F_{gt}(u,v) - F_{rec}(u,v)|^{\alpha}, \tag{5}$$

where $F_{gt}$ represents 2D discrete Fourier transform of ground truth image (represents $x_{gt}$ here), and $F_f$ represents 2D discrete Fourier transform of reconstructed image (represents $x_{rec}$ here); $\alpha$ denotes a frequency distance coefficient to make the distance correlation adjustable, and the analysis of frequency-distance coefficients $\alpha$ is provided in Section 4.3. The matrix $w(u, v)$ represents the dynamic weight for the spatial frequency at coordinate $(u, v)$, which is defined as:

$$w(u,v) = |F_{gt}(u,v) - F_{rec}(u,v)|. \tag{6}$$

We introduce the weight factor $\lambda$ to balance the spatial and frequency loss, and the total loss of our MDR-Net is:

$$\mathcal{L} = \mathcal{L}_{SSIM} + \lambda\mathcal{L}_{FFL}. \tag{7}$$

The analysis of the weight factor $\lambda$ is provided in Section 4.3.

# 4 Experiments

## 4.1 Datasets and Settings

**fastMRI dataset.** We conduct experiments on the single-coil knee track of the fastMRI dataset [Zbontar *et al.*, 2018]. The dataset consists of raw k-space data from 1372 knee MRI exams, which are obtained by four different MRI devices. Two types of commonly used MRI sequences are provided by the FastMRI dataset: a Proton Density (PD) weighted sequence and a Proton Density with Fat Saturation (PDFS). We used the same training, validation and testing split as in the original dataset. The training, validation, and testing sets consisted of 973, 199 and 108 volumes, respectively. The single-coil k-space data were retrospectively undersampled

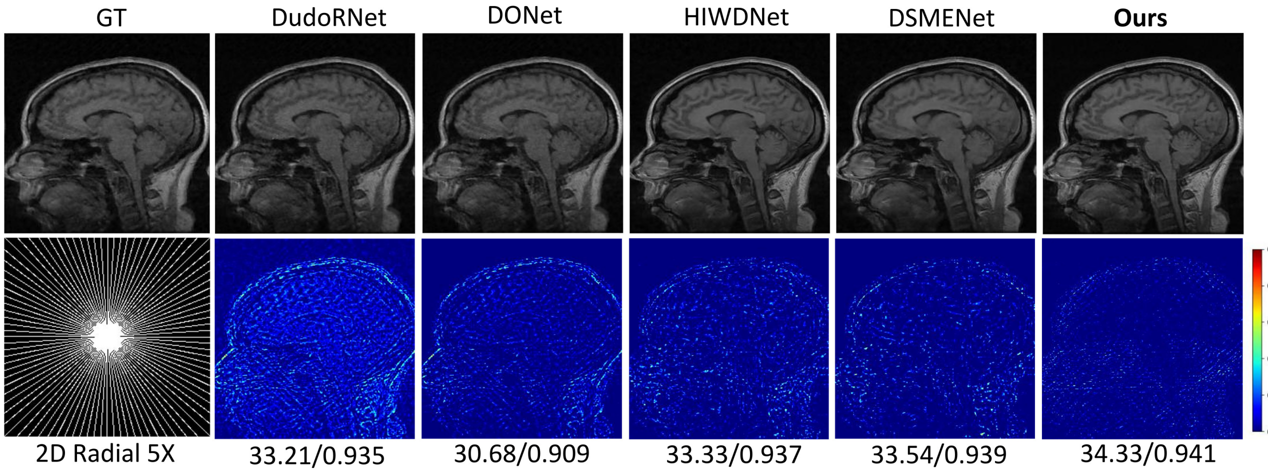


Figure 6: Visual comparisons of different methods on the CC-359 dataset with reduction factor RF=5.

using 1D Cartesian random sampling masks, which are based on code released with the fastMRI dataset.

**CC359 dataset.** The CC-359 Brain dataset[Souza *et al.*, 2018] consists of T1-weighted MR brain images with the size of $256 \times 256$, which were collected from various vendors using scanners at both 1.5T and 3T. It includes a total of 45 volumes, with 25 designated for training, 10 for validation, and 10 for testing. In the experiments, MDR-Net is trained on the full training volumes and evaluated on the full validation volumes for comparison.

**Implementation Details.** The hyper-parameter setting of the network is as follows: the entry channel number MDR-Net is $c = 32$, and the number of recurrent blocks is $N = 4$. All experiments are implemented using the Pytorch platform on two NVIDIA GeForce GTX 3090 with 24GB GPU memory. Our network is trained with an RMSProp optimizer. The initial learning rate is $10^{-3}$ and reduce to $10^{-4}$ after 40 epochs. The batch size is set as 1, and the network is trained for 50 epochs to ensure convergence. NMSE, PSNR and SSIM are used for quantitative evaluation.

## 4.2 Comparisons with the State-of-the-art

**Quantitative Comparison.** We compare our proposed MDR-Net with ten state-of-the-art methods, including one conventional method (Zero-Filled) and nine CNNs-based methods (U-Net [Ronneberger *et al.*, 2015], DCCNN [Schlemper *et al.*, 2017], KIKI-Net [Eo *et al.*, 2018], CDDN [Zheng *et al.*, 2019], MD-Recon-Net [Ran *et al.*, 2020], DudoRNet [Zhou and Zhou, 2020], DONet[Feng *et al.*, 2021], DSMENet [Wang *et al.*, 2022] and HIWDNet [Tong *et al.*, 2022]). To ensure a fair comparison, we conducted all experiments using the same train, validation, and test sets, as well as the same computing environment. Additionally, we used the released code of each competitor with the default settings. This allows us to compare the performance of each method under the same conditions. Quantitative results on FastMRI and CC359 datasets in terms of SSIM, PSNR and NMSE are reported in Table 1. We can see that our MDR-Net significantly outperforms other methods.

| Methods | SSIM↑ | PSNR↑ | NMSE↓ |
|---|---|---|---|
| (A) Net-baseline | 0.7678 | 32.8 | 0.0299 |
| (B) Net-MDLB | 0.7761 | 33.4 | 0.0278 |
| (C) Net-FL | 0.7705 | 33.0 | 0.0302 |
| (D) Net-Rec | 0.7738 | 33.0 | 0.0304 |
| (E) Net-MDLB-Rec | 0.7786 | 33.6 | 0.0279 |
| (F) Net-MDLB-FL | 0.7773 | 33.5 | 0.0278 |
| (G) Net-FL-Rec | 0.7768 | 33.5 | 0.0281 |
| (H) Net-MDLB-FL-Rec | 0.7807 | 33.7 | 0.0271 |

Table 2: Quantitative evaluations for MDLB, FL and Rec components in MDR-Net.

| Metric | w/o FT&IFT | 2-FB | 2-SB | w/o Interaction | w/o MDTA | MDL block |
|---|---|---|---|---|---|---|
| SSIM ↑ | 0.7651 | 0.7781 | 0.7788 | 0.7792 | 0.7797 | **0.7807** |
| PSNR ↑ | 33.0 | 32.6 | 33.2 | 33.4 | 33.6 | **33.7** |

Table 3: Ablation study of investigating the MDL block. 2-SB represents both branches set to the spatial branches, and 2-FB represents both branches set to the frequency branches.

**Visual Comparison.** As shown in Figures 5 and 6, these qualitative results show that our reconstructed MRI is the closest and highest correlation to the ground truth. Besides, we perform our MDR-Net on different sampling patterns and draw the corresponding error maps between reconstructed images and ground truth. As a result, our MDR-Net generates more visually pleasant results than previous methods, especially in the reconstruction of high-frequency structural content, which benefits from feature-level and frequency-level dual-domain learning.

## 4.3 Ablation Study

In this study, we evaluated three key components of MDR-Net on the fastMRI dataset with a reduction factor RF=4, including multi-domain learning block (MDLB), frequency loss (FL) and recurrent learning (Rec). We used the MDR-Net without these three components as a baseline, and set the value of $N_{rec}$ in Rec to 5. The results of this component analysis are shown in Table 2. We found that the MDL block (B)

| Metric | $N_{rec}$=1 | $N_{rec}$=2 | $N_{rec}$=3 | $N_{rec}$=4 | $N_{rec}$=5 | $N_{rec}$=6 | $N_{rec}$=7 |
|---|---|---|---|---|---|---|---|
| SSIM ↑ | 0.7769 | 0.7783 | 0.7794 | 0.7804 | **0.7807** | 0.7805 | 0.7801 |
| PSNR ↑ | 32.9 | 33.3 | 33.5 | 33.6 | **33.7** | 33.6 | 33.5 |

Table 4: The effect of increasing the number of recurrent learning($N_{rec}$) in our MDR-Net.

| Metric | $\lambda = 0$ | $\lambda = 0.2$ | $\lambda = 0.4$ | $\lambda = 0.6$ | $\lambda = 0.8$ | $\lambda = 1$ |
|---|---|---|---|---|---|---|
| SSIM ↑ | 0.7781 | 0.7788 | 0.7792 | 0.7797 | **0.7807** | 0.7793 |
| PSNR ↑ | 33.2 | 33.3 | 33.5 | 33.6 | **33.7** | 33.6 |
| NMSE ↓ | 0.0283 | 0.0281 | 0.0277 | **0.0270** | 0.0271 | 0.0273 |

Table 5: Performance comparisons of different loss weight factor.

increased the PSNR value by 0.6 dB compared to the baseline (A), a more significant improvement than either frequency learning (C) or recurrent learning (D) alone. The combination of the MDL block and recurrent learning (E) produced the largest boost in performance, while all three components together (H) yielded the best reconstruction results overall. These findings demonstrate that all three components contribute to the enhanced performance of the MDR-Net. Next, we will separately provide the ablation studies of each component mentioned above.

**Ablation Study of MDLB.** We validate the effectiveness of the design of the MDL block in Table 3. As can be seen, the performance decreases significantly when removing both FT and IFT operations. Furthermore, both replacing the spatial branch with the frequency branch or replacing the frequency branch with the spatial branch results in a significant performance drop. While the interaction between these two branches can improve performance remarkably, demonstrating the effectiveness of integrating these two complementary representations. Moreover, the introduced MDTA provides a favorable gain of 0.2dB over the baseline.

**The Number of Recurrent Learning.** The Number of Recurrent Learning. An important factor of MDR-Net is the number of recurrent learning. We compare the results of different recurrent numbers in Table 4. We can observe that both SSIM and PSNR values tend to be saturated when N is four and reaches the highest when N is 5.

**Loss Weight Factor.** The parameter $\lambda$ in Eq.(7) is used to balance the importance of spatial and frequency loss. We studied how the model's performance changes with different values of $\lambda$ and present the results in Table 5. When $\lambda = 0$, the model only focuses on the spatial domain loss which is not optimal, showing that frequency-level supervision is crucial. As the frequency loss increases, the model's performance improves. The best results in terms of PSNR and SSIM are achieved when $\lambda = 0.8$.

**Frequency Loss.** Figure 7 illustrates the difference in frequency spectra reconstructed with and without the frequency loss. The corresponding log frequency distance(LFD) is also shown in the same figure. Without frequency supervision, the reconstructed spectra exhibit an aliasing artifact. However, when using our proposed frequency loss, the frequency spectra are reconstructed more accurately and with a lower LFD, resulting in a closer match to the true frequency statistics. Additionally, the fine-grained spectrum supervision in frequency

| Metric | $\alpha = 0.1$ | $\alpha = 0.3$ | $\alpha = 0.5$ | $\alpha = 1$ | $\alpha = 2$ | $\alpha = 3$ |
|---|---|---|---|---|---|---|
| LFD ↓ | 14.8633 | 14.3792 | 13.9825 | 13.6571 | **13.3238** | 15.0863 |
| SSIM ↑ | 0.7739 | 0.7752 | 0.7782 | 0.7795 | **0.7807** | 0.7792 |
| PSNR ↑ | 33.1 | 33.2 | 33.5 | **33.8** | 33.7 | 33.5 |

Table 6: Model performance comparison using different coefficients to calculate the spectrum distances in frequency domain.

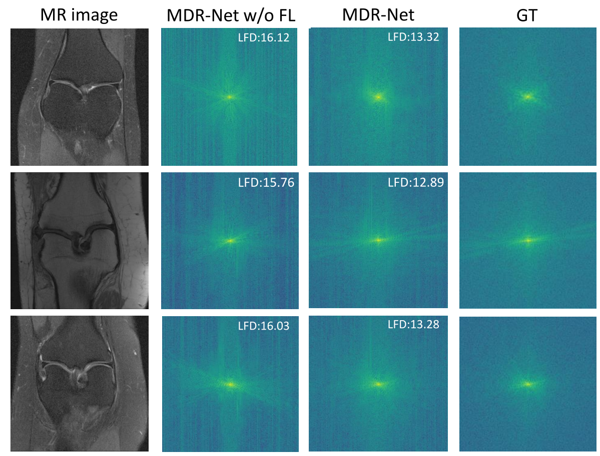


Figure 7: Frequency spectrum visualization with or without (w/o) FDL. The metric LFD is used to measure the frequency similarity.

loss helps to preserve high-frequency details that are difficult to synthesize.

**Frequency Distance Coefficient.** The frequency distance coefficient $\alpha$ in Eq.(5) is introduced to emphasize the frequency parts of the reconstructed MR image. The higher the value of $\alpha$, the more the network will penalize underfitting frequencies. Table6 shows the model performance for different values of $\alpha$. The highest PSNR is achieved when $\alpha = 1$, while the best SSIM and LFD performance is obtained when $\alpha = 2$. Lowering $\alpha$ results in weaker frequency penalty and lower performance, while increasing it leads to more stringent frequency domain supervision, which can distort the reconstructed MR image. Therefore, we set $\alpha = 2$ as the default value to balance structural similarity and perceptual quality.

## 5 Conclusion

The investigation of CNNs for MR image reconstruction has arrived at a serious bottleneck as networks with better performance usually adopt a dual-domain learning strategy generally while could not recover the detailed structures in the reconstructed MR images. To solve this problem, we proposed a Multi-Domain Recurrent Network(MDR-Net) to restore both image and k-space signal simultaneously with multi-domain interactions. Experiments demonstrated that while previous MR image reconstruction methods on single-domain and dual-domain have limited capability of reducing aliasing artifacts in the image domain, our MDR-Net can effectively restore the reconstruction with complementary features. Future work includes exploring MDR-Net on multi-coil MR data and the application to other image reconstruction tasks, such as Photoacoustic Tomography images reconstruction.

## Acknowledgements

This work was supported in part by the CAS Project for Young Scientists in Basic Research under Grant YSBR-067, in part by the National Natural Science Foundation of China under Grant 32271431.

## Contribution Statement

Jinbao Wei and Zhijie Wang contributed equally to this work.

## References

[Aggarwal *et al.*, 2018] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394–405, 2018.

[Eo *et al.*, 2018] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80(5):2188–2201, 2018.

[Fan *et al.*, 2018] Zhiwen Fan, Liyan Sun, Xinghao Ding, Yue Huang, Congbo Cai, and John Paisley. A segmentation-aware deep fusion network for compressed sensing mri. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 55–70, 2018.

[Feng *et al.*, 2021] Chun-Mei Feng, Zhanyuan Yang, Geng Chen, Yong Xu, and Ling Shao. Dual-octave convolution for accelerated parallel mr image reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 116–124, 2021.

[Fritsche *et al.*, 2019] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3599–3608. IEEE, 2019.

[Han *et al.*, 2019] Yoseob Han, Leonard Sunwoo, and Jong Chul Ye. k-space deep learning for accelerated mri. *IEEE transactions on medical imaging*, 2019.

[Huang *et al.*, 2022] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, pages 163–180. Springer, 2022.

[Jiang *et al.*, 2021] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13919–13929, 2021.

[Katznelson, 2004] Yitzhak Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, 2004.

[Liang *et al.*, 2020] Dong Liang, Jing Cheng, Ziwen Ke, and Leslie Ying. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Processing Magazine*, 37(1):141–151, 2020.

[Liu, 2021] Risheng Liu. Optimization learning: perspective, method, and applications. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 5146–5168, 2021.

[Lustig *et al.*, 2007] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[Mardani *et al.*, 2018] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasanawala, Greg Zaharchuk, Lei Xing, and John M Pauly. Deep generative adversarial neural networks for compressive sensing mri. *IEEE transactions on medical imaging*, 38(1):167–179, 2018.

[Quan *et al.*, 2018] Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. *IEEE transactions on medical imaging*, 37(6):1488–1497, 2018.

[Rahaman *et al.*, 2019] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019.

[Ran *et al.*, 2020] Maosong Ran, Wenjun Xia, Yongqiang Huang, Zexin Lu, Peng Bao, Yan Liu, Huaiqiang Sun, Jiliu Zhou, and Yi Zhang. Md-recon-net: A parallel dual-domain convolutional neural network for compressed sensing mri. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):120–135, 2020.

[Ravishankar and Bresler, 2010] Saiprasad Ravishankar and Yoram Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE transactions on medical imaging*, 30(5):1028–1041, 2010.

[Ronen *et al.*, 2019] Basri Ronen, David Jacobs, Yoni Kasten, and Shira Kritchman. The convergence rate of neural networks for learned functions of different frequencies. *Advances in Neural Information Processing Systems*, 32, 2019.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[Sandino *et al.*, 2020] Christopher M Sandino, Joseph Y Cheng, Feiyu Chen, Morteza Mardani, John M Pauly, and Shreyas S Vasanawala. Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging. *IEEE signal processing magazine*, 37(1):117–127, 2020.

[Schlemper *et al.*, 2017] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2017.

[Souza *et al.*, 2018] Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018.

[Souza *et al.*, 2019] Roberto Souza, R Marc Lebel, and Richard Frayne. A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. In *International Conference on Medical Imaging with Deep Learning*, pages 437–446. PMLR, 2019.

[Sriram *et al.*, 2020] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated mri reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 64–73. Springer, 2020.

[Tong *et al.*, 2022] Chuan Tong, Yanwei Pang, and Yueze Wang. Hiwdnet: A hybrid image-wavelet domain network for fast magnetic resonance image reconstruction. *Computers in Biology and Medicine*, 151:105947, 2022.

[Wang *et al.*, 2016] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517. IEEE, 2016.

[Wang *et al.*, 2020a] 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.

[Wang *et al.*, 2020b] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. Cnn-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8695–8704, 2020.

[Wang *et al.*, 2022] Yueze Wang, Yanwei Pang, and Chuan Tong. Dsmenet: Detail and structure mutually enhancing network for under-sampled mri reconstruction. *Computers in Biology and Medicine*, page 106204, 2022.

[Wei *et al.*, 2021] Yunxuan Wei, Shuhang Gu, Yawei Li, Radu Timofte, Longcun Jin, and Hengjie Song. Unsupervised real-world image super resolution via domain-distance aware training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13385–13394, 2021.

[Wu *et al.*, 2019] Yan Wu, Yajun Ma, Jing Liu, Jiang Du, and Lei Xing. Self-attention convolutional neural network for improved mr image reconstruction. *Information sciences*, 490:317–328, 2019.

[Xu *et al.*, 2019] Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. Training behavior of deep neural network in frequency domain. In *International Conference on Neural Information Processing*, pages 264–274. Springer, 2019.

[Yang *et al.*, 2017] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, et al. Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging*, 37(6):1310–1321, 2017.

[Zamir *et al.*, 2022] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

[Zbontar *et al.*, 2018] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.

[Zhan *et al.*, 2015] Zhifang Zhan, Jian-Feng Cai, Di Guo, Yunsong Liu, Zhong Chen, and Xiaobo Qu. Fast multi-class dictionaries learning with geometrical directions in mri reconstruction. *IEEE Transactions on biomedical engineering*, 63(9):1850–1861, 2015.

[Zhang *et al.*, 2019] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images. In *2019 IEEE international workshop on information forensics and security (WIFS)*, pages 1–6. IEEE, 2019.

[Zhang *et al.*, 2021] Yulun Zhang, Kai Li, Kunpeng Li, and Yun Fu. Mr image super-resolution with squeeze and excitation reasoning attention network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13425–13434, 2021.

[Zhao *et al.*, 2019] Xiaole Zhao, Yulun Zhang, Tao Zhang, and Xueming Zou. Channel splitting network for single mr image super-resolution. *IEEE transactions on image processing*, 28(11):5649–5662, 2019.

[Zheng *et al.*, 2019] Hao Zheng, Faming Fang, and Guixu Zhang. Cascaded dilated dense network with two-step data consistency for mri reconstruction. *Advances in Neural Information Processing Systems*, 32, 2019.

[Zhou and Zhou, 2020] Bo Zhou and S Kevin Zhou. Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4273–4282, 2020.

[Zhu *et al.*, 2018] 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, 2018.