

SS-JIRCS: Self-Supervised Joint Image Reconstruction and Coil Sensitivity Calibration in Parallel MRI without Ground Truth

Weijie Gan* Yuyang Hu* Cihat Eldeniz Jiaming Liu
Yasheng Chen Hongyu An Ulugbek S. Kamilov

Washington University in St. Louis, MO, USA

{weijie.gan, h.yuyang, cihat.eldeniz, jiaming.liu, yasheng.chen, hongyuan, kamilov}@wustl.edu

Abstract

Parallel magnetic resonance imaging (MRI) is a widely-used technique that accelerates data collection by making use of the spatial encoding provided by multiple receiver coils. A key issue in parallel MRI is the estimation of coil sensitivity maps (CSMs) that are used for reconstructing a single high-quality image. This paper addresses this issue by developing SS-JIRCS, a new self-supervised model-based deep-learning (DL) method for image reconstruction that is equipped with automated CSM calibration. Our deep network consists of three types of modules: data-consistency, regularization, and CSM calibration. Unlike traditional supervised DL methods, these modules are directly trained on undersampled and noisy k -space data rather than on fully sampled high-quality ground truth. We present empirical results on simulated data that show the potential of the proposed method for achieving better performance than several baseline methods.

1. Introduction

Magnetic resonance imaging (MRI) is one of the leading diagnostic modalities in radiology. It is well known that the data acquisition in MRI is relatively slow compared to other popular diagnostic modalities such as computed tomography (CT). As a consequence, there has been broad interest in techniques for improving the speed of MRI data acquisition. Parallel MRI (PMRI) is one of the most widely used acceleration strategies [9, 34, 30, 38] and relies on the spatial encoding provided by multiple receiver coils to reduce the amount of data that is acquired. In order to combine the data collected by multiple coils, PMRI requires the calibration of coil sensitivities. Calibration can be performed either in k -space [34, 9] or in the image space using coil sensitivity maps (CSMs) [30, 38]. Compressed sensing (CS)

*These authors contributed equally.

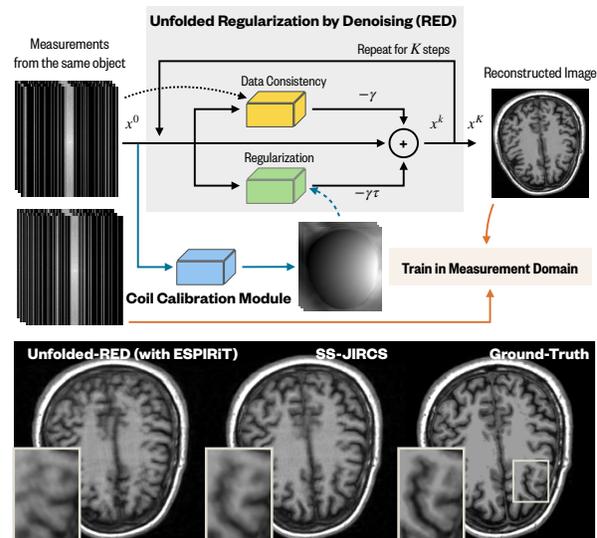


Figure 1. The conceptual illustration of the proposed self-supervised learning framework, SS-JIRCS, that can form images from *uncalibrated* multicoil measurements by leveraging a coil sensitivity calibration module. This visual comparison shows the improvement provided by the proposed method over a baseline method that relies on pre-estimated *coil sensitivity maps (CSMs)*.

MRI is a complementary technique that is used for further accelerating data collection by exploiting prior knowledge (sparsity, low-rankness) on the unknown image [26, 44, 22].

Over the past few years, *deep learning (DL)* has gained popularity for image reconstruction in CS-MRI due to its excellent performance [28, 40, 17]. Recent work has shown the potential of jointly estimating high-quality images and CSMs in an end-to-end manner [4, 15, 36]. However, these methods require fully sampled *ground-truth* data for the *supervised training* of the corresponding deep neural networks, making their application challenging when ground truth is unavailable. On the other hand, there has also been broad interest in developing self-supervised DL methods

that rely exclusively on the information available in the under-sampled measurements [43, 3, 21, 6, 8, 24]. To the best of our knowledge, the potential of self-supervised DL was never investigated in the context of joint image reconstruction and coil sensitivity estimation.

In this paper, we address this gap by proposing a novel self-supervised image reconstruction method, *SS-JIRCS*, for parallel MRI that is equipped with automatic CSM calibration. The key contributions of this work are as follows:

- We develop a model-based DL architecture based on the unfolded *regularization by denoising (RED)* [31] algorithm equipped with a dedicated CSM calibration module. Thus, the full network consists of three complementary modules for (a) data-consistency, (b) regularization, and (c) CSM calibration.
- We develop a self-supervised training strategy inspired from *Artifact2Artifact(A2A)* [24], which is an extension of the well-known *Noise2Noise (N2N)* [21] framework. Our training data corresponds to a set of under-sampled and noisy k-space measurements, without any fully sampled ground truth.
- We present a set of numerical results on simulated MRI data that show the potential improvements due to the proposed method over several baseline methods.

2. Background

2.1. Inverse Problem Formulation

The measurement model in PMRI can be formulated as a linear system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} \quad \text{with} \quad \mathbf{H} = \mathbf{P}\mathbf{F}\mathbf{S}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^n$ is the unknown image, $\mathbf{y} \in \mathbb{C}^{mc}$ is the corresponding multicoil measurement, $\mathbf{F} \in \mathbb{C}^{nc \times nc}$ is the Fourier transform operator, $\mathbf{P} \in \mathbb{C}^{mc \times nc}$ is the k-space sampling operator, and $\mathbf{e} \in \mathbb{C}^{mc}$ is a noise vector. The matrix $\mathbf{S} \in \mathbb{C}^{nc \times n}$ represents the *unknown* sensitivity profiles of the receiver coils [30], where the constant c denotes the total number of coils. Note that \mathbf{S} varies for each scan, since it depends on the interaction of the coils with the anatomy being imaged.

When the matrix \mathbf{S} is known, image reconstruction can be formulated as regularized optimization

$$\arg \min_{\mathbf{x} \in \mathbb{C}^n} g(\mathbf{x}) + h(\mathbf{x}), \quad (2)$$

where g is the data fidelity term that quantifies consistency with the observed data \mathbf{y} and h is a regularizer that encodes prior knowledge on \mathbf{x} . For example, widely used functions

in PMRI for g and h are, respectively, the least-squares and the total variation

$$g(\mathbf{x}) = \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{and} \quad h(\mathbf{x}) = \tau \|\mathbf{D}\mathbf{x}\|_1, \quad (3)$$

where $\tau > 0$ controls the regularization strength and \mathbf{D} is the discrete gradient [33].

Deep learning (DL) has gained popularity over the past few years in MRI reconstruction due to its excellent performance [28, 40, 17, 28]. Traditional DL seeks to train a *deep neural network (DNN)*, such as U-net [32], to learn a regularized inversion of \mathbf{H} by mapping the corrupted images $\{\mathbf{H}_i^\dagger \mathbf{y}_i\}_{i=1}^N$ [20, 41] or the raw measurements $\{\mathbf{y}_i\}_{i=1}^N$ [47, 12] to their desired fully sampled ground-truth $\{\mathbf{x}_i\}_{i=1}^N$ [22]. Here, $N > 0$ refers to the number of training samples, and \mathbf{H}^\dagger is an approximate inversion of \mathbf{H} like the zero-filled inverse Fourier transform.

Model-based DL methods seek to integrate DL with model-based optimization. *Plug-and-play priors (PnP)* [39] and *regularization by denoisers (RED)* [31] denote a family of methods that can leverage pre-trained deep denoisers as imaging priors [2, 29, 24, 35, 23, 7]. *Deep unfolding* is a related class of methods that interpret the iterations of regularized optimization as layers of a deep neural network and trains the resulting architecture in an end-to-end fashion [11, 22, 5, 1, 10, 11]. Different deep unfolding architectures can be obtained by using various optimization/reconstruction algorithms. In this paper, we will rely on an unfolded variant of the gradient-based RED algorithm as the basis of our image reconstruction method.

2.2. Reconstruction using Pre-Calibrated CSMs

There are two widely-used image formation approaches in PMRI (see recent review [17]): (a) those that formulate reconstruction as a k-space interpolation problem [34, 9], and (b) those that seek to explicitly characterize \mathbf{S} as a set of CSMs. In (b), \mathbf{S} is first pre-calibrated and then used to solve the inverse problem (1) [30, 38]. Our work adopts Strategy (b), which will be the focus of the subsequent discussion.

Pre-calibrated CSMs can either be obtained by doing a separate calibration scan [45] or estimated directly from the central k-space regions of a fully sampled scan. The latter approach is adopted in the widely-used ESPIRiT method [38]. There are a number of issues and challenges with the pre-calibration approaches [37, 45]. One issue is that the inconsistencies between the calibration scan and the accelerated scan can result in imaging artifacts. Another issue is that the estimated CSMs may not be sufficiently accurate for high levels of k-space subsampling. Additionally, a calibration scan extends the total scan time.

2.3. Joint Reconstruction and CSM Calibration

Traditionally, optimization-based methods for joint image reconstruction and CSM calibration treat \mathbf{S} as another

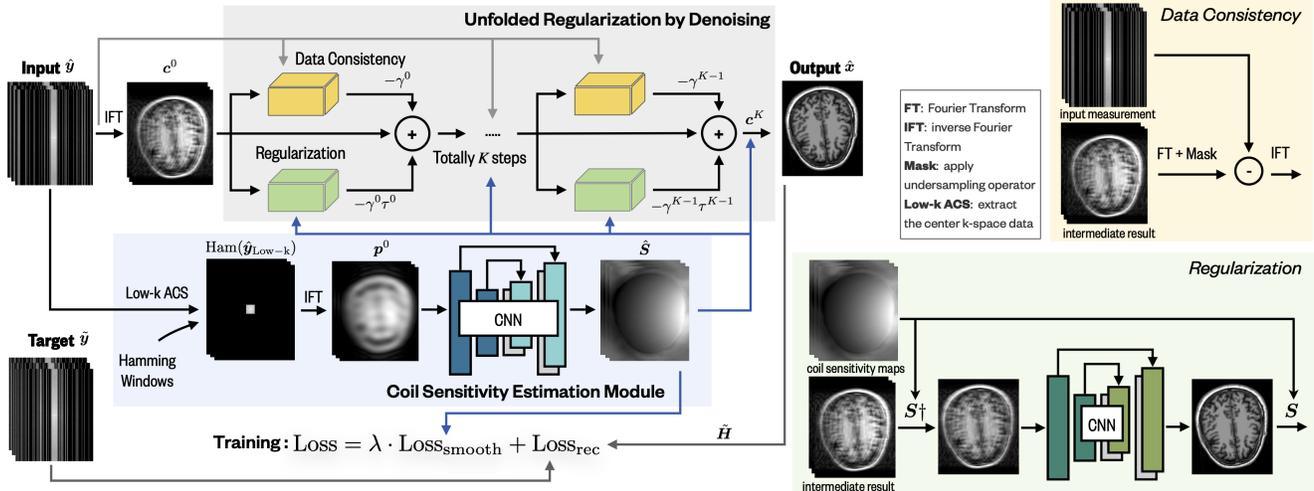


Figure 2. The proposed method consists of an unfolded regularization by denoising (U-RED) reconstruction module and a coil sensitivity calibration module, mapping multicoil undersampled measurements to a single high-quality image and coil sensitivity maps, respectively. The network is trained directly on raw k-space measurements where the input and the target measurement correspond to the same object.

unknown variable in (2) and alternate between updating the image and updating the coil sensitivities [37, 45]. Deep unfolding has recently been proposed to perform joint estimation of image and CSMs without any pre-calibration procedure [4, 15, 36]. The concept behind these methods is to model CSM calibration as a trainable DNN module that can be optimized simultaneously with other learnable parameters in the deep network. The inputs to the CSM calibration modules could be the original undersampled measurements [36] or the intermediate results available at different layers of the deep unfolded networks [4, 15].

Our work contributes to this area by proposing a self-supervised approach for joint reconstruction and CSM calibration that requires no fully-sampled ground truth.

2.4. Self-supervised Image Reconstruction

There is a growing interest in DL-based image reconstruction to reduce the dependence of training on high-quality ground-truth data [3, 21, 18]. Our work is inspired by a widely adopted framework, *Noise2Noise* (N2N) [21], where a DNN R_θ is trained on a set of noisy images $\{\hat{x}_{i,j}\}$ with j indexing different realizations of the same underlying image i [13, 42, 14, 43, 6, 24, 8]. This framework has been investigated in PMRI reconstruction [43, 6, 24]. For instance, *Artifact2Artifact* (A2A) [24] has shown excellent performance on using multiple noisy and artifact-corrupted images $\{\hat{x}_{i,j}\}$ obtained directly from sparsely sampled MR measurements. In A2A, i, j denotes the j th MRI acquisition of Subject i such that each acquisition consists of an independent undersampling pattern and noise realization. The

DNN R_θ can be trained by minimizing a loss function

$$\arg \min_{\theta} \sum_{i,j,j'} \mathcal{L}(\hat{x}_{i,j'}, R_\theta(\hat{x}_{i,j})), \quad (4)$$

where popular choices of \mathcal{L} include the ℓ_1 norm and the ℓ_2 norm. Another related work is [43] that showed the potential of training a *deep unfolded* network without ground-truth by dividing a single acquisition into two subsets and using both subsampled sets of measurements as training targets.

While the concept of N2N enables the training of the DNN for PMRI without any fully sampled data, to the best of our knowledge, the prior work is based on using pre-calibrated CSMs. Our method, to be introduced in the next section, does not require pre-scan calibration, but rather seeks to use the N2N/A2A framework for joint reconstruction and CSM calibration without a ground truth.

3. Method

Our framework takes multicoil undersampled measurements as its input and produces the reconstructed images and CSMs as the output. As illustrated in Fig. 2, the proposed framework consists of two modules: (a) a **CSM estimation module** that uses information extracted from the raw measurements, and (b) an **unfolded MRI reconstruction module** that forms reconstructed images from the input measurements and the estimated CSMs. Our training procedure uses a set of multicoil undersampled measurement pairs $\{(\hat{y}_i, \tilde{y}_i)\}_i^N$ where the measurements in the same pair are acquired from the same object

$$\hat{y}_i = \hat{H}_i x_i + \hat{e}_i \quad \text{and} \quad \tilde{y}_i = \tilde{H}_i x_i + \tilde{e}_i. \quad (5)$$

The number $N \geq 1$ denotes the total number of training samples. The measurements $\hat{\mathbf{y}}_i$ and $\tilde{\mathbf{y}}_i$ can correspond to two subsets extracted from a single acquisition [43] or two separate MRI acquisitions. Note that our training procedure does not require any ground-truth images or known CSMs.

3.1. Coil Sensitivity Estimation Module

Let $\hat{\mathbf{y}}$ be an input measurement and $\hat{\mathbf{P}}$ denote the corresponding sampling matrix. The coil sensitivity estimation module forms CSMs from the *uncalibrated* multicoil measurements by performing three steps: (a) a small central region of k-space is extracted by applying the Hamming window: $\text{Ham}(\hat{\mathbf{y}}_{\text{Low-k}})$, where $\hat{\mathbf{y}}_{\text{Low-k}}$ denotes the central region without the Hamming window; (b) $\text{Ham}(\hat{\mathbf{y}}_{\text{Low-k}})$ is mapped back to the image domain by applying the zero-filled inverse Fourier transform $\mathbf{p}^0 = \mathbf{F}^\dagger \text{Ham}(\hat{\mathbf{y}}_{\text{Low-k}})$; (c) \mathbf{p}^0 is fed into a DNN P_φ with trainable parameters $\varphi \in \mathbb{R}^q$ to obtain estimated CSMs: $\hat{\mathbf{S}} = \text{P}_\varphi(\mathbf{p}^0)$.

We implemented P_φ as a convolutional neural network that operates on the real-valued multichannel data. Since \mathbf{p}^0 is complex valued, we first reshape it by splitting its real and imaginary parts and concatenating the resulting two multicoil data into an image with the number of channels being $2c$. We reformat the output of P_φ into a complex-valued matrix $\hat{\mathbf{S}}$ with the same dimensions as the original \mathbf{p}^0 .

3.2. Unfolded Regularization by Denoising

Our image reconstruction module, based on the *unfolded RED (U-RED)* [25], iteratively refines the image by integrating information from DNN R_θ with learnable parameters $\theta \in \mathbb{R}^p$ and imposing consistency between the predicted and the raw measurements via ∇g . Let $\hat{\mathbf{c}}^0 = \mathbf{F}^\dagger \hat{\mathbf{y}}$ represent the initial image, and $K \geq 1$ be the total number of steps. U-RED can be formulated as

$$\hat{\mathbf{c}}^{k+1} = \hat{\mathbf{c}}^k - \underbrace{\gamma^k (\nabla g(\hat{\mathbf{c}}^k, \mathbf{y}))}_{\text{Data Consistency}} + \underbrace{\tau^k \hat{\mathbf{S}} \text{R}_\theta^k(\hat{\mathbf{S}}^\dagger \hat{\mathbf{c}}^k)}_{\text{Regularization}}, \quad (6)$$

where γ^k and τ^k are learnable parameters, $\hat{\mathbf{c}}^k$ represents the intermediate *multicoil* images in the k th step, and

$$\nabla g(\hat{\mathbf{c}}^k, \mathbf{y}) = \mathbf{F}^\dagger (\hat{\mathbf{P}} \mathbf{F} \hat{\mathbf{c}}^k - \mathbf{y}). \quad (7)$$

The DNN R_θ^k in (6) takes single-coil images as inputs, requiring $\hat{\mathbf{S}}^\dagger$ and $\hat{\mathbf{S}}$ that fuse multiple images into a single image and expand a single image into multiple images, respectively. The final reconstructed image $\hat{\mathbf{x}}$ can be obtained from the output of the last step: $\hat{\mathbf{x}} = \hat{\mathbf{S}}^\dagger \hat{\mathbf{c}}^K$. Note that, unlike the previous unfolded methods [43] that consider pre-calibrated $\hat{\mathbf{S}}$, our method trains a network to calibrate $\hat{\mathbf{S}}$ simultaneously with reconstruction.

3.3. Training Procedure

We use standard stochastic gradient method to jointly optimize $\{\theta_k\}_k^K$ and φ by minimizing a weighted sum loss function

$$\text{Loss} = \text{Loss}_{\text{rec}} + \lambda \cdot \text{Loss}_{\text{smooth}}, \quad (8)$$

where λ is a regularization parameter.

Loss_{rec} seeks to map each $\hat{\mathbf{y}}_i$ and the corresponding $\tilde{\mathbf{y}}_i$ to each other. The key idea here is to map the reconstructed images back to the k-space domain by applying the forward operator of the training target. For example, one can map $\hat{\mathbf{x}}_i$ back to the k-space domain by applying the forward operator $\tilde{\mathbf{H}}_i$ then penalize the discrepancy between the resulting measurements $\tilde{\mathbf{H}}_i \hat{\mathbf{x}}_i$ and raw measurements $\tilde{\mathbf{y}}_i$. Here, the CSMs $\tilde{\mathbf{S}}_i$ in $\tilde{\mathbf{H}}_i$ are estimated by the coil sensitivity estimation module after feeding $\tilde{\mathbf{y}}_i$ as the input. The formulation of Loss_{rec} is

$$\text{Loss}_{\text{rec}} = \frac{1}{N} \sum_i^N \mathcal{L}_{\text{rec}}(\tilde{\mathbf{H}}_i \hat{\mathbf{x}}_i, \tilde{\mathbf{y}}_i) + \mathcal{L}_{\text{rec}}(\hat{\mathbf{H}}_i \tilde{\mathbf{x}}_i, \hat{\mathbf{y}}_i), \quad (9)$$

where $\tilde{\mathbf{x}}_i$ is the reconstructed image when $\tilde{\mathbf{y}}_i$ is the input measurement, and \mathcal{L}_{rec} denotes the ℓ_2 -norm. Note that N2N/A2A can be seen as a special case of (9) that assumes the CSMs are pre-calibrated and not parameterized. During minimization, Loss_{rec} enforces the accuracy between the predicted and the raw measurements, but it can also generate non-smooth CSMs that are not physically realistic and cause overfitting. Therefore, we include $\text{Loss}_{\text{smooth}}$ to impose smoothness on the estimated CSMs

$$\text{Loss}_{\text{smooth}} = \frac{1}{N} \sum_i^N \left\| D \hat{\mathbf{S}}_i \right\|_2^2 + \left\| D \tilde{\mathbf{S}}_i \right\|_2^2. \quad (10)$$

4. Numerical Validation

4.1. Data Preparation

We used the T1-weighted MR brain acquisitions of 60 subjects obtained from the open dataset OASIS-3 [19] as the raw ground-truth for simulating measurements. These 60 subjects were split into 48, 6, and 6 for training, validation, and testing, respectively. For each subject, we extracted the middle 20 to 50 (depending on the shape of the brain) out of the 256 slices on the transverse plane, containing the most relevant regions of the brain. Those 2D slices correspond to the set of $\{\mathbf{x}_i\}$ depicted in (5). We synthesized CSMs by using the SigPy [27]. The pre-defined parameters of this function were the relative radius $r = 1.5$ and the number of coils $\text{nzz} = 8$. In order to obtain undersampled measurements, we simulated Cartesian sampling operator of the k-space data. We set the sampling rate to 12.5%, 16.67% and

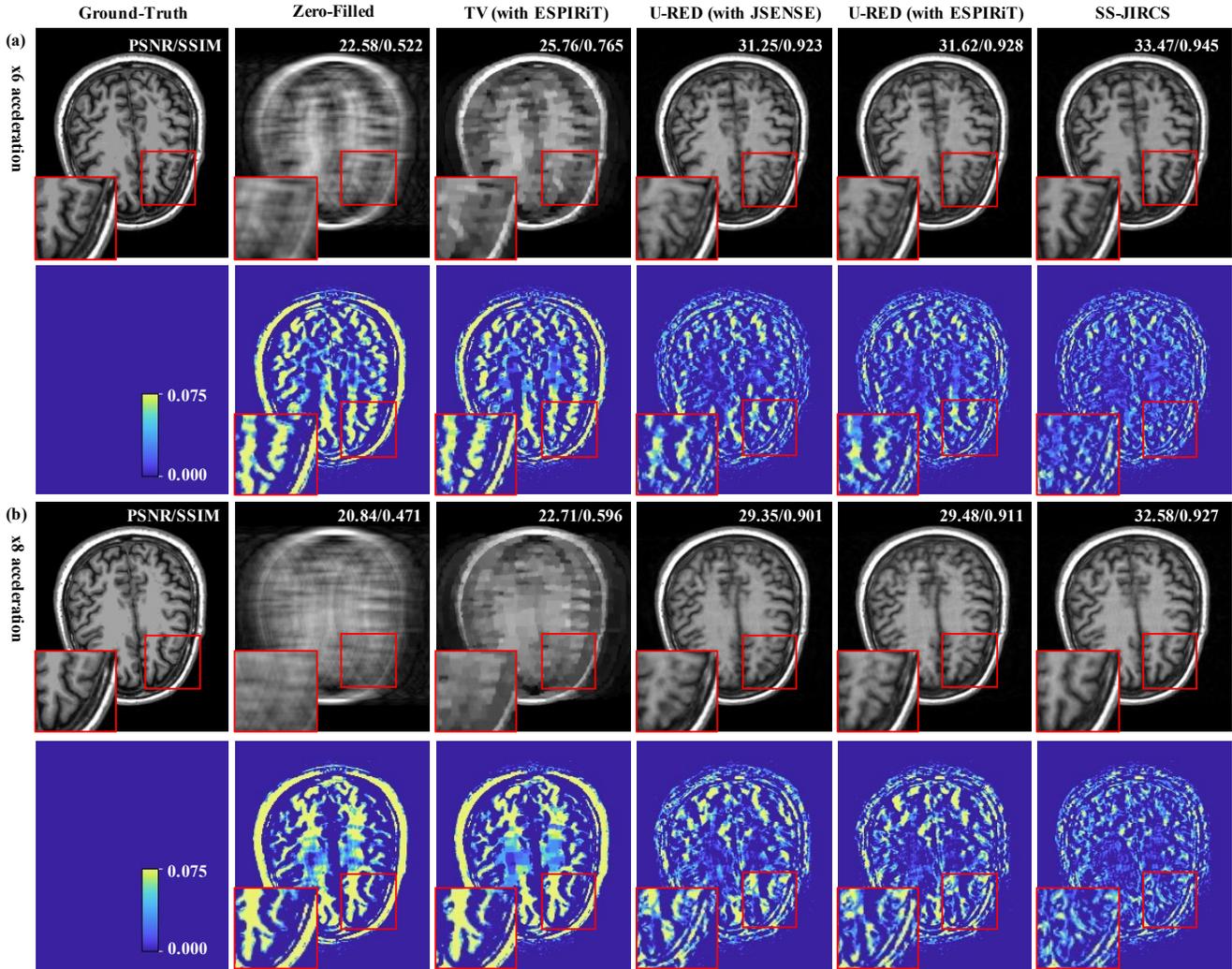


Figure 3. Illustration of image reconstruction from undersampled measurements at two different subsampling rates with an input SNR of 40dB. *TV (with ESPIRiT)* refers to a TV-regularized optimization method that uses CSMs pre-calibrated via ESPIRiT [38]. While both *U-RED (with JSENSE)* and *U-RED (with ESPIRiT)* achieve better performance than *TV (with ESPIRiT)*, the proposed method achieves the best performance compared to the baseline methods by jointly performing image reconstruction and CSM calibration in an end-to-end fashion.

Schemes	PSNR (dB)						SSIM					
	25%		16.67%		12.5%		25%		16.67%		12.5%	
Sampling Rate												
Measurement SNR	30 dB	40 dB										
Zero-Filled	24.17	24.38	22.37	22.56	21.65	21.68	0.520	0.538	0.498	0.509	0.454	0.464
TV (with ESPIRiT)	28.68	28.95	24.90	25.14	23.01	23.21	0.885	0.886	0.701	0.712	0.569	0.573
U-RED (with JSENSE)	34.17	34.45	30.39	30.99	28.89	29.10	0.948	0.957	0.921	0.926	0.883	0.891
U-RED (with ESPIRiT)	34.39	34.75	30.91	31.56	29.04	29.29	0.961	0.962	0.922	0.931	0.875	0.896
SS-JIRCS	35.10	35.51	31.62	32.37	30.56	30.97	0.966	0.966	0.930	0.939	0.918	0.922

Table 1. Average quantitative results in the testing set over several undersampling rates and measurement SNRs. Note that the proposed method jointly performs image reconstruction and CSM estimation, while other methods use pre-acquired CSMs.

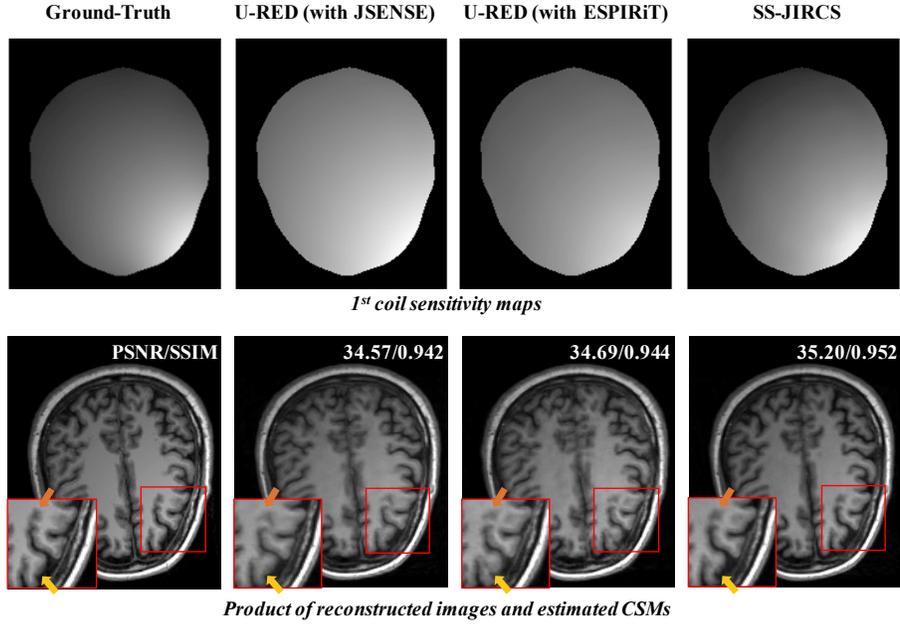


Figure 4. Illustration of the 1st CSM estimated at a sampling rate of 16.67% and a level of measurement noise corresponding to 30dB SNR. In the top row, we show the estimated CSMs. In the bottom row, we show the product of the CSM with the image reconstructed by the corresponding method. We highlighted the visual difference in the regions of interest by using yellow and orange arrows. This figure demonstrates the ability of the proposed method to obtain better CSMs in terms of visual consistency and image quality.

25% of the full sampling rate for the complete k-space data (corresponding to $8\times$, $6\times$ and $4\times$ acceleration, in respective order) and added measurement noise corresponding to input *signal-to-noise ratios* (SNRs) of 30 dB and 40dB.

4.2. Implementation

The DNN architectures for P_φ and $\{R_\theta^k\}$ are customized from U-net [32]. We have experimented with different values of the regularization parameter λ in (8) and the unfolded step K . The best results were obtained when λ is 0.002 and K is 5. We used Adam [16] as the optimizer with its learning rate being 0.001 for the initial 20 epochs and being 0.0001 for the rest epochs. We performed all our experiments on a machine equipped with an Intel Xeon Gold 6130 Processor and an NVIDIA GeForce RTX 2080 Ti GPU.

4.3. Evaluation

We implemented two widely used quantitative metrics, *peak signal-to-noise ratio* (PSNR), measured in decibels (dB) and *structural similarity index* (SSIM), relative to the ground-truth images used to synthesize the measurements.

4.4. Baseline Methods

We experimented with the following baseline methods

- *Zero-Filled*: The raw multicoil and undersampled measurements are directly back-projected to the image domain by applying zero-filled inverse Fourier

transform followed by the root-sum-of-squares operation [46].

- *TV (with ESPIRiT)*: A two-step model-based optimization where the CSMs are pre-calibrated by using ESPIRiT [38] and the total variation (TV) reconstruction method in (3) is applied. The regularization parameter τ is optimized via the grid-search.
- *U-RED (with JSENSE)*: Pre-calibrated CSMs using JSENSE [45] are integrated into the U-RED algorithm.
- *U-RED (with ESPIRiT)*: Similar to *U-RED (with JSENSE)* except that ESPIRiT [38] is used instead of JSENSE to pre-calibrate CSMs. Note that this baseline method is conceptually similar to the method [43], except it uses a different unfolding architecture.

Note that the final two methods are identical to the proposed method, except for the fact that they use CSMs estimated using existing algorithms.

4.5. Results

Fig. 3 illustrates the results of image reconstruction for undersampling rates 16.67% (top) and 12.5% (bottom) at a measurement noise level corresponding to an SNR of 40 dB. *Zero-Filled* images contain blurring and ghosting artifacts, since the raw measurements are directly mapped to the image domain without imposing any regularization. *TV (with*

Schemes	PSNR (dB)						SSIM					
	25%		16.67%		12.5%		25%		16.67%		12.5%	
Sampling Rate	30 dB	40 dB										
Measurement SNR	30 dB	40 dB										
$\hat{\mathbf{y}}_{\text{ACS}}$	34.81	35.15	31.28	31.59	30.08	30.41	0.965	0.965	0.929	0.935	0.910	0.911
$\hat{\mathbf{y}}_{\text{Low-k}}$	34.59	34.88	30.99	31.27	29.79	30.01	0.964	0.964	0.928	0.929	0.907	0.910
$\hat{\mathbf{y}}_{\text{Low-k}} + \text{Loss}_{\text{Smooth}}$	34.38	34.73	30.89	31.55	28.88	29.01	0.961	0.961	0.927	0.928	0.877	0.894
Ham($\hat{\mathbf{y}}_{\text{Low-k}}$)	34.90	35.27	31.35	31.71	30.27	30.66	0.937	0.940	0.930	0.939	0.907	0.915
SS-JIRCS: Ham($\hat{\mathbf{y}}_{\text{Low-k}}$) + Loss_{Smooth}	35.10	35.51	31.62	32.37	30.56	30.97	0.966	0.966	0.930	0.939	0.918	0.922

Table 2. Quantitative results from an ablation study showing the influence of different steps of our CSM estimation pipeline, such as the usage of the low-frequency region of the k-space and the inclusion of the smooth regularization in (10). The table shows the benefit of using both the Hamming-windowed low-frequency region and the smoothness regularization on the final image quality.

ESPIRiT) provides better results, but also leads to the loss of detail due to the well-known “staircasing” artifacts. Deep unfolding methods making use of the pre-calibrated CSMs, U-RED (with *ESPIRiT*) and U-RED (with *JSENSE*), lead to significantly better performance but still suffer from some blurring. Overall, the proposed method achieves the best performance compared to all of the baseline methods. The improvement of the proposed method over U-RED (with *ESPIRiT*), where the only difference is due to the different calibration of the CSMs, demonstrates the benefit of the proposed joint reconstruction and CSMs calibration.

We summarize the quantitative results on the testing set in Table 1. The results show that the performance gains due to the proposed method are maintained across different undersampling rates and noise levels.

Fig. 4 illustrates the CSMs obtained by different methods at the undersampling rate of 16.67% and at a measurement noise level corresponding to 30dB SNR. We highlighted the visual difference in the regions of interest by using yellow and orange arrows. In order to evaluate the quality of the estimated CSMs, we first compute the product of the 1st CSMs (*i.e.*, CSMs from the first channel) obtained using different methods with the corresponding reconstructed images. The resulting single-coil images are then quantitatively evaluated using PSNR and SSIM values relative to the ground-truth images multiplied by the 1st ground-truth CSMs. For instance, let \mathcal{S}_1 be the 1st synthetic CSMs, \mathbf{x} be a slice of the ground-truth volume, $\hat{\mathcal{S}}_1$ be the corresponding estimated CSM, and $\hat{\mathbf{x}}$ be the corresponding reconstructed image. We compared the quantitative metrics of $\hat{\mathcal{S}}_1 \hat{\mathbf{x}}$ in reference to $\mathcal{S}_1 \mathbf{x}$. Fig. 4 demonstrates that the proposed method can estimate better CSMs than several baseline methods in terms of visual consistency and image quality.

Table 2 presents the results of an ablation study showing the influence of different steps of our CSM estimation

pipeline and the effectiveness of $\text{Loss}_{\text{Smooth}}$ in the proposed loss function (8). In Step (a) of the proposed CSM estimation module, we extracted the Hamming-windowed low-frequency region of the k-space. We also experimented with other data extraction schemes in k-space. The vector $\hat{\mathbf{y}}_{\text{ACS}}$ denotes the auto-calibration signal (ACS) region of the k-space and $\hat{\mathbf{y}}_{\text{Low-k}}$ denotes the low-frequency region of the k-space before the application of the Hamming windows. ACS refers to a fully sampled region at the center of k-space collected along with the undersampled data, which contains both low- and high-frequency information. In Table 2, we experimented with various combinations of the extracted k-space data and the inclusion or exclusion of $\text{Loss}_{\text{Smooth}}$. For example, $\hat{\mathbf{y}}_{\text{ACS}}$ corresponds to an experimental setup where we extracted the ACS region of the k-space and excluded $\text{Loss}_{\text{Smooth}}$ from the loss function. $\hat{\mathbf{y}}_{\text{Low-k}}$ is similar to $\hat{\mathbf{y}}_{\text{ACS}}$ except that we extracted the low-frequency region of the k-space. Table 2 shows the benefit of using both the Hamming windowed low-frequency region of the k-space and the smoothness regularization.

5. Conclusion

We propose a model-based deep learning method for jointly reconstructing images calibrating CSMs for parallel MRI. The proposed framework consists of a reconstruction module, based on the unfolded regularization by denoising (U-RED) method, that enforces data consistency and imposes smoothness, and a coil sensitivity estimator that estimates CSMs directly from the raw measurements. We train our network in a self-supervised manner by exclusively using undersampled and uncalibrated measurements without fully sampled ground-truth. Our empirical results show that the proposed method equipped with the joint CSM estimator can lead to better image quality compared to several baseline methods that use pre-calibrated CSMs.

6. Acknowledgments

Research reported in this publication was supported by the NSF CAREER award under CCF-2043134 and the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH).

References

- [1] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Trans. Med. Imaging*, 38(2):394–405, Feb. 2019. 2
- [2] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter. Plug-and-Play Methods for Magnetic Resonance Imaging: Using Denoisers for Image Recovery. *IEEE Signal Process. Mag.*, page 12, 2020. 2
- [3] M. Akçakaya, B. Yaman, H. Chung, and J. C. Ye. Unsupervised Deep Learning Methods for Biological Image Reconstruction. *arXiv:2105.08040 [physics]*, May 2021. 2, 3
- [4] M. Arvinte, S. Vishwanath, A. H. Tewfik, and J. I. Tamir. Deep J-Sense: Accelerated MRI reconstruction via unrolled alternating optimization. *arXiv:2103.02087 [cs, eess]*, Apr. 2021. 1, 3
- [5] J. Duan, J. Schlemper, C. Qin, C. Ouyang, W. Bai, C. Biffi, G. Bello, B. Statton, D. P. O’Regan, and D. Rueckert. VS-Net: Variable Splitting Network for Accelerated Parallel MRI Reconstruction. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, volume 11767, pages 713–722, Cham, 2019. Springer International Publishing. 2
- [6] C. Eldeniz, W. Gan, S. Chen, T. J. Fraum, D. R. Ludwig, Y. Yan, J. Liu, T. Vahle, U. Krishnamurthy, U. S. Kamilov, and H. An. Phase2Phase: Respiratory Motion-Resolved Reconstruction of Free-Breathing Magnetic Resonance Imaging Using Deep Learning Without a Ground Truth for Improved Liver Imaging. *Invest. Radiol.*, Publish Ahead of Print, May 2021. 2, 3
- [7] W. Gan, C. Eldeniz, J. Liu, S. Chen, H. An, and U. S. Kamilov. Image reconstruction for mri using deep cnn priors trained without groundtruth. In *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 475–479, 2020. 2
- [8] W. Gan, Y. Sun, C. Eldeniz, J. Liu, H. An, and U. S. Kamilov. Deep Image Reconstruction Using Unregistered Measurements Without Groundtruth. In *Proc. Int. Symp. Biomedical Imaging*, pages 1531–1534, Nice, France, Apr. 2021. IEEE. 2, 3
- [9] 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). *Magn. Reson. Med.*, 47(6):1202–1210, June 2002. 1, 2
- [10] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*, 79(6):3055–3071, 2018. 2
- [11] K. Hammernik, J. Schlemper, C. Qin, J. Duan, R. M. Summers, and D. Rueckert. Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination. *Magn. Reson. Med.*, page mrm.28827, June 2021. 2
- [12] Y. Han, L. Sunwoo, and J. C. Ye. K-Space Deep Learning for Accelerated MRI. *IEEE Trans. Med. Imaging*, 39(2):377–386, Feb. 2020. 2
- [13] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg. Noise2Inverse: Self-Supervised Deep Convolutional Denoising for Tomography. *IEEE Trans. Comput. Imaging*, 6:1320–1335, 2020. 3
- [14] Z. Jiang, Z. Huang, B. Qiu, X. Meng, Y. You, X. Liu, M. Geng, G. Liu, C. Zhou, K. Yang, A. Maier, Q. Ren, and Y. Lu. Weakly Supervised Deep Learning-Based Optical Coherence Tomography Angiography. *IEEE Trans. Med. Imaging*, 40(2):688–698, Feb. 2021. 3
- [15] Y. Jun, H. Shin, T. Eo, and D. Hwang. Joint deep model-based MR image and coil sensitivity reconstruction network (joint-icnet) for fast MRI. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 5270–5279, 2021. 1, 3
- [16] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, Jan. 2017. 6
- [17] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akçakaya. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal Process. Mag.*, 37(1):128–140, 2020. 1, 2
- [18] A. Krull, T.-O. Buchholz, and F. Jug. Noise2Void - Learning Denoising From Single Noisy Images. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 2124–2132, Long Beach, CA, USA, June 2019. IEEE. 3
- [19] P. J. LaMontagne, T. LS Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko, et al. OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *MedRxiv 2019.12.13.19014902*, 2019. 4
- [20] D. Lee, J. Yoo, and J. C. Ye. Deep residual learning for compressed sensing MRI. In *Proc. Int. Symp. Biomedical Imaging*, pages 15–18, Melbourne, Australia, Apr. 2017. IEEE. 2
- [21] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2Noise: Learning image restoration without clean data. In *Proc. Int. Conf. Mach. Learn.*, 2018. 2, 3
- [22] D. Liang, J. Cheng, Z. Ke, and L. Ying. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Process. Mag.*, 37(1):141–151, Jan. 2020. 1, 2
- [23] J. Liu, M. S. Asif, B. Wohlberg, and U. S. Kamilov. Recovery analysis for plug-and-play priors using the restricted eigenvalue condition. *arXiv:2106.03668 [cs.CV]*, 2021. 2
- [24] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov. RARE: Image Reconstruction Using Deep Priors Learned Without Groundtruth. *IEEE J. Sel. Top. Signal Process.*, 14(6):1088–1099, Oct. 2020. 2, 3
- [25] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov. SGD-Net: Efficient Model-Based Deep Learning

- With Theoretical Guarantees. *IEEE Trans. Comput. Imaging*, 7:598–610, 2021. 4
- [26] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, 2007. 1
- [27] F. Ong and M. Lustig. SigPy: A python package for high performance iterative reconstruction. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 4819, Montréal, QC, 2019. 4
- [28] G. Ongie, A. Jalal, C. A Metzler, R. G Baraniuk, A. G Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory*, 1(1):39–56, 2020. 1, 2
- [29] A. Pour Yazdanpanah, O. Afacan, and S. Warfield. Deep Plug-and-Play Prior for Parallel MRI Reconstruction. In *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, pages 3952–3958, Seoul, Korea (South), Oct. 2019. IEEE. 2
- [30] K. P Pruessmann, M. Weiger, M. B Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42(5):952–962, Nov. 1999. 1, 2
- [31] Y. Romano, M. Elad, and P. Milanfar. The Little Engine That Could: Regularization by Denoising (RED). *SIAM J. Imaging Sci.*, 10(4):1804–1844, Jan. 2017. 2
- [32] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015. 2, 6
- [33] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, Nov. 1992. 2
- [34] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.*, 38(4):591–603, Oct. 1997. 1, 2
- [35] G. Song, Y. Sun, J. Liu, Z. Wang, and U. S. Kamilov. A New Recurrent Plug-and-Play Prior Based on the Multiple Self-Similarity Network. *IEEE Signal Process. Lett.*, 27:451–455, 2020. 2
- [36] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson. End-to-end variational networks for accelerated MRI reconstruction. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, pages 64–73, 2020. 1, 3
- [37] M. Uecker, T. Hohage, K. T. Block, and J. Frahm. Image reconstruction by regularized nonlinear inversion-Joint estimation of coil sensitivities and image content. *Magn. Reson. Med.*, 60(3):674–682, Sept. 2008. 2, 3
- [38] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn. Reson. Med.*, 71(3):990–1001, Mar. 2014. 1, 2, 5, 6
- [39] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. Plug-and-Play priors for model based reconstruction. In *Proc. IEEE Glob. Conf. Signal Process. Inf. Process.*, pages 945–948, Austin, TX, USA, Dec. 2013. IEEE. 2
- [40] G. Wang, J. C. Ye, and B. De Man. Deep learning for tomographic image reconstruction. *Nat. Mach. Intell.*, 2(12):737–748, Dec. 2020. 1, 2
- [41] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *Proc. Int. Symp. Biomedical Imaging*, pages 514–517, Prague, Czech Republic, Apr. 2016. IEEE. 2
- [42] D. Wu, K. Gong, K. Kim, X. Li, and Q. Li. Consensus Neural Network for Medical Imaging Denoising with Only Noisy Training Samples. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, volume 11767, pages 741–749, Cham, 2019. Springer International Publishing. 3
- [43] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Ugurbil, and M. Akçakaya. Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data. In *Proc. Int. Symp. Biomedical Imaging*, pages 921–925, 2020. 2, 3, 4, 6
- [44] J. C. Ye. Compressed sensing MRI: A review from signal processing perspective. *BMC Biomed. Eng.*, 1(1):8, Dec. 2019. 1
- [45] L. Ying and J. Sheng. Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). *Magn. Reson. Med.*, 57(6):1196–1202, June 2007. 2, 3, 6
- [46] 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:1811.08839 [cs.CV]*, 2018. 6
- [47] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, Mar. 2018. 2