

# PUERT: Probabilistic Under-sampling and Explicable Reconstruction Network for CS-MRI

Jingfen Xie, Jian Zhang, Yongbing Zhang, Xiangyang Ji

**Abstract**—Compressed Sensing MRI (CS-MRI) aims at reconstructing de-aliased images from sub-Nyquist sampling  $k$ -space data to accelerate MR Imaging, thus presenting two basic issues, *i.e.*, where to sample and how to reconstruct. To deal with both problems simultaneously, we propose a novel end-to-end Probabilistic Under-sampling and Explicable Reconstruction neTwork, dubbed PUERT, to jointly optimize the sampling pattern and the reconstruction network. Instead of learning a deterministic mask, the proposed sampling subnet explores an optimal probabilistic sub-sampling pattern, which describes independent Bernoulli random variables at each possible sampling point, thus retaining robustness and stochastics for a more reliable CS reconstruction. A dynamic gradient estimation strategy is further introduced to gradually approximate the binarization function in backward propagation, which efficiently preserves the gradient information and further improves the reconstruction quality. Moreover, in our reconstruction subnet, we adopt a model-based network design scheme with high efficiency and interpretability, which is shown to assist in further exploitation for the sampling subnet. Extensive experiments on two widely used MRI datasets demonstrate that our proposed PUERT not only achieves state-of-the-art results in terms of both quantitative metrics and visual quality but also yields a sub-sampling pattern and a reconstruction model that are both customized to training data.<sup>1</sup>

**Index Terms**—Compressed sensing MRI, deep unfolding network, joint learning, probabilistic under-sampling

## I. INTRODUCTION

MAGNETIC Resonance Imaging (MRI) is a widely-used biomedical imaging technology that enjoys superior benefits of good soft-tissue contrast, non-ionizing radiation, and the availability of multiple tissue contrasts. A main challenge lies in how to reduce the long scan time so as to improve accessibility and decrease costs. One solution is to accelerate MRI via Compressed Sensing (CS) [1], [2]. In Compressed Sensing MRI (CS-MRI), sub-Nyquist sampling [1] is utilized to get under-sampled  $k$ -space data, *i.e.*, part of the Fourier transform of the image, following a predetermined sampling mask. Then, given a sub-sampled set of measurements, a CS-MRI reconstruction algorithm with high quality and a fast speed is expected to reconstruct the full-resolution MRI

without aliased artifacts. Thus, there exist two important issues within CS-MRI, *i.e.*, where to sample and how to reconstruct.

For sampling schemes, some popular patterns in CS-MRI include Cartesian [3] with skipped lines, Random Uniform [2] and Variable Density (VD) [4], thanks to their simplicity and good performance when coupled with reconstruction methods. These sampling schemes mostly follow variable-density probability density functions, based on the empirical observation that lower frequencies should be sampled more densely than high frequencies to promote image recovery. Another common Poisson disc sampling strategy [5] adopts sampling locations separated by a minimum distance [6] in addition to following a density, thus further exploiting redundancies in parallel MRI. However, these masks are designed heuristically and independently, lacking the ability to adapt to specific data and recovery methods, leaving much room for further improvement.

For MRI reconstruction, there exist various works improving it from many aspects. On one hand, traditional model-based MRI restoration has been widely studied [1], [3], [7]–[14]. These methods usually adopt iterative optimization, resulting in over-smoothed recovery and long consuming time. On the other hand, data-driven methods have been introduced as a promising alternative [15]–[23] with high quality and a fast speed. In [16], a widely-used convolutional neural network called U-Net [17] is used to reconstruct MR images. In [18], a cascaded CNN with a data consistency layer is presented to further ensure measurement fidelity. Most recently, some Deep Unfolding Networks (DUNs) [24]–[30] are developed to integrate the interpretability of traditional model-based approaches and the efficiency of data-driven methods, thus yielding a better recovery performance. Note that DUNs are not limited to CS-MRI, but also widely studied in CS reconstruction [31]–[35]. As an instance, a state-of-the-art method ISTA-Net [24] unfolds the traditional iterative shrinkage-thresholding algorithm (ISTA) update steps to a network with a fixed number of stages, each corresponding to one iteration in ISTA. Besides, recent works have also focused on exploring different training objectives such as adversarial loss [36]–[38] and cyclic loss [39], [40] to enhance perceptual recovery quality.

The above-mentioned works regard sub-sampling and reconstruction as two independent problems, in consideration of generality and simplicity. However, they ignore the fact that, in general, the optimal under-sampling pattern depends on the specific MRI anatomy and reconstruction method, thus requiring more customization.

In this paper, we jointly deal with the above two issues, *i.e.*, under-sampling and reconstruction, and propose a novel end-to-end Probabilistic Under-sampling and Explicable Re-

J. Xie and J. Zhang are with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China. J. Zhang is also with the Peng Cheng Laboratory, Shenzhen 518055 China. (E-mail: xiejf@stu.pku.edu.cn; zhangjian.sz@pku.edu.cn)

Y. Zhang is with the the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China. (E-mail: ybzhang08@hit.edu.cn)

X. Ji is with the Department of Automation, Tsinghua University, Beijing 100084, China. (E-mail: xyji@tsinghua.edu.cn)

<sup>1</sup>For reproducible research, the source codes and training models of our PUERT are available at <https://github.com/chuan1093/PUERT>.

construction neTwork, dubbed PUERT and pronounced like Pu'er Tea, to achieve an efficient combination of sub-sampling learning and reconstruction network training. Specifically, considering the stochastic strategies of compressed sensing, we develop a sampling subnet that explores an optimal Bernoulli probabilistic sampling pattern instead of a deterministic mask, thus achieving robustness and stochastics for a more reliable reconstruction. Then, a dynamic gradient estimation strategy is proposed to gradually approximate the binarization function in backward propagation, which efficiently retains the gradient information and further improves the recovery quality. Moreover, we adopt an efficient reconstruction subnet unfolded by the traditional ISTA algorithm, and also emphasize the superiority of DUN in facilitating exploration and efficient training of the sampling subnet.

Overall, the main contributions of this paper are four-fold:

- A novel end-to-end Probabilistic Under-sampling and Explicable Reconstruction neTwork, dubbed PUERT, for CS-MRI is proposed, which implements an efficient combination of sampling mask learning and reconstruction network training.
- The sampling subnet explores an optimal probabilistic sampling pattern to retain robustness and stochastics, and introduces a dynamic gradient estimation strategy to enable efficient training and promote network performance.
- The end-to-end reconstruction subnet adopts an explicable ISTA-unfolding network with high reconstruction quality and fast speed, which is also shown to facilitate further exploration for the sampling subnet.
- Experiments show that our proposed PUERT not only performs favorably against state-of-the-arts in terms of both quantitative metrics and visual quality but also yields a sub-sampling pattern and a reconstruction model that are both customized to training data.

## II. RELATED WORK

Existing sampling pattern optimization schemes consist of two classes: 1) active algorithms [41], [42] using a sampling subnet to predict the *next*  $k$ -space sample location based on current recovery from reconstruction subnet, and 2) non-active methods directly predicting the *whole* mask in every epoch. Here, we focus on the latter kind and further divide them into nested optimization methods and end-to-end learning methods.

### A. Nested Optimization Methods

Nested optimization methods formulate the sampling matrix learning problem as two nested optimization issues. The outer issue is to find an optimal sub-sampling mask or trajectory that behaves best under the current reconstruction algorithm, which is then taken as input for the inner problem to further optimize the reconstruction method. Several heuristic or greedy algorithms [43]–[51] have been proposed to tackle the above nested pair of problems, and has been demonstrated to achieve better reconstructions than conventional methods [50], [52] in an empirical study [53]. Most recently, in [54], continuous optimization methods are applied to a formulated bilevel optimization problem, with the aim of learning sparse sampling

patterns for MRI within a supervised learning framework for a given variational reconstruction method. However, the above algorithms suffer from high computational complexity when dealing with large-scale training data, and often lack the flexibility to explore different types of sampling patterns.

### B. End-to-end Learning Methods

Fueled by the rise of deep learning, some end-to-end data-driven sampling optimization methods [55]–[59] are proposed to achieve computational efficiency and improve reconstruction. PILOT [57] introduces hardware constraints into the learning pipeline and explores an optimal physically viable  $k$ -space trajectory that is enabled by interpolation on the discrete  $k$ -space grid. However, the learned trajectory significantly depends on the initialization and lacks exploration. J-MoDL [59] proposes a multichannel sampling model consisting of a non-uniform Fourier operator with continuously defined sampling locations to promote the optimization process without approximations. However, it constrains the sampling mask as the tensor product of two 1D sampling patterns and restricts its performance. Note that, due to learning a deterministic mask, the above methods inhibit flexibility and randomness, and are against the stochastic strategy of CS theory [60], [61].

LOUPE [56] assumes that each binary sampling location is an independent Bernoulli random variable and learns a probabilistic sampling pattern instead of a deterministic mask. In this aspect, both LOUPE and our PUERT have similar inspirations, but in fact, our PUERT enjoys three superior advantages compared to LOUPE. First, due to the relaxation of the binarization operation, LOUPE not only pays a performance penalty, but also needs to retrain the reconstruction model with the learned binary mask, whereas our PUERT directly uses the binarization function in the forward pass and introduces an efficient gradient estimation strategy for the backward pass, which overcomes the above two shortcomings and is shown to promote the recovery quality. Second, LOUPE mainly adopts a U-Net network and does not point out the superiority of DUNs against non-DUNs on further exploitation of the sampling subnet. However, we emphasize that, thanks to DUN's intrinsic structural feature of fully utilizing knowledge on the mask in each stage, it's of great significance to use DUNs to facilitate exploration and training of the sampling subnet. Third, when testing, LOUPE directly samples from the learned probabilistic pattern and gets a mask with an inexact sampling ratio, whereas in PUERT, the vanilla binarization function is replaced by a greedy version, thus achieving precise control of sampling ratio and promoting fair comparisons.

## III. PROPOSED METHOD

In this section, we elaborate on the design of our proposed PUERT. We first formulate the problem in Section III-A. Then, as shown in Fig. 1, our PUERT consists of a sampling subnet and a reconstruction subnet, which are described in Section III-B and Section III-C respectively. Finally, we describe the details about parameters and loss function in Section III-D.

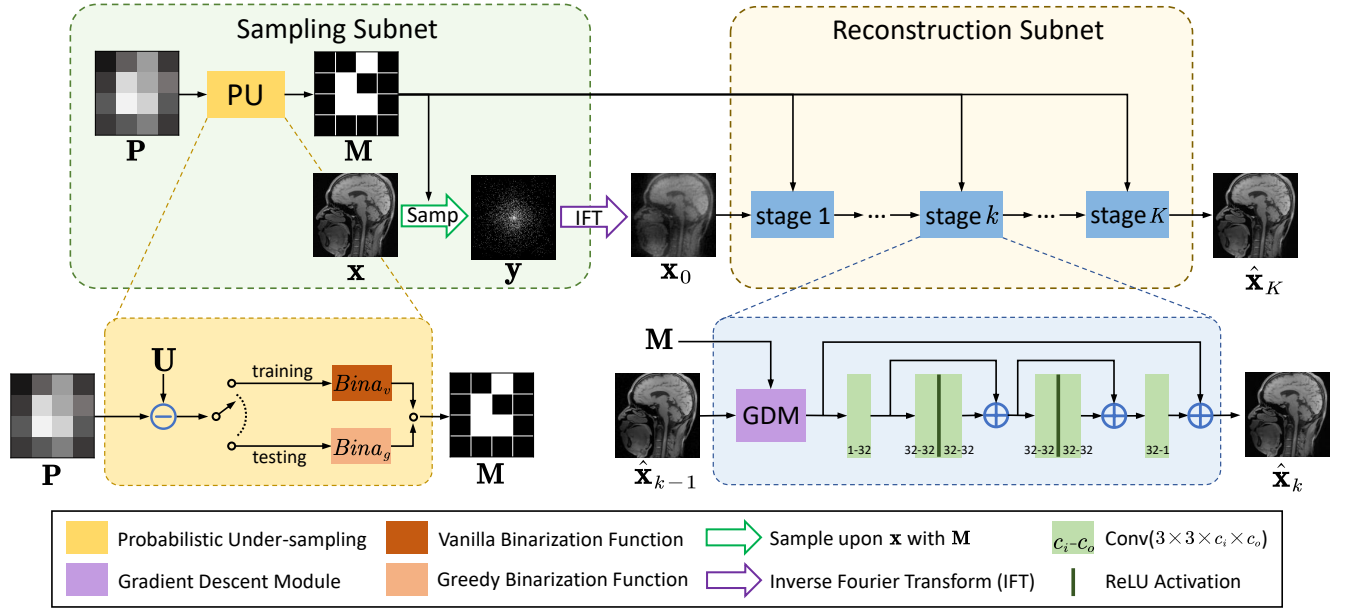


Fig. 1. Our proposed PUERT consists of two subnets: a Sampling Subnet and a Reconstruction Subnet. In the Sampling Subnet, we learn a probabilistic sampling pattern  $\mathbf{P}$  to generate the binary mask  $\mathbf{M}$ . Each value in  $\mathbf{P}$  stands for the probability of the corresponding location in  $\mathbf{M}$  of taking value 1. We generate  $\mathbf{M}$  via adopting a uniform-distributed  $\mathbf{U}$  and a binarization function to implement Bernoulli sampling. Note that, during testing, we use a greedy binarization function to implement precise control of the sampling ratio. In the Reconstruction Subnet, we cast the traditional ISTA method into a deep unfolding network containing  $K$  stacked stages. Each stage ( $k$ -th) inputs  $\hat{\mathbf{x}}_{k-1}$ , outputs  $\hat{\mathbf{x}}_k$ , and is composed of a Gradient Descent Module (GDM) and a Proximal Mapping Module (PMM). We implement PMM with 2 residual blocks, two convolutions and a long residual connection with input.

#### A. Problem Formulation

In CS-MRI, the acquisition process can be formulated as:

$$\mathbf{y} = \mathbf{M} \odot \mathbf{F}\mathbf{x} + \varepsilon, \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^{N_x \times N_y}$  and  $\mathbf{y} \in \mathbb{C}^{N_x \times N_y}$  denote the original fully-sampled image to be reconstructed and the under-sampled  $k$ -space observation, respectively.  $\varepsilon \in \mathbb{C}^{N_x \times N_y}$  is the noise generated during acquisition.  $\mathbf{F}$  and  $\odot$  denote the Fourier Transform (FT) and the element-wise multiplication operation, respectively.  $\mathbf{M} \in \mathbb{R}^{N_x \times N_y}$  is a *binary* sampling mask matrix composed of 1 and 0, which separately represent whether to sample or not at the corresponding  $k$ -space position. Note that the CS sampling ratio, denoted by  $\alpha$ , is defined as the ratio of value 1 in  $\mathbf{M}$ , i.e.,  $\alpha = \frac{\|\mathbf{M}\|_0}{N_x \times N_y}$ .

To tackle the ill-posed inversion of reconstructing  $\mathbf{x}$  from sub-sampled data  $\mathbf{y}$  under a given mask  $\mathbf{M}$ , traditional CS-MRI reconstruction methods usually reconstruct the original image  $\mathbf{x}$  by solving the following optimization problem:

$$\arg \min_{\mathbf{x}} \|\mathbf{M} \odot \mathbf{F}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \psi(\mathbf{x}), \quad (2)$$

where  $\psi(\mathbf{x})$  denotes an image prior-regularized term and  $\lambda$  is a weight to balance between fidelity and regularization. Despite the broad task defined above, note that we only consider partial Fourier reconstruction here, excluding parallel imaging [6], [62], [63] and structured low-rank matrix methods [64]–[66].

In pursuit of superior recovery performance, we propose to concurrently explore an optimal sampling pattern while learning reconstruction network parameters. To this end, we take  $\mathbf{M}$  as learnable parameters and construct a novel end-to-end Probabilistic Under-sampling and Explicable Reconstruction neTwork (PUERT), which is composed of a sampling subnet and a reconstruction subnet, as illustrated in Fig. 1.

#### B. Sampling Subnet

In the Sampling Subnet (**SampNet**), we first explore a Probabilistic Under-sampling (PU) scheme to preserve robustness and stochastics, and then adopt a Dynamic Gradient Estimation (DGE) strategy to enable efficient training and promote network performance.

**Probabilistic Under-sampling (PU):** Building on stochastic strategies and the robustness requirement of compressed sensing, stochastic sub-sampling patterns are able to create noise-like artifacts that are relatively easier to remove [2]. Therefore, we propose to directly optimize a probabilistic sampling pattern rather than a fixed binary mask. Furthermore, considering that adopting a classic sampling pattern (e.g., Gaussian [4] or Uniform [2] distribution) exhibits great limitation and lacks adaptability, we propose to learn independent Bernoulli random variables for each  $k$ -space point, thus learning a sampling pattern highly customized to specific training data and recovery methods.

Concretely, a learnable probabilistic sampling pattern  $\mathbf{P} \in \mathbb{R}^{N_x \times N_y}$  is introduced and each value  $\mathbf{P}_{i,j} \in [0, 1]$  stands for the probability of  $\mathbf{M}_{i,j}$  taking value 1, i.e.,  $\mathbf{M}_{i,j} \sim \mathcal{B}(\mathbf{P}_{i,j})$ , where  $\mathcal{B}(z)$  denotes a Bernoulli random variable with parameter  $z$ . In order to draw a sample from  $\mathbf{P}$  and generate a binary mask  $\mathbf{M}$ , we introduce a matrix  $\mathbf{U} \in \mathbb{R}^{N_x \times N_y}$  with each element independently drawn from a Uniform distribution on  $[0, 1]$ , i.e.,  $\mathbf{U}_{i,j} \sim \mathcal{U}(0, 1)$ , and then binarize the difference between  $\mathbf{P}_{i,j}$  and  $\mathbf{U}_{i,j}$  by a function  $Bina$ . So the generation of  $\mathbf{M}$  is formulated as:

$$\mathbf{M}_{i,j} = Bina(\mathbf{P}_{i,j} - \mathbf{U}_{i,j}), \quad (3)$$

where  $Bina$  is implemented by a vanilla binarization function

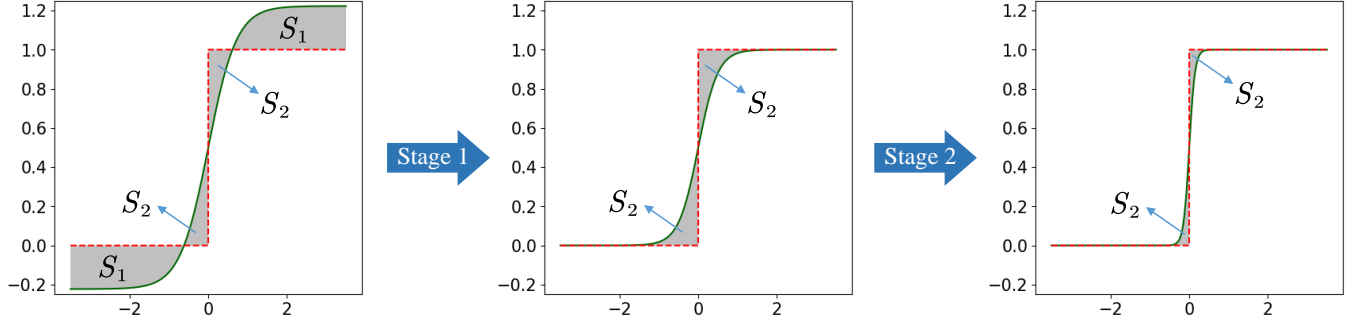


Fig. 2. Our proposed Dynamic Gradient Estimation consists of two stages. The error of gradient estimation is denoted by the area of gray shades. During Stage 1, in order to retain fast updating ability, we keep the derivative value close to one and reduce  $S_1$  by decreasing the output range. While during Stage 2, more attention is paid to achieving accurate gradients and we increase the slope of the function to make  $S_2$  shrink.

$Bina_v$  in the **training** process:

$$Bina_v(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad (4)$$

Note that  $\mathbf{U}$  is not fixed, but randomly generated every time  $\mathbf{P}$  generates  $\mathbf{M}$ . In this way, the sampling subnet is able to learn a probabilistic sampling pattern  $\mathbf{P}$  that expresses belief or importance across all  $k$ -space locations, thus gaining robustness and stochastics for a more reliable CS-MRI reconstruction. Note that the use of  $\mathbf{U}$  is similar to the re-parameterization trick used in VAE [67], which generates non-uniform random numbers by transforming some base distribution, so as to recast the statistical expression and tackle the stochastic node.

Another critical issue is how to control the sparsity of the generated mask  $\mathbf{M}$  at a given target sampling ratio  $\alpha$ . To this end, we first introduce a rescale operator to adjust the average value of  $\mathbf{P}$ , and then adopt a greedy binarization operator specifically for testing to implement a totally accurate control.

To be concrete, before sampling out  $\mathbf{M}$  as Eq. (3), the probabilistic sampling pattern  $\mathbf{P}$  is rescaled as follows:

$$Res(\mathbf{P}_{i,j}) = \begin{cases} \frac{\alpha}{\bar{p}} \mathbf{P}_{i,j}, & \text{if } \bar{p} \geq \alpha \\ 1 - \frac{1-\alpha}{1-\bar{p}} (1 - \mathbf{P}_{i,j}), & \text{otherwise.} \end{cases} \quad (5)$$

Here,  $\bar{p}$  stands for the average value of  $\mathbf{P}$ , i.e.,  $\bar{p} = \frac{\sum_{i,j} \mathbf{P}_{i,j}}{N_x \times N_y}$ . It can be proven that Eq. (5) yields  $\mathbf{P}$  with its average value rescaled to the given CS ratio  $\alpha$ . With this rule, the ratio of the generated binary mask  $\mathbf{M}$  would be *close* to the target  $\alpha$ .

During the testing process, in order to achieve an *accurate* control of the ratio of  $\mathbf{M}$ , a greedy binarization function  $Bina_g$  is further utilized to replace the vanilla one  $Bina_v$  in Eq. (3). Specifically, the greedy binarization function  $Bina_g$  used during **testing** is defined as follows:

$$Bina_g(x; \Omega, \alpha) = \begin{cases} 1, & x \geq b(\Omega, \alpha) \\ 0, & x < b(\Omega, \alpha), \end{cases} \quad (6)$$

where  $\Omega$  represents the set  $\{\mathbf{P}_{i,j} - \mathbf{U}_{i,j}\}$  and  $b(\Omega, \alpha)$  stands for the  $(\alpha \times |\Omega|)$ -th largest number in the set  $\Omega$ . In this way, our method precisely controls the ratio of  $\mathbf{M}$  and implements a fair comparison with other methods. Note that since the learned probabilistic pattern  $\mathbf{P}$  represents importance across all  $k$ -space locations, our value-maximization design as Eq. (6) for the testing process is actually cooperative with the training

period and exhibits nice rationality. Besides, thanks to the adoption of  $\mathbf{U}$  in Eq. (6), our method outputs a sampling mask with randomness rather than a fixed mask.

**Dynamic Gradient Estimation (DGE):** In order to make the above sampling subnet trainable, inspired by recent works in Binarized Neural Networks (BNNs) [68], [69], a dynamic gradient estimation strategy is further introduced for the vanilla binarization function  $Bina_v$  in Eq. (4). To be specific, we adopt the following dynamic function  $g$  as a progressive approximation of  $Bina_v$  during backward propagation:

$$g(x) = \frac{k \tanh(2tx) + 1}{2}, \quad (7)$$

whose derivative is used as the backward gradient of  $Bina_v$ . Here  $t$  and  $k$  are control variables and change as follows during the training period:

$$t = T_{\min} 10^{\frac{i}{N_e} \times \lg \frac{T_{\max}}{T_{\min}}}, \quad k = \max\left(\frac{1}{t}, 1\right), \quad (8)$$

where  $i$  is the current epoch and  $N_e$  is the number of epochs,  $T_{\min} = 10^{-1}$  and  $T_{\max} = 10^1$ . Note that  $t$  and  $k$  controls the slope and output range of function  $g$ , respectively.

Essentially, this is a progressive two-stage strategy to approximate the vanilla binarization function  $Bina_v$ . As shown in Fig. 2, during Stage 1, i.e., the early stage of training, we focus more on the updating ability and speed of the backward propagation algorithm. Therefore, the gradient estimation function's derivative value is kept close to one, and then the output range is reduced progressively from a large scope to  $[0, 1]$ . As for Stage 2, we lay more importance on retaining accurate gradients for parameters around zero. Consequently, we keep the function output range as  $[0, 1]$  and gradually increase the slope so as to push the estimation curve to the shape of the binarization function  $Bina_v$ , thus achieving the consistency of forward and backward propagation during the late stage of training. Based on the above two-stage scheme, our proposed DGE gradually approximates the binarization function in backward propagation and reasonably updates all parameters, which is shown to efficiently retain the gradient information and promote network performance.

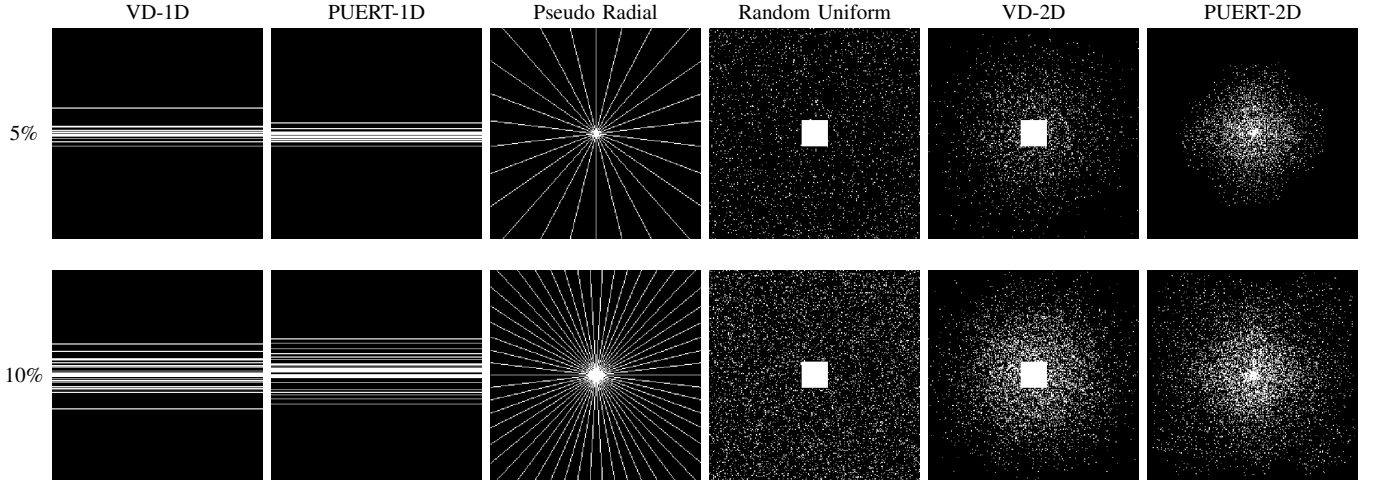


Fig. 3. Visual reconstruction comparisons of different masks under RecNet trained on Brain dataset for two different ratios  $\alpha = 5\%$  and  $\alpha = 10\%$ . VD is short for Variable Density. Compared with VD-2D masks, the learned 2D masks are similar for lower frequencies, but different for high frequency values, which demonstrates that our PUERT enjoys the ability to adaptively optimize sub-sampling patterns for specific ratios.

### C. Reconstruction Subnet

In the Reconstruction Subnet (**RecNet**), without losing generality, we construct a deep unfolding network by casting the traditional ISTA method into a deep network form, so as to achieve simplicity, effectiveness and interpretability.

The classic CS-MRI optimization problem Eq. (2) can be efficiently solved with ISTA by iterating the following two update steps:

$$\mathbf{r}_k = \hat{\mathbf{x}}_{k-1} - \rho \mathbf{F}^H (\mathbf{M} \odot \mathbf{F} \hat{\mathbf{x}}_{k-1} - \mathbf{y}), \quad (9)$$

$$\hat{\mathbf{x}}_k = \text{prox}_{\lambda\psi}(\mathbf{r}_k). \quad (10)$$

Here,  $k$  denotes the iteration index,  $\rho$  is the step size,  $\mathbf{F}^H$  denotes the Inverse Fourier Transform (IFT) and the proximal mapping operator of regularizer  $\psi$  is defined as  $\text{prox}_{\lambda\psi}(\mathbf{r}) = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{r}\|_2^2 + \lambda\psi(\mathbf{x})$ . Eq. (9) and Eq. (10) are usually called gradient descent step (GDS) and proximal mapping step (PMS), respectively.

Following [24], our ISTA-unfolding reconstruction subnet consists of  $K$  stages and each stage corresponds to one iteration in ISTA. Concretely, each stage is composed of a gradient descent module (GDM) and a proximal mapping module (PMM), which correspond to the above two update steps Eq. (9) and Eq. (10), respectively.

For GDM, to preserve the ISTA structure while increasing network flexibility, the step size  $\rho$  is allowed to vary across iterations, and GDS, *i.e.*, Eq. (9) is casted as follows:

$$\mathbf{r}_k = \hat{\mathbf{x}}_{k-1} - \rho_k \mathbf{F}^H (\mathbf{M} \odot \mathbf{F} \hat{\mathbf{x}}_{k-1} - \mathbf{y}). \quad (11)$$

As for PMM in each stage, we propose a simple yet effective module as shown in Fig. 1, which can be formulated as:

$$\hat{\mathbf{x}}_k = \mathbf{r}_k + \mathcal{H}_k^{\text{rec}}(\mathcal{H}_k^{\text{RB},2}(\mathcal{H}_k^{\text{RB},1}(\mathcal{H}_k^{\text{ext}}(\mathbf{r}_k)))). \quad (12)$$

Here, PMM consists of 2 residual blocks (RBs)  $\mathcal{H}_k^{\text{RB},1}$  and  $\mathcal{H}_k^{\text{RB},2}$ , two convolution layers  $\mathcal{H}_k^{\text{ext}}$ ,  $\mathcal{H}_k^{\text{rec}}$ , which devote to extracting the image features and reconstruction, respectively, and a long residual connection with input  $\mathbf{r}_k$ .

### D. Network Parameters and Loss Function

Given the training dataset with full-resolution images  $\{\mathbf{x}^{(i)}\}_{i=1}^{N_D}$ , the sampling subnet first uses a learnable sampling mask  $\mathbf{M}$  to get simulation measurements  $\mathbf{y}^{(i)} = \mathbf{M} \odot \mathbf{F} \mathbf{x}^{(i)}$ , and then, with initialization  $\mathbf{x}_0^{(i)} = \mathbf{F}^H \mathbf{y}^{(i)}$  as input, the reconstruction subnet outputs the recovery results, aiming to reduce the discrepancy between  $\mathbf{x}^{(i)}$  and  $\hat{\mathbf{x}}_K^{(i)}$ . Therefore, we use the following end-to-end loss function to train our PUERT:

$$\mathcal{L}(\Theta) = \frac{1}{N_D N} \sum_{i=1}^{N_D} \|\hat{\mathbf{x}}_K^{(i)} - \mathbf{x}^{(i)}\|_2^2, \quad (13)$$

where  $N_D$  is the number of training images and  $N$  is the size of each image  $\mathbf{x}^{(i)}$ , *i.e.*,  $N = N_x \times N_y$ .  $\Theta$  denotes the learnable parameter set in the two subnets of PUERT. Note that, in the sampling subnet, the learnable parameters are not directly the probabilistic sampling pattern  $\mathbf{P}$ , but an unconstrained image  $\mathbf{O}$  with the same size, such that  $\mathbf{P}_{i,j} = \sigma(\mathbf{O}_{i,j})$  to realize  $\mathbf{P}_{i,j} \in (0, 1)$ . Here,  $\sigma$  denotes the sigmoid function with the slope set to be 5. It is also worth emphasizing that, thanks to the rescale operator as Eq. (5), a loss term to constrain the sparsity ratio of the learned sampling mask is unnecessary.

## IV. EXPERIMENTS

Our PUERT is implemented in PyTorch [70] on one NVIDIA Tesla V100. We utilize Adam [71] optimization with a learning rate of 0.0001 (3000 epochs) and a batch size of 8. The default value of stage number  $K$  is 9. Following common practices in previous works, we use two widely used benchmark datasets: Brain [25] and FastMRI [72], which contain 100 brain and 4501 knee MR images for training, and 50 brain and 657 knee MR images for testing, respectively. Brain dataset is based on ground-truth images and follows the simulation process stated in Section III-D. As for FastMRI dataset, the simulation is similar, except that the fully-sampled ground truth is in  $k$ -space (emulated single coil [73]) instead of the image domain, and thus, is relatively more realistic. The recovered results are evaluated with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [74].



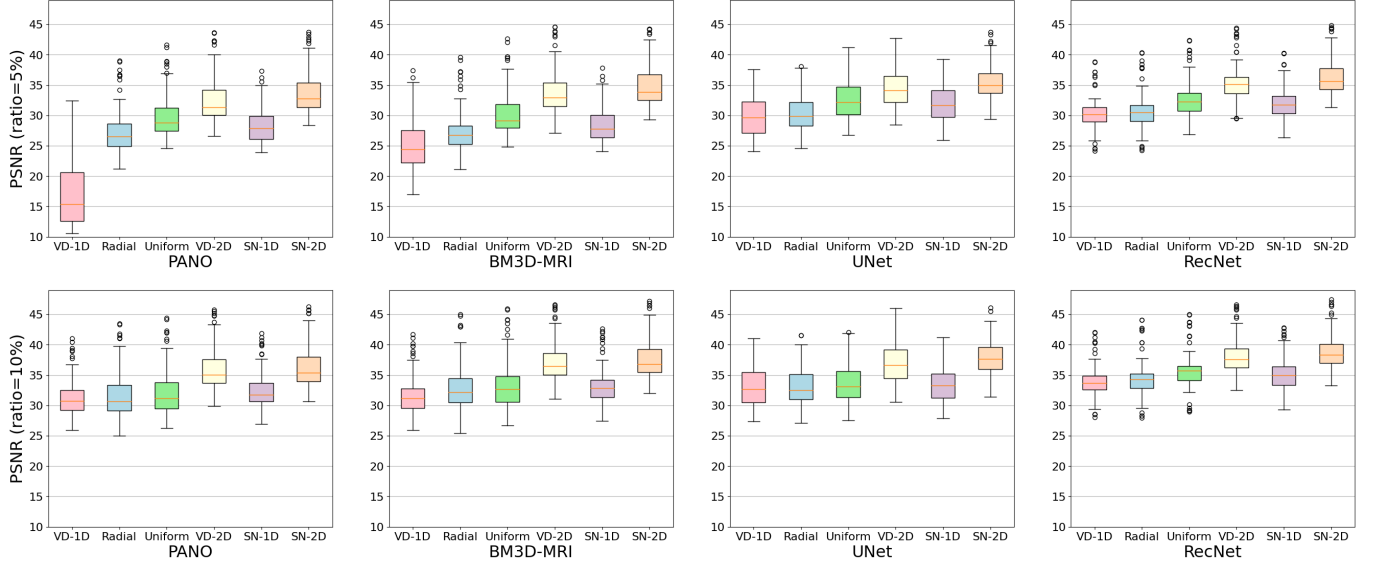


Fig. 4. Box plots for PSNR values of reconstructions obtained with four reconstruction methods, six different sampling schemes and two sub-sampling ratios on Brain dataset. SN-1D and SN-2D denote our proposed SampNet under 1D and 2D settings, respectively. Our SampNet yields attractively better results than competing masks when combined with any of the four reconstruction methods and under any of the two CS ratios.

TABLE I

AVERAGE PSNR COMPARISONS OF VARIOUS SAMPLING SCHEMES, WITH FOUR RECOVERY METHODS UNDER DIFFERENT CS RATIOS FOR BOTH 1D AND 2D SETTINGS ON BRAIN DATASET. OUR PROPOSED SAMPNET YIELDS BEST RESULTS AMONG ALL COMPETING MASKS UNDER ALL CONDITIONS.

Ratios	Methods	1D Settings		2D Settings			
		VD-1D	SampNet	Radial	Uniform	VD-2D	SampNet
5%	PANO [75]	17.51	28.53	27.70	30.26	33.04	34.14
	BM3D-MRI [76]	25.43	28.72	27.78	30.77	34.26	35.27
	U-Net [72]	30.03	32.09	30.49	32.91	34.66	35.65
	RecNet	30.64	32.26	31.02	33.00	35.79	36.74
10%	PANO [75]	31.66	32.79	31.93	32.69	36.26	36.75
	BM3D-MRI [76]	32.06	33.65	33.19	33.83	37.58	38.09
	U-Net [72]	33.32	33.68	33.45	33.82	37.31	38.03
	RecNet	34.22	35.51	34.88	36.03	38.45	39.13

#### A. Comparisons with Classic Masks under Multiple Reconstruction Methods

To validate the effectiveness of our proposed sampling pattern optimization scheme, *i.e.*, our SampNet, we compare our learnable masks with four types of classic fixed masks under the conditions of four reconstruction methods.

- For *classic fixed masks*, we consider four categories that are widely used in the literature: Cartesian [3] with skipped lines (dubbed VD-1D), pseudo Radial [25], Random Uniform [2] and Variable Density under 2D (dubbed VD-2D) [4] masks, where the first category is 1D sub-sampling mask and the last three are 2D masks. Note that a fixed calibration region of size  $32 \times 32$  is adopted in the center of the  $k$ -space for Uniform and VD-2D masks in order to yield better recovery performance.
- For *reconstruction methods*, we consider two traditional model-based methods, *i.e.*, PANO [75] and BM3D-MRI [76], and one widely used deep learning model called U-Net [72]. Our reconstruction subnet (RecNet) is also extracted as a recovery model for fair comparisons.

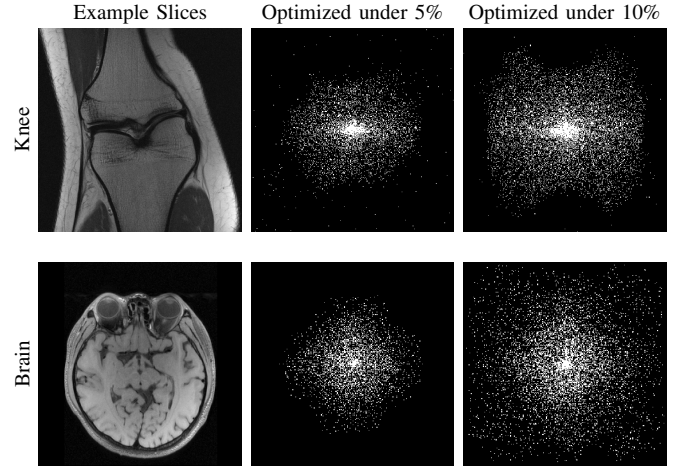


Fig. 5. Visual comparisons of PUERT-optimized sampling masks for the knee and brain anatomies (under  $\alpha = 5\%$  and  $\alpha = 10\%$ ). For the knee anatomy, more attention has been paid to lateral frequencies, whereas for the brain, the learned masks are more radially symmetric. This comparison result highlights the importance of adapting the learnable under-sampling pattern to the anatomy and data at hand.

In this experiment, the aforementioned four classic masks as well as our proposed SampNet, including 1D and 2D versions, are experimented under the above four reconstruction methods, thus leading to 24 possible combinations. Note that when combining the two traditional methods PANO and BM3D-MRI with our SampNet, we use the *fixed* sampling masks learned by our PUERT to conduct the reconstruction. While for U-Net, it is combined with our proposed SampNet to *jointly* optimize the parameters for under-sampling and reconstruction. Note that for the 1D setting, our SampNet adopts a Bernoulli probabilistic pattern  $\mathbf{P} \in \mathbb{R}^{N_x}$  and generates a binary vector  $\mathbf{V} \in \mathbb{R}^{N_x}$ , *i.e.*,  $\mathbf{V}_i \sim \mathcal{B}(\mathbf{P}_i)$ , which is then expanded on the column dimension to a 1D sampling mask  $\mathbf{M} \in \mathbb{R}^{N_x \times N_y}$ .

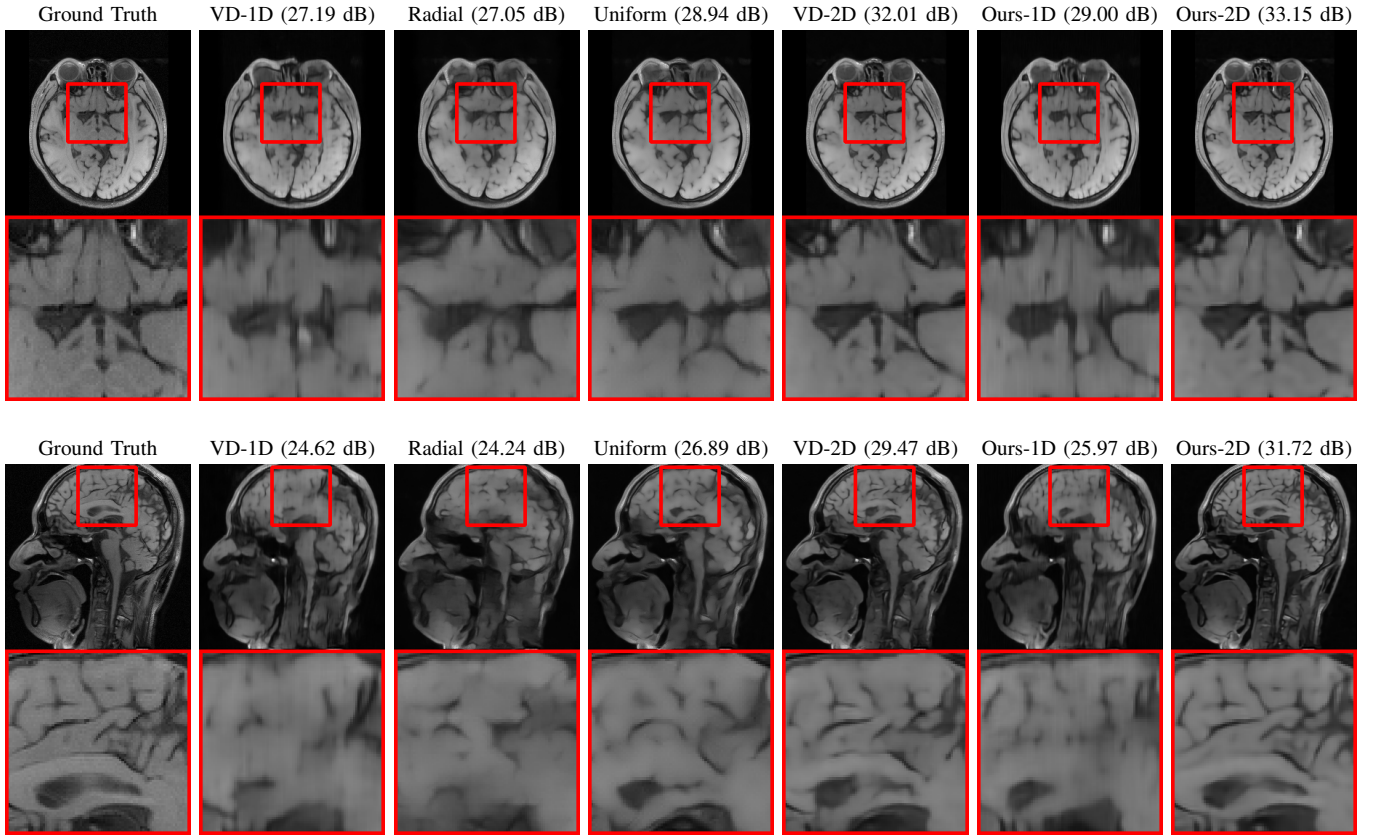


Fig. 6. Visual reconstruction comparisons of different masks under RecNet at ratio 5% for 1D and 2D settings on Brain dataset. The proposed PUERT is able to recover more details and much sharper edges than other competing masks, even under the aggressive 5% sampling ratio.

Fig. 3 shows the visual comparisons among the classic masks and our optimized sub-sampling matrices trained on Brain dataset for two different ratios  $\alpha = 5\%$  and  $\alpha = 10\%$ . One can see that the learned 2D masks are similar to VD-2D masks, both showing a strong preference for lower frequencies and exhibiting a denser sampling pattern closer to the origin of  $k$ -space. While for high frequency values, compared with VD-2D masks, our optimized masks show a relatively smaller density when  $\alpha = 5\%$  but larger when  $\alpha = 10\%$ , which demonstrates that our PUERT enjoys the ability to adaptively optimize sub-sampling patterns for specific ratios.

Furthermore, the optimized masks by our PUERT for the knee and brain anatomies are compared in Fig. 5. We observed that, for the knee MR images, more attention has been paid to lateral frequencies rather than ventral frequencies, owing to the unique asymmetric feature of the knee anatomy, where there is dramatically more tissue contrast in the lateral direction. The masks learned on Brain dataset, on the other hand, exhibit a more radially symmetric feature. This comparison highlights the importance of adapting the learnable under-sampling pattern to the anatomy and data at hand.

Fig. 4 shows box plots for PSNR values of reconstructions obtained with four reconstruction methods, six different sampling schemes, and two sub-sampling ratios on Brain dataset. One can intuitively observe that overall our proposed SampNet yields attractively better results than competing masks when combined with any of the four reconstruction methods and under any of the two CS ratios, which demonstrates the

superiority of our proposed sub-sampling learning scheme. It is worth emphasizing that U-Net with our proposed SampNet achieves higher PSNR than competing classic masks, thus validating that the proposed PUERT enjoys the generality of being extended to other reconstruction networks.

Table I lists the concrete PSNR values of Fig. 4. We can observe that, benefiting from our efficient probabilistic under-sampling scheme and dynamic gradient estimation strategy, the proposed SampNet achieves the highest PSNR results among all competing masks. Taking RecNet under ratio 10% for 2D setting as an example, our SampNet achieves remarkable 0.68 dB PSNR gains over the state-of-the-art classic VD-2D masks. It is also noteworthy that, when equipped with our SampNet, U-Net still performs worse than our RecNet, thus demonstrating that, compared with the classic data-driven reconstruction model, a deep unfolding network is more suitable to facilitate exploration and efficient training of our SampNet. We attribute such superiority to its intrinsic structural feature of fully utilizing knowledge on the sampling mask in each stage of the deep unfolding network.

Fig. 6 further shows the visual reconstruction comparisons of different masks under RecNet at ratio 5% for 1D and 2D settings on Brain dataset. As can be intuitively appreciated from the two test images, the PUERT is able to consistently recover more details and much sharper edges than other competing masks, even under the aggressive 5% sampling ratio, thus validating the effectiveness, efficiency and practicability of our proposed PUERT.

TABLE II  
AVERAGE PSNR PERFORMANCE COMPARISONS WITH VARIOUS STATE-OF-THE-ART METHODS WITH DIFFERENT CS RATIOS ON BRAIN AND FASTMRI DATASETS. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN RED AND BLUE COLORS, RESPECTIVELY.

Datasets	Methods	Masks	Ratios			Avg.	GPU Time
			5%	10%	15%		
Brain	Zero-filled	Fixed (Radial)	24.22/0.5140	26.81/0.6030	28.80/0.6713	26.61/0.5961	0.0038s
	UNet-DC [16]		30.07/0.7540	32.83/0.8221	34.75/0.8639	32.55/0.8133	0.0199s
	ADMMNet [25]		30.13/0.7958	34.46/0.8972	36.83/0.9306	33.81/0.8745	0.0192s
	ISTA-Net+ [24]		30.64/0.8176	34.73/0.9052	37.07/0.9343	34.15/0.8857	0.0277s
	RDN [20]		31.04/0.8056	34.62/0.8887	36.85/0.9221	34.17/0.8721	0.0310s
	CDDN [19]		31.01/0.8163	35.21/0.9074	37.35/0.9353	34.52/0.8863	0.0483s
	LOUPE-1D [56]	Learned	32.11/0.8647	35.16/0.9311	36.51/0.9447	34.59/0.9135	0.0536s
	LOUPE-2D [56]		<b>35.34/0.9245</b>	<b>36.88/0.9381</b>	<b>38.30/0.9473</b>	<b>36.84/0.9366</b>	0.0699s
	PUERT-1D (Ours)		32.26/0.8839	35.51/0.9285	37.44/0.9484	35.07/0.9203	0.0214s
	PUERT-2D (Ours)		<b>36.74/0.9409</b>	<b>39.13/0.9551</b>	<b>40.84/0.9636</b>	<b>38.90/0.9532</b>	0.0578s
FastMRI	Zero-filled	Fixed (Cartesian)	25.29/0.5759	26.31/0.6205	27.50/0.6565	26.37/0.6176	0.0045s
	UNet-DC [16]		27.25/0.6347	29.19/0.7048	30.60/0.7458	29.01/0.6951	0.0226s
	ADMMNet [25]		27.32/0.6372	29.05/0.6995	30.85/0.7506	29.07/0.6958	0.0227s
	ISTA-Net+ [24]		27.50/0.6436	29.82/0.7210	31.57/0.7657	29.63/0.7101	0.0241s
	RDN [20]		27.70/0.6487	30.19/0.7290	31.78/0.7721	29.89/0.7166	0.0252s
	CDDN [19]		27.81/0.6507	30.32/0.7272	31.92/0.7715	30.02/0.7165	0.0723s
	LOUPE-1D [56]	Learned	30.68/0.7217	31.15/0.7689	33.06/0.8211	31.63/0.7706	0.0487s
	LOUPE-2D [56]		<b>31.94/0.7530</b>	<b>33.43/0.8033</b>	<b>34.73/0.8409</b>	<b>33.37/0.7991</b>	0.0623s
	PUERT-1D (Ours)		30.71/0.7170	32.57/0.7712	33.75/0.8110	32.34/0.7664	0.0301s
	PUERT-2D (Ours)		<b>32.87/0.7557</b>	<b>33.96/0.7991</b>	<b>35.16/0.8412</b>	<b>34.00/0.7987</b>	0.0694s

### B. Comparisons with State-of-the-Art Methods

We compare our proposed PUERT (including 1D and 2D versions) with six representative state-of-the-art methods, namely UNet-DC [16], ADMMNet [25], ISTA-Net [24], RDN [20], CDDN [19] and LOUPE [56]. The first five methods are reconstruction networks trained under some *fixed* sampling mask, while the last method jointly optimizes the *learnable* sampling pattern and the reconstruction network parameters. Following [25] and [72], for those trained under some fixed mask, we adopt Radial masks (2D) and Cartesian masks with skipped lines (1D) on Brain and FastMRI datasets respectively. Note that LOUPE is also able to implement both 1D and 2D sub-sampling optimization schemes.

The average PSNR/SSIM performance reconstructions of various methods on two datasets with respect to three CS ratios are summarized in Table II. One can observe that methods with learnable sampling patterns, especially our PUERT, overall yield higher PSNR than those with fixed masks, which verifies the superiority of sampling pattern optimization schemes. Concretely, for Brain dataset, our PUERT-1D and PUERT-2D achieve on average 0.55 dB and 4.38 dB PSNR gain over the state-of-the-art method CDDN respectively. It is noteworthy that the Radial masks used for Brain dataset belong to a 2D setting, whereas our PUERT under the 1D setting still enjoys superiority by a large margin. As for FastMRI dataset, our PUERT-1D and PUERT-2D achieve on average 2.32 dB and 3.98 dB PSNR gain over the state-of-the-art method CDDN respectively.

When compared with the sampling pattern optimization scheme LOUPE, one can observe that, on FastMRI dataset, LOUPE-2D has a slightly better SSIM while PUERT-2D obtains a remarkably higher PSNR on average. As for Brain dataset, the proposed PUERT-2D consistently produces higher PSNR and SSIM results than LOUPE-2D across three CS ratios. The performance advantage of PUERT under 2D is

also shown under 1D. The above results verify the superiority of the proposed PUERT against LOUPE. Note that, LOUPE mainly adopts the U-Net architecture as the recovery network, it also shows an example of reconstructing with a DUN called CascadeNet [18], [56], with an observation that the performance metrics are very close between different reconstruction networks. However, we highlight the importance of adopting DUNs to facilitate exploration and efficient training of the sampling subnet, owing to the intrinsic structural feature of fully utilizing knowledge on the sampling mask. More discussions on the importance of DUN are shown in Section IV-E.

As shown in Table II, six representative state-of-the-art methods are based on neural networks, and thus achieve a real-time reconstruction speed, with the inference GPU time less than 0.1s. Among the five methods trained under some fixed mask, CDDN needs the largest amount of inference time, owing to the dense blocks and dilated convolutions adopted in the network. Besides, compared with LOUPE, our PUERT achieves a comparable inference speed on average, but with higher PSNR in reconstruction accuracy, thus demonstrating the efficiency and superiority of the proposed PUERT.

Furthermore, visual comparisons of all the competing methods on FastMRI dataset are shown in Fig. 7 with CS ratio  $\alpha = 10\%$ . Obviously, the proposed PUERT is able to produce more faithful and clearer results than the other competitive methods. Overall, the above experiments on two widely used MRI datasets demonstrate that our proposed PUERT performs favorably against state-of-the-arts in terms of both quantitative metrics and visual quality.

### C. Ablation on Probabilistic Under-Sampling

In this section, we present an ablation study on Probabilistic Under-Sampling (PU) in our sampling subnet, so as to emphasize the importance of learning a probabilistic sampling pattern rather than a deterministic mask.



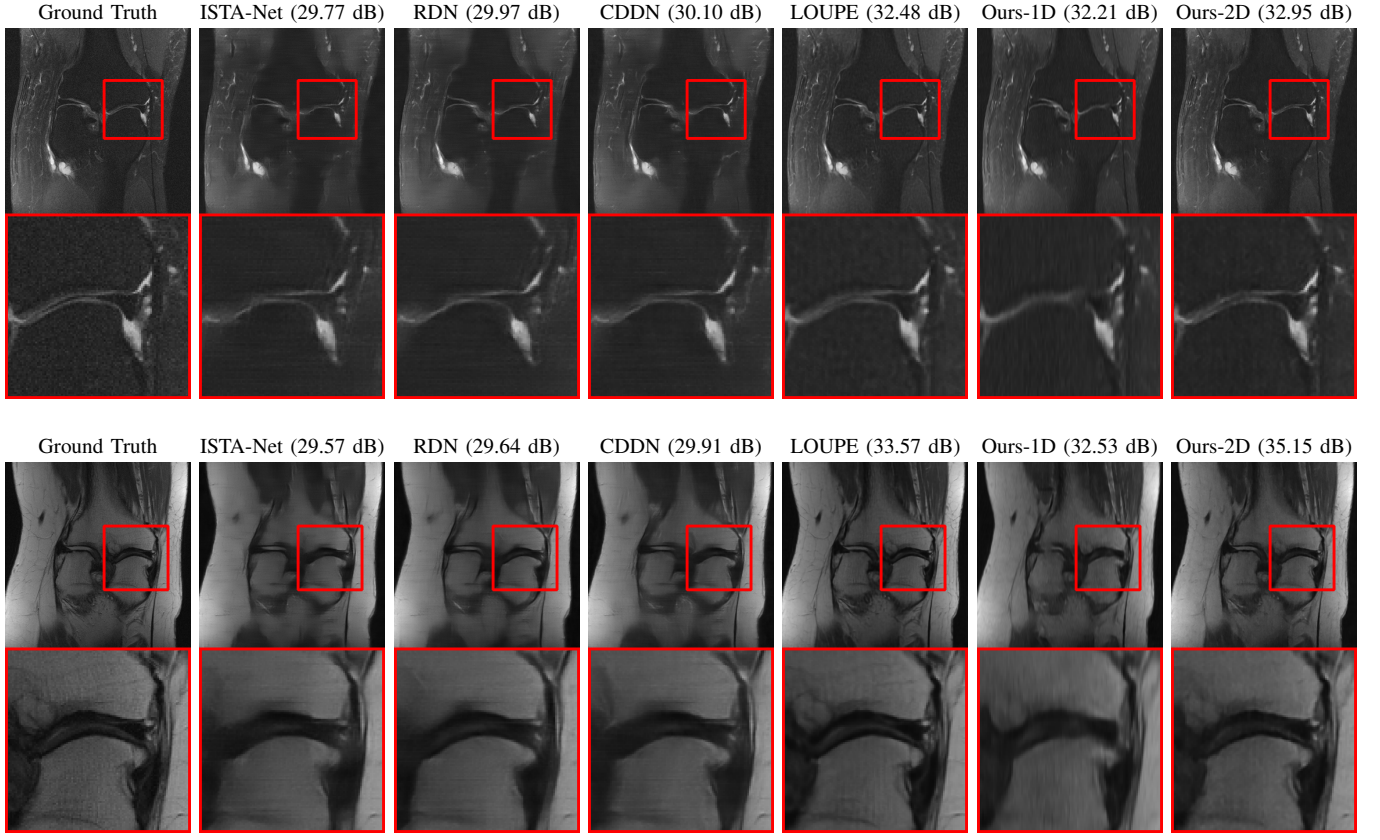


Fig. 7. Visual reconstruction comparisons with various state-of-the-art methods on FastMRI dataset with CS ratio  $\alpha = 10\%$ . Obviously, the proposed PUERT is able to produce more faithful and clearer results than the other competitive methods.

TABLE III  
AVERAGE PSNR RESULTS FOR ABLATION STUDY ON PU WHEN RATIO IS 5% AND 10% UNDER 1D AND 2D SETTINGS ON BRAIN DATASET.

Methods	PU	U (test)	1D setting		2D setting	
			5%	10%	5%	10%
case (a)	✗	✗	31.02	35.06	36.60	39.09
case (b)	✗	✗	31.08	34.91	36.51	38.91
case (c)	✓	✗	32.18	35.49	36.67	39.14
PUERT	✓	✓	32.26	35.51	36.74	39.13

First, we consider two possible mask learning schemes uncorrelated with probability (without PU), named case (a) and (b), so as to demonstrate the superiority of PU during *training*. Then, we consider testing without the Uniform distribution U, called case (c), to investigate the contribution of adopting U during *testing*. Table III shows the PSNR comparisons for the above three cases and our PUERT when ratio is 5% and 10% under 1D and 2D settings on Brain dataset. Fig. 8 further illustrates their progression curves of test PSNR when ratio is 10% under the 1D setting. We detail the above three cases and their results in the following three paragraphs.

In case (a), the sampling subnet directly binarizes the unconstrained parameter matrix  $\mathbf{O}$  into the sampling mask  $\mathbf{M}$  via  $\mathbf{M}_{i,j} = \text{Bina}(\mathbf{O}_{i,j})$  as a replacement of Eq. (3). Here, a L2 loss term to constrain the sparsity ratio of the learned sampling mask is necessary. However, such loss term still can not stabilize the ratio to the target setup, and therefore, in pursuit of fair comparisons, we have to test case (a) via

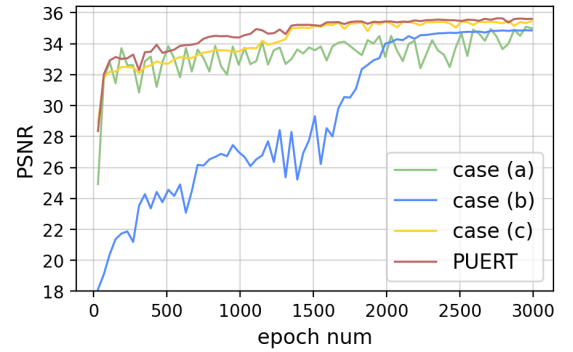


Fig. 8. The progression curves of test PSNR results for ablation study on PU when ratio is 10% under the 1D setting on Brain dataset.

greedy binarization Eq. (6) but with  $\Omega$  being the set  $\{\mathbf{O}_{i,j}\}$ . Table III shows that the final PSNR result of case (a) is inferior against PUERT under all conditions. This is mainly due to: 1) the limited optimization space without PU, 2) the mismatch between training and testing, and 3) the fluctuation caused by the poor ratio control, as shown in Fig. 8.

In order to eliminate the influence of the latter two limitations in case (a), we design case (b) to adopt greedy binarization during both training and testing, via Eq. (6) with  $\Omega$  being the set  $\{\mathbf{O}_{i,j}\}$ . Table III reports that case (b) achieves PSNR scores comparable with case (a), but still lower than PUERT. Fig. 8 further shows that, during the later period of training, the curve stability of case (b) is better than case (a).

TABLE IV  
AVERAGE PSNR RESULTS FOR ABLATION STUDY ON DGE WHEN RATIO IS 5% AND 10% UNDER 1D AND 2D SETTINGS ON BRAIN DATASET.

Methods	Update Speed	Estimation Accuracy	1D setting		2D setting	
			5%	10%	5%	10%
STE	✓	✗	30.22	34.38	36.25	38.06
Sigmoid	✗	✓	31.08	34.93	36.64	39.04
DGE	✓	✓	32.26	35.51	36.74	39.13

However, in the earlier training period, the PSNR upgrade of case (b) is obviously slow. We attribute such slow update speed to the greedy binarization used for training, which adopts a movable and relative boundary, rather than a fixed boundary adopted in vanilla binarization (*i.e.*, the constant 1).

With the aim of investigating the contribution of adopting  $\mathbf{U}$  during *testing*, we consider case (c) which tests PUERT without the Uniform distribution  $\mathbf{U}$ , *i.e.*, testing via Eq. (6) but with  $\Omega$  being the set  $\{\mathbf{P}_{i,j}\}$ . Without the randomness and exploration introduced by  $\mathbf{U}$ , case (c) completely depends on the values in  $\mathbf{P}$ , and thus, with an earlier  $\mathbf{P}$  which has not yet fully learned the probability value, case (c) achieves obviously lower PSNR than PUERT in the earlier period, as shown in Fig. 8. Table III also reports that case (c) achieves slightly lower PSNR than PUERT. Note that, in actual scanning, we might use a fixed mask as case (c), but this experiment still verifies the contribution of adopting randomness.

Overall, the above results corroborate the importance of learning a probabilistic sampling pattern rather than a deterministic mask, thus demonstrating the superiority of our design on Probabilistic Under-Sampling (PU) in the sampling subnet.

#### D. Ablation on Dynamic Gradient Estimation

In this section, we present an ablation study on our proposed Dynamic Gradient Estimation (DGE) in our sampling subnet, so as to underscore the great significance of two-step gradient estimation. Note that, in LOUPE, the binarization operator is directly relaxed to the sigmoid function to enable optimizing the network by back-propagation. However, this relaxation not only causes a network performance penalty, but also leads to a non-binary output mask, which means that, after one has obtained the optimized non-binary mask, the reconstruction network needs to be retrained with the learned binary mask. To overcome the above two limitations, our PUERT chooses to still adopt the binarization in the forward pass, but develops an efficient gradient estimator for the backward pass.

Such gradient estimators have been widely investigated in Binarized Neural Networks (BNNs) [68], [77]. One typical and simple estimator is Straight-Through Estimator (STE), which was first proposed by Hinton [78] to train networks with binary activations (*i.e.*, binary neuron). In STE, the values are passed through a binarization layer that evaluates the sign in the forward pass and performs the identity function during the backward pass. Adopting STE as an ablation study, we replace the dynamic gradient estimation function  $g(x)$  in Eq. (7) simply by the identity function  $h(x) = x$  and conduct an experiment. Besides, we also consider an STE variant proposed in [79], which uses the sigmoid function  $\sigma(x) = \frac{1}{1+e^{-x}}$  as a replacement of the identity function.

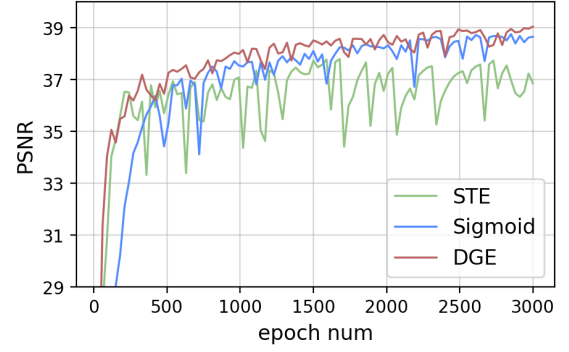


Fig. 9. The progression curves of test PSNR results for ablation study on DGE when ratio is 10% under the 2D setting on Brain dataset.

Table III reports the PSNR comparisons for 1) STE, 2) Sigmoid, and 3) our DGE, when ratio is 5% and 10% under 1D and 2D settings on Brain dataset, which shows that DGE outperforms the other two methods across all situations. Fig. 9 further provides the progression curves of test PSNR results for the above three methods, when ratio is 10% under the 2D setting. One can intuitively observe that, STE converges faster than Sigmoid during the earlier training process, whereas lacks the ability to stably and consistently improve the reconstruction accuracy (resembling Sigmoid) during the later training period. Such different performances are the result of their different focuses. Concretely, STE directly passes the gradients through the binarization operator so as to guarantee the updating speed, while Sigmoid estimates the gradients with high similarity to the binarization operator to ensure estimation accuracy. As a combination of the above two focuses, our proposed two-step DGE is able to dynamically estimate the gradients and emphasize different priorities at the right time, thus implementing both fast convergence speed and high reconstruction accuracy, as can be appreciated from Fig. 9.

#### E. Discussions on the Importance of DUN

In the previous statements, we have emphasized to adopt Deep Unfolding Networks (DUNs) for our PUERT in order to fully utilize the knowledge on the learned sampling mask. In this section, two experiments are further conducted to investigate the contribution of adopting DUN in our PUERT.

Firstly, considering a DUN and a non-DUN with similar performances, we respectively equip them with our sampling subnet (SampNet), so as to compare the different performance gains brought by SampNet. Concretely, we select U-Net [72] as an instance of non-DUNs, and choose our RecNet with 4 stages, named RecNet<sub>s4</sub>, to represent DUNs. Table V provides the PSNR results of U-Net with and without SampNet, as well as RecNet<sub>s4</sub> with and without SampNet, when ratio is 5% and 10% under 1D and 2D settings on Brain dataset. It can be observed that, without SampNet, U-Net and RecNet<sub>s4</sub> achieve similar PSNR results on average. However, when equipped with SampNet, RecNet<sub>s4</sub> realizes remarkably higher PSNR increase than U-Net (1.46 dB *v.s.* 1.03 dB on average). This result confirms the superiority of DUNs against non-DUNs on fully exploiting the knowledge on the learned mask.

TABLE V  
AVERAGE PSNR RESULTS FOR DISCUSSIONS ON DUN WHEN RATIO IS 5% AND 10% UNDER 1D AND 2D SETTINGS ON BRAIN DATASET.

Methods	1D setting		2D setting		Average
	5%	10%	5%	10%	
U-Net w/o. SampNet	30.03	33.32	34.66	37.31	33.83
U-Net w. SampNet	32.09	33.68	35.65	38.03	34.86
PSNR Increase	2.06	0.36	0.99	0.72	1.03
RecNet <sub>s4</sub> w/o. SampNet	28.76	33.36	34.58	37.50	33.55
RecNet <sub>s4</sub> w. SampNet	30.62	34.52	36.22	38.67	35.01
PSNR Increase	1.86	1.16	1.64	1.17	1.46
PUERT	32.26	35.51	36.74	39.13	35.91
PUERT <sup>+</sup>	32.30	35.58	36.89	39.44	36.05
PSNR Increase	0.04	0.07	0.15	0.31	0.14

Secondly, in order to validate the promising superiority of DUNs on further exploiting the information in SampNet and improving the performance, we propose an advanced version of PUERT, dubbed PUERT<sup>+</sup>, which integrates the probabilistic sampling pattern  $\mathbf{P}$  in each stage of RecNet. Specifically, for each stage, we add a module called GDM<sup>P</sup>, which is the same as GDM in Eq. (11) except for replacing  $\mathbf{M}$  with  $\mathbf{P}$ , formulated as:

$$\mathbf{r}_k^{\mathbf{P}} = \hat{\mathbf{x}}_{k-1} - \rho_k \mathbf{F}^H (\mathbf{P} \odot \mathbf{F} \hat{\mathbf{x}}_{k-1} - \mathbf{y}). \quad (14)$$

The input of each PMM is correspondingly modified to the concatenation of  $\mathbf{r}_k^{\mathbf{P}}$  and  $\mathbf{r}_k$ , formulated as:

$$\hat{\mathbf{x}}_k = \mathbf{r}_k + \mathcal{H}_k^{\text{rec}}(\mathcal{H}_k^{\text{RB},2}(\mathcal{H}_k^{\text{RB},1}(\mathcal{H}_k^{\text{ext}}([\mathbf{r}_k, \mathbf{r}_k^{\mathbf{P}}])))), \quad (15)$$

where  $[\cdot]$  is the concatenation operator. In this way, each stage can fully use the information of not only the sampling mask  $\mathbf{M}$  but also the probabilistic sampling pattern  $\mathbf{P}$  from SampNet. Compared to the *hard-sampling* GDM with the binary mask  $\mathbf{M}$ , the newly added module GDM<sup>P</sup> with the non-binary  $\mathbf{P}$  can be regarded as a *soft-sampling* version. Table V provides the PSNR comparisons between PUERT and PUERT<sup>+</sup>, when ratio is 5% and 10% under 1D and 2D on Brain dataset. One can clearly see that PUERT<sup>+</sup> obtains consistently higher scores than PUERT across all situations, with average PSNR increased from 35.91 dB to 36.05 dB. The 2D version achieves higher PSNR improvements than 1D, due to the utilization of more information. Note that, taking the 2D version as an example, PUERT<sup>+</sup> only adds 2K learnable parameters, compared to PUERT with 404K parameters. Above results verify the promising superiority of DUNs on further exploiting the information in SampNet and improving the performance. More elaborate and efficient designs on further exploitation of SampNet are considered as our important future direction.

## V. DISCUSSIONS AND LIMITATIONS

One weakness of PUERT is the assumption of the sampling model, in which all  $k$ -space locations are independent. Although, due to the control of sampling ratio, our network would learn the relative importance among  $k$ -space locations and obtain the dependency implicitly, we must admit that our PUERT is incapable of directly and explicitly learning the dependency among  $k$ -space locations. Since such dependency

is useful [5] and promising to further improve the optimised sampling mask, we leave it as an important future direction.

Note that our SampNet can also be trivially generalized to other neural networks, and we consider the exploration of other elaborate architectural designs for PUERT as an important future direction. However, PUERT does not support sampling pattern learning for traditional model-based reconstruction methods, since we need to optimize the learnable probabilistic sampling pattern  $\mathbf{P}$  in an end-to-end manner. Considering that traditional methods still enjoy great advantages (*e.g.*, fast adaptation to multiple masks and ratios, strong interpretability, and no training requirements), we consider the exploration of extending our sampling pattern learning method to traditional reconstruction methods as a future research direction.

In addition, there is still a certain distance from practical application. In actual scanning, one should also consider how to implement a specific under-sampling pattern in an MR pulse sequence, *e.g.*, the constraints of hardware requirements [57] and the design of viable trajectories. And we leave the analysis on extending our PUERT to be totally applicable as an important direction for future research. Besides, our experiment does not consider the simulation of noise. However, in practical applications, noise is inevitable in the measurement process [80], and we consider the extension of PUERT to handling noisy  $k$ -space data as an important future direction.

Also note that our current PUERT is restricted to the single coil CS-MRI reconstruction. However, accelerated parallel imaging [62], [81], [82] is remarkably promising to achieve higher degrees of acceleration. And we consider the combination of PUERT with multi-coil imaging as an important area of research. There exists some literature to explore data-driven learning of sampling patterns in accelerated parallel MRI. [83] employs a combinatorial method that lets the data decide sampling masks matched to specific parallel MRI decoders in use. [84] proposes a learning approach that alternates between improving the sampling pattern, using bias-accelerated subset selection, and improving parameters of the variational networks. Insights from these studies are promising to help inspire our relevant future research on parallel imaging.

## VI. CONCLUSIONS

In this paper, we simultaneously deal with two problems in CS-MRI, *i.e.*, under-sampling and reconstruction, and propose a novel end-to-end Probabilistic Under-sampling and Explicable Reconstruction neTwork, dubbed PUERT, to achieve an efficient combination of sub-sampling learning and reconstruction network training. Based on extensive experiments on two widely used MRI datasets, we have validated that our proposed PUERT performs favorably against state-of-the-art methods in terms of both quantitative metrics and visual quality, and achieves remarkable results even under challenging low sampling ratios. Besides, our detailed ablation studies confirm the importance of three components in our proposed PUERT, *i.e.*, Probabilistic Under-sampling (PU), Dynamic Gradient Estimation (DGE) and Deep Unfolding Network (DUN), where we highly emphasize to adopt the DUN in PUERT so as to fully explore the information from SampNet. In addition,

an enhanced version of PUERT, dubbed PUERT<sup>+</sup>, is also developed as an attempt to implement further exploitation of SampNet and obtain performance improvements.

## REFERENCES

- [1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [2] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008.
- [3] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing MRI with random encoding," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 893–903, 2010.
- [4] Z. Wang and G. R. Arce, "Variable density compressed image sampling," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 264–270, 2009.
- [5] E. Levine, B. Daniel, S. Vasanawala, B. Hargreaves, and M. Saranathan, "3D Cartesian MRI with compressed sensing and variable view sharing using complementary poisson-disc sampling," *Magnetic Resonance in Medicine*, vol. 77, no. 5, pp. 1774–1785, 2017.
- [6] M. Lustig and J. M. Pauly, "SPiRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–471, 2010.
- [7] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 767–776, 2010.
- [8] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [9] X. Qu, D. Guo, B. Ning, Y. Hou, Y. Lin, S. Cai, and Z. Chen, "Undersampled MRI reconstruction with patch-based directional wavelets," *Magnetic Resonance Imaging*, vol. 30, no. 7, pp. 964–977, 2012.
- [10] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 288–297, 2010.
- [11] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint," *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [12] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic  $\ell_0$ -minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2008.
- [13] D. Liang, B. Liu, J. Wang, and L. Ying, "Accelerating sense using compressed sensing," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1574–1584, 2009.
- [14] C. Zhao, J. Zhang, R. Wang, and W. Gao, "Cream: Cnn-regularized admm framework for compressive-sensed image reconstruction," *IEEE Access*, vol. 6, pp. 76838–76853, 2018.
- [15] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [16] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, 2018.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- [18] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2017.
- [19] H. Zheng, F. Fang, and G. Zhang, "Cascaded dilated dense network with two-step data consistency for MRI reconstruction," *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [20] L. Sun, Z. Fan, Y. Huang, X. Ding, and J. Paisley, "Compressed sensing MRI using a recursive dilated network," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
- [21] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "Rare: Image reconstruction using deep priors learned without groundtruth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, 2020.
- [22] R. Souza, Y. Beauferris, W. Loos, R. M. Lebel, and R. Frayne, "Enhanced deep-learning-based magnetic resonance image reconstruction by leveraging prior subject-specific brain imaging: Proof-of-concept using a cohort of presumed normal subjects," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1126–1136, 2020.
- [23] Y. Chen, C.-B. Schönlieb, P. Liò, T. Leiner, P. L. Dragotti, G. Wang, D. Rueckert, D. Firmin, and G. Yang, "AI-based reconstruction for fast MRI—a systematic review and meta-analysis," *Proceedings of the IEEE*, vol. 110, no. 2, pp. 224–245, 2022.
- [24] J. Zhang and B. Ghanem, "ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [25] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2016.
- [26] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: model-based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2018.
- [27] R. Liu, Y. Zhang, S. Cheng, Z. Luo, and X. Fan, "A deep framework assembling principled modules for CS-MRI: unrolling perspective, convergence behaviors, and practical modeling," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4150–4163, 2020.
- [28] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 280–290, 2018.
- [29] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, "Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1280–1291, 2020.
- [30] J. Song, B. Chen, and J. Zhang, "Memory-augmented deep unfolding network for compressive sensing," in *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM)*, 2021.
- [31] J. Zhang, C. Zhao, and W. Gao, "Optimization-inspired compact deep compressive sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 765–774, 2020.
- [32] D. You, J. Zhang, J. Xie, B. Chen, and S. Ma, "COAST: controllable arbitrary-sampling network for compressive sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 6066–6080, 2021.
- [33] D. You, J. Xie, and J. Zhang, "ISTA-Net<sup>++</sup>: flexible deep unfolding network for compressive sensing," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021.
- [34] Z. Wu, J. Zhang, and C. Mou, "Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [35] Z. Wu, Z. Zhang, J. Sing, and M. Zhang, "Spatial-temporal synergic prior driven unfolding network for snapshot compressive imaging," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021.
- [36] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [37] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [38] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo et al., "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.
- [39] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.
- [40] G. Li, J. Lv, X. Tong, C. Wang, and G. Yang, "High-resolution pelvic MRI reconstruction using a generative adversarial network with attention and cyclic loss," *IEEE Access*, vol. 9, pp. 105 951–105 964, 2021.
- [41] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drodzdzal, "Reducing uncertainty in undersampled MRI reconstruction with active acquisition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [42] K. H. Jin, M. Unser, and K. M. Yi, "Self-supervised deep active accelerated MRI," *arXiv preprint arXiv:1901.04547*, 2019.



- [43] Y. Gao and S. J. Reeves, "Optimal k-space sampling in mrsi for images with a limited region of support," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1168–1178, 2000.
- [44] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf, "Optimization of k-space trajectories for compressed sensing by bayesian experimental design," *Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 116–126, 2010.
- [45] A. Chakrabarti, "Learning sensor multiplexing design through back-propagation," *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2016.
- [46] R. Horstmeyer, R. Y. Chen, B. Kappes, and B. Judkewitz, "Convolutional neural networks that teach microscopes how to image," *arXiv preprint arXiv:1709.07223*, 2017.
- [47] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Cukur, J. Scarlett, and V. Cevher, "Learning-based compressive MRI," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1394–1406, 2018.
- [48] Y. F. Cheng, M. Strachan, Z. Weiss, M. Deb, D. Carone, and V. Ganapati, "Illumination pattern design with deep learning for single-shot fourier ptychographic microscopy," *Optics Express*, vol. 27, no. 2, pp. 644–656, 2019.
- [49] J. P. Haldar and D. Kim, "Oedipus: An experiment design framework for sparsity-constrained MRI," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1545–1558, 2019.
- [50] A. Muthumbi, A. Chaware, K. Kim, K. C. Zhou, P. C. Konda, R. Chen, B. Judkewitz, A. Erdmann, B. Kappes, and R. Horstmeyer, "Learned sensing: jointly optimized microscope hardware for accurate image classification," *Biomedical Optics Express*, vol. 10, no. 12, pp. 6351–6369, 2019.
- [51] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [52] F. Knoll, C. Clason, C. Diwok, and R. Stollberger, "Adapted random sampling patterns for accelerated MRI," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 24, no. 1, pp. 43–50, 2011.
- [53] F. Zijlstra, M. A. Viergever, and P. R. Seevinck, "Evaluation of variable density and data-driven k-space undersampling for compressed sensing magnetic resonance imaging," *Investigative Radiology*, vol. 51, no. 6, pp. 410–419, 2016.
- [54] F. Sherry, M. Benning, J. C. De los Reyes, M. J. Graves, G. Maierhofer, G. Williams, C.-B. Schönlieb, and M. J. Ehrhardt, "Learning the sampling pattern for MRI," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4310–4321, 2020.
- [55] C. D. Bahadir, A. V. Dalca, and M. R. Sabuncu, "Learning-based optimization of the under-sampling pattern in MRI," in *Proceedings of International Conference on Information Processing in Medical Imaging (IPMI)*, 2019.
- [56] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in MRI," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1139–1152, 2020.
- [57] T. Weiss, O. Senouf, S. Vedula, O. Michailovich, M. Zibulevsky, A. Bronstein *et al.*, "Pilot: Physics-informed learned optimized trajectories for accelerated mri," *Machine Learning for Biomedical Imaging*, vol. 1, no. April 2021 issue, pp. 1–10, 2021.
- [58] I. A. Huijben, B. S. Veeling, and R. J. van Sloun, "Learning sampling and model-based signal recovery for compressed sensing MRI," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [59] H. K. Aggarwal and M. Jacob, "J-ModL: joint model-based deep learning for optimized sampling and reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1151–1162, 2020.
- [60] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [61] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [62] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [63] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (grappa)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [64] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magnetic Resonance in Medicine*, vol. 72, no. 4, pp. 959–970, 2014.
- [65] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (LO-RAKS) for constrained MRI," *IEEE Transactions on Medical Imaging*, vol. 33, no. 3, pp. 668–681, 2013.
- [66] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank hankel matrix," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 480–495, 2016.
- [67] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [68] T. Simons and D.-J. Lee, "A review of binarized neural networks," *Electronics*, vol. 8, no. 6, p. 661, 2019.
- [69] H. Qin, R. Gong, X. Liu, M. Shen, Z. Wei, F. Yu, and J. Song, "Forward and backward information retention for accurate binary neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [70] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: an imperative style, high-performance deep learning library," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [71] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2015.
- [72] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno *et al.*, "FastMRI: an open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.
- [73] M. Tygert and J. Zbontar, "Simulating single-coil MRI from the responses of multiple coils," *Communications in Applied Mathematics and Computational Science*, vol. 15, no. 2, pp. 115–127, 2020.
- [74] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [75] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Medical Image Analysis*, vol. 18, no. 6, pp. 843–856, 2014.
- [76] E. M. Eksioğlu, "Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI," *Journal of Mathematical Imaging and Vision*, vol. 56, no. 3, pp. 430–440, 2016.
- [77] P. Yin, J. Lyu, S. Zhang, S. Osher, Y. Qi, and J. Xin, "Understanding straight-through estimator in training activation quantized neural nets," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.
- [78] G. Hinton, *Neural networks for machine learning, video lectures*. Coursera, 2012.
- [79] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," *arXiv preprint arXiv:1308.3432*, 2013.
- [80] H. Gudbjartsson and S. Patz, "The rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.
- [81] J. Lv, G. Li, X. Tong, W. Chen, J. Huang, C. Wang, and G. Yang, "Transfer learning enhanced generative adversarial networks for multi-channel MRI reconstruction," *Computers in Biology and Medicine*, vol. 134, p. 104504, 2021.
- [82] J. Lv, C. Wang, and G. Yang, "PIC-GAN: A parallel imaging coupled generative adversarial network for accelerated multi-channel MRI reconstruction," *Diagnostics*, vol. 11, no. 1, p. 61, 2021.
- [83] B. Gözcü, T. Sanchez, and V. Cevher, "Rethinking sampling in parallel MRI: A data-driven approach," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019.
- [84] M. V. Zibetti, F. Knoll, and R. R. Regatte, "Alternating learning approach for variational networks and undersampling pattern in parallel MRI applications," *arXiv preprint arXiv:2110.14703*, 2021.