

Encoding Sampling Pattern for Robust and Generalized MRI Reconstruction

Lina SUN^{a,1}, Hong WANG^{b,1}, Qi XIE^{a,*}, Yefeng ZHENG^c, Deyu MENG^a

^a*School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China*

^b*School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China*

^c*Medical Artificial Intelligence Lab, Westlake University, Hangzhou, 310030, China*

Abstract

Against the magnetic resonance imaging (MRI) reconstruction task, current deep learning based methods have achieved promising performance. Nevertheless, most of them are confronted with two main problems: 1) For most current MRI reconstruction methods, the down-sampling pattern is generally preset in advance, which makes it hard to flexibly handle the complicated real scenarios where the training data and the testing data are obtained under different sampling settings, thus constraining the model generalization capability. 2) They have not fully incorporated the physical imaging mechanism between the down-sampling pattern estimation and high-resolution MRI reconstruction into deep network design for this specific task. To alleviate these issues, we propose a model-driven MRI reconstruction network called MXNet, which considers the relationship between the undersampling pattern and imaging by encoding the mask into the network. Specifically, based on the MR physical imaging process, we first jointly optimize the down-sampling pattern and MRI reconstruction network. Then, based on the proposed optimization algorithm and the deep unfolding technique, we correspondingly construct the deep network where the physical imaging mechanism for MRI reconstruction is fully embedded into the entire learning process. Based on different settings between training data and testing data, with both consistent and inconsistent down-sampling patterns, extensive experiments comprehensively substant-

* Corresponding author
Email address: xie.qi@mail.xjtu.edu.cn (Qi XIE)

¹ The two authors contribute equally to this work.

tiate the effectiveness of our proposed MXNet in detail reconstruction as well as its fine generality. Moreover, we provide detailed model analysis and validate that our proposed framework shows fine generality and it can still accomplish superior performance when the downsampling mask is accurately available. The code is available at <https://github.com/sunliyangna0705/MXNet>.

Keywords: MR imaging, reconstruction, model drift, sampling pattern, encoder mask

1. Introduction

Magnetic resonance imaging (MRI) aims to adopt magnetic fields and computer-generated radio waves to produce detailed images of internal organs and tissues. As a commonly adopted medical imaging technology, it possesses specific merits such as non-invasiveness, high soft tissue contrast, and multi-directional acquisition, and thus plays an important role in disease diagnosis. Considering MRI scanning is time-consuming and causes motion artifacts and discomfort to patients, how to achieve fast MRI with favorable image quality has long been an imperative and profoundly impactful matter in the field.

Specifically, for scanned objects, MR images are obtained by inverse Fourier transform on the domain k-space data collected from the MR scanner. High-resolution MR images can be obtained by fully sampling the k-space data, which however is time-consuming. To speed up the data acquisition process for single coil MR images, the most widely adopted method is to down-sample the k-space data. Such a degradation process can be mathematically formulated as [1]:

$$\mathbf{Y} = \mathbf{M} \odot \mathbf{F}\mathbf{X} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{X} \in \mathbb{C}^{H \times W}$ is the to-be-reconstructed MR image; \mathbf{F} is the Fourier transform operator; $\mathbf{M} \in \mathbb{R}^{H \times W}$ is a binary point-wise down-sampling matrix (mask) which specifies the sampling pattern executed in the k-space domain; specifically, $M_{ij} \in \{0, 1\}, i = 1, 2, \dots, H, j = 1, 2, \dots, W$; $\mathbf{Y} \in \mathbb{C}^{H \times W}$ is the under-sampled k-space data; \odot represents the dot product; $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ denotes the random Gaussian noise with noise

level σ . Obviously, to reconstruct an image from under-sampled one is an NP-hard problem, which always violates the Nyquist sampling criterion.

With the rapid development of deep learning, researchers have proposed diverse approaches to deal with the fast MRI reconstruction task which can be broadly categorized into two types: data-driven and model-driven methods. Specifically, the former aims to utilize a large amount of training data to automatically learn a mapping function from under-sampled data to fully-sampled data, typically using k-space data or MR images. This is achieved through different network architectures, including U-Net, generative adversarial network[2], ResNet[3], and CasCade-Net[4]. Although achieving promising performance, most of these methods are generally heuristically constructed based on off-the-shelf network blocks without fully considering the physical degradation mechanism in Eq. (1).

To alleviate this issue, the model-driven research line for fast MRI reconstruction has emerged, which mainly focuses on integrating the physical imaging constraint into network designs via the deep unfolding technique. Some excellent works have been proposed, such as ADMMCS-Net[5], convolution de-aliasing network[6], variational network[7], model-based deep learning (MoDL)[8], and neural proximal gradient descent. Transformer-based models[9][10] have made some progress in various medical imaging tasks because transformers can learn global information by modeling long-range dependency. The advent of Mamba[11][12] has also provided new insights into medical image reconstruction. Attributed to the involved optimization-inspired unrolling design scheme, these model-driven methods generally have a clearer working mechanism than the purely data-driven MRI reconstruction methods. Although the networks have achieved great results, they heavily rely on mask prior that need to be input into the network. In clinical settings, the down-sampling pattern may be unavailable due to scanner proprietary protocols, automatically selected sampling modes (preset in rapid imaging protocols), randomized PRM-based patterns[13], complex non-Cartesian trajectories, or data distortion during acquisition caused by coil sensitivity, magnetic field inhomogeneity, or motion artifacts. Besides, in practice, we generally need to address the MR data, which are obtained by different downsampling imaging protocols, especially in the cross-device and cross-hospital scene. Directly tak-

ing some preset sampling mask as the network input would cause the model drift[14] and limit the model generalization performance in a common cross-distribution scenario where the training-testing data have different sampling patterns[15][16]. Transfer learning[17][15] can address such challenges, which utilize large-scale publicly available datasets to train neural networks and then fine-tune them with limited data. However, some approaches still rely on predefined sampling patterns. Moreover, these general-purpose methods exhibit insufficient integration of MRI-specific physical characteristics tailored for image reconstruction tasks. Therefore, it is necessary and meaningful to develop a more general MR reconstruction algorithm that can not only address the case with an accurately known downsampling pattern but also the case with an unknown sampling mask. Recent research has explored the joint optimization of sampling patterns and reconstruction models to find optimal sampling pattern. In the compressed sensing framework, the quality of MRI reconstruction can be improved by introducing learning-based sampling pattern optimization methods[18][19][20]. These methods typically employ different regularizers, such as TV (Total Variation), sparsity in the wavelet domain, or pre-trained diffusion models. For instance, LOUPE[13] assumed each point as an independent Bernoulli random variable with learnable parameters in k-space (probabilistic mask) and optimized sampling patterns and high-resolution MR images simultaneously. J-MoDL[21] considered a multichannel forward model, learning a common down-sampling pattern and a neural network reconstructor, but the learned pattern was fixed to the training set. The J-MoDL method applied to multi-contrast MRI[22] expands its universality. Very recently, a data-driven sampler used a CNN to generate object-specific, adaptive sampling patterns from limited low-frequency k-space data, improving reconstruction quality[23]. These methods seek optimal down-sampling for MR images. In contrast, the current work aims to develop a network that can predict a likely true mask for high-quality image reconstruction.

In this paper, we attempt to encode the sampling mask into the deep reconstruction network with the guidance of the physical imaging mechanism in Eq. (1) and construct a model-driven MR reconstruction network, called MXNet, as shown in Fig.1. The key characteristics of MXNet are: 1) It does not require pre-collecting downsampling

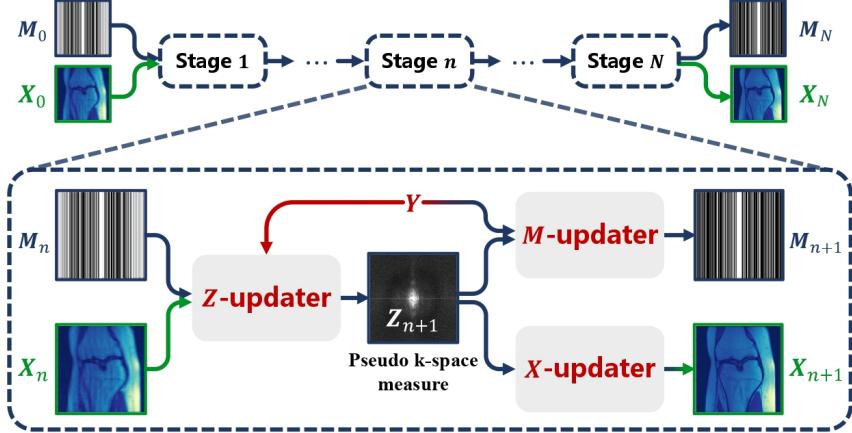


Figure 1: Illustration of the overall framework of the proposed MXNet with N stages. It outputs the reconstructed image and learned sampling pattern mask. Each stage contains three parts: Z -updater, M -updater, and X -updater which are separately used to update Z_{n+1} , M_{n+1} , and X_{n+1} based on the algorithm in Eqs. (15) (16) and (17).

mask; 2) when the sampling mask is accurately available, our proposed framework still applies (see Sec. 5.4); 3) It evidently alleviates the model drift issue and achieves promising generalization performance in the inconsistent training-testing scenario with different sampling masks. Specifically, our contribution can be summarized as follows:

1) **Model-Driven Joint Optimization of M and X .** Based on the physical degradation process of Eq. (1), we construct the optimization model for the joint reconstruction of down-sampling pattern M and MR image X . We adopt the variable splitting technique and the coordinate descent technique to derive an optimization algorithm that alternately updates the unknown variables M and X . It is worth mentioning that our proposed model can be easily degraded to the non-blind case where the sampling pattern M is accurately known.

2) **Network Design with Physical Interpretability.** Using the proposed optimization algorithm as a foundation, we iteratively unfold optimization steps into network modules, effectively transforming them into deep network operations to construct the entire network framework. As shown in Fig. 1, the proposed MXNet is composed of N stages. For every stage, the network structure consists of three subnetworks with one-

to-one step correspondence to achieve the joint learning of \mathbf{M} and \mathbf{X} . As observed, the mutual optimization of \mathbf{M} and \mathbf{X} always proceeds under the guidance of the physical model (1). Compared to the existing heuristic usage of \mathbf{M} , which usually directly concatenates it with MR image[21], our design has evidently clearer physical interpretability. Such a clear network design facilitates general readers able to intuitively understand the working mechanism of MXNet by visualizing the model components extracted at every stage.

3) Fine Generality. We encode the mask into the network and use data generated from different masks to stimulate the model’s generalization and robustness for the MRI reconstruction problem. Based on different experimental settings, including consistent and inconsistent training-testing sampling patterns, we comprehensively validate the effectiveness of our proposed MXNet as well as its fine model generalization. Attributed to the physics-driven adaptive design, our method can always extract rational down-sampling masks, and using predicted sample masks achieves similar performance to using known masks.

2. MRI reconstruction model

For the fast MRI reconstruction task, we first construct the optimization model for the joint recovery of down-sampling pattern \mathbf{M} and MR image \mathbf{X} and then derive the corresponding optimization algorithm.

2.1. Model Formulation

As Eq. (1), given the under-sampled measure \mathbf{Y} , our goal is to estimate the MR image \mathbf{X} with unknown sampling pattern \mathbf{M} . Generally, this can be formulated as the following optimization problem:

$$\min_{\mathbf{M}, \mathbf{X}} \|\mathbf{M} \odot \mathbf{F}\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda_1 \mathcal{R}_1(\mathbf{M}) + \lambda_2 \mathcal{R}_2(\mathbf{X}), \quad (2)$$

where $\mathbf{M}_{ij} \in \{0, 1\}$, λ_1 and λ_2 denote trade-off regularization parameters, and $\|\cdot\|_F$ is the Frobenius norm. The first item in Eq. (2) reflects the physical generation mechanism of accelerated reconstruction of MR images, which provides explicit guidance in

the iterative updating process of \mathbf{M} and X . $\mathcal{R}_1(\cdot)$ and $\mathcal{R}_2(\cdot)$ represent the regularization terms and deliver the prior information of X and \mathbf{M} , respectively.

In this paper, we specifically introduce a constraint for \mathbf{M} , i.e., $(1 - \mathbf{M}) \odot \mathbf{Y} = 0$. This reflects the fact the k-space data not sampled is 0 in \mathbf{Y} , i.e., $1 - \mathbf{M}_{ij} = 1 \rightarrow \mathbf{Y}_{ij} = 0$ and $\mathbf{Y}_{ij} \neq 0 \rightarrow 1 - \mathbf{M}_{ij} = 0$. Besides, we also introduce a pseudo k-space measure \mathbf{Z} , satisfied $\mathbf{Z} = \mathbf{F}X$. Then Eq. (2) can be transformed into the optimization problem:

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{M}, X} \|\mathbf{M} \odot \mathbf{Z} - \mathbf{Y}\|_F^2 + \lambda_1 \mathcal{R}_1(\mathbf{M}) + \lambda_2 \mathcal{R}_2(X) \\ & \text{s.t. } \mathbf{Z} - \mathbf{F}X = 0, \quad (1 - \mathbf{M}) \odot \mathbf{Y} = 0. \end{aligned} \quad (3)$$

By converting the optimization problem (3) into an unconstrained optimization form, we can obtain that:

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{M}, X} \|\mathbf{M} \odot \mathbf{Z} - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{F}X - \mathbf{Z}\|_F^2 \\ & + \beta \|(1 - \mathbf{M}) \odot \mathbf{Y}\|_F^2 + \lambda_1 \mathcal{R}_1(\mathbf{M}) + \lambda_2 \mathcal{R}_2(X), \end{aligned} \quad (4)$$

where α and β denote penalty parameters. Similarly to existing model-driven methods, such as [24] [25], we rely on the flexible representation of CNN to automatically learn the implicit priors $\mathcal{R}_1(\cdot)$ and $\mathcal{R}_2(\cdot)$ from training data. More details are presented in Sec. 3.

Especially, for the non-blind MRI reconstruction with the known down-sampling pattern \mathbf{M} , Eq. (4) is simplified as:

$$\min_{\mathbf{Z}, X} \|\mathbf{M} \odot \mathbf{Z} - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{F}X - \mathbf{Z}\|_F^2 + \lambda_2 \mathcal{R}_2(X). \quad (5)$$

When $\alpha = +\infty$, Eq. (5) corresponds to a standard compressed sensing model, which is the basis of many MRI reconstruction methods, such as [26, 27]. As seen, they are the special cases of our proposed reconstruction model.

2.2. Model Optimization

To solve the problem (4), we propose to adopt the coordinate descent algorithm to alternately update these unknown variables, including the auxiliary variable \mathbf{Z} , down-sampling pattern \mathbf{M} and the MR image X . Specifically, at the n -th iterative stage, the updating rules are derived as follows:

1) Updating \mathbf{Z} : With X and \mathbf{M} fixed, the auxiliary variable \mathbf{Z} can be updated by solving the sub-problem about \mathbf{Z} as:

$$\min_{\mathbf{Z}} \|\mathbf{M}_n \odot \mathbf{Z} - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{F}X_n - \mathbf{Z}\|_F^2. \quad (6)$$

Clearly, we can directly derive the analytical solution as:

$$\mathbf{Z}_{n+1} = \frac{\alpha \mathbf{F}X_n + \mathbf{M}_n \odot \mathbf{Y}}{\alpha + \mathbf{M}_n^2}, \quad (7)$$

where the division symbol denotes an element-wise operation.

2) Updating \mathbf{M} : With \mathbf{Z} and X fixed, the down-sampling pattern \mathbf{M} can be updated by solving the following problem:

$$\min_{\mathbf{M}} \|\mathbf{M} \odot \mathbf{Z}_{n+1} - \mathbf{Y}\|_F^2 + \beta \|(1 - \mathbf{M}) \odot \mathbf{Y}\|_F^2 + \lambda_1 \mathcal{R}_1(\mathbf{M}). \quad (8)$$

By completing the square for the variable \mathbf{M} , we can rewrite this formula as:

$$\min_{\mathbf{M}} \left\| \sqrt{W} \odot \left(\mathbf{M} - \frac{\mathbf{Z}_{n+1} \odot \mathbf{Y} + \beta \mathbf{Y}^2 + \omega}{W + \omega} \right) \right\|_F^2 + \lambda_1 \mathcal{R}_1(\mathbf{M}), \quad (9)$$

where $W = \mathbf{Z}_{n+1}^2 + \beta \mathbf{Y}^2$, ω is a small number added on the denominator² to prevent it from being 0. Furthermore, since \mathcal{R}_1 is usually set as an element-wise regularization (such as L_2 norm, L_1 norm and $L_{1/2}$ norm[28]), then Eq. (9) can be equivalently rewritten as the following form:

$$\min_{\mathbf{M}} \left\| \mathbf{M} - \frac{\mathbf{Z}_{n+1} \odot \mathbf{Y} + \beta \mathbf{Y}^2 + \omega}{\mathbf{Z}_{n+1}^2 + \beta \mathbf{Y}^2 + \omega} \right\|_F^2 + \frac{\lambda_1}{W} \mathcal{R}_1(\mathbf{M}). \quad (10)$$

By employing the coordinate descent algorithm [29], we can obtain the solution of Eq. (8) as:

$$\mathbf{M}_{n+1} = \text{Prox}_{\lambda_1/W} \left(\frac{\beta \mathbf{Y}^2 + \mathbf{Z}_{n+1} \odot \mathbf{Y} + \omega}{\beta \mathbf{Y}^2 + \mathbf{Z}_{n+1}^2 + \omega} \right), \quad (11)$$

where $\text{Prox}_{\lambda_1/W}(\cdot)$ denotes the proximal operator, which is related to the regularization term. As [24], we will implement this operator via deep networks, as described in Sec. 3.

² We add ω to both the numerator and denominator, which is essentially equivalent to introducing a squared regularization term with respect to \mathbf{M} .

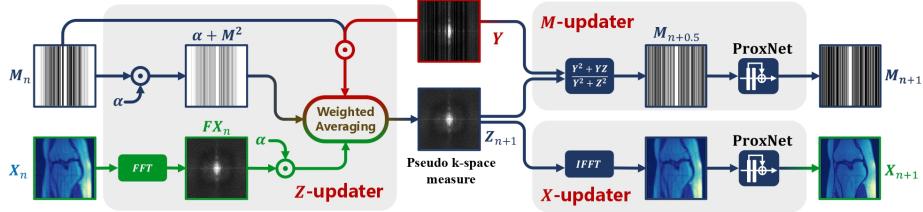


Figure 2: Illustration of the n -th stage of the proposed MXNet. It is composed of the step-by-step computations of these three updaters (Z -updater, M -updater, and X -updater), corresponding to Eqs. (15), (16) and (17). The inputs consist of a learned sampling mask M_n and the reconstructed image X_n from the previous stage. In the Z -updater, the pseudo k-space measure Z_{n+1} is obtained through the weighted averaging of FX_n and Y . In M -updater and X -updater, M_{n+1} and X_{n+1} are separately computed via a ProxNet network.

3) Updating X : With Z and M fixed, the updating of X is accomplished by addressing the optimization sub-problem as:

$$\min_X \alpha \|FX - Z_{n+1}\|_F^2 + \lambda_2 \mathcal{R}_2(X). \quad (12)$$

Since F is a unitary transformation, Eq. (12) is equivalent to

$$\min_X \alpha \|X - F^{-1}Z_{n+1}\|_F^2 + \lambda_2 \mathcal{R}_2(X), \quad (13)$$

where F^{-1} represents the inverse Fourier transform operator. This is also a standard least squares regularization problem, whose close-form solution is:

$$X_{n+1} = \text{Prox}_{\lambda_2/\alpha}(F^{-1}Z_{n+1}), \quad (14)$$

where $\text{Prox}_{\lambda_2/\alpha}(\cdot)$ represents the proximal operator, which is related to the regularization term $\mathcal{R}_2(\cdot)$.

As seen, for solving Eq. (4), the entire iterative process sequentially consists of Eqs. (7), (11) and (14), with simple operators, which makes it possible to unfold them into the corresponding network modules as analyzed below.

3. Model-Driven MRI Reconstruction Network

Motivated by the great success of the deep unfolding techniques in various tasks, like hyperspectral fusion [30] and CT metal artifact reduction[31], in the section, we

aim to exploit this technique and carefully construct a model-driven MRI reconstruction framework by unfolding the iterative steps (7) (11) and (14) into the corresponding network modules, called MXNet. For every network stage, the network structure sequentially consists of \mathbf{Z} -updater, \mathbf{M} -updater, and \mathbf{X} -updater for accomplishing the iterative optimization of variables \mathbf{Z} , \mathbf{M} and \mathbf{X} , respectively. Specifically, the corresponding operations are as follows:

$$\mathbf{Z}\text{-updater} : \mathbf{Z}_{n+1} = \frac{\alpha \mathbf{F} \mathbf{X}_n + \mathbf{M}_n \odot \mathbf{Y}}{\alpha + \mathbf{M}_n^2}, \quad (15)$$

$$\mathbf{M}\text{-updater} : \begin{cases} \mathbf{M}_{n+0.5} = \left(\frac{\beta \mathbf{Y}^2 + \mathbf{Y} \odot \mathbf{Z}_{n+1} + \omega}{\beta \mathbf{Y}^2 + \mathbf{Z}_{n+1}^2 + \omega} \right), \\ \mathbf{M}_{n+1} = \text{ProxNet}_{\theta_m^{(n)}}(\mathbf{M}_{n+0.5}) \end{cases}, \quad (16)$$

$$\mathbf{X}\text{-updater} : \mathbf{X}_{n+1} = \text{ProxNet}_{\theta_x^{(n)}}(\mathbf{F}^{-1} \mathbf{Z}_{n+1}), \quad (17)$$

where $\text{ProxNet}_{\theta_m^{(n)}}$ and $\text{ProxNet}_{\theta_x^{(n)}}$ are shallow ResNets with the parameters $\theta_m^{(n)}$ and $\theta_x^{(n)}$ at the iterative stage n , which are adopted to automatically learn the proximal functions. According to Eqs. (15) (16) and (17), we can easily construct the network structure at each stage, as shown in Fig. 2.

Remark to physical interpretability: It should be noted that the proposed network is not only consistent with the iteration steps of the optimization algorithm, but also with clear physical interpretability in its modules. Specifically, as shown in Fig. 2, we can find the following insightful conclusions about the proposed network modules.

- (I) The \mathbf{Z} -updater is particularly reasonable in human intuition. Actually, we can easily observe that the outputted \mathbf{Z}_{n+1} is achieved by the weighted averaging of $\mathbf{F} \mathbf{X}_n$ and \mathbf{Y} . For any position (i, j) , when $(\mathbf{M}_n)_{ij} = 0$, then the k-space measure is not observed in \mathbf{Y} , then $(\mathbf{Z}_{n+1})_{ij} = (\mathbf{F} \mathbf{X}_n)_{ij}$; when $(\mathbf{M}_n)_{ij} \neq 0$, and then $(\mathbf{Z}_{n+1})_{ij}$ would be the average of \mathbf{Y}_{ij} and $(\mathbf{F} \mathbf{X}_n)_{ij}$, with weights $(\mathbf{M}_n)_{ij}$ and α , respectively.

- (II) The \mathbf{M} -updater is also reasonable, where \mathbf{M}_{n+1} is achieved by mathematically comparing \mathbf{Y} and \mathbf{Z}_{n+1} and extracting the mask information from \mathbf{Y} . Specifically, by

Eq. (16), we can find that $\mathbf{M}_{n+0.5}$ approximately satisfies the following ruler for \mathbf{M} :

$$\mathbf{Y}_{ij} = 0 \& \mathbf{Z}_{ij} \neq 0 \rightarrow \mathbf{M}_{ij} = 0, \mathbf{Y}_{ij} = 0 \& \mathbf{Z}_{ij} = 0 \rightarrow \mathbf{M}_{ij} = 1 \text{ and } \mathbf{Y}_{ij} \neq 0 \rightarrow \mathbf{M}_{ij} = 1.$$

For this reason, we can obtain more information about the sample pattern. This greatly helps us to robustly update \mathbf{M} .

(III) The inherent physical degradation process (1) is explicitly embedded in X -updater and Z -updater, which finely guides the iterative updating of the sampling pattern \mathbf{M} , the k-space data \mathbf{Z} , and the MR image \mathbf{X} . In this physics-driven manner, the mask is rationally encoded into the entire optimization process for assisting in the reconstruction of MR image.

Network details. Similar to [24][31], we consist the proximal network $\text{ProxNet}_{\theta_x^{(n)}}$ with two Resblocks, each formed by a sequence of Convolution+ReLU+Convolution+Skip connection. $\text{ProxNet}_{\theta_m^{(n)}}$ has the same structure with $\text{ProxNet}_{\theta_x^{(n)}}$, while at the final iterative stage an extra sigmoid activation layer is adopted for constraining the optimized \mathbf{M} to the range $[0, 1]$. Compared to the strict binary constraint $\mathbf{M}_{ij} \in \{0, 1\}$, such a relaxation processing has been shown to facilitate a more effective information optimization in previous research[32]. Besides, the initialization of \mathbf{Z} , \mathbf{M} and \mathbf{X} are necessary for the iteration of network stages. In our experiments, \mathbf{Z}_0 is easily initialized by adopting Fourier transformation on the under-sampled measure \mathbf{Y} , and \mathbf{M}_0 and \mathbf{X}_0 are readily derived from \mathbf{Z}_0 , according to formulas (16) (17).

All the parameters, including step size, penalty parameter, and network parameters are set as learnable parameter, and automatically learn from training data in an end-to-end manner. We adopt the following loss to supervise the joint reconstruction of \mathbf{Z}_n and \mathbf{X}_n :

$$\mathcal{L} = \sum_{n=1}^{N_D} \frac{\|\mathbf{X}_n - \mathbf{X}_{gt}\|_F^2}{\|\mathbf{X}_{gt}\|_F^2} + \frac{\|\mathbf{Z}_n - \mathbf{Z}_{gt}\|_F^2}{\|\mathbf{Z}_{gt}\|_F^2} \quad (18)$$

where \mathbf{X}_n and \mathbf{Z}_n represent the recovered results in the image domain and frequency domain for N_D training data, respectively. \mathbf{X}_{gt} and \mathbf{Z}_{gt} denote the ground truth MR image and k-space data.

4. Experimental Detail

Datasets. We comprehensively evaluate the effectiveness of the proposed MXNet based on two publicly available datasets, *i.e.*, the Brain Dataset³ and the Knee Dataset⁴. Specifically, the Brain Dataset [5] consists of 200 fully-sampled MR images with 256×256 pixels collected from 10 volunteers. The imaging sequence parameters are given as: T1-weighted images: TE=2.29 ms, TR=1370 ms, FOV=220×220 mm², slice thickness=1 mm; and PD-weighted images: TE=10 ms, TR=1500 ms, FOV=220×220 mm, slice thickness=1 mm. According to the ratio 2:1:1, 200 images are randomly split into training, validation, and testing, respectively. The Knee Dataset is from the NYU FastMRI Dataset [33]. Following [34][10], we filter out 973 and 199 PDWI knee volumes for training and testing containing 34,732 and 7135 fully-sampled MRI images, respectively. The Knee Dataset is acquired under the Coronal PD-weighted Protocol as: TE =27-32 ms, TR = 2750-3000 ms, FOV = 140×140 mm², and slice thickness = 3 mm. Every MR image is cropped to the central region of 320 × 320 pixels.

For each dataset, given the aforementioned fully-sampled MR image X_{gt} , following the publicly available simulation procedure⁵, we can generate the accelerated Cartesian down-sampling matrix \mathbf{M} and then synthesize the corresponding under-sampled k-space data \mathbf{Y} based on the degradation process in Eq. (1). More details about \mathbf{M} and the noise level σ are described in the following experiments where we consider several different settings for comprehensive comparisons.

Implementation Details. Our proposed MXNet is implemented based on the PyTorch framework with two NVIDIA Tesla V100 GPUs with 32G memory. The iterative stages N is 10 and the ResBlocks in every proximal network is 2. Adam optimizer is utilized to train the network and the batch size is 1. For the Brain and Knee Dataset, the total numbers of training epochs are 1000 and 30, respectively. The initial learning rate is 0.0003 and it is multiplied by 0.5 at epochs [40, 80] for the Brain Dataset.

Comparison Baselines. We evaluate our MXNet by comparing it with two cat-

³ <https://github.com/yangyan92/Deep-ADMM-Net>

⁴ <https://fastmri.med.nyu.edu/>

⁵ <https://github.com/facebookresearch/fastMRI>

egories of baselines: (1) methods using fixed down-sampling patterns, which we call “Fixed”, including ADMMCS-Net[5] ,and MD-Rec-Net[35] and (2) methods handling flexible down-sampling patterns, which we call “Flexible”, including MoDL[8] and KTMR[36]. The method using fixed down-sampling is quite sensitive to changes in the mask, while the second method is insensitive to the different down-sampling patterns.

Performance Metrics. Following [7] [13], we use the commonly-adopted metrics for the quantitative evaluation on the MR reconstruction performance, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and normalized root mean square error (NRMSE)[5].

5. Experimental Results

In this section, to comprehensively validate the effectiveness of our proposed method under varying down-sampling patterns M and noise levels σ , we consider several experimental settings, including consistent training-testing cases and inconsistent training-testing cases.

5.1. Consistent Training-Testing Setting

In the consistent training-testing setting, for every dataset, we adopt the degradation process (1) with the same down-sampling pattern M to generate the training and testing set. Here M is generated based on the $\times s$ accelerated Cartesian down-sampling setting where s is the down-sampling factor.

Quantitative Evaluation. Table 1 evaluates all competing methods on different down-sampling factors s ($s = 4, 8$) for the Brain Dataset. The consistent part in Table 3 shows the results of the Knee Dataset for $s = 4$. Due to its time-consuming computation procedure, we did not include ADMMCS-Net results for the Knee Dataset. The experiments of both datasets are executed on noise-free ($\sigma = 0$) and noisy settings ($\sigma = 0.03$). The best results are marked in bold font and the second ones are underlined.

Table 1: (Consistent training-testing case): The average PSNR (dB), SSIM(%) and NRMSE(%) with standard deviation of all comparison methods on Brain Dataset under different down-sampling factors and noise levels σ .

Method		Brain Dataset ($\sigma=0$)								
		×4			×8			×10		
		PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
Fixed	Zero-Filling	28.34±2.03	68.01±6.36	24.16±5.35	23.61±2.13	52.75±8.88	43.53±4.53	22.32±1.93	48.78±8.05	53.83±7.48
	ADMMCS-Net	37.15±2.30	88.96±5.36	7.89±3.98	25.91±2.23	62.91±7.14	25.21±5.43	24.21±2.19	58.58±7.55	29.55± 6.29
	MD-Rec-Net	34.20±2.38	85.03±3.62	11.59±3.55	25.95±2.69	64.93 ±7.58	29.15±6.38	25.67 ±2.64	65.28±6.98	31.13±5.69
Flexible	MoDL	36.24± 2.59	86.78±4.08	9.36±2.69	26.81±2.35	63.38±7.33	28.23±6.21	25.81±2.58	60.93±6.32	32.44 ±5.89
	KTMR	33.54±2.08	85.42±3.53	12.75±3.41	29.03±2.14	71.92±10.85	22.30±6.58	28.17±2.24	70.12±11.87	24.14±7.64
	MXNet	37.74±2.40	90.43±2.82	7.79±2.06	29.56±2.45	71.21±6.79	20.28±6.22	29.07±2.54	69.77±7.07	21.61±6.86
Brain Dataset ($\sigma=0.03$)										
Method		×4			×8			×10		
		PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
Fixed	Zero-Filling	25.95±1.18	48.49±3.24	32.74±7.17	22.64±1.69	35.36±3.52	50.33±6.32	21.58±1.63	32.87±3.18	60.13±9.21
	ADMMCS-Net	32.76±1.96	76.26±5.25	13.35±4.35	24.59±2.01	56.89±6.79	28.78±5.97	22.87 ±1.68	52.47 ±5.91	33.64 ±6.89
	MD-Rec-Net	31.30±1.69	80.59 ±5.67	15.62±2.03	25.98±1.95	66.00±5.81	49.05±6.02	25.83 ±2.09	66.04 ±4.36	32.43±7.31
Flexible	MoDL	30.24±1.63	69.34±5.35	18.83±3.41	20.59±2.35	20.39±5.89	67.38±6.28	20.38 ±1.33	19.86±5.68	66.22±7.36
	KTMR	32.33±1.73	80.84±6.37	14.48±3.26	28.74±1.92	72.79±9.99	23.69±6.56	28.07 ±1.99	71.29±10.69	21.59±5.56
	MXNet	33.89±1.25	80.28±6.45	11.91±2.77	28.79±2.20	67.39±6.67	22.15±6.33	28.21 ±2.30	66.55±6.73	23.70±7.06

It is observed that: 1) Compared to other baselines, our method MXNet consistently achieves higher PSNR and SSIM scores and obtain lower NRMSE on the two benchmark datasets under different settings, *i.e.*, down-sampling factors and noise levels; 2) It is expected that compared to the noiseless case ($\sigma = 0$), the involvement of the random noise ($\sigma = 0.03$) adversely affects the accurate estimation of k-space data and almost all the comparison methods present a performance drop for MR reconstruction.

Visual Evaluation. Fig. 3 presents the reconstruction results and the corresponding error maps as well as PSNR/SSIM on an image slice randomly selected from the Brain Dataset under different down-sampling factors in the noise-free case. As seen, the traditional Zero-Filling method contains severe artifacts. MoDL, ADMMCS-Net, and MD-Rec-Net also blur image details evidently. While KTMR will introduce new detailed information while removing artifacts, which is not favorable for clinical diagnosis. In contrast, from the error maps, our proposed method consistently achieves better image fidelity and obtains higher PSNR/SSIM scores under different down-sampling factors, and the corresponding error maps show fewer artifacts.

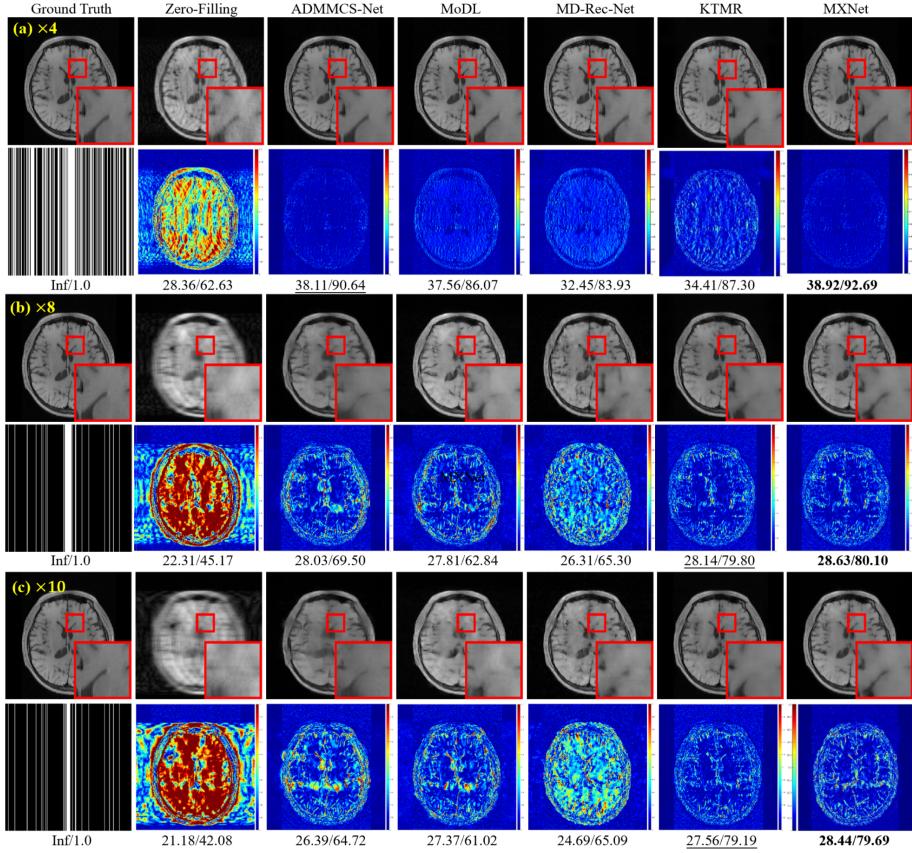


Figure 3: (Consistent training-testing case): Performance comparisons on the Brain Dataset under different down-sampling factors with the noise-free setting ($\sigma = 0$). Specifically, for every down-sampling factor, the first row shows the reconstructed image and the second row represents the error maps between its reconstructed result and the ground truth. The first column is the ground truth image (upper) and down-sampling pattern (lower). PSNR/SSIM is listed below for easy reference.

Fig. 4 presents the comparisons on the noisy case for Brain Dataset. It is evident that deep learning based methods ADMMCS-Net, MoDL, MD-Rec-Net, and KTMR are more sensitive to noise and they demonstrate obvious blurring effects. The noise consistently causes the restored MR images to exhibit varying degrees of detail loss and over-smoothness. Nevertheless, it is noticed that our proposed method exhibits better detail preservation, resulting in higher PSNR/SSIM scores. Similar trends can be observed on the Knee Dataset as shown in Fig. 5. It should be noted that for the setting with $\times 10$, the reconstructed images of all the competing methods have obvious blurring effects with loss of textual tissue features. This implies that the existing methods still need further improvement to be effectively used in this high-ratio acceleration situation.

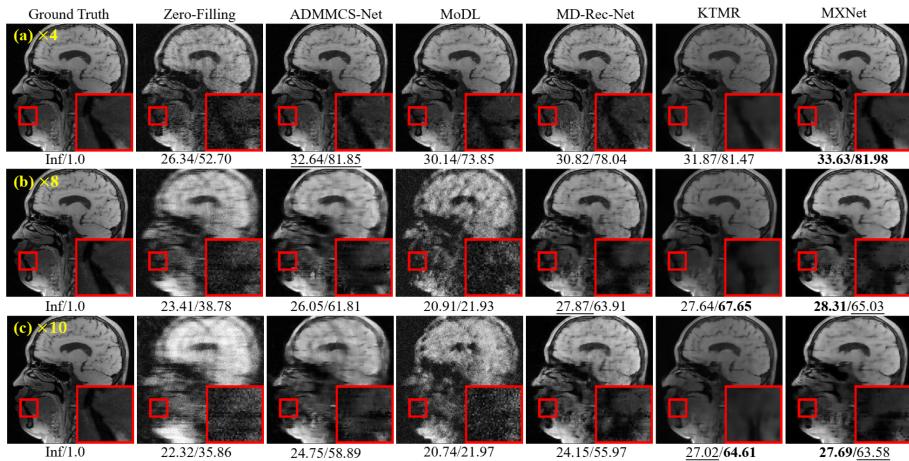


Figure 4: (Consistent training-testing case): Comparison on Brain Dataset under different down-sampling factors with the noisy setting ($\sigma = 0.03$).

5.2. Inconsistent Training-Testing Setting

In clinical applications, to obtain a comprehensive and accurate representation of anatomy, multi-source MRI with different sampling protocols is usually adopted [37]. Motivated by this practical necessity, we further simulate a more complicated setting that for each fully-sampled MR image from either the training set or the testing set, the corresponding degraded k-space data is generated based on different $\times s$ accelerated

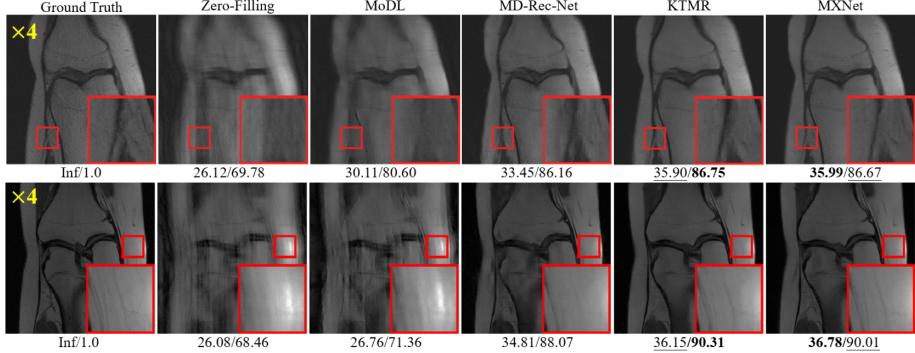


Figure 5: (Consistent training-testing case): Comparison on Knee Dataset under $\times 4$ down-sampling factor with $\sigma = 0$ (top) and $\sigma = 0.03$ (bottom).

Cartesian down-sampling patterns \mathbf{M} for Eq. (1). Similar to Sec. 5.1, for this inconsistent training-testing scenario, we also consider two different noise-level settings, *i.e.*, noise-free case with $\sigma = 0$, and noisy case with $\sigma = 0.03$.

Quantitative Evaluation. Table 2 lists the results of different comparison methods under this inconsistent training-testing setting with different noise levels. Compared with Table 1 where the training set and testing set have the same simulation process, it is observed that under such a more challenging cross-domain testing setting, the performance of almost all comparing methods deteriorates dramatically, especially on the Brain Dataset. However, the performance of our proposed MXNet fluctuates less and shows strong generalization among all competing methods. This is mainly attributed to the careful model-driven network design, which is helpful for rationally regularizing the mask learning and finely boosting the joint estimation of MR images, thus alleviating the generalization issue in such cross-domain scenarios.

Visual Evaluation. Fig. 6 displays the visual reconstruction of all comparison methods on Brain Dataset with the noisy setting $\sigma = 0.03$. Fig. 7 shows that visual comparison on Knee Dataset $\times 4$ down-sampling setting under the noisy setting $\sigma = 0, 0.03$. As seen, our proposed MXNet accomplishes a better reconstruction effect with higher structural fidelity under different down-sampling rates. For the noisy inconsistent training-testing setting, the existence of the simulation noise negatively impairs the MR image restoration. However, our proposed MXNet still perform competitively.

Table 2: (Inconsistent training-testing case): The average PSNR (dB), SSIM(%), and NRMSE(%) of all comparison methods on the Brain Dataset under different down-sampling factors and noise levels σ .

Brain Dataset ($\sigma=0$)										
Method		$\times 4$			$\times 8$			$\times 10$		
		PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
Fixed	Zero-Filling	19.46 \pm 2.07	51.55 \pm 6.43	51.19 \pm 5.31	22.43 \pm 2.11	46.65 \pm 8.91	36.99 \pm 4.95	21.91 \pm 2.07	44.55 \pm 8.87	39.86 \pm 5.82
	ADMMCS-Net	<u>34.58\pm 2.52</u>	<u>85.14\pm 3.81</u>	<u>11.10\pm 3.35</u>	24.53 \pm 2.09	60.03 \pm 7.03	28.95 \pm 5.96	22.51 \pm 2.13	54.42 \pm 7.62	33.63 \pm 6.15
	MD-Rec-Net	30.71 \pm 2.71	78.94 \pm 5.16	17.38 \pm 5.26	24.61 \pm 2.16	62.07 \pm 7.21	32.69 \pm 7.65	22.54 \pm 3.25	59.51 \pm 8.34	49.74 \pm 5.49
Flexible	MoDL	32.46 \pm 2.51	84.40 \pm 4.85	12.52 \pm 4.16	25.14 \pm 2.18	59.71 \pm 6.22	34.26 \pm 5.89	24.18 \pm 3.02	58.34 \pm 7.98	38.80 \pm 6.33
	KTMR	30.06 \pm 2.01	80.54 \pm 5.14	19.64 \pm 4.91	<u>27.44\pm 1.79</u>	<u>63.37\pm 6.44</u>	30.39 \pm 6.25	<u>27.06\pm 1.65</u>	<u>62.69\pm 4.88</u>	26.85\pm 7.22
	MXNet	36.18\pm 2.67	89.11\pm 3.69	9.49\pm 3.01	28.09\pm 2.34	68.49\pm 6.39	23.77\pm 5.78	27.29\pm 2.24	66.18\pm 7.43	<u>26.03\pm 6.60</u>
Brain Dataset ($\sigma=0.03$)										
Method		$\times 4$			$\times 8$			$\times 10$		
		PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
Fixed	Zero-Filling	19.56 \pm 1.54	39.16 \pm 3.01	52.86 \pm 7.03	21.65 \pm 1.74	31.26 \pm 3.67	42.27 \pm 6.61	21.23 \pm 1.75	30.32 \pm 3.60	45.01 \pm 7.46
	ADMMCS-Net	<u>30.69\pm 1.62</u>	71.29 \pm 4.68	<u>16.74\pm 4.17</u>	23.01 \pm 1.74	52.87 \pm 6.30	32.72 \pm 6.26	21.62 \pm 1.87	49.54 \pm 7.05	36.91 \pm 6.62
	MD-Rec-Net	30.57 \pm 2.70	<u>75.83\pm 5.06</u>	17.77 \pm 5.01	24.04 \pm 3.34	60.79 \pm 9.58	52.15 \pm 14.9	23.41 \pm 2.44	60.14 \pm 7.58	35.19 \pm 5.20
Flexible	MoDL	27.62 \pm 1.21	60.80 \pm 4.41	25.89 \pm 6.86	20.27 \pm 2.03	18.25 \pm 7.31	67.97 \pm 5.69	20.12 \pm 2.10	18.48 \pm 1.96	67.92 \pm 5.67
	KTMR	30.06 \pm 1.87	74.13 \pm 8.74	17.79 \pm 4.81	<u>27.44\pm 1.79</u>	<u>63.37\pm 6.44</u>	<u>30.39\pm 6.25</u>	<u>26.03\pm 1.53</u>	<u>61.19\pm 5.95</u>	<u>29.76\pm 6.20</u>
	MXNet	32.62\pm 1.62	76.63\pm 4.44	13.86\pm 3.21	<u>27.38 \pm 2.33</u>	64.04\pm 6.34	25.60 \pm 5.98	26.96\pm 2.14	63.63\pm 7.23	27.35\pm 6.92

Table 3: (Consistent and Inconsistent training-testing) The average PSNR (dB), SSIM(%), and NRMSE(%) of all comparison methods on the Knee Dataset under $\times 4$ down-sampling factors and noise levels σ .

Method	Consistent case						Inconsistent case					
	$\sigma=0$			$\sigma=0.03$			$\sigma=0$			$\sigma=0.03$		
	PSNR	SSIM	NRMSE									
Zero-Filling	27.32 \pm 2.33	62.45 \pm 7.45	28.01 \pm 5.83	22.92 \pm 3.21	61.02 \pm 7.56	48.03 \pm 5.21	26.36 \pm 2.89	65.73 \pm 8.90	33.11 \pm 5.67	29.48 \pm 3.02	65.37 \pm 5.37	22.15 \pm 4.23
MoDL	29.34 \pm 2.88	66.44 \pm 5.23	22.64 \pm 5.01	27.74 \pm 2.21	63.37 \pm 7.21	26.53 \pm 3.79	29.78 \pm 2.53	65.74 \pm 6.54	21.59 \pm 4.71	29.76 \pm 2.88	65.91 \pm 6.21	21.58 \pm 4.03
MD-Rec-Net	29.46 \pm 2.32	68.26 \pm 6.11	22.66 \pm 4.32	29.85 \pm 1.89	<u>68.25\pm 5.39</u>	21.70 \pm 4.53	30.34 \pm 2.36	69.42\pm 7.32	20.68 \pm 5.86	30.68 \pm 2.45	69.29\pm 6.89	20.11 \pm 5.76
KTMR	<u>30.75\pm 2.01</u>	70.21\pm 6.83	18.35\pm 4.89	<u>30.47\pm 2.71</u>	69.21\pm 6.38	19.20\pm 4.23	<u>30.76\pm 2.47</u>	<u>68.33\pm 5.47</u>	<u>19.51\pm 3.78</u>	<u>31.23\pm 2.71</u>	<u>67.54\pm 4.89</u>	<u>19.13\pm 3.71</u>
MXNet	30.82\pm 2.47	<u>68.27\pm 5.79</u>	<u>19.77\pm 5.31</u>	30.66\pm 2.23	67.96 \pm 5.37	<u>20.05\pm 4.87</u>	31.58\pm 2.70	<u>68.83\pm 5.44</u>	<u>18.49\pm 5.33</u>	<u>31.44\pm 2.61</u>	<u>68.52\pm 4.91</u>	18.71\pm 5.11

5.3. Statistical Analysis

For all the above experiments results in the Sec. 5.1 and Sec. 5.2, we perform statistical testings on the PSNR, SSIM and NRMSE metrics following the approach in [38]. The p-values in all experimental settings are less than 0.05. Under the 5% significance level, the PSNR, SSIM and NRMSE metrics of MXNet are statistically significant. This confirms that the improvements achieved by MXNet are not due to

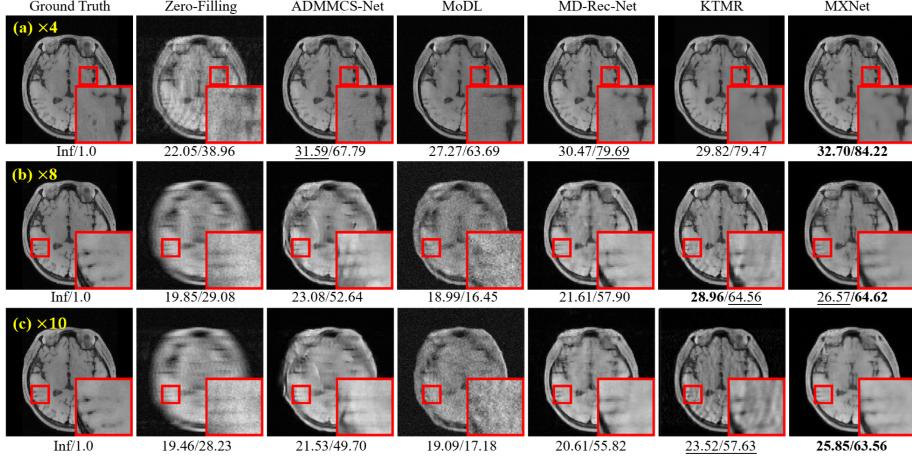


Figure 6: (Inconsistent training-testing case): Comparison on Brain Dataset under different down-sampling

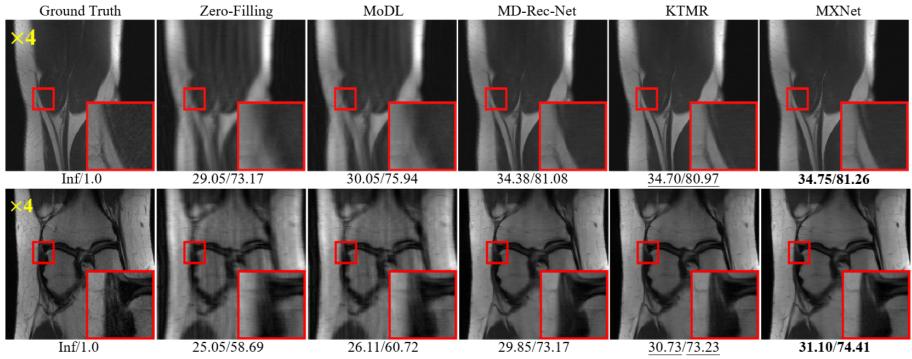


Figure 7: (Inconsistent training-testing case): Comparison on Knee Dataset with $\sigma = 0$ (upper) and $\sigma = 0.03$ (lower).

random chance but statistically robust.

5.4. Ablation Study

Efficacy of Model-Driven Down-Sampling Mask Learning. To fully validate the rationality of our proposed model-driven down-sampling mask learning, we add the comparison with a degraded version, named CXNet, where a conventional convolutional network replaces the proposed M -updater. That is, for CXNet, the down-sampling mask M is directly estimated by feeding Z_n into the CNN network, without the iterative procedure derived in Eq. (16). The design of the CNN network is identical to ProxNet $_{\theta_m^{(n)}}$ to guarantee a fair comparison. Furthermore, we also conduct additional

Table 4: Average PSNR (dB), SSIM(%) and NRMSE(%) of CXNet and MXNet under inconsistent training-testing setting.

Methods	Brain Dataset ($\times 4$)			Knee Dataset ($\times 4$)		
	PSNR \uparrow	SSIM \uparrow	NRMSE \downarrow	PSNR \uparrow	SSIM \uparrow	NRMSE \downarrow
Inconsistent training-testing: noise level $\sigma = 0$						
CXNet	34.69 \pm 2.35	86.65 \pm 4.32	11.30 \pm 2.31	31.40 \pm 3.22	68.45 \pm 5.23	18.74 \pm 3.45
MXNet	36.18 \pm 2.14	89.11 \pm 3.85	9.49 \pm 2.85	31.58 \pm 3.03	68.83\pm5.16	18.49 \pm 3.04
XNet	36.20\pm1.87	89.32\pm3.75	9.44\pm2.02	31.62\pm2.89	68.87\pm5.32	18.43\pm2.87
Inconsistent training-testing: noise level $\sigma = 0.03$						
CXNet	31.88 \pm 2.80	74.21 \pm 5.03	15.09 \pm 5.01	31.41 \pm 4.36	68.45 \pm 6.07	18.75 \pm 6.21
MXNet	32.62 \pm 1.93	76.63 \pm 4.82	13.86 \pm 4.75	31.44 \pm 4.17	68.52 \pm 5.83	18.71 \pm 5.70
XNet	32.68\pm2.01	77.05\pm4.37	13.72\pm5.21	31.62\pm3.78	68.95\pm5.76	18.43\pm5.32

comparative method named XNet, where the down-sampling pattern M is known and inputted into the network. In other words, we no longer need M -updater to update M in XNet.

In Table 4, the quantitative results of CXNet, MXNet and XNet are presented, particularly under the inconsistent training-testing setting described in Sec. 5.2 with $\times 4$ Cartesian sampling pattern. The comparison demonstrates that our MXNet consistently surpasses CXNet in performance under noise levels $\sigma = 0, 0.03$. This performance evaluation was conducted on both the Brain Dataset and Knee Dataset, reinforcing the effectiveness of MXNet in these varied contexts. Additionally, just as expected, XNet consistently achieves higher PSNR and SSIM scores and obtains lower NRMSE on the two benchmark datasets. However, our proposed MXNet and XNet still achieve competing performance. Also, MXNet shows good robustness on the proposed down-sampling mask estimation. Although MXNet does not preset the mask as a known input, it always obtains a quite competitive reconstruction performance, which is only slightly inferior to XNet.

Fig. 8 displays the iterative reconstruction processes of MXNet under the inconsistent training-testing setting on the Brain Dataset with noise level $\sigma = 0.03$. It is clearly observed that the down-sampling mask extracted by M -update in the proposed MXNet can be gradually ameliorated and approaches the ground truth, which in turn enhances the image reconstruction quality achieved by the X -update module.

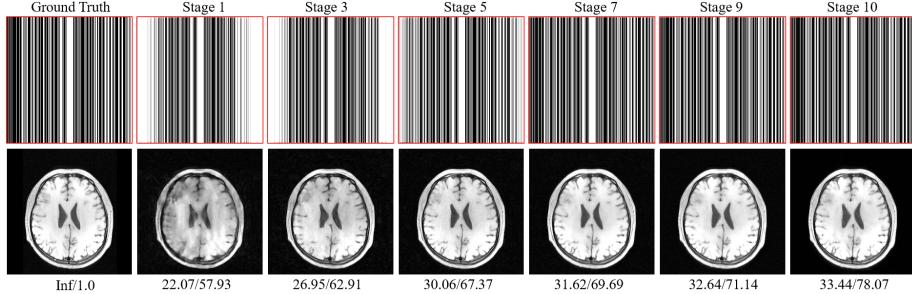


Figure 8: Down-sampling pattern M_n and the corresponding MR image X_n reconstructed by MXNet at different stages. PSNR/SSIM is provided below every image for quantitative evaluation.

Table 5: Effect of the total stage number N on the performance of the proposed MXNet on Brain Dataset.

Stage N	PSNR↑	SSIM↑	NRMSE↓	Inference Time
$N = 2$	29.87 ± 3.24	71.67 ± 3.58	19.02 ± 2.71	0.09s
$N = 4$	30.77 ± 2.75	72.69 ± 4.26	17.00 ± 3.32	0.10s
$N = 6$	29.26 ± 2.63	69.93 ± 5.23	20.56 ± 2.06	0.12s
$N = 8$	32.30 ± 1.01	75.28 ± 4.81	14.40 ± 3.78	0.14s
$N = 10$	32.62 ± 1.63	76.63 ± 4.71	13.86 ± 2.41	0.16s
$N = 12$	32.52 ± 1.87	76.01 ± 5.02	14.05 ± 2.33	0.17s

This finely substantiates the effectiveness and the rationality of the proposed deep unfolding framework. It is the explicit embedding of the physical imaging mechanism defined in Eq. (1) that helps MXNet be optimized towards to the right direction, thus achieving relatively accurate extraction of the mask and then rationally boosting the fine estimation of MR images.

The Impact of the Number of Stages N . Based on the noisy inconsistent training-testing setting on Brain Dataset with $\times 4$ down-sampling as described in Sec. 5.2, we investigate the impact of the number of iterative stages N on the performance of the proposed MXNet. As reported in Table 5, with varying N from 2 to 10, the reconstruction performance generally can be improved at the cost of increasing inference time. Here the times are averagely computed on one image with 256×256 pixels based on an NVIDIA Tesla V100 GPU. Besides, when we increase N from 10 to 12, there is a slight decrease in performance. To strike a balance between network performance and inference time, we set N to 10 in all comparison experiments, as deeper network structures can hinder effective gradient propagation.

Table 6: Results of our proposed MXNet and XNet on Brain Dataset under different parameter settings for $\text{ProxNet}_{\theta_m}$ and $\text{ProxNet}_{\theta_x}$ across different iterative stages.

Method	$\times 4$			$\times 8$			$\times 10$			Parameters
	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	
MXNet(Shared)	32.62 \pm 2.34	76.63 \pm 4.31	13.86 \pm 3.26	27.38 \pm 1.74	64.04 \pm 5.23	25.60 \pm 3.31	26.96 \pm 1.78	63.63\pm6.23	27.35 \pm 3.11	315.5k
MXNet(Unshared)	32.67\pm2.41	77.67\pm3.86	13.83\pm3.71	27.95\pm2.02	66.14\pm4.93	24.01\pm2.83	27.33\pm2.03	63.06 \pm 4.79	25.79\pm2.78	1,730.6k
XNet(Shared)	32.68 \pm 2.14	77.05 \pm 4.23	13.72\pm2.89	28.02 \pm 2.31	65.58 \pm 4.86	23.97 \pm 2.75	27.14 \pm 1.69	64.10\pm4.96	26.66 \pm 2.83	315.3k
XNet(Unshared)	32.69\pm1.79	77.75\pm4.40	13.85 \pm 3.02	28.04\pm2.28	66.28\pm5.30	23.96\pm2.34	27.44\pm1.45	63.91 \pm 5.02	25.63\pm3.21	1,729.0k

ProxNet $_{\theta_m^{(n)}}$ and ProxNet $_{\theta_x^{(n)}}$ across Different Stages. Here we study the proximal networks ProxNet $_{\theta_m^{(n)}}$ and ProxNet $_{\theta_x^{(n)}}$ by separately setting their involved network parameters $\theta_m^{(n)}$ and $\theta_x^{(n)}$ to be shared or unshared between different stages. Table 6 reports the results of MXNet and XNet on the Brain Dataset under the noisy inconsistent setting with different down-sampling factors. As observed, with the non-shared setting, both MXNet and XNet can generally achieve better reconstruction performance with higher PSNR/SSIM scores and lower NRMSE scores, which outperform the corresponding version with the shared setting, respectively. However, such performance improvement comes at the expense of increased network parameters. Considering the performance and computation cost, in all experiments, we choose the parameter-shared setting for different iterative stages by default.

6. Conclusion and Future Work

In this paper, against the MR reconstruction task with an unknown down-sampling pattern, we have built a novel model-driven blind MR reconstruction network, called MXNet. Through model visualization, we have validated the clear working mechanism underlying the proposed MXNet. Comprehensive experiments conducted under different settings substantiated the superiority and robustness of our proposed method beyond the current representative MR reconstruction methods as well as its fine generalization capability in the cross-domain scenario.

Following most of the existing work, in this paper, we only evaluate the effectiveness of the proposed model-driven framework in the single-coil scenario. Actually, the proposed model-driven paradigm together with its mask learning design can be easily

extended to the multi-coil scene. Besides, combining MRI reconstruction with super-resolution to accelerate MR imaging is a new trend[39].

The combination with the work[21][13][23] on sampling pattern learning is worthy of considering in future work. With the proposed framework, the sampling pattern design method can be further optimized from multiple aspects, including embedding physical imaging mechanisms, adding the anatomical structure information of MRI, and introducing the multi-modal image information [9][40].

References

- [1] A. Pal, Y. Rathi, A review and experimental evaluation of deep learning methods for MRI reconstruction, *The Tournal of Machine Learning for Biomedical Imaging* 1 (2022).
- [2] G. Yang, et al., DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction, *IEEE Transactions on Medical Imaging* 37 (6) (2017) 1310–1321.
- [3] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [4] Y. Zhang, J. Lyu, X. Bi, A dual-task dual-domain model for blind mri reconstruction, *Computerized Medical Imaging and Graphics* 89 (2021) 101862.
- [5] Y. Yang, et al., ADMM-CSNet: A deep learning approach for image compressive sensing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (3) (2018) 521–538.
- [6] Y. Chen, et al., Model-based convolutional de-aliasing network learning for parallel MR imaging, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III* 22, Springer, 2019, pp. 30–38.

- [7] A. Sriram, et al., End-to-end variational networks for accelerated MRI reconstruction, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23, Springer, 2020, pp. 64–73.
- [8] H. K. Aggarwal, M. P. Mani, M. Jacob, MoDL: Model-based deep learning architecture for inverse problems, *IEEE Transactions on Medical Imaging* 38 (2) (2018) 394–405.
- [9] C.-M. Feng, H. Fu, S. Yuan, Y. Xu, Multi-contrast mri super-resolution via a multi-stage integration network, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24, Springer, 2021, pp. 140–149.
- [10] C.-M. Feng, Y. Yan, G. Chen, Y. Xu, Y. Hu, L. Shao, H. Fu, Multimodal transformer for accelerated mr imaging, *IEEE Transactions on Medical Imaging* 42 (10) (2022) 2804–2816.
- [11] B. Kabas, F. Arslan, V. A. Nezhad, S. Ozturk, E. U. Saritas, T. Çukur, Physics-driven autoregressive state space models for medical image reconstruction, arXiv preprint arXiv:2412.09331 (2024).
- [12] Y. Korkmaz, V. M. Patel, Mambarecon: Mri reconstruction with structured state space models, arXiv preprint arXiv:2409.12401 (2024).
- [13] C. D. Bahadir, et al., Deep-learning-based optimization of the under-sampling pattern in MRI, *IEEE Transactions on Computational Imaging* 6 (2020) 1139–1152.
- [14] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, J. Tamir, Robust compressed sensing mri with deep generative priors, *Advances in Neural Information Processing Systems* 34 (2021) 14938–14954.
- [15] D. Gilton, G. Ongie, R. Willett, Model adaptation for inverse problems in imaging, *IEEE Transactions on Computational Imaging* 7 (2021) 661–674.

- [16] M. Bento, et al., Deep learning in large and multi-site structural brain mr imaging datasets, *Frontiers in Neuroinformatics* 15 (2022) 805669.
- [17] S. U. H. Dar, M. Özbey, A. B. Çatlı, T. Çukur, A transfer-learning approach for accelerated mri using deep neural networks, *Magnetic Resonance in Medicine* 84 (2) (2020) 663–685.
- [18] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. İlıcak, T. Cukur, J. Scarlett, V. Cevher, Learning-based compressive mri, *IEEE Transactions on Medical Imaging* 37 (6) (2018) 1394–1406.
- [19] F. Sherry, M. Benning, J. C. De los Reyes, M. J. Graves, G. Maierhofer, G. Williams, C.-B. Schönlieb, M. J. Ehrhardt, Learning the sampling pattern for mri, *IEEE Transactions on Medical Imaging* 39 (12) (2020) 4310–4321.
- [20] S. Ravula, B. Levac, A. Jalal, J. I. Tamir, A. G. Dimakis, Optimizing sampling patterns for compressed sensing mri with diffusion generative models, arXiv preprint arXiv:2306.03284 (2023).
- [21] H. K. Aggarwal, M. Jacob, J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction, *IEEE Journal of Selected Topics in Signal Processing* 14 (6) (2020) 1151–1162.
- [22] S. Seo, H. M. Luu, S. H. Choi, S.-H. Park, Simultaneously optimizing sampling pattern for joint acceleration of multi-contrast mri using model-based deep learning, *Medical Physics* 49 (9) (2022) 5964–5980.
- [23] Z. Huang, S. Ravishankar, Single-pass Object-adaptive data undersampling and reconstruction for MRI, *IEEE Transactions on Computational Imaging* 8 (2022) 333–345.
- [24] J. Fu, et al., KXNet: A model-driven deep neural network for blind super-resolution, in: European Conference on Computer Vision, Springer, 2022, pp. 235–253.

- [25] H. Wang, Q. Xie, Q. Zhao, D. Meng, A model-driven deep neural network for single image rain removal, in: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2020, pp. 3103–3112.
- [26] A. Majumdar, Compressed Sensing for Magnetic Resonance Image Reconstruction, Cambridge University Press, 2015.
- [27] B. Deka, S. Datta, Compressed Sensing Magnetic Resonance Image Reconstruction Algorithms, Springer Series on Bio-and Neurosystems (2019).
- [28] Z. Xu, et al., L 1/2 regularization, *Science China Information Sciences* 53 (2010) 1159–1169.
- [29] S. J. Wright, Coordinate descent algorithms, *Mathematical Programming* 151 (1) (2015) 3–34.
- [30] Q. Xie, et al., MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (3) (2020) 1457–1473.
- [31] H. Wang, et al., Dicdnet: deep interpretable convolutional dictionary network for metal artifact reduction in ct images, *IEEE Transactions on Medical Imaging* 41 (4) (2021) 869–880.
- [32] T. Roubíček, Relaxation in optimization theory and variational calculus, Vol. 4, Walter de Gruyter GmbH & Co KG, 2020.
- [33] J. Zbontar, et al., fastMRI: An open dataset and benchmarks for accelerated MRI, arXiv preprint arXiv:1811.08839 (2018).
- [34] K. Xuan, S. Sun, Z. Xue, Q. Wang, S. Liao, Learning mri k-space subsampling pattern using progressive weight pruning, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23, Springer, 2020, pp. 178–187.

- [35] M. Ran, et al., MD-Recon-Net: a parallel dual-domain convolutional neural network for compressed sensing MRI, *IEEE Transactions on Radiation and Plasma Medical Sciences* 5 (1) (2020) 120–135.
- [36] 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* 233 (2023) 107452.
- [37] K. H. Kim, et al., Varying undersampling directions for accelerating multiple acquisition magnetic resonance imaging, *NMR in Biomedicine* 35 (4) (2022) e4572.
- [38] C. Ye, X. Li, J. Chen, A deep network for tissue microstructure estimation using modified lstm units, *Medical Image Analysis* 55 (2019) 49–64.
- [39] C.-M. Feng, Y. Yan, H. Fu, L. Chen, Y. Xu, Task transformer network for joint mri reconstruction and super-resolution, in: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI* 24, Springer, 2021, pp. 307–317.
- [40] C.-M. Feng, Y. Yan, G. Chen, Y. Xu, Y. Hu, L. Shao, H. Fu, Multimodal transformer for accelerated mr imaging, *IEEE Transactions on Medical Imaging* 42 (10) (2022) 2804–2816.