

# SRM-Net: Joint Sampling and Reconstruction and Mapping Network for Accelerated 3T Brain Multi-Parametric MR Imaging

Yuxuan Liu , Kaicong Sun , Haikun Qi , Junwei Yang , Xiaopeng Zong, Yongsheng Pan , Yuning Gu , Sifan He, Han Zhang , Yu Zhang , and Dinggang Shen , Fellow, IEEE

**Abstract**—Multi-parametric magnetic resonance imaging (MRI) can provide complementary quantitative information by generating multi-parametric maps and is becoming a promising imaging technique for advanced medical diagnosis. However, multi-parametric MRI requires longer acquisition time than normal MRI scanning. The existing reconstruction methods for accelerated multi-parametric MRI suffer from suboptimal performance due to stage-wise optimization, and inefficient utilization of intra- and inter-contrast information. To address these challenges, we propose an all-in-one joint Sampling, Reconstruction, and Mapping network, dubbed as SRM-Net, for multi-parametric MRI reconstruction on multi-coil and multi-contrast MR images. Specifically, our model consists of three modules including sampling, reconstruction, and mapping. In the sampling module, we introduce a sampling scheme to generate individually-optimized sampling pattern across multi-contrast images. In the reconstruction module, we adopt a spatio-temporal attention mechanism, which is embedded in a dual-domain-based unrolling framework, to better exploit inter- and intra-contrast correlations. In the mapping module, we employ multi-layer perceptron to model

complex nonlinear mapping. Integrating Sampling, Reconstruction, and Mapping, our SRM-Net enables the end-to-end learning paradigm. Experimental results show that our SRM-Net generates superior multi-parametric maps including T1, T2\*, and PD for brain on 3T MR scanner compared to state-of-the-art methods, and meanwhile provides promising intermediate weighted MR images.

**Index Terms**—Multi-parametric MRI, parametric mapping network, quantitative MRI, sampling pattern optimization, spatio-temporal attention.

## I. INTRODUCTION

MULTI-parametric magnetic resonance imaging (MRI) is an imaging technique for quantitative MRI (qMRI) [1]. Largely different from the typically-used qualitative MRI in which only a single weighted MR image is reconstructed, multi-contrast images are acquired during multi-parametric MR imaging. Thus, multiple quantitative parametric maps such as T1 [2], T2\* [3] and Proton Density (PD) [4] can be simultaneously estimated. The parametric maps can provide quantitative information about tissue characteristics and compositions, which enables and facilitates quantitative clinical diagnosis. However, the acquisition time of multi-parametric MRI is much longer than qualitative MRI since acquisition of multiple contrast images with varying repetition time (TR), echo time (TE), and flip angles (FA) is needed. As we know that fast and multi-modality MRI acquisition is important for clinical study [5], [6], [7] and disease diagnosis [8], [9], it is essential to develop a reconstruction method that supports accelerated multi-parametric MR imaging.

The evolution of deep learning (DL) leads to huge development in various medical image applications [10], [11], [12], [13], [14], [15]. The existing DL-based qMRI methods usually fall into two categories [16]. The first category [17], [18], [19], [20] splits qMRI reconstruction into two steps, i.e., 1) multi-contrast image reconstruction from the under-sampled k-space, and 2) multi-parametric map estimation based on physical models. In the image reconstruction step, many CNN-based models [21], [22], [23], [24] have been employed. With regard to the map estimation step, empirical physical models are utilized [2], [4], [25], [26], [27] to fit the parametric mapping. However, these methods solve the two steps separately, and the second step highly relies on the quality and fidelity of the first step, i.e.,

Received 18 January 2024; revised 4 July 2024 and 20 October 2024; accepted 10 December 2024. Date of publication 30 December 2024; date of current version 12 May 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62131015, Grant 62250710165, Grant 62301318, and Grant 82394432, in part by Shanghai Pujiang Program under Grant 22PJ1410800, in part by the China Ministry of Science and Technology under Grant STI2030-Major Projects-2022ZD0209000 and Grant STI2030-Major Projects-2022ZD0213100, in part by Shanghai Municipal Central Guided Local Science and Technology Development Fund under Grant YDZX20233100001001, and in part by The Key R&D Program of Guangdong Province, China under Grant 2023B0303040001. (*Corresponding authors:* Kaicong Sun; Dinggang Shen.)

Yuxuan Liu, Haikun Qi, Junwei Yang, Xiaopeng Zong, Yongsheng Pan, Yuning Gu, Sifan He, and Han Zhang are with the School of Biomedical Engineering and State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, China.

Yu Zhang is with the School of Southern Medical University, China.

Kaicong Sun is with the School of Biomedical Engineering and State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China (e-mail: sunkc@shanghaitech.edu.cn).

Dinggang Shen is with the School of Biomedical Engineering and State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China, and with the Shanghai United Imaging Intelligence Company Ltd., Shanghai 201210, China, and also with the Shanghai Clinical Research and Trial Center, Shanghai 201210, China (e-mail: dinggang.shen@gmail.com).

Code is available at [https://github.com/aloneForLiu/fast\\_mri](https://github.com/aloneForLiu/fast_mri).

Digital Object Identifier 10.1109/TBME.2024.3523480

reconstruction of multi-contrast images. Thus, these methods usually suffer from sub-optimal map estimation, and have difficulty in dealing with reconstructed MR images contaminated by residual artifacts [16].

The second category [28], [29], [30], [31], [32], [33] performs the two steps of multi-contrast MR image reconstruction and parametric map estimation by a single network, i.e., estimating parametric maps directly from the under-sampled k-space data. MANTIS [31] employs U-Net to approximate the parametric map based on under-sampled data. Although these methods employ end-to-end learning scheme, they discard deep supervision of the intermediate reconstructed image, which not only increases the mapping difficulty, but also decreases the model interpretability. A recent work JoJoNet [34] proposes a framework for multi-contrast MR imaging by simultaneously optimizing sampling, reconstruction, and downstream task such as parametric map estimation. However, JoJoNet uses recurrent U-Net for multi-contrast reconstruction, which cannot jointly exploit all the multi-contrast information, especially for the reconstruction of early-stage contrast images. Moreover, JoJoNet uses a stack of fully connected layers as a mapping function, which cannot approximate complex nonlinear mapping between the reconstructed MR images and the parametric maps. Besides, Magnetic Resonance Fingerprinting (MRF) is also a multi-parametric Magnetic Resonance Imaging (MRI) technique [35], [36]. Specifically, the collected data (fingerprints) for individual voxels is compared with simulated fingerprints for MRF sequence. The best match for the voxel fingerprint is selected and the tissue properties that were used to generate the simulated fingerprint are identified as the pixel-wise maps.

Besides MRI reconstruction, sampling pattern optimization also plays a crucial role for accelerated MRI. Recent attempts have used deep neural networks [37], [38], [39], [40], [41], [42] for sampling pattern optimization. Some approaches employ reinforcement learning [37] or Monte Carlo Tree Search (MCTS) [38] to learn the policy for generating sampling pattern. PILOT [39] considers hardware constraints, such as peak currents and maximum slew rates of magnetic gradients, during sampling optimization. Nevertheless, PILOT is tailored for single-contrast sampling. Shor et al. [43] have extended PILOT to dynamic MRI. However, both methods are built on U-Net, which largely limits their reconstruction performance. LOUPE [40] employs an end-to-end learning scheme by relaxing the binarization of the probabilistic mask with a Sigmoid function to enable backpropagation. However, LOUPE is designed for single-contrast single-coil imaging, which is suboptimal for multi-contrast MR imaging. Following LOUPE, a LOUPE-based work [41] utilizes a straight-through (ST) estimator to calculate the gradient for threshold operations within a neural network. The use of ST is to push the sampling probability towards 0 or 1 during the training, instead of using decimals, to avoid performance degradation in the inference phase due to mismatch. However, similar to LOUPE and PILOT, it is a method dedicated for single-contrast sampling. Directly applying it to multi-contrast MR imaging could lead to suboptimum.

In fact, some practical limitation should be considered when dealing with sampling pattern optimization for multi-contrast

images. For instance, it is important to limit the change of acceleration factor among different TE images within one TR, while changes across TR and FA are allowed [27]. However, to our best knowledge, there is no existing work considering this practical constraint for multi-contrast sampling optimization. In addition, accurate coil sensitivity map estimation also plays a significant role for MRI reconstruction. In [44], the authors propose a deep unrolling network that jointly performs sensitivity map estimation and MRI reconstruction, which shows promising performance improvement. Considering the above challenges and insights, in this work, we propose SRM-Net, a joint optimization network that integrates sampling, reconstruction, and mapping into all-in-one framework. Our network **not only** supports end-to-end learning for multi-parametric map estimation, **but also** provides multi-contrast MR reconstruction. The main contributions of our work can be summarized as follows:

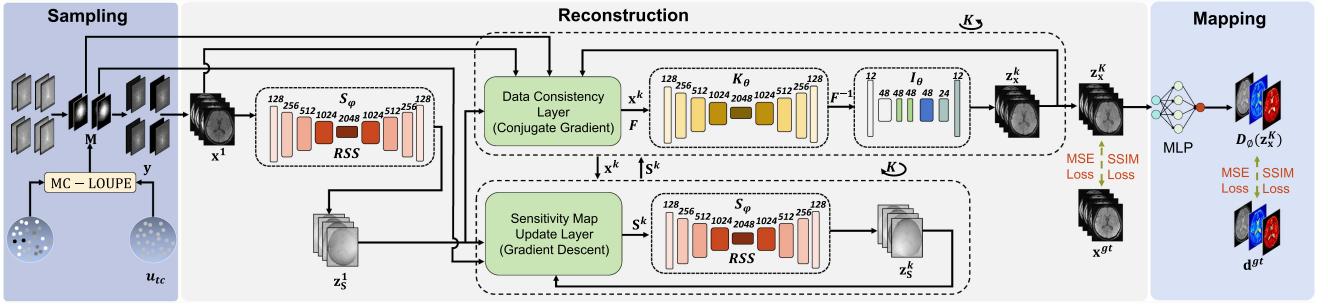
- We propose an all-in-one network called SRM-Net to simultaneously optimize multi-contrast sampling patterns, image reconstruction, and multi-parametric mapping.
- We introduce a sampling scheme, dubbed as multi-contrast LOUPE (MC-LOUPE), allowing multi-contrast images to have diverse sampling patterns while maintaining similar acceleration factors.
- We perform coil sensitivity map estimation and multi-contrast MRI reconstruction simultaneously based on deep unrolling. Besides, we adopt a spatio-temporal attention mechanism to more efficiently exploit the intra- and inter-contrast information.
- To enable end-to-end learning, we propose to utilize modality-specific multilayer perceptron (MLP) to approximate complex nonlinear mapping between multi-contrast images and multi-parametric maps, instead of using pre-defined physical models.
- Extensive experiments demonstrate that our SRM-Net achieves promising quantitative and qualitative improvement compared to state-of-the-art methods for brain multi-parametric mapping including T1, T2\*, and PD using gradient echo sequence.

## II. BACKGROUND

Given the acquired multi-contrast multi-coil k-space data defined as  $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_T} \mid \mathbf{y}_i \in \mathbb{C}^{N_C \times N}, i \in [1, 2, \dots, N_T]\}$ , where  $N = N_u \times N_v$  represents the number of pixels,  $N_T$  denotes the amount of contrasts, and  $N_C$  is the number of coils. For individual contrast image, the same sampling mask is used for all the coils. we can formulate the multi-contrast MR imaging model as

$$\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{b}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^{N_T \times N}$  denotes the latent multi-contrast MR images and  $\mathbf{b} \in \mathbb{C}^{N_T \times N_C \times N}$  is an additive noise. The operator  $\mathcal{A}$  acts as the system matrix which is usually defined as  $\mathcal{A} := \mathbf{M}\mathbf{F}\mathbf{s}$  where  $\mathbf{M} = \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{N_T}\}$  is a collection of multi-contrast sampling masks with diagonal matrix  $\mathbf{M}_i \in \{0, 1\}^{N \times N}$ ,  $\mathbf{F}$  is the



**Fig. 1.** Schematic illustration of our proposed SRM-Net for multi-parametric map estimation based on multi-contrast multi-coil MR imaging. SRM-Net consists of three cascaded modules, i.e., sampling mask optimization (Sampling), multi-contrast MRI reconstruction (Reconstruction), and multi-parametric map estimation (Mapping). Detailed description of each module is given in the Methods section.

2D discrete Fourier transform, and  $\mathbf{S}$  is the set of coil sensitivity maps.

Since accelerated MRI reconstruction is an ill-posed problem, regularization on the latent image  $\mathbf{x}$  is required to constrain the solution space. Typically, the optimization problem can be formulated as

$$\arg \min_{\mathbf{x}} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad (2)$$

with  $\lambda \in \mathbb{R}$  being the weight to balance the data fidelity term and the regularization term  $\mathcal{R}(\mathbf{x})$ . Based on the reconstructed multi-contrast images  $\mathbf{x}$ , multi-parametric maps can be approximated by parameter fitting described as below

$$\mathbf{d} = \mathcal{G}(\mathbf{x}), \quad (3)$$

where  $\mathbf{d} \in \mathbb{R}^N$  and  $\mathcal{G} : \mathbb{C}^{NT \times N} \rightarrow \mathbb{R}^N$  respectively denote the parametric map, e.g., T1 map, and the corresponding fitting function. Based on (2) and (3), we can yield the reconstructed multi-contrast images and the multi-parametric maps, respectively. However, solving them separately may lead to suboptimal parametric maps. To overcome this issue, we combine (2) and (3), and yield the following equation:

$$\arg \min_{\mathbf{x}, \mathbf{d}} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_x \mathcal{R}(\mathbf{x}) + \lambda_d \|\mathbf{d} - \mathcal{G}(\mathbf{x})\|_2^2. \quad (4)$$

When using network  $\mathcal{R}_\theta$  to approximate regularized image and network  $\mathcal{D}_\phi$  to approximate the optimal parameter mapping  $\mathcal{G}$ , (4) can be expressed as:

$$\arg \min_{\mathbf{x}, \mathbf{d}} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_x \|\mathbf{x} - \mathcal{R}_\theta(\mathbf{x})\|_2^2 + \lambda_d \|\mathbf{d} - \mathcal{D}_\phi(\mathbf{x})\|_2^2. \quad (5)$$

### III. METHODS

Our framework consists of three modules, namely sampling, reconstruction, and mapping as shown in Fig. 1, which cover the process from data acquisition to multi-parametric map estimation. The whole pipeline is optimized in an end-to-end manner. Details are given in the following sections.

#### A. Sampling

Sampling mask optimization has attracted increasing interest in recent years, and is especially vital for multi-contrast MR

imaging, since multi-contrast images share anatomical structures and complementary information among different contrasts should be sampled. LOUPE [40] is a representative deep learning-based method for probabilistic mask optimization. Our sampling scheme is based on LOUPE, but extended to multi-contrast multi-coil scenario, dubbed as multi-contrast LOUPE (MC-LOUPE). Specifically, we allow multi-contrast images to have different sampling masks to encourage sampling diversity. Besides, we enforce multi-contrast images from the same TR (repetition time) to have similar acceleration factor to match practical requirement.

Without loss of generality, we consider a multi-flip-angle (FA) and multi-TR scanning sequence. We define  $t$  as the index of the TR with  $t \in [1, \dots, T]$ , where  $T$  denotes the total number of TR. The probabilistic mask of the  $c$ th contrast (echo) image in the  $t$ th TR is denoted as  $\mathbf{p}_{tc}$ , where  $c \in [1, \dots, C]$  and  $C$  is the total number of images within each TR. Since each pixel has the sampling probability within  $[0, 1]$ , we employ the Sigmoid function to represent  $\mathbf{p}_{tc} = \frac{1}{1 + e^{-h\Omega}}$ , where  $\Omega \in \mathbb{R}^N$  and  $h$  determines the slope of the Sigmoid function. Thus, each element  $n$  in the binary sampling mask  $\mathbf{M}_{tc}$  is a Bernoulli variable with  $M_{tcn} \sim \text{Ber}(p_{tcn})$ . Similar to LOUPE, to ease the implementation of Bernoulli, another set of independent realizations are drawn from the uniformly distributed variables  $\mathbf{u}_{tc}$ , which has the same size as  $\mathbf{p}_{tc}$  and  $u_{tcn} \sim U(0, 1)$ , serving as the threshold. Thus, binarized sampling mask  $\mathbf{M}_{tc}$  can be realized by  $\mathbb{1}_{[u_{tc}, 1]}(\mathbf{p}_{tc})$  with  $\mathbb{1}$  being the indicator function.

We consider two constraints on the probabilistic mask  $\mathbf{p}$  during optimization. First, the probabilistic mask  $\mathbf{p}$  should retain a predefined acceleration factor  $R$ , which indicates that  $\mathbf{p}$  should satisfy

$$\sum_t \sum_c \sum_n p_{tcn} / (TCN) = \alpha, \quad (6)$$

where  $\alpha$  is defined as sparsity coefficient with  $\alpha = 1/R$ . For each probabilistic mask  $\mathbf{p}_{tc}$ ,  $\alpha_{tc} = \sum_n p_{tcn} / N$ . Besides, for multi-contrast imaging, we prefer to have diverse probabilistic masks for different contrast images to enrich acquired information and avoid repetitive sampling. However, optimizing  $\mathbf{p}_{tc}$  solely constrained by (6) is not adequate. Concerning practical limitation during scanning, large deviation of  $\alpha_{tc}$  among contrast

images within one TR is not desirable, since usually same amount of k-lines are acquired sequentially across different contrasts [27]. Concerning the above design rules, we propose two differentiable update schemes. Specifically, to regularize the maps  $\mathbf{p}$  to retain the overall sparsity coefficient, we employ the following adjustment:

$$Adj(\mathbf{p}) = \begin{cases} \frac{\alpha}{\sum \bar{\mathbf{p}}_t} \mathbf{p}, & \sum \bar{\mathbf{p}}_t \geq \alpha, \\ \frac{1-\sum \bar{\mathbf{p}}_t}{1-\alpha} \mathbf{p}, & \sum \bar{\mathbf{p}}_t < \alpha, \end{cases} \quad (7)$$

where  $\bar{\mathbf{p}}_t = \|\mathbf{p}_t\|_1/(CN)$  with  $\|\cdot\|_1$  is the L1-norm, and  $\mathbf{p} = \{\mathbf{p}_t | t \in [1, \dots, T]\}$  being the set of probabilistic maps. Similarly, we impose adjustment on the map  $\mathbf{p}_{tc}$  within one TR as below:

$$Adj(\mathbf{p}_{tc}) = \begin{cases} \frac{\bar{\mathbf{p}}_t}{\bar{\mathbf{p}}_{tc}} \mathbf{p}_{tc}, & \bar{\mathbf{p}}_{tc} \geq \bar{\mathbf{p}}_t, \\ \frac{1-\bar{\mathbf{p}}_t}{1-\bar{\mathbf{p}}_{tc}} \mathbf{p}_{tc}, & \bar{\mathbf{p}}_{tc} < \bar{\mathbf{p}}_t, \end{cases} \quad (8)$$

where  $\bar{\mathbf{p}}_{tc} = \|\mathbf{p}_{tc}\|_1/N$ . In such a way, different contrast images within one TR can be regularized to have similar  $\alpha_{tc}$ . In other words, we allow different TR to have different sparsity coefficients.

During the training phase, we integrate (8) into the sampling mask  $\mathbf{M}_{tc}$  by

$$\mathbf{M}_{tc} = \frac{1}{1 + e^{-h(Adj(\mathbf{p}_{tc}) - \mathbf{u}_{tc})}}, \quad (9)$$

while in the inference phase, the sampling mask  $\mathbf{M}_{tc}$  is binarized by

$$\mathbf{M}_{tc} = \begin{cases} 1, & Adj(\mathbf{p}_{tc}) \geq \mathbf{u}_{tc}, \\ 0, & Adj(\mathbf{p}_{tc}) < \mathbf{u}_{tc}. \end{cases} \quad (10)$$

## B. Reconstruction

In the last section, we formulate the update of sampling mask  $\mathbf{M}$ . In fact, for multi-coil based parallel imaging, coil sensitivity maps  $\mathbf{S}$  are also unknown, and estimation of  $\mathbf{S}$  plays a critical role for MRI reconstruction. Similar to sampling mask  $\mathbf{M}$ ,  $\mathbf{S}$  is also embedded in the system matrix  $\mathcal{A}$ . In this work, we synchronously reconstruct the MR image  $\mathbf{x}$  and estimate the coil sensitivity maps  $\mathbf{S}$ . To be specific, we adopt network  $\mathcal{S}_\varphi$  to estimate  $\mathbf{S}$  and network  $\mathcal{R}_\theta$  to reconstruct  $\mathbf{x}$ . Therefore, we extend (5) to the following equation:

$$\begin{aligned} \arg \min_{\mathbf{x}, \mathbf{p}, \mathbf{S}, \mathbf{d}} & \|\mathcal{A}_{\mathbf{p}\mathbf{S}}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_S \|\mathbf{S} - \mathcal{F}_\varphi(\mathbf{x})\|_2^2 \\ & + \lambda_x \|\mathbf{x} - \mathcal{R}_\theta(\mathbf{x})\|_2^2 + \lambda_d \|\mathbf{d} - \mathcal{D}_\phi(\mathbf{x})\|_2^2. \end{aligned} \quad (11)$$

where  $\mathcal{F}_\varphi(\cdot) := \mathcal{S}_\varphi(\cdot)/RSS(\mathcal{S}_\varphi(\cdot))$  and  $RSS$  represents root-sum-squares. We employ the unrolling framework, and in each unrolling iteration we update  $\mathbf{x}$  and  $\mathbf{S}$  alternately. To this end, we introduce auxiliary variables  $\mathbf{z}_x$  and  $\mathbf{z}_S$  and reformulate (11) as

$$\begin{aligned} \mathbf{S}^{k+1} &= \arg \min_{\mathbf{S}} \|\mathcal{A}_{\mathbf{p}\mathbf{S}}\mathbf{x}^k - \mathbf{y}\|_2^2 + \lambda_S \|\mathbf{S} - \mathbf{z}_S^k\|_2^2 \\ \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} \|\mathcal{A}_{\mathbf{p}\mathbf{S}}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_x \|\mathbf{x} - \mathbf{z}_x^k\|_2^2, \end{aligned} \quad (12)$$

where the auxiliary variables  $\mathbf{z}_x$  and  $\mathbf{z}_S$  are estimated by the networks  $\mathbf{z}_x^{k+1} = \mathcal{R}_\theta(\mathbf{x}^{k+1})$  and  $\mathbf{z}_S^{k+1} = \mathcal{F}_\varphi(\mathbf{S}^{k+1})$ , respectively,

and  $k$  denotes the unrolling index. When taking the gradient of (12) with respect to image  $\mathbf{x}$  and sensitivity map  $\mathbf{S}$ , we have

$$\begin{aligned} \nabla_{\mathbf{S}} &= \mathcal{B}^H (\mathcal{B}\mathbf{S}\mathbf{x}^k - \mathbf{y})\mathbf{x}^{kH} + \lambda_S (\mathbf{S} - \mathbf{z}_S) \\ \nabla_{\mathbf{x}} &= \mathcal{A}_{\mathbf{p}\mathbf{S}}^H (\mathcal{A}_{\mathbf{p}\mathbf{S}}\mathbf{x} - \mathbf{y}) + \lambda_x (\mathbf{x} - \mathbf{z}_x), \end{aligned} \quad (13)$$

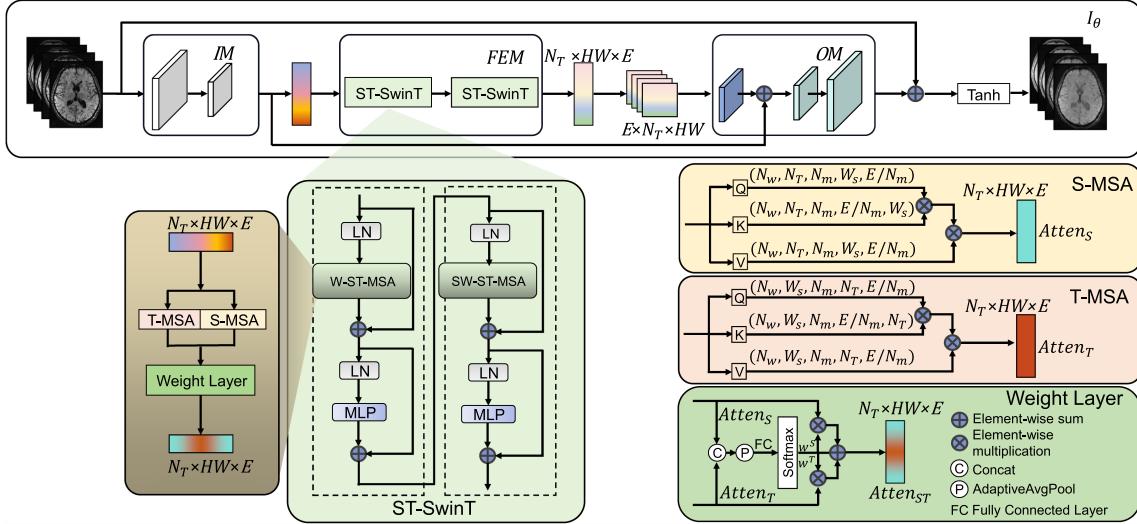
where  $\mathcal{B} := \mathbf{M}\mathcal{F}$ , and  $H$  represents the adjoint operator. Based on the gradient, we can resolve  $\mathbf{x}$  by the conjugate gradient (CG) algorithm and  $\mathbf{S}$  by the gradient descent algorithm. The update of reconstructed image  $\mathbf{x}$  and sensitivity map  $\mathbf{S}$  are respectively performed by the Data Consistency Layer and Sensitivity Map Update (SMU) layer as shown in Fig. 1. The estimated  $\mathbf{x}$  and  $\mathbf{S}$  are further refined by  $\mathcal{R}_\theta(\mathbf{x})$  and  $\mathcal{F}_\varphi(\mathbf{S})$  to obtain  $\mathbf{z}_x$  and  $\mathbf{z}_S$ , respectively.

In this work, we employ a U-shaped network as  $\mathcal{S}_\varphi$  to estimate the sensitivity map. Regarding network  $\mathcal{R}_\theta$ , to better exploit correspondence across contrasts and across coils, we employ a vision transformer based network  $\mathcal{I}_\theta$  in the image domain. Since each component in the k-space has global receptive field, we use a CNN-based network  $\mathcal{K}_\theta$  in the k-space domain to take advantage of the inductive bias of CNN. Formally, we have  $\mathcal{R}_\theta(\cdot) := \mathcal{I}_\theta(F^{-1}\mathcal{K}_\theta F(\cdot))$ , with  $F$  and  $F^{-1}$  being the Fourier and inverse Fourier transform operators, respectively. In fact,  $F$  is formulated by a matrix and denoted as  $\mathbf{F}$  in (1).

Specifically, we adopt U-Net as the backbone of  $\mathcal{K}_\theta$ , and concatenate the real and imaginary parts of the multi-contrast k-space data along the channel dimension.  $\mathcal{K}_\theta$  comprises four pooling and up-sampling layers, and each 2D convolutional layer has a filter size of  $3 \times 3$  followed by an instance normalization (IN) layer and a Leaky ReLU with a negative slope of 0.2. The output of  $\mathcal{K}_\theta$  is forwarded to the cascaded  $\mathcal{I}_\theta$ . To exploit multi-contrast information more efficiently, we introduce spatio-temporal Swin Transformer (ST-SwinT) in  $\mathcal{I}_\theta$ . The structure of  $\mathcal{I}_\theta$  is shown in Fig. 2, which includes an input module ( $IM$ ), a feature extraction module ( $FEM$ ), and an output module ( $OM$ ). Since performing self-attention directly in the original image dimension is computationally expensive, we first reduce the dimension through  $IM$ , which consists of two cascaded convolutions with a stride size of 2. These extracted features are then passed to  $FEM$ , which is composed of two cascaded ST-SwinT blocks. Finally, the feature maps are passed to  $OM$ , which consists of a residual convolutional layer and two deconvolutional layers to restore the original image size. Formally, the structure of  $\mathcal{I}_\theta$  can be described as

$$\begin{aligned} \mathbf{F}_{IM} &= IM(F^{-1}\mathcal{K}_\theta F(\mathbf{x})) \\ \mathbf{F}_{FEM} &= FEM(\mathbf{F}_{IM}) \\ \mathbf{F}_{OM} &= OM(\mathbf{F}_{FEM} + \mathbf{F}_{IM}), \end{aligned}$$

where  $\mathbf{F}_{IM} \in \mathbb{R}^{E \times N_T \times H \times W}$ ,  $\mathbf{F}_{FEM} \in \mathbb{R}^{E \times N_T \times H \times W}$ , and  $\mathbf{F}_{OM} \in \mathbb{R}^{2 \times N_T \times N}$  are the outputs of the individual modules and  $E$  denotes the embedding dimension. Finally, the output of  $\mathcal{I}_\theta$  is added with the input and forwarded to the hyperbolic tangent activation function (Tanh) to obtain the output  $\mathbf{z} = \text{Tanh}(\mathbf{F}_{OM} + F^{-1}\mathcal{K}_\theta F(\mathbf{x}))$ . The ST-SwinT block within  $\mathbf{F}_{FEM}$  is built on the backbone of Swin Transformer [45]. The structure of ST-SwinT is demonstrated in Fig. 2 and can be



**Fig. 2.** Structure of the proposed  $\mathcal{I}_\theta$  with the spatio-temporal Swin Transformer (ST – SwinT). ST – SwinT is composed of spatial attention S – MSA, temporal attention T – MSA, and fused attention in Weight Layer.

formally expressed as

$$\begin{aligned}\mathbf{F}_{X_1} &= W - ST - MSA(LN(\mathbf{F}_{IM})) + \mathbf{F}_{IM} \\ \mathbf{F}_{X_2} &= MLP(LN(\mathbf{F}_{X_1})) + \mathbf{F}_{X_1} \\ \mathbf{F}_{X_3} &= SW - ST - MSA(LN(\mathbf{F}_{X_2})) + \mathbf{F}_{X_2} \\ \mathbf{F}_{IM} &= MLP(LN(\mathbf{F}_{X_3})) + \mathbf{F}_{X_3},\end{aligned}$$

where  $W - ST - MSA$  represents the spatio-temporal multi-head attention, and  $MLP$  and  $LN$  denote the multilayer perceptron and layer normalization, respectively. Specifically, ( $S$ )  $W - ST - MSA$  consists of spatial-attention  $S - MSA$  and temporal-attention  $T - MSA$ . The spatial-attention and temporal-attention account for intra- and inter-contrast information, respectively. Formally, the spatial-attention  $Atten_S$  and the temporal-attention  $Atten_T$  can be expressed as

$$\begin{aligned}Atten_S &= \text{Softmax}(\mathbf{Q}_S \mathbf{K}_S^T / \sqrt{d} + \mathbf{B}_S) \mathbf{V}_S \\ Atten_T &= \text{Softmax}(\mathbf{Q}_T \mathbf{K}_T^T / \sqrt{d}) \mathbf{V}_T.\end{aligned}$$

The calculated  $Atten_S$  and  $Atten_T$  are integrated in the proposed Weight Layer to obtain  $Atten_{ST}$ . To be specific,  $S - MSA$  and  $T - MSA$  are concatenated and passed to an Adaptive AvgPool layer. An FC layer and Softmax are employed to change the embedding dimension from  $2E$  to  $2$ . The first scalar serves as the spatial-attention weight  $w^S$ , while the second one serves as the weight  $w^T$  for temporal-attention. Therefore, we can obtain the corresponding  $w^S$  and  $w^T$  for each contrast image. We multiply these two attention weights with the associated attention map and perform element-wise sum to yield the combined attention map  $Atten_{ST}$ . In the experiments, the embedding has a length of  $E = 48$ , the window size is  $W_s = 8 \times 8$ , and the length of window sequence  $N_w$  is set as  $6$ . In the spatio-temporal Swin transformer layers, the feature map with the size of  $N_T \times HW \times E$  is divided into  $N_T \times N_w$  non-overlapping windows with the size of  $W_s \times E$ . The estimated image is then

### Algorithm 1: SRM-Net.

**Input:** Initialized probabilistic masks  $p_{tc}$  and variables  $u_{tc}$

- 1: Calculate the sampling mask  $\mathbf{M}_{tc}$  by MC-LOUPE
- 2: Obtain  $\mathbf{y}$  through  $\mathbf{M}_{tc}$
- 3: **for**  $k = 1 \rightarrow K$  **do**
- 4:   Calculate  $\mathbf{z}_x^k = \mathcal{R}_\theta(\mathbf{x}^k)$  and  $\mathbf{z}_S^k = \mathcal{F}_\varphi(\mathbf{S}^k)$
- 5:   Calculate  $\mathbf{S}^{k+1}$  and  $\mathbf{x}^{k+1}$  by (12)
- 6: **end for**
- 7: Calculate reconstruction loss by (14)
- 8: **for**  $i = 1 \rightarrow m$  **do**
- 9:   Calculate map using mapping network  $\mathcal{D}_\phi^i(\mathbf{z}_x^K)$
- 10:   Calculate mapping loss by (15)
- 11: **end for**
- 12: Calculate total loss  $\mathcal{L}_{overall}$  by (16)

**Output:**  $\mathbf{z}_x^K, \mathbf{z}_S^K, \mathcal{D}_\phi^i(\mathbf{z}_x^K), \mathcal{L}_{overall}$

passed to DC layer, and the unrolled update continues until the last iteration  $K$ .

The loss of the reconstruction network contains two parts, including a structure similarity loss measured by structural similarity index measure (SSIM) and a pixelwise loss measured by mean squared error (MSE), as formulated below:

$$\mathcal{L}_R = \rho_1 \|\mathbf{z}_x^K - \mathbf{x}^{gt}\|_2^2 + (1 - SSIM(\mathbf{z}_x^K, \mathbf{x}^{gt})), \quad (14)$$

where  $\mathbf{x}^{gt}$  represents the ground-truth (GT) multi-contrast images, and  $\rho_1$  is a weighting factor to balance the two terms in  $\mathcal{L}_R$ .

### C. Parametric Mapping

The parametric maps are usually estimated by Bloch equations. However, the performance of physical-model based map estimation is highly dependent on the quality and fidelity of the reconstructed multi-contrast images. More importantly, it hampers the use of the end-to-end learning paradigm. To tackle

these issues, we resort to deep neural networks  $\mathcal{D}_\phi$  for map estimation, which can model highly nonlinear mappings. The mapping module is trained simultaneously with the sampling and reconstruction modules in an end-to-end manner, rather than considered as downstream task, since our goal is to estimate multi-parametric maps based on the reconstructed multi-contrast images.

Specifically,  $\mathcal{D}_\phi$  is built on multilayer perceptron (MLP) to approximate the nonlinear mapping between the reconstructed multi-contrast images and the corresponding parametric maps. Particularly, each parameter mapping network  $\mathcal{D}_\phi$  consists of four MLPs, and each MLP contains an activation layer, two FC layer, one GeLU [46], and one LN. The number of input channels is set as 24 and the output channels is set as 96. In this work, we target at three common parametric maps, i.e., PD Map, T1 Map, and T2\* Map. Each parametric map is estimated by an individual  $\mathcal{D}_\phi$ . Similar to the reconstruction loss, the mapping loss for the  $i$ th map is formulated as

$$\mathcal{L}_M^i = \rho_2 \|\mathcal{D}_\phi^i(\mathbf{z}_x^K) - \mathbf{d}_i^{gt}\|_2^2 + (1 - SSIM(\mathcal{D}_\phi^i(\mathbf{z}_x^K), \mathbf{d}_i^{gt})). \quad (15)$$

where  $\mathbf{d}_i^{gt}$  represents the GT of the  $i$ th parametric map, and  $\rho_i$  denotes the trade-off weight. It is worth noting that the reference PD, T1, and T2\* maps are estimated by physical model-based least squares fitting on the GT multi-contrast images. More detailed descriptions of the Bloch functions can be found in [27].

To jointly optimize multi-contrast image reconstruction and multi-parametric map estimation, we combine the reconstruction loss (as formulated in (14)) and the mapping loss (as expressed in (15)), and yield the overall loss for our SRM-Net as below:

$$\mathcal{L}_{overall} = \mathcal{L}_R + \sum_i^m \mathcal{L}_M^i, \quad (16)$$

where  $m$  is the number of estimated parametric maps. In this study, we set  $m = 3$  in the following experiments. The whole SRM-Net is summarized and formulated in Algorithm 1.

## IV. EXPERIMENTS

### A. Experimental Settings

**Datasets:** In the experiments, we have performed retrospective study on in-house brain and public knee data. Specifically, we use the public fastMRI knee dataset [47] and three in-house multi-echo multi-coil MRI datasets. The complex-valued knee dataset is composed of 266 paired PD (Proton density) and PDFS (Proton density with fat saturation) images with a matrix size of  $320 \times 320 \times 15 \times 33$ , where 33 represents the number of 2D slice and 15 indicates the number of coils for each slice. The voxel size is  $1 \times 1 \times 1 \text{ mm}^3$ . We select 3120/620/620 slices for training, validation, and testing, respectively.

All the complex-valued multi-echo multi-coil in-house datasets are acquired by MULTIPLEX (MTP) [27] on a 3 T scanner (uMR 890). Specifically, MTP is a gradient echo pulse sequence (GRE) with dual-TR, dual-flip angle (FA), and multi-echoes. For convenience, we name the in-house datasets as In-house I, In-house II, and In-house

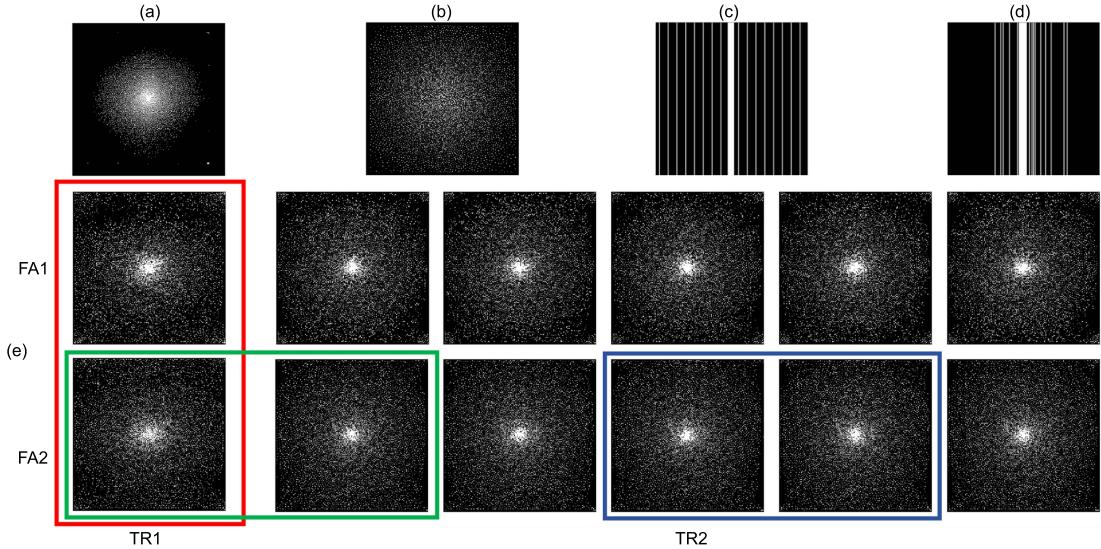
III, respectively. Specifically, In-house I contains 20 subjects. Each scan contains 16 contrast images and 3 parametric maps using a 17-channel head coil. The MRI parameters are as follows: FA1/FA2=4/16, TR1/TR2=35/40 ms, and 8 contrast images per FA with TE=2.99 ms for TR1 and TE=2.99/7.00/9.66/13.67/16.33/20.34/23.00 ms for TR2. The fully-sampled scan time is about 10 minutes. We select 70 slices in each scan with a voxel size of  $1 \times 1 \times 2 \text{ mm}^3$  and slice matrix size= $200 \times 224 \times 70$ .

In-house II includes 18 subjects. Each contains 12 contrasts and 3 parametric maps using a 64-channel head coil. The MRI scanning parameters are as follows: FA1/FA2=4/16, TR1/TR2=37.2/40 ms, and 6 contrast images per FA with TE=3.86 ms for TR1 and TE=3.86/8.95/14.04/19.13/24.22 ms for TR2. The fully-sampled scan time is about 15 minutes. We select 64 slices from each scan with a voxel size of  $0.823 \times 0.823 \times 2 \text{ mm}^3$  and slice matrix size= $231 \times 272 \times 64$ . We perform 2D image reconstruction for each voxel position along the readout (RO) direction, i.e., the two phase encoding (PE) directions are under-sampled. The central 60 slices along RO direction are extracted from each subject. 720/240/240 (In-house I) and 720/180/180 (In-house II) multi-echo slices are used for training, validation, and testing, respectively. For In-house III, each subject is scanned twice on the same machine on two separate days (with 7 days apart) following parameters: TR1/TR2 = 9.09/38.61 ms; TE = 3.68 ms for the short TR and TE = 3.68/9.69/13.33/19.34/22.98/28.99 for the long TR; Flip angle = 4 and 16, resulting in 12 different TR/TE/flip angle combinations. In-house III is used as unseen data and utilized to evaluate the reliability and generalizability of our model.

**Training Details:** We adopt Adam as the optimizer with  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$  on PyTorch backend. We set the weight  $\rho_1$  as 1000 and  $\rho_2$  as 500. The model is trained for 150 epochs with an initial learning rate of  $10^{-3}$  and the mini-batch size of 2 on two NVIDIA A100 GPUs. We adjust the learning rate by decaying it to 10% when eight consecutive epochs do not show performance improvement. The learnable regularization weights  $\lambda_x, \lambda_d, \lambda_S$  are initialized as 0.3, 1, and 0.1, respectively. The number of iterations in the reconstruction network is set as  $K = 5$  and each iteration shares the same weights. In the following experiments, for the sake of convenience, we denote our proposed sampling, reconstruction, and mapping modules as S, R, and M, respectively. The magnitude images and parameter maps generated by the vendor-provided software on the scanner are used as the GT.

### B. Evaluation on Sampling Module

To evaluate the effectiveness of the proposed sampling scheme (MC-LOUPE), we compare with the state-of-the-art learning-based sampling algorithm (LOUPE), and the commonly used 1D equispaced/random Cartesian and variable density (VD). In Fig. 3, we demonstrate different sampling patterns for acceleration of  $R = 8$ . The first row demonstrates the sampling masks obtained by (a) LOUPE, (b) VD, (c) 1D Equispaced, and (d) 1D Random. Since all the contrast images employ the same sampling mask, we demonstrate only one particular for



**Fig. 3.** Evaluation of sampling patterns on 12 multi-contrast images from In-house II (dual-repetition time (TR), dual flip angle (FA)) at acceleration factor of 8. (a) Optimized sampling mask by LOUPE [40]; (b) Sampling mask of variable density; (c) Sampling mask of 1D equispaced Cartesian; (d) Sampling mask of 1D random Cartesian; (e) Sampling masks by our MC-LOUPE. For (a)-(d), all the 12 contrast images share the same sampling masks. For (e), two rows represent sampling masks for 12 TE across two flip angles (FA). The first column represents the masks in TR1, and the remaining columns represent the ones in TR2.

one typical contrast. It seems that the sampling mask generated by LOUPE is more concentrated in the low-frequency region than VD. Our MC-LOUPE is illustrated in Fig. 3(e). We can see that MC-LOUPE provides different sampling masks for different contrast images. Specifically, according to our scanning protocol described in Section IV-A, the first column (masked by red box) represents the sampling masks for acquisition in TR1 at FA1 and FA2, respectively, while the rest columns represent those acquired in TR2 at FA1 and FA2. When observing the red box, we can see that our MC-LOUPE generates different sampling masks at different FA under the same TR. The green box shows that different sampling patterns are obtained in different TRs under the same FA. In addition, the blue box demonstrates that although using the same FA in the same TR, different TEs can have different masks.

For quantitative assessment, we employ the same reconstruction module R and mapping module M for these investigated sampling schemes. We summarize the results of In-house I for acceleration factors of 4 and 8 in Table I. We can see that our MC-LOUPE, denoted as S, outperforms the other sampling schemes for both the reconstructed multi-contrast images and estimated multi-parametric maps by a large margin in terms of PSNR, SSIM, NRMSE, and HFEN. This is mainly because our MC-LOUPE allows individual sampling across multi-contrast images based on overall optimization, while other methods use the same sampling mask for all the contrast images. The significant improvement by MC-LOUPE indicates the importance of using optimized sampling mask for multi-contrast MRI reconstruction and multi-parametric mapping.

### C. Evaluation on Reconstruction Module

In this experiment, we evaluate the effectiveness of the proposed reconstruction module. Specifically, we replace our

reconstruction module R by other representative models, i.e., DC-UNet [21], MoDL [22], and KIU-Net [23], and fine-tune all the S, M modules to fit the individual R network. Moreover, we compare with different variants of R by removing the temporal attention or spatial attention, denoted as S-SwinT and T-SwinT, respectively. We demonstrate the results in Table I. It is shown that our reconstruction module outperforms the other variants quantitatively by a large margin. This indicates that our proposed reconstruction module R plays a critical role in our SRM-Net. Moreover, it also verifies the superiority of our proposed spatio-temporal attention mechanism (ST-SwinT).

In addition, we also demonstrate the update of the sensitivity maps and reconstructed image through 5 unrolled iterations in Fig. 5(a). We can see that there exist noticeable changes in each iteration. Fig. 5(b) shows that the estimated sensitivity maps are smoother than the ones of E2EVarNet. Due to space limitation, only 18 coil sensitivity maps out of 64 are shown.

### D. Evaluation on Mapping Module

One contribution of our work is the replacement of physical model [27] by a stack of MLP for multi-parametric maps estimation. To evaluate the effectiveness of our mapping model, similar to the sampling and reconstruction modules, we replace our mapping module M by physical model PM. We list the results in Table I. As we can see, MLP obtains noticeable performance gain compared to PM-based mapping in terms of PSNR, SSIM, NRMSE, and HFEN. The superior performance of our mapping module may arise from the fact that physical models assume “perfect” quality of reconstructed images and are vulnerable to undersampling artifacts, while our end-to-end trained mapping network tends to be robust to the reconstructed image quality, and hence provides more plausible mappings.

TABLE I

QUANTITATIVE EVALUATION ON RECONSTRUCTED CONTRAST IMAGES FROM IN-HOUSE I AND PUBLIC KNEE DATASETS, AND ESTIMATED PD, T1, AND T2\* MAPS BY DIFFERENT METHODS IN TERMS OF PSNR, SSIM, NRMSE AND HFEN. HIGH-FREQUENCY ERROR NORM (HFEN) IS USED TO MEASURE THE DIFFERENCE BETWEEN TWO IMAGES, EMPHASIZING THE HIGH-FREQUENCY COMPONENTS. S, R, M REPRESENT THE SAMPLING, RECONSTRUCTION, AND MAPPING MODULE, RESPECTIVELY. “—” DENOTES VALUES ARE NOT AVAILABLE. PM INDICATES PHYSICAL MODELS. BEST AND SECOND BEST ARE IN BOLD AND UNDERLINED, RESPECTIVELY. AF REPRESENTS THE ACCELERATION FACTOR AND RI REPRESENTS THE RECONSTRUCTED IMAGE

AF	Methods			Evaluation metric: PSNR ↑ / SSIM ↑ / NRMSE ↓ / HFEN ↓			
	Sampling	Reconstruction	Mapping	RI	PD Map	T1 Map	T2* Map
<i>R = 4</i> (Brain)	Equispaced	R	M	36.64 / 0.9773 / 0.0143 / 0.9932	27.75 / 0.9349 / 0.0414 / 0.8643	26.15 / 0.8273 / 0.0531 / 0.9412	33.69 / 0.8420 / 0.0210 / 1.1932
	Random	R	M	37.35 / 0.9816 / 0.0133 / 0.9822	26.06 / 0.9238 / 0.0509 / 0.9523	25.95 / 0.8223 / 0.0547 / 0.9442	34.38 / 0.8741 / 0.0194 / 1.1726
	VD	R	M	39.35 / 0.9872 / 0.0111 / 0.8832	28.10 / 0.9550 / 0.0398 / 0.8566	27.11 / 0.8525 / 0.0487 / 0.9292	34.33 / 0.8736 / 0.0193 / 1.1722
	LOUPE [40]	R	M	38.89 / 0.9863 / 0.0113 / 0.8943	27.54 / 0.9596 / 0.0423 / 0.8623	26.66 / 0.8435 / 0.0504 / 0.9327	33.65 / 0.8605 / 0.0211 / 1.1945
	S	DC-UNet [21]	M	38.16 / 0.9710 / 0.0124 / 0.9423	27.33 / 0.9416 / 0.0434 / 0.8666	26.61 / 0.8376 / 0.0497 / 0.9309	34.20 / 0.8691 / 0.0197 / 1.1811
	S	MoDL [22]	M	37.56 / 0.9615 / 0.0133 / 0.9578	28.20 / 0.9532 / 0.0390 / 0.8536	26.62 / 0.8426 / 0.0501 / 0.9329	34.18 / 0.8631 / 0.0198 / 1.1822
	S	KIU-Net [23]	M	38.42 / 0.9815 / 0.0121 / 0.9338	27.49 / 0.9523 / 0.0426 / 0.8632	26.04 / 0.8353 / 0.0538 / 0.9442	34.90 / 0.8760 / 0.0182 / 1.1487
	S	S-SwinT	M	39.82 / 0.9883 / 0.0102 / 0.8221	28.72 / <b>0.9619</b> / 0.0369 / 0.8371	27.66 / 0.8710 / 0.0447 / 0.9131	34.35 / 0.8772 / 0.0194 / 1.1779
	S	T-SwinT	M	40.03 / 0.9846 / <u>0.0100</u> / <u>0.8117</u>	28.40 / 0.9617 / 0.0378 / 0.8428	27.78 / 0.8689 / 0.0421 / 0.9023	34.66 / 0.8793 / 0.0191 / 1.1766
	S	R	PM [27]	<u>40.03</u> / 0.9846 / <u>0.0100</u> / <u>0.8117</u>	27.34 / 0.9483 / 0.0438 / 0.8651	26.44 / 0.8517 / 0.0494 / 0.9319	33.20 / 0.8589 / 0.0213 / 1.1962
	S	R	M	<b>40.28</b> / <b>0.9898</b> / <b>0.0097</b> / <b>0.8022</b>	<b>29.04</b> / <b>0.9619</b> / <b>0.0355</b> / <b>0.8215</b>	<b>27.95</b> / <b>0.8720</b> / <b>0.0416</b> / <b>0.8931</b>	<b>34.91</b> / <b>0.8813</b> / <b>0.0181</b> / <b>1.1457</b>
<i>R = 8</i> (Brain)	Equispaced	R	M	34.46 / 0.9649 / 0.0164 / 0.1032	25.23 / 0.8754 / 0.0554 / 0.9733	24.66 / 0.7781 / 0.0618 / 0.9855	32.63 / 0.8501 / 0.0244 / 1.2474
	Random	R	M	35.98 / 0.9743 / 0.0157 / 1.0127	26.28 / 0.9095 / 0.0488 / 0.8973	25.37 / 0.8047 / 0.0576 / 0.9627	32.56 / 0.8585 / 0.0248 / 1.2508
	VD	R	M	33.96 / 0.9567 / 0.0173 / 1.0628	24.64 / 0.9077 / 0.0603 / 1.0123	25.79 / 0.8159 / 0.0546 / 0.9487	33.34 / 0.8438 / 0.0210 / 1.1935
	LOUPE [40]	R	M	37.15 / 0.9790 / 0.0139 / 0.9747	26.74 / 0.9432 / 0.0462 / 0.9143	26.40 / 0.8322 / 0.0513 / 0.9331	33.29 / 0.8616 / 0.0226 / 1.2242
	S	DC-UNet [21]	M	37.23 / 0.9603 / 0.0138 / 0.9751	27.11 / 0.9516 / 0.0444 / 0.8978	<b>27.02</b> / 0.8451 / 0.0473 / 0.9258	33.16 / 0.8568 / 0.0226 / 1.2247
	S	MoDL [22]	M	36.36 / 0.9631 / 0.0152 / 1.0011	25.35 / 0.9368 / 0.0543 / 0.9538	26.77 / 0.8471 / 0.0478 / 0.9267	32.90 / 0.8513 / 0.0233 / 1.2271
	S	KIU-Net [23]	M	37.78 / 0.9771 / 0.0130 / 0.9447	27.52 / 0.9399 / 0.0423 / 0.8858	26.28 / 0.8349 / 0.0522 / 0.9379	33.35 / 0.8669 / 0.0208 / 1.1898
	S	S-SwinT	M	<b>38.69</b> / <b>0.9857</b> / 0.0116 / 0.9169	28.39 / <b>0.9562</b> / 0.0384 / 0.8568	26.54 / 0.8396 / 0.0504 / 0.9336	33.54 / 0.8665 / 0.0204 / 1.1865
	S	T-SwinT	M	38.63 / 0.9766 / 0.0117 / 0.9179	27.77 / 0.9516 / 0.0411 / 0.8675	26.71 / 0.8436 / 0.0494 / 0.9309	<b>33.58</b> / <b>0.8693</b> / 0.0205 / <b>1.1844</b>
	S	R	PM [27]	—	26.47 / 0.9414 / 0.0488 / 0.9179	25.84 / 0.8398 / 0.0524 / 0.9372	32.88 / 0.8565 / 0.0236 / 1.2299
	S	R	M	<b>39.42</b> / <b>0.9879</b> / <b>0.0107</b> / <b>0.8567</b>	<b>28.45</b> / <b>0.9581</b> / <b>0.0380</b> / <b>0.8487</b>	<b>27.14</b> / <b>0.8504</b> / <b>0.0469</b> / <b>0.9225</b>	<b>33.58</b> / <b>0.8692</b> / <b>0.0203</b> / <b>1.1879</b>
<i>R = 4</i> (Knee)	Equispaced	R	M	31.69 / 0.9488 / 0.0252 / 1.0262	- / - / - / -	- / - / - / -	- / - / - / -
	Random	R	M	32.89 / 0.9582 / 0.0219 / 0.9836	- / - / - / -	- / - / - / -	- / - / - / -
	VD	R	M	<b>33.52</b> / 0.9684 / <b>0.0205</b> / <b>0.9607</b>	- / - / - / -	- / - / - / -	- / - / - / -
	LOUPE [40]	R	M	33.15 / 0.9657 / 0.0210 / 0.9741	- / - / - / -	- / - / - / -	- / - / - / -
	S	R	M	<b>37.21</b> / <b>0.9715</b> / <b>0.0140</b> / <b>0.8432</b>	- / - / - / -	- / - / - / -	- / - / - / -
<i>R = 8</i> (Knee)	Equispaced	R	M	29.97 / 0.9480 / 0.0320 / 1.1266	- / - / - / -	- / - / - / -	- / - / - / -
	Random	R	M	32.58 / 0.9537 / 0.0230 / 0.9936	- / - / - / -	- / - / - / -	- / - / - / -
	VD	R	M	32.46 / 0.9497 / 0.0231 / 0.9979	- / - / - / -	- / - / - / -	- / - / - / -
	LOUPE [40]	R	M	<b>32.88</b> / 0.9598 / 0.0218 / 0.9841	- / - / - / -	- / - / - / -	- / - / - / -
	S	R	M	<b>36.06</b> / <b>0.9695</b> / <b>0.0160</b> / <b>0.8944</b>	- / - / - / -	- / - / - / -	- / - / - / -

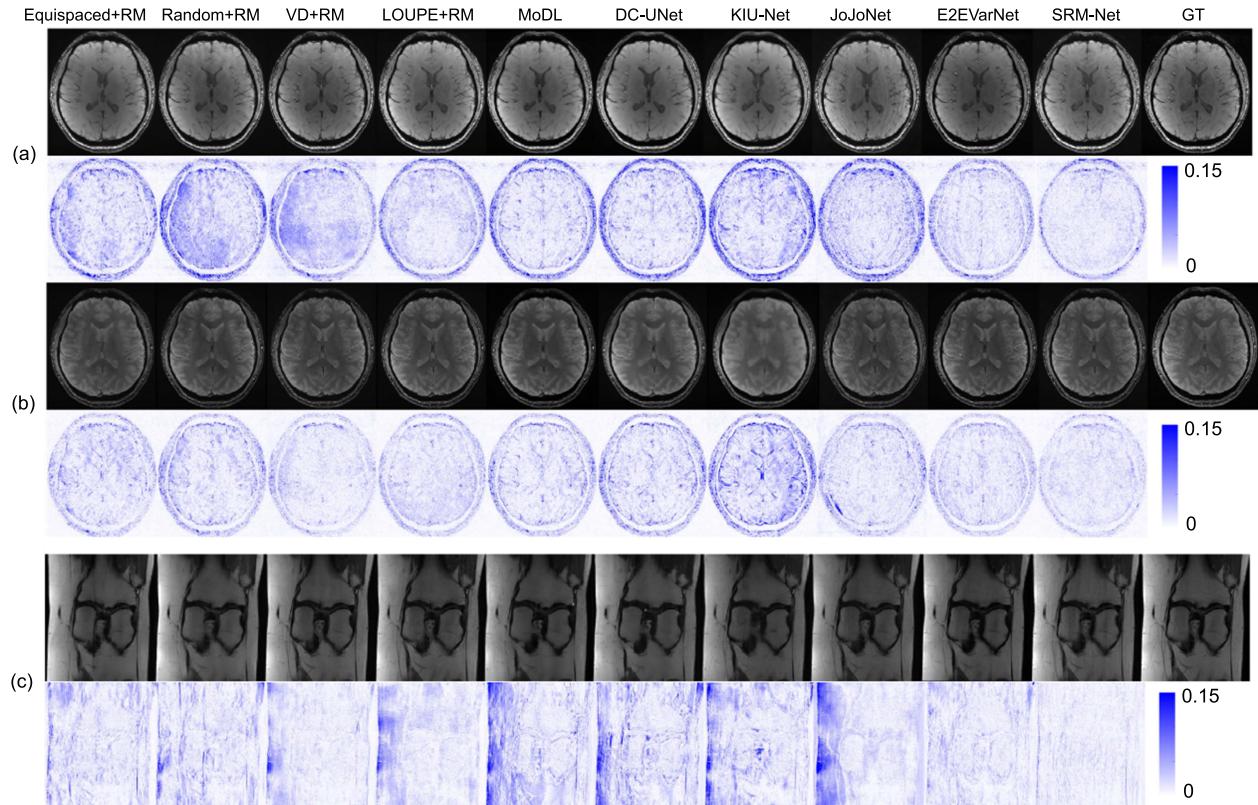
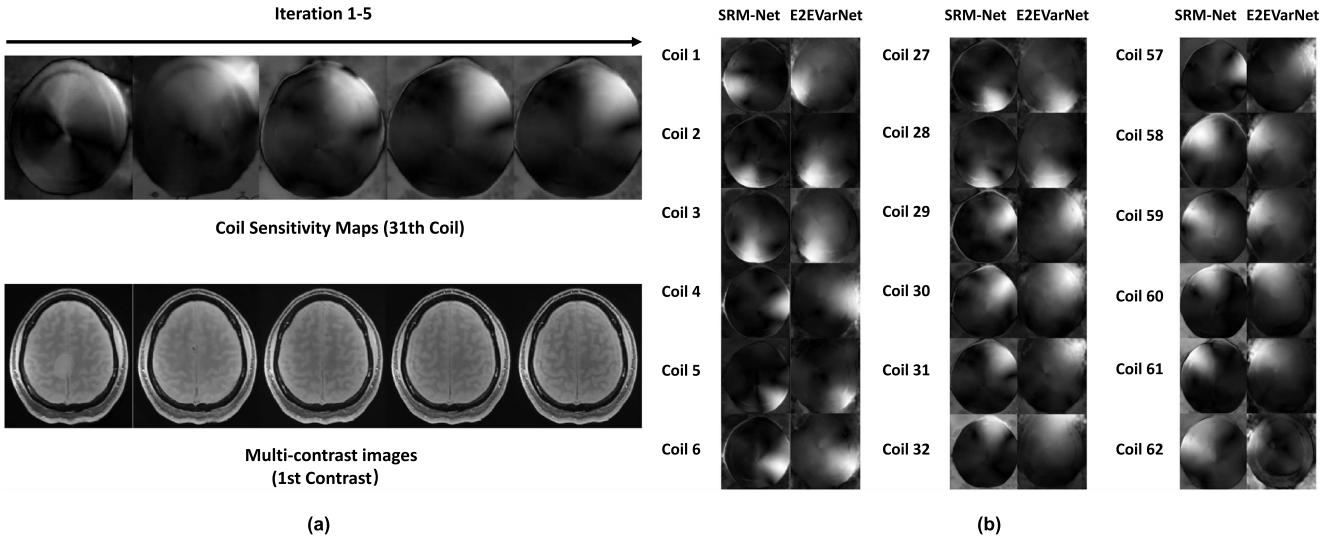


Fig. 4. Reconstructed multi-contrast images by different methods under acceleration factor of *R* = 4 on the brain and knee datasets, where (a) and (b) are from in-house II dataset, and (c) is from the public fastMRI dataset. The blue images represent error maps.



**Fig. 5.** Estimated coil sensitivity maps and reconstructed images. **(a)** Changes of sensitivity maps and reconstructed image through 5 unrolled iterations; **(b)** Estimated sensitivity maps by our SRM-Net and E2EVarNet with the acceleration factor of 8 on the In-house II dataset.

### E. Comparison With Representative Methods

Besides evaluation of the proposed individual modules, herein we also compare SRM-Net with the other representative methods including the conventional method based on total variation (TV) [48] and deep learning-based methods DC-UNet [21], MoDL [22], and KIU-Net [23], MANTIS [31], and a recent method for multi-contrast MRI named JoJoNet [34] quantitatively and qualitatively. We demonstrate visual comparison for reconstructed images and estimated maps in Figs. 4 and 6, respectively. In Fig. 4, from error maps, we can see that our model generates the reconstructed images with least errors. In Fig. 6, we illustrate the estimated PD, T1, and T2<sup>\*</sup> maps by different methods, and show the regions of interest (ROIs) in the close-up views. It can be seen that the proposed SRM-Net, which combines the reconstruction loss and mapping loss via end-to-end learning, obtains more plausibly estimated maps especially in the regions containing details. Moreover, we can observe that the parametric maps generated by MANTIS are relatively smooth, while the estimated maps by our method are sharper and much closer to the GT. Additionally, we summarize quantitative comparison results in Table II. It is shown that the proposed SRM-Net outperforms the other methods dramatically for both multi-contrast image reconstruction and multi-parametric map estimation under different acceleration factors.

Compared to the recent method JoJoNet, our SRM-Net achieves much superior performance quantitatively and qualitatively. The reasons are mainly threefolds. 1) JoJoNet is based on recurrent U-Net and each recurrent cell accounts for an individual contrast image. This recurrent structure hinders joint use of multi-contrast information for the reconstruction of early stage contrast images; 2) JoJoNet uses a stack of fully connected layers as regression function for map estimation, which actually cannot model nonlinear mapping between the contrast images and the corresponding parametric maps, while our SRM-Net uses MLPs to enable complex nonlinear mapping; 3) Our SRM-Net *not only*

reconstructs contrast images based on dual-domain (image and k-space) information and temporal-spatial attention, *but also* simultaneously optimizes coil sensitivity maps, which further enhances reconstruction performance and hence the estimated maps.

### F. Impact of Binarization of Learned Fractional Mask

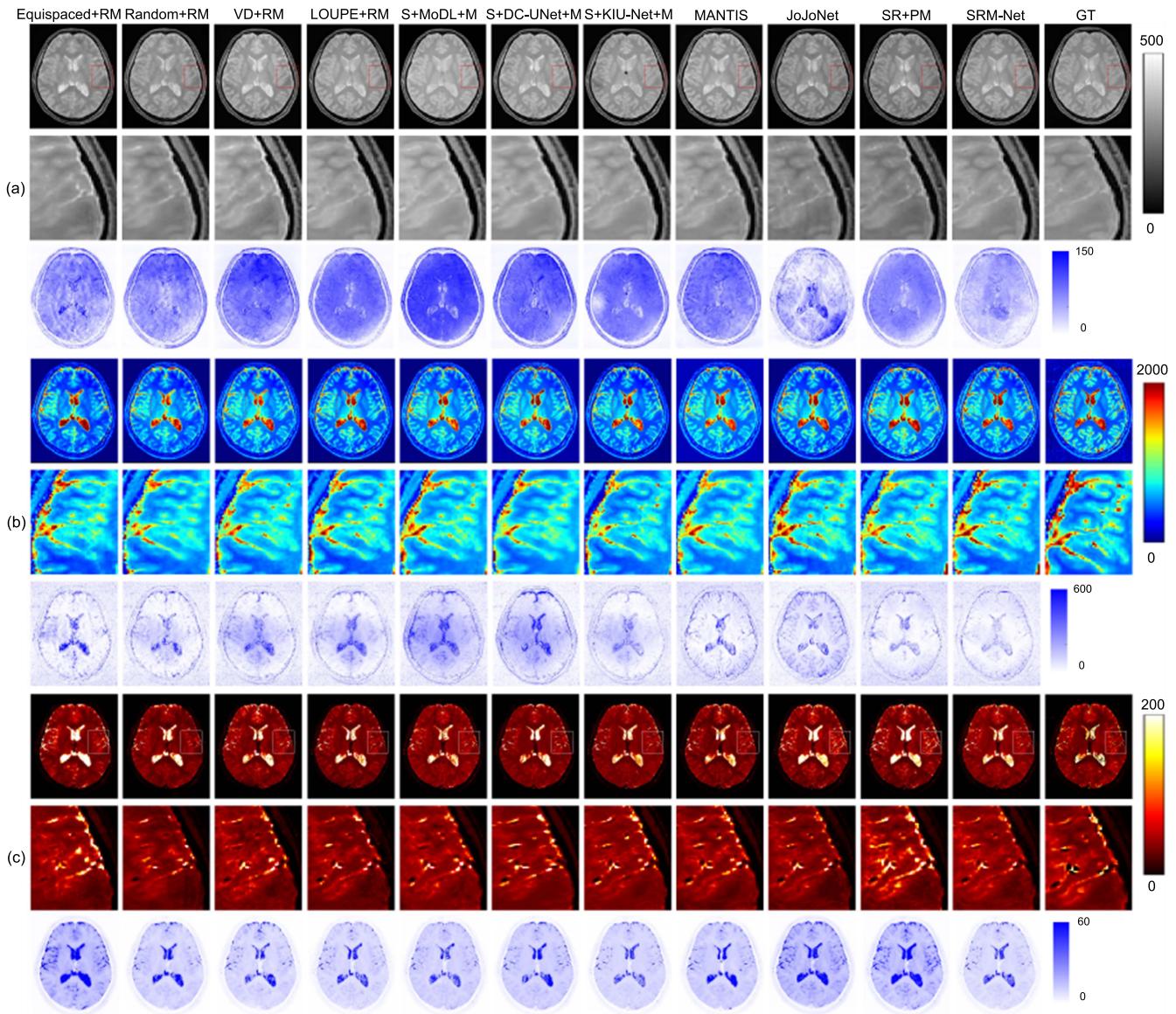
In Fig. 7, we demonstrate the learned fractional sampling mask and the binarized counterpart for inference. The values towards 1 are marked in red and the values towards 0 are marked in blue. To better show the values in the mask, we also plot the histograms. We can see that the values of the optimized fractional mask tend to converge to 0 or 1 due to the hyperparameters set in (9). Consequently, there is negligible difference between using the binarized mask and the fractional one in the reconstruction performance in terms of PSNR, SSIM, NRMSE, and HFEN.

### G. Impact of Learnable Regularization Weights

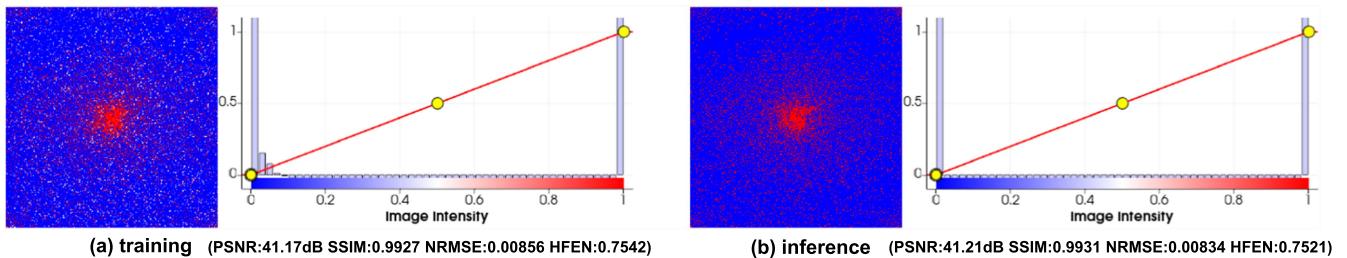
In this section, we evaluate the impact of the regularization weights of  $\lambda_x$ ,  $\lambda_S$ , and  $\lambda_d$  on the reconstructed images and estimated PD map in terms of average PSNR. We perform a grid search of these hyperparameters and choose the best of them based on the validation set. We demonstrate the results in Fig. 8. According to the experimental performance, we empirically select  $\lambda_x = 0.3$ ,  $\lambda_S = 0.1$ , and  $\lambda_d = 1$ .

### H. Analysis of Spatial- and Temporal-Attention Maps

For in-depth analysis of the spatio-temporal attention, we demonstrate the attention weights obtained by *S*-MSA and *T*-MSA in Fig. 9. The left panel exhibits the spatial-attention and the right panel shows the temporal-attention. For spatial attention, there are 16 subfigures. Each subfigure represents the attention for an individual contrast. To obtain each subfigure, we calculate the attention between patches within each Swin window and



**Fig. 6.** Visual comparison among different representative methods for (a) PD map, (b) T1 map, and (c) T2\* map on In-house I under acceleration factor of  $R = 4$ . ROI of different maps are illustrated in close-up views. The third row of (a), (b), and (c) represent the error maps.



**Fig. 7.** Visualization and analysis of the impact of sampling mask binarization on the reconstruction performance in terms of PSNR, SSIM, NRMSE, and HFEN by MC-LOUPE. Left: Optimized fractional sampling mask during training with the histogram (bluish colors represent values towards 0 and reddish ones are towards 1); Right: Used binary sampling mask during inference with the histogram.

TABLE II

QUANTITATIVE EVALUATION FROM IN-HOUSE I AND PUBLIC KNEE DATASETS ON RECONSTRUCTED CONTRAST IMAGES, PD, T1, AND T2\* MAPS UNDER ACCELERATION FACTORS OF  $R = 4$  AND  $R = 8$  IN TERMS OF PSNR, SSIM, NRMSE AND HFEN ON IN-HOUSE I AND PUBLIC KNEE DATASETS. HIGH-FREQUENCY ERROR NORM (HFEN) IS USED TO MEASURE THE DIFFERENCE BETWEEN TWO IMAGES, EMPHASIZING THE HIGH-FREQUENCY COMPONENTS. “–” DENOTES VALUES ARE NOT AVAILABLE. BEST AND SECOND BEST ARE IN BOLD AND UNDERLINED, RESPECTIVELY. AF REPRESENTS THE ACCELERATION FACTOR AND RI REPRESENTS THE REconstructed IMAGE

AF	Methods	PSNR ↑				SSIM ↑				NRMSE ↓				HFEN ↓			
		RI	PD Map	T1 Map	T2* Map	RI	PD Map	T1 Map	T2* Map	RI	PD Map	T1 Map	T2* Map	RI	PD Map	T1 Map	T2* Map
$R = 4$ (Brain)	Zero-Filled	27.94	20.16	19.18	25.44	0.8888	0.7548	0.6856	0.7845	0.0442	0.1033	0.1123	0.0646	1.2859	1.2577	1.4667	2.3468
	TV [48]	31.73	–	–	–	0.9268	–	–	–	0.0271	–	–	–	1.1233	–	–	–
	DC-UNet [21]	36.00	–	–	–	0.9618	–	–	–	0.0154	–	–	–	1.0089	–	–	–
	MoDL [22]	37.18	–	–	–	0.9795	–	–	–	0.0138	–	–	–	0.9749	–	–	–
	KIU-Net [23]	36.44	–	–	–	0.9424	–	–	–	0.0142	–	–	–	0.9876	–	–	–
	E2EVarNet [49]	38.98	–	–	–	0.9844	–	–	–	0.0117	–	–	–	0.8966	–	–	–
	MANTIS [31]	–	26.99	25.80	33.07	–	0.9233	0.7702	0.8610	–	0.0449	0.0533	0.0230	–	0.9076	0.9357	1.2104
	JoJoNet [34]	38.59	27.66	27.18	33.97	0.9825	0.9553	0.8499	0.8679	0.0126	0.0418	0.0464	0.0204	0.9121	0.8687	0.9304	1.1886
$R = 8$ (Brain)	Zero-Filled	22.13	18.44	17.66	22.15	0.7458	0.6437	0.6253	0.7454	0.0863	0.1366	0.1453	0.0813	1.4459	1.7577	2.1267	4.1755
	TV [48]	30.43	–	–	–	0.9135	–	–	–	0.0310	–	–	–	1.1561	–	–	–
	DC-UNet [21]	34.62	–	–	–	0.9603	–	–	–	0.0186	–	–	–	1.0287	–	–	–
	MoDL [22]	34.93	–	–	–	0.9705	–	–	–	0.0179	–	–	–	1.0213	–	–	–
	KIU-Net [23]	34.47	–	–	–	0.9453	–	–	–	0.0189	–	–	–	1.0327	–	–	–
	E2EVarNet [49]	35.98	–	–	–	0.9648	–	–	–	0.0159	–	–	–	1.0136	–	–	–
	MANTIS [31]	–	25.33	25.51	32.66	–	0.8845	0.7576	0.8531	–	0.0542	0.0548	0.0240	–	0.9707	0.9581	1.2478
	JoJoNet [34]	36.74	26.74	26.54	33.06	0.9756	0.9245	0.8307	0.8579	0.0146	0.0463	0.0489	0.0223	0.9827	0.9145	0.9344	1.2213
$R = 4$ (Knee)	SRM-Net	<b>40.28</b>	<b>29.04</b>	<b>27.95</b>	<b>34.91</b>	<b>0.9898</b>	<b>0.9619</b>	<b>0.8720</b>	<b>0.8813</b>	<b>0.0097</b>	<b>0.0355</b>	<b>0.0416</b>	<b>0.0181</b>	<b>0.8022</b>	<b>0.8215</b>	<b>0.8931</b>	<b>1.1457</b>
	TV [48]	29.66	–	–	–	0.9365	–	–	–	0.0282	–	–	–	1.0719	–	–	–
	DC-UNet [21]	32.75	–	–	–	0.9547	–	–	–	0.0223	–	–	–	0.9871	–	–	–
	MoDL [22]	34.02	–	–	–	0.9607	–	–	–	0.0189	–	–	–	0.9466	–	–	–
	KIU-Net [23]	32.14	–	–	–	0.9609	–	–	–	0.0237	–	–	–	1.0047	–	–	–
	E2EVarNet [49]	36.02	–	–	–	0.9697	–	–	–	0.0161	–	–	–	0.8971	–	–	–
	JoJoNet [34]	34.31	–	–	–	0.9699	–	–	–	0.0182	–	–	–	0.9421	–	–	–
	SRM-Net	<b>37.21</b>	–	–	–	<b>0.9715</b>	–	–	–	<b>0.0140</b>	–	–	–	<b>0.8432</b>	–	–	–
$R = 8$ (Knee)	TV [48]	28.82	–	–	–	0.9252	–	–	–	0.0324	–	–	–	1.0947	–	–	–
	DC-UNet [21]	31.76	–	–	–	0.9506	–	–	–	0.0259	–	–	–	1.0245	–	–	–
	MoDL [22]	32.23	–	–	–	0.9484	–	–	–	0.0234	–	–	–	0.9971	–	–	–
	KIU-Net [23]	31.54	–	–	–	0.9493	–	–	–	0.0254	–	–	–	1.0366	–	–	–
	E2EVarNet [49]	35.45	–	–	–	0.9679	–	–	–	0.0171	–	–	–	0.9133	–	–	–
	JoJoNet [34]	31.05	–	–	–	0.9444	–	–	–	0.0268	–	–	–	1.0677	–	–	–
	SRM-Net	<b>36.06</b>	–	–	–	<b>0.9695</b>	–	–	–	<b>0.0160</b>	–	–	–	<b>0.8944</b>	–	–	–

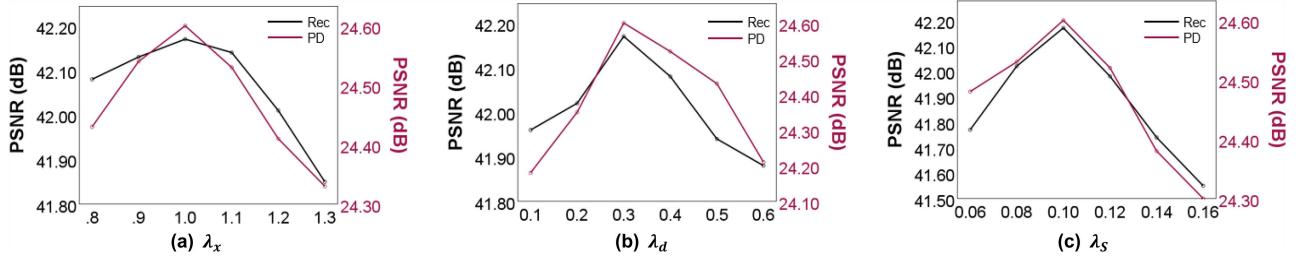


Fig. 8. The impact of learnable regularization weights  $\lambda_x$ ,  $\lambda_d$ ,  $\lambda_s$  on the reconstructed images and estimated PD map on In-house II with AF=4.

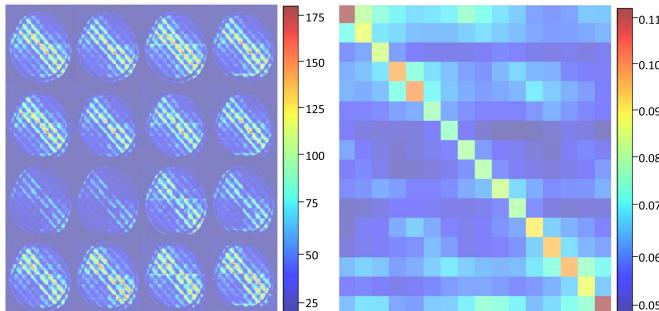


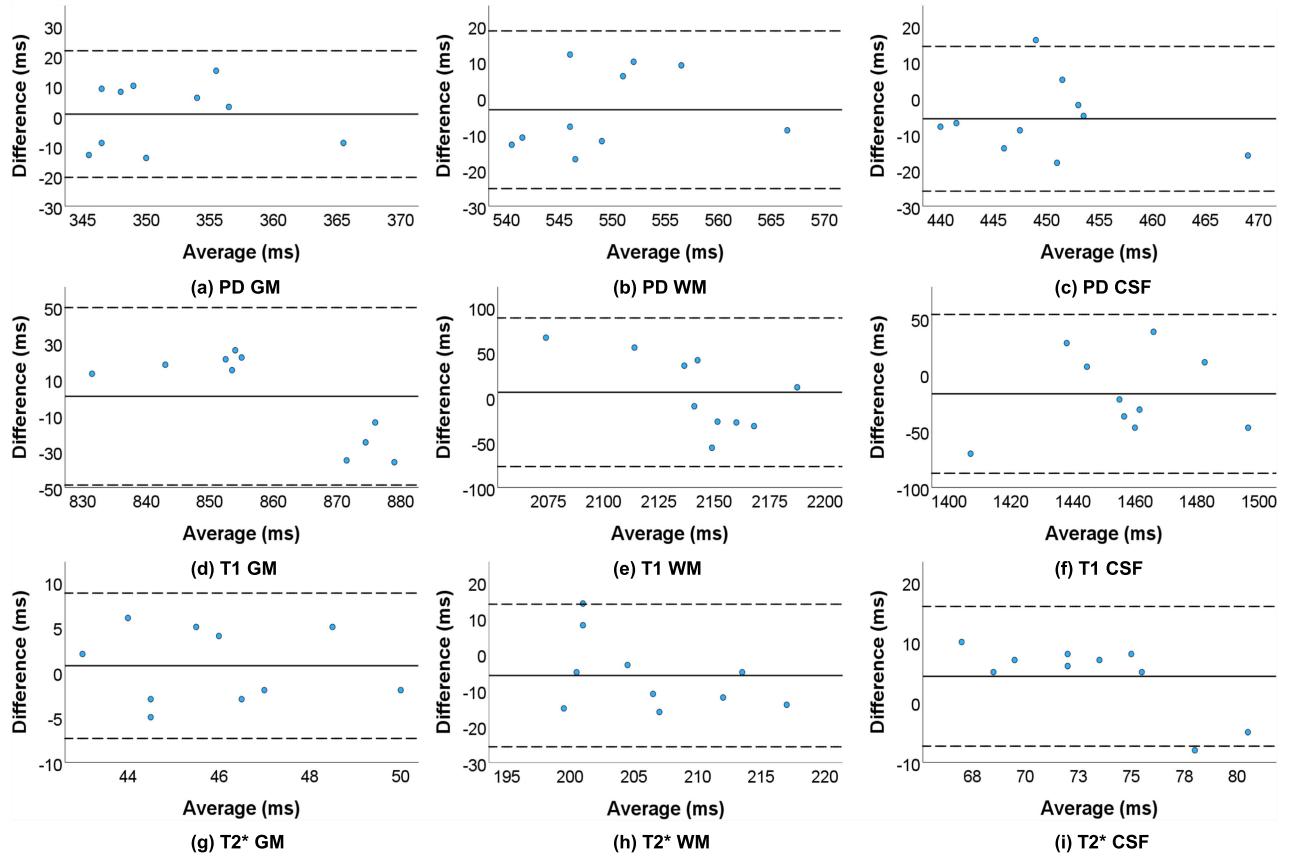
Fig. 9. Visualization of the spatial- and temporal-attention. Left: Spatial attention for 16 contrast images; Right:  $16 \times 16$  temporal attention.

then average the attentions across windows. We can see that the attention weights have higher values in diagonal regions, which indicates that patches within window are more related

to the adjacent ones. With regard to the temporal attention, we calculate the attention across 16 contrasts for each element of the feature map, and then take the average attention across the feature map, thus we obtain a  $16 \times 16$  temporal weight. It is shown that different contrast images have weak correspondence between each other, indicating that complementary patterns are collected among contrasts during sampling.

### I. Reliability Analysis

To evaluate the reliability of our model, we apply the pre-trained model based on In-house II directly on In-house III (10 subjects, with each subject scanned twice). We split these 20 scans into two sets, with each set containing only one scan (not two scans) from each individual subject. For multi-contrast images reconstructed by our pretrained model on in-House II, the first and second sets have an average PSNR of 42.33 dB



**Fig. 10.** Bland-Altman analysis on the mean PD, T1, and T2\* values for gray matter, white matter, and cerebrospinal fluid based on two measurements for each of the 10 subjects on In-house III with the acceleration factor of 4. From top to bottom are PD, T1, and T2\*.

and 42.45 dB, and SSIM of 0.9923 and 0.9926, respectively, under acceleration factor of 4. With acceleration of  $R = 8$ , the average PSNR is 40.12dB and 40.18dB, and SSIM is 0.9879 and 0.9881 for the first and second set, respectively. We measure the intra-class correlation coefficients (ICC) of the two sets of data. The ICC of multi-parametric maps are 0.733, 0.706, and 0.713 for PD, T1, and T2\* maps, respectively, with acceleration factor of 4, while 0.724, 0.704, and 0.692 with acceleration factor of 8. The ICC for reconstructed images is 0.742 for acceleration factor of 4, and 0.737 for acceleration factor of 8. Besides, when comparing the performance of our model on unseen In-house III with In-house II, our model reveals good generalizability on unseen data with different image size.

Furthermore, we perform Bland-Altman analysis on T1, T2\*, and PD for gray matter, white matter, and cerebrospinal fluid on in-house III. The Bland-Altman plots are generated using the SPSS tool and shown in Fig. 10. It can be observed that the coefficient of repeatability (CR) for different tissues in most cases are small, indicating good consistency for T1, T2\*, and PD maps.

### J. Potential Applications and Limitations

The proposed method can accelerate the MULTIPLEX sequence and reduce the scan time of 10 mins by a factor of 4 but preserving the quality of the estimated multi-parametric maps. It

shows the potential to provide multiple quantitative maps within one acquisition. Although our approach has achieved promising performance for multi-parametric mapping on in-house data for brain, there are still several limitations that we will address in our feature work. First, our model is evaluated on the in-house data acquired by u890 MR scanner using MULTIPLEX sequence [27]. Since MULTIPLEX sequence cannot provide T2 map due to sequence design, we have not evaluated the estimation of T2 map in our current experiments. Second, the goal of our model is to estimate multi-parametric maps for accelerated scans. Therefore, we perform retrospective analysis on fully sampled scans using the MULTIPLEX sequence, where the ground-truth quantitative maps are generated by the scanner using fully sampled images. The validation of the ground-truth quantitative maps is already demonstrated in the original paper of MULTIPLEX [27] and is out of the scope of our work. Furthermore, the current model is evaluated only on healthy subjects. Clinical evaluation on patients would be our future work. Lastly, we have only evaluated quantitative maps for brain. Other organs such as heart and abdominal regions may require a different imaging sequence and need to be explored in the future.

### K. Discussion

In our SRM-Net, the proposed sampling method MC-LOUPE provides different sampling masks for different echo times. This

sampling scheme enables complementary acquisition and will not impose additional slew rate constraints as explained below. In conventional multi-echo gradient echo sequence, the gradient pulses for phase and partition encoding are applied prior to the first readout gradient. To accommodate different sampling patterns for different echo times, additional gradient pulses need to be applied to move the sampled k-space coordinates between consecutive readout gradients. The time integral of the gradient pulse is equal to the distance between the sampled k-space coordinates for consecutive TE. By choosing nearby k-space coordinates for consecutive TE in the same TR, the distance will be much smaller than the maximum distance travelled in the conventional gradient echo sequence which is equal to the maximum magnitude of the sampled k-space coordinate.

## V. CONCLUSION

In this work, we present an all-in-one network for accelerated multi-parametric MR imaging, dubbed as SRM-Net, covering sampling mask optimization, multi-contrast image reconstruction, and multi-parametric map estimation based on end-to-end learning paradigm. SRM-Net enables simultaneous multi-parametric map estimation in single accelerated scan. Specifically, a dedicated multi-contrast sampling scheme (MC-LOUPE) is proposed, which supports diverse sampling pattern among different contrast images to boost sampling efficiency. Combining MC-LOUPE with the proposed spatio-temporal attention based reconstruction module, SRM-Net is able to explore the sampling space and also exploit the acquired inter- and intra-contrast images more effectively. Moreover, within the reconstruction module, we alternatively update reconstructed image and coil sensitivity map for more plausible image reconstruction. Lastly, to make the sampling and reconstruction modules better serve parametric map estimation, we utilize a stack of MLP to imitate complex nonlinear mapping between the reconstructed contrast images and the parametric maps, instead of using handcrafted physical models. Extensive experiments on both fastMRI knee dataset and three in-house brain datasets show that our SRM-Net outperforms state-of-the-art methods for both multi-contrast image reconstruction and multi-parametric map estimation at different acceleration factors. Moreover, analysis on reproducibility based on ICC and Bland-Altman reveals reliability of our SRM-Net.

## REFERENCES

- [1] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, no. 5394, pp. 190–191, 1973.
- [2] H.-L. M. Cheng and G. A. Wright, "Rapid high-resolution T<sub>1</sub> mapping by variable flip angles: Accurate and precise measurements in the presence of radiofrequency field inhomogeneity," *Magn. Reson. Med.*, vol. 55, no. 3, pp. 566–574, 2006.
- [3] G. H. Welsch et al., "T2 and T2\* mapping," *Curr. Radiol. Rep.*, vol. 2, pp. 1–9, 2014.
- [4] D. Cordes, "A new algebraic method for quantitative proton density mapping using multi-channel coil data," *Med. Image Anal.*, vol. 40, pp. 154–171, 2017.
- [5] G. Li et al., "Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age," *Cereb. Cortex*, vol. 24, no. 5, pp. 1289–1300, 2014.
- [6] L. Wang et al., "Automatic segmentation of neonatal images using convex optimization and coupled level sets," *NeuroImage*, vol. 58, no. 3, pp. 805–817, 2011.
- [7] F. Shi et al., "Altered structural connectivity in neonates at genetic risk for schizophrenia: A combined study using morphological and white matter networks," *NeuroImage*, vol. 62, no. 3, pp. 1622–1633, 2012.
- [8] M. Liu et al., "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.
- [9] Y. Fan et al., "Multivariate examination of brain abnormality using both structural and functional MRI," *NeuroImage*, vol. 36, no. 4, pp. 1189–1199, 2007.
- [10] J. Zhang et al., "A generalized dual-domain generative framework with hierarchical consistency for medical image reconstruction and synthesis," *Commun. Eng.*, vol. 2, no. 1, 2023, Art. no. 72.
- [11] K. Sun et al., "Achieving multi-modal brain disease diagnosis performance using only single-modal images through generative AI," *Commun. Eng.*, vol. 3, no. 1, 2024, Art. no. 96.
- [12] M. Liu et al., "View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data," *Med. Image Anal.*, vol. 36, pp. 123–134, 2017.
- [13] M. Liu et al., "Inherent structure-based multiview learning with multi-template feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1473–1482, Jul. 2016.
- [14] C.-Y. Lee et al., "Resting-state multi-spectrum functional connectivity networks for identification of MCI patients," *PLoS One*, vol. 7, no. 5, 2012, Art. no. e37828.
- [15] G. Li et al., "Construction of 4D high-definition cortical surface atlases of infants: Methods and applications," *Med. Image Anal.*, vol. 25, no. 1, pp. 22–36, 2015.
- [16] L. Feng, D. Ma, and F. Liu, "Rapid MR relaxometry using deep learning: An overview of current techniques and emerging trends," *NMR Biomed.*, vol. 35, no. 4, 2022, Art. no. e4416.
- [17] E. Z. Chen et al., "Accelerating 3D multiplex MRI reconstruction with deep learning," 2021, *arXiv:2105.08163*.
- [18] C. Zhang et al., "A unified model for reconstruction and R2\* mapping of accelerated 7T data using the quantitative recurrent inference machine," *NeuroImage*, vol. 264, 2022, Art. no. 119680.
- [19] M. V. W. Zibetti et al., "Comparing learned variational networks and compressed sensing for T1ρ mapping of knee cartilage," in *Proc. ISMRM Workshop Mach. Learn.*, 2018.
- [20] A. Chaudhari et al., "Deep learning super-resolution enables rapid simultaneous morphological and quantitative magnetic resonance imaging," in *Machine Learning for Medical Image Reconstruction*. Berlin, Germany: Springer, 2018, pp. 3–11.
- [21] L. Sun et al., "Joint CS-MRI reconstruction and segmentation with a unified deep network," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer, 2019, pp. 492–504.
- [22] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.
- [23] R. Souza et al., "Dual-domain cascade of U-Nets for multi-channel magnetic resonance image reconstruction," *Magn. Reson. Imag.*, vol. 71, pp. 140–153, 2020.
- [24] K. Hammernik et al., "Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 98–114, Jan. 2023.
- [25] S. C. Deoni, "High-resolution T1 mapping of the brain at 3T with driven equilibrium single pulse observation of T1 with high-speed incorporation of RF field inhomogeneities (despot1-HiFi)," *J. Mag. Reson. Imag.*, vol. 26, no. 4, pp. 1106–1111, 2007.
- [26] Y. Ye et al., "A multi-dimensional integration (MDI) strategy for MR T2\* mapping," *NMR Biomed.*, vol. 34, no. 7, 2021, Art. no. e4529.
- [27] Y. Ye et al., "Multi-parametric mr imaging with flexible design (multiplex)," *Magn. Reson. Med.*, vol. 87, no. 2, pp. 658–673, 2022.
- [28] K. T. Block, M. Uecker, and J. Frahm, "Model-based iterative reconstruction for radial fast spin-echo MRI," *IEEE Trans. Med. Imag.*, vol. 28, no. 11, pp. 1759–1769, Nov. 2009.
- [29] C. Cai et al., "Single-shot T2 mapping using overlapping-echo detachment planar imaging and a deep convolutional neural network," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2202–2214, 2018.
- [30] Y. Jun et al., "Deep model-based magnetic resonance parameter mapping network (dopamine) for fast T1 mapping using variable flip angle method," *Med. Image Anal.*, vol. 70, 2021, Art. no. 102017.

- [31] F. Liu, L. Feng, and R. Kijowski, "Mantis: Model-augmented neural network with incoherent k-space sampling for efficient MR parameter mapping," *Magn. Reson. Med.*, vol. 82, no. 1, pp. 174–188, 2019.
- [32] F. Liu et al., "Magnetic resonance parameter mapping using model-guided self-supervised deep learning," *Magn. Reson. Med.*, vol. 85, no. 6, pp. 3211–3226, 2021.
- [33] G. Wang et al., "Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3089–3099, Oct. 2020.
- [34] L. Zhao et al., "JoJoNet: Joint-contrast and joint-sampling-and-reconstruction network for multi-contrast MRI," 2022, *arXiv:2210.12548*.
- [35] D. Ma et al., "Magnetic resonance fingerprinting," *Nature*, vol. 495, no. 7440, pp. 187–192, 2013.
- [36] Y. Gu et al., "Deep learning-assisted preclinical MR fingerprinting for sub-millimeter T1 and T2 mapping of entire macaque brain," *Magn. Reson. Med.*, vol. 91, no. 3, pp. 1149–1164, 2024.
- [37] L. Pineda et al., "Active MR k-space sampling with reinforcement learning," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2020, pp. 23–33.
- [38] T. Bakker, H. van Hoof, and M. Welling, "Experimental design for MRI by greedy policy search," in *Proc. Int. Conf. Adv. Neural Inf. Process.*, 2020, vol. 33, pp. 18954–18966.
- [39] T. Weiss et al., "PILOT: Physics-informed learned optimized trajectories for accelerated MRI," *Mach. Learn. Biomed. Imag.*, vol. 1, pp. 1–23, Apr. 2021.
- [40] C. D. Bahadir et al., "Deep-learning-based optimization of the undersampling pattern in MRI," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1139–1152, 2020.
- [41] J. Zhang et al., "Extending loupe for k-space under-sampling pattern optimization in multi-coil MRI," in *Machine Learning for Medical Image Reconstruction*. Berlin, Germany: Springer, 2020, pp. 91–101.
- [42] H. K. Aggarwal and M. Jacob, "J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1151–1162, Oct. 2020.
- [43] T. Shor et al., "Multi pilot: Learned feasible multiple acquisition trajectories for dynamic MRI," 2023, *arXiv:2303.07150*.
- [44] K. Sun, Q. Wang, and D. Shen, "Joint cross-attention network with deep modality prior for fast MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 43, no. 1, pp. 558–569, Jan. 2024.
- [45] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.
- [46] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
- [47] J. Zbontar et al., "fastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.
- [48] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [49] A. Sriram et al., "End-to-end variational networks for accelerated MRI reconstruction," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2020, pp. 64–73.