A Faithful Deep Sensitivity Estimation for Accelerated Magnetic Resonance Imaging

Zi Wang, Haoming Fang, Chen Qian, Boxuan Shi, Lijun Bao, Liu Hong Zhu, Jianjun Zhou, Wenping Wei, Jiazhong Lin, Di Guo, and Xiaobo Qu, Senior Member, IEEE

Abstract—Magnetic resonance imaging (MRI) is an essential diagnostic tool that suffers from prolonged scan time. To alleviate this limitation, advanced fast MRI technology attracts extensive research interests. Recent deep learning has shown its great potential in improving image quality and reconstruction speed. Faithful coil sensitivity estimation is vital for MRI reconstruction. However, most deep learning methods still rely on pre-estimated sensitivity maps and ignore their inaccuracy, resulting in the significant quality degradation of reconstructed images. In this work, we propose a Joint Deep Sensitivity estimation and Image reconstruction network, called JDSI. During the image artifacts removal, it gradually provides more faithful sensitivity maps with high-frequency information, leading to improved image reconstructions. To understand the behavior of the network, the mutual promotion of sensitivity estimation and image reconstruction is revealed through the visualization of network intermediate results. Results on in vivo datasets and radiologist reader study demonstrate that, for both calibration-based and calibrationless reconstruction, the proposed JDSI achieves the state-of-the-art performance visually and quantitatively, especially when the acceleration factor is high. Additionally, JDSI owns nice robustness to patients and autocalibration signals.

Index Terms—Deep learning, parallel imaging, image reconstruction, magnetic resonance imaging.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is a leading non-invasive and non-radioactive diagnostic modality in modern medical science [1]. However, MRI scans usually require a long acquisition time, and thus fast MRI technology attracts extensive research interests [2]. Parallel imaging [3], [4] is a widely used multi-coil acceleration technique in clinical systems, while sparse sampling [2], [5] provides a complementary scheme for further obtaining a higher acceleration factor (AF).

Over the past two decades, coil correlation modeling has always been a key issue in advanced MRI reconstructions, which can be categorized into explicit and implicit sensitivity estimations. The former typically pre-acquires [3] or pre-estimates (e.g., using ESPiRiT [6]) the sensitivity maps from the autocalibration signal (ACS), and further applies sparse regularizations [2], [7], [8], [9], [10], [11] to coil-combined images to reduce the reconstruction error. Moreover, sensitivity maps could also be regularized using polynomial functions (JSENSE [12]), so as to continuously update the sensitivity estimation with the image reconstruction. For implicit methods, two main strategies include the k-space kernel learning (e.g., GRAPPA [4] and SPIRiT [13]) and structured low-rank (e.g., SAK [14] and P-LORAKS [15]), which can be applied efficiently in the calibrationless reconstruction. Recently, based on their good reconstruction performance and complementarity, the two approaches have been combined for further improvement [16], [17].

At present, deep learning has shown its great potential in fast MRI to improve image quality and reconstruction speed [18], [19], [20], [21]. Like conventional methods, there are also implicit and explicit strategies for deep learning when dealing with multi-coil data. Several implicit methods without sensitivity map estimation utilize multiple convolutional layers to achieve k-space interpolation (e.g., DOTA [22], HDSLR [23], and ODLS [24]) and is robust to the number of ACS lines. Whereas most existing deep learning methods like MoDL [25] involve explicit sensitivity maps pre-estimated using ESPiRiT.
Fig. 1. Patient brain reconstructions under a 1D Cartesian undersampling pattern with AF = 6. First row: (a) Ground truth of 1st coil which is generated via dividing the individual fully sampled coil image by the SoS image of fully sampled multi-coil images. Estimated sensitivity maps using (b) JSENSE [12], (c) ESPiRiT [6], (d) U-Net [31], and (e)–(f) the proposed method. Second row: (a)–(e) Reconstructed images using SENSE-based reconstruction [11] and (f) the proposed joint deep reconstruction. Third row: (a)–(f) Magnified lesions and the corresponding error maps. Note: PSNR(dB) is listed for each reconstruction.

[6], and they can work efficiently under the moderate AF [26], [27], [28], [29].

When the AF is high or ACS lines are limited, the conventional sensitivity estimation will be inaccurate, and the error may affect the image reconstruction and bring significant image degradation [12], [14]. To overcome this problem, one interesting attempt is to estimate sensitivity maps through a deep network (DeepSENSE [30]) but its reconstruction performance is limited by the subsequent SENSE reconstruction. To avoid pre-estimating, another work employs a U-Net as the sensitivity estimator and inserts it at the beginning of the network (E2E-VarNet [31]), but ignores the importance of sensitivity refinements. To solve this problem, some joint image and sensitivity reconstruction networks are proposed to update sensitivity maps with conventional linear operations (Joint-ICNet [32]) or compact kernels (Deep-JSENSE [33]). Their promising results arouse our interest, however, they are hard to obtain faithful sensitivity maps since they only use the ACS region to get low-resolution sensitivity maps and lose the abundant high-frequency information. Thus, we found that they still need to be improved to achieve better network interpretability (Section III-C), lower reconstruction error (Section IV-C), and greater robustness to patients (Section IV-D).

A faithful sensitivity estimation is important for high-quality MRI reconstruction using explicit methods [6], [12], [30]. In this paper, the ground truth of coil sensitivity maps is obtained via dividing the individual fully sampled coil image by the sum of squares (SoS) of fully sampled multi-coil images. With these ground truth sensitivity maps with high-frequency information, Fig. 1(a) illustrates that an almost error-free lesion reconstruction can be achieved. In contrast, the estimated sensitivity maps which are much closer to the ground truth have better SENSE-based reconstructions [11] (Fig. 1(b)–(e)). And our proposed joint deep learning can further suppress artifacts and reconstruct the lesion margin clearly (Fig. 1(f)). This example clearly shows that a better sensitivity map with high-frequency information leads to a lower reconstruction error and better lesion details preservation, no matter for using the conventional SENSE or deep learning reconstructions.

Our joint deep network, called JDSI, is proposed to simultaneously achieve the faithful estimation of sensitivity maps and reconstruction of high-quality images. The main contributions are summarized as follows:

1) The designed network estimates sensitivity maps using all measured k-space and explicitly constrains these maps in the loss function. It provides faithful sensitivity maps with high-frequency information.
2) We explain our network by visualizing the intermediate sensitivity maps and images. It is found that the alternating sensitivity estimation and image artifacts removal promote each other, leading to high-quality reconstructions.
3) Extensive results on in vivo brain datasets show that, for both calibration-based and calibrationless reconstruction, the proposed JDSI provides superior reconstruction performance.
4) The reader study shows the overall image quality of JDSI steps into the excellent level under three experienced radiologists’ evaluations, indicating its reliability in clinical diagnosis even under highly accelerated scanning (AF = 6).

II. PROBLEM FORMULATION

Magnetic resonance image reconstruction problem is to recover a desired image \( \mathbf{x} \in \mathbb{C}^{N} \) from the measured multi-coil k-space \( \mathbf{y} \in \mathbb{C}^{JN} \) which is undersampled by an operator \( \mathcal{U} \) with non-acquired positions zero-filled. The forward model can be
defined as
\[
y = UF Sx + \varepsilon \tag{1}
\]
where \( S = [S_1; \ldots; S_j; \ldots; S_J] \in \mathbb{C}^{JN \times N} \) is the set of sensitivity maps. \( N \) represents the dimension of the vectorized image and \( J \) is the number of coils. \( S_j \in \mathbb{C}^{N \times N} \) is a diagonal matrix which denotes the sensitivity map of the \( j \)-th coil. \( F \) is the Fourier transform, \( \varepsilon \in \mathbb{C}^{JN} \) is the additive Gaussian noise.

For this ill-posed problem in (1), assuming that the sensitivity maps \( S \) are pre-acquired or pre-calculated from the ACS region, it can be solved using the following optimization model with a regularization term \( R_1(\cdot) \), e.g., the \( l_1 \)-norm in image sparsity [2, 7, 8, 9, 10, 11]:
\[
\min_{x} \frac{1}{2} \| y - UF Sx \|_2^2 + R_1(x). \tag{2}
\]

However, the pre-estimated sensitivity maps may not be accurate due to inconsistent pre-scan or limited ACS lines [12]. Then, inaccuracy sensitivity estimation will affect the image reconstruction and result in artifacts [23, 30].

To improve the sensitivity estimation, a joint sensitivity estimation and image reconstruction model is proposed [12]. It simultaneously updates the sensitivity maps and image. After introducing another regularization term of the sensitivity maps \( R_2(\cdot) \), e.g., polynomial regularization [12], (2) is extended as:
\[
\min_{x, S} \frac{1}{2} \| y - UF Sx \|_2^2 + R_1(x) + R_2(S). \tag{3}
\]

To solve the optimization problem in (3), the conventional reconstruction methods update \( S \) and \( x \) alternatively, until the convergence [12]. The reconstruction process is time-consuming due to the usage of an iterative optimization algorithm, while the regularization constraints and their parameters need to be manually determined.

### III. PROPOSED METHOD

In this section, we propose a joint deep network whose architecture is inspired by the alternating iterative idea of the conventional reconstruction. Then, the detailed network implementation is presented.

#### A. JDSI: Joint Deep Sensitivity Estimation and Image Reconstruction Network

Here, we propose a Joint Deep Sensitivity estimation and Image reconstruction network (JDSI) to alternate between updating \( S \) and \( x \), using a deep learning scheme. Their updates correspond to two network modules, i.e., the sensitivity estimation module \( S \) and image reconstruction module \( I \). An initialization module is also integrated at the network beginning to provide proper initial image and sensitivity maps. Besides, to ensure the network reconstructions are aligned with the acquired data, the data consistency module is added. It is noted that all sensitivity maps are normalized to satisfy \( \sum_{j=1}^{J} S_j^* S_j = I \), where \( I \) is an identity matrix and the superscript * represents the adjoint operation.

Fig. 2 gives the network architecture of our proposed JDSI. Specifically, our network firstly estimates initial sensitivity maps from the undersampled multi-coil k-space and provides the initial coil-combined image. Second, the sensitivity estimation module is utilized to refine sensitivity maps from the whole undersampled k-space. Then, the refined sensitivity maps are provided to the image reconstruction module to update the coil-combined image. This iterative process continues until the reconstruction converges.
multi-coil k-space. Third, the undersampling artifacts are removed through the image reconstruction module. Fourth, reconstructed images and sensitivity maps are projected back to the k-space and then forced to maintain the data consistency to the measured k-space. After repeating the last three modules $K$ times, the final reconstructed image can be obtained.

1) Initialization Module: Fig. 2(b) shows the initialization module that aims to initialize sensitivity maps and the coil-combined image. We employ two sub-networks (i.e., $\mathcal{E}$, $\mathcal{D}$), to obtain initial sensitivity maps $S^{(0)}$ from undersampled multi-coil image $x_{IL} = \mathcal{F}^t u^t y$. The overall process is written as

$$S^{(0)} = \mathcal{N}(x_{IL}; \theta_{I}),$$

where $\theta_{I}$ is the network parameters for the mapping $\mathcal{N}$. The superscript $*$ represents the adjoint operation. Two sub-networks are defined as

$$\begin{cases} S_\mathcal{E} = \mathcal{E}(x_{IL}; \theta_\mathcal{E}) \\ S^{(0)} = \mathcal{D}(S_\mathcal{E}, x_{IL}; \theta_\mathcal{D}) \end{cases},$$

where $\theta_\mathcal{E}$, $\theta_\mathcal{D}$ are the network parameters for the mapping $\mathcal{E}$, $\mathcal{D}$, respectively. Through an encoder-decoder network $\mathcal{E}$, the undersampled image $x_{IL}$ is obtained with just low-resolution information, then sensitivity maps $S_\mathcal{E}$ are generated via dividing the individual coil images by the SoS image [31]. Unlike some existing methods [6], [30] that only use the ACS region to get low-resolution sensitivity maps, we exploit the whole acquired k-space [12] to capture more high-frequency information. After that, a denoiser $D$ is designed to remove the residual noises and artifacts, then the initial sensitivity maps $S^{(0)}$ are obtained.

Once the initial sensitivity maps $S^{(0)}$ are generated, the initial coil-combined image can be obtained as $x^{(0)} = (S^{(0)})^* x_{IL}$, which commonly has strong artifacts.

2) Sensitivity Estimation Module: Fig. 2(c) shows the sensitivity estimation module that corresponds to update $S$, which inputs are the fusion of $x^{(k-1)}$ and $S^{(k-1)}$:

$$S^{(k)} = S(S^{(k-1)}, x^{(k-1)}; \theta_S),$$

where $k$ denotes the number of network phase, and $\theta_S$ is the network parameters for the mapping $S$. The mapping is a multi-layer convolutional network that refines the sensitivity maps estimation using the whole k-space rather than only the ACS region.

In this way, more high-frequency information can be retained to further approximate the ground truth sensitivity maps. It prevents the errors of the estimated sensitivities from propagating to the final reconstruction.

3) Image Reconstruction Module: Fig. 2(d) shows the image reconstruction module that corresponds to update $x$. Since this module is designed for artifacts removal, here, we utilize the efficient deep thresholding network [24], [27], [34], [35]:

$$\tilde{x}^{(k)} = \mathcal{I}(x^{(k-1)}; \theta_\mathcal{I}),$$

where $\theta_\mathcal{I}$ is the network parameters for the mapping $\mathcal{I}$. The mapping involves two multi-layer convolutional networks, and an element-wise soft-thresholding $soft(t; \rho) = \max\{|t| - \rho\},$

![Image](Fig. 3. Quantitative comparison of T2 FLAIR brain reconstructions using different network configurations. (a) is the PSNR and training time from different number of network phase. When network phase is set to 6, it consumes the maximum capacity of our GPU (16 GB memory). (b)–(d) (Or (e)–(g)) are results from different number of convolutional layers (or filters) in the initialization module (denoiser part), sensitivity estimation module, and image reconstruction module, respectively. Note: The black arrow denotes the chosen setting in the paper. The 1D Cartesian undersampling pattern with AF = 8 is used. The means are computed over all test data.)
contains 64 filters, followed by batch normalization and ReLU. (3) Image reconstruction module $I$ involves an element-wise soft-thresholding and two 4-layer convolutional networks. Each layer contains 32 filters, followed by batch normalization and ReLU. Besides, the size of all convolutional filters in our network is $3 \times 3$, and the input and output channels of all sub-networks are adjusted according to the number of data coils. Fig. 3(b)–(g) show that our choices are efficient and they provide good reconstruction performance.

The network weights are Xavier initialized and trained for 200 epochs with the Adam optimizer. Their initial learning rate is set to 0.001 with an exponential decay of 0.99. The batch size is 2. The loss function of the proposed JDSI is defined as:

$$
\mathcal{L} = (\mathcal{L}_{\text{coil}} + \gamma \mathcal{L}_{\text{combine}} + \alpha \mathcal{L}_{\text{sensitivity}} = \mathcal{L}_{\text{image}} + \alpha \mathcal{L}_{\text{sensitivity}}
$$

$$
= \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \left( \left\| x_t^{ref} - S_t^{(K)} x_t^{(K)} \right\|_2^2 + \gamma \left\| S_t^{ref} x_t^{ref} - x_t^{(K)} \right\|_2^2 \right) + \alpha \left\| S_t^{ref} - S_t^{(K)} \right\|_2^2 
$$

where $T$ is the number of training samples, $\gamma, \alpha$ are set as 0.1 and 0.1, respectively. $\mathcal{L}_{\text{coil}}, \mathcal{L}_{\text{combine}}, \mathcal{L}_{\text{sensitivity}}$ are the multi-coil image loss, combined-coil image loss, and sensitivity map loss, respectively. The former two can be further defined as image loss $\mathcal{L}_{\text{image}}, x_t^{ref}$ is the fully sampled multi-coil image of the $t$th training sample. $S_t^{ref}$ denote the ground truth sensitivity maps.

The proposed network was implemented in Tensorflow 2.2.0 on a server equipped with dual Intel Xeon Silver 4210 CPUs, 128 GB RAM, and the Nvidia Tesla T4 GPU (16 GB memory). The typical training of JDSI took about 30 hours.

### C. Network Interpretability

To demonstrate the network interpretability of JDSI, we visualized the intermediate sensitivity maps and the corresponding reconstructed images in Fig. 4.

With the increase of the network phase, the alternating sensitivity estimation and image de-aliasing promote each other. The sensitivity maps gradually approximate the ground truth, while image artifacts are gradually removed, leading to a higher PSNR. This enables JDSI to obtain faithful sensitivity maps and benefits high-quality image reconstruction.

In summary, these observations reveal the mutual promotion of sensitivity estimation and image reconstruction. Besides, like iterative methods, JDSI gradually removes image artifacts with the increase of network phases. Different from the pure end-to-end deep learning methods, it provides a good interpretation to understand the behavior of the network.

## IV. EXPERIMENTAL RESULTS

### A. Datasets and Evaluation Criteria

In our paper, T2 FLAIR and T2 weighted brain datasets from fastMRI database [37] were used. They were also widely used in many existing deep learning [31], [32], due to the importance of these two contrasts in clinical diagnosis. Specifically, T2 weighted is commonly used to observe tissue lesions, while T2 FLAIR suppresses the cerebrospinal fluid to more clearly display the lesions nearby [38].

For the T2 FLAIR dataset, we randomly used 32 subjects, where each subject contained about 16 slices of size $320 \times 320$ and coil 16. 26 subjects were used for training and the remaining for test. For the T2 weighted dataset, we randomly used 37 subjects, where each subject contained about 20 slices of size $320 \times 320$ and coil 16. 30 subjects were used for training and the remaining for test. All datasets were fully sampled, and they were retrospectively undersampled for training and test. The number of our method used training and test data was equal to all compared deep learning methods.

To quantitatively evaluate the reconstruction performance, we utilized three evaluation criteria: The relative $l_2$ norm error (RLNE) [9], peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) [39]. The lower RLNE, higher PSNR, and higher SSIM indicate lower reconstruction error, less image distortions, and better details preservation, respectively.

### B. Ablation Studies

#### 1) Initialization Module: It should be especially prudent for a reconstruction method that involves sensitivity maps, and thus requires a reliable estimation as the initial input. Here, we evaluated the importance of the proposed initialization module as mentioned in (5). We compared the proposed method (Full module) with other two schemes: No module and partial module. The former means that sensitivity maps are initialized via dividing the individual zero-filling coil image by the SoS of zero-filling multi-coil images. The latter represents that sensitivity maps are initialized via only $E$ in (5). The network architectures of these two are consistent with JDSI, except for the initialization module.

Fig. 5 shows that both no and partial modules provide inferior sensitivity maps with obvious artifacts, while the proposed full
module has better initial maps and directly leads to a much better reconstruction image in the 1st network phase (i.e., outperforms others by about 2.49—4.33 dB in PSNR). The gains continue until the end, resulting in the best reconstruction of the network with the full module. Consistent quantitative results of more slices are summarized in Table I. These results demonstrate that the proposed module can generate the best initial sensitivity maps. It helps the network reconstruct a proper image from the beginning and benefits the final reconstructions indirectly.

2) Joint Loss Function: Then, we evaluated the importance of the proposed joint loss function, which not only minimizes the error of images, but also explicitly constrains sensitivity maps in the loss function.

Considering the commonly used image loss $L_{\text{image}}$ as the basic loss function. In Table II, we find that the reconstruction performance can be improved by adding an explicit constraint $L_{\text{sensitivity}}$ on sensitivity maps to minimize the error of them (i.e., joint loss function in (10)). Take PSNR as an example, the improvement is 0.36 dB.

These results imply that, involving the sensitivity loss in training is an effective strategy to improve reconstructions.

3) Joint Network Architecture: Finally, we evaluated the importance of the proposed joint network architecture, which simultaneously updates sensitivity maps during image reconstruction.

Consider a network using fixed sensitivity maps as the basic network architecture. In Table III, it is found that by further inserting the sensitivity estimation module as shown in (6), the joint network can dramatically boost the quality of reconstructed images, e.g., PSNR improves by 2.20 dB.

These results show that simultaneous sensitivity update and image reconstruction as proposed is a major source of image quality improvement.

C. Comparison With State-of-the-Art Methods

1) Calibration-Based Reconstruction: To validate the advantages of the proposed JDSI, we first compared it with the state-of-the-art methods under common calibration-based undersampling patterns. These patterns have fully sampled central ACS regions for sensitivity encoding or k-space kernel learning, to model the relationship between multiple coils.

For the comparison study, the conventional parallel imaging method JSENSE [12] was used as a reconstruction baseline (Only supports 1D undersampling). We also compared the state-of-the-art sparse reconstruction method pFISTA-SENSE [11], and two calibration-based deep learning methods including MoDL [25] and Joint-ICNet [32]. JSENSE is a classical joint sensitivity estimation and image reconstruction method. pFISTA-SENSE and MoDL need pre-estimate sensitivity maps using ESPRIT [6]. Joint-ICNet is a joint method but updates sensitivity maps using a conventional linear operation. Parameters of all conventional methods were optimized to obtain the lowest RLNE. The network phase of Joint-ICNet was set as 5 due to GPU memory limitation (16 GB). Other deep learning methods were executed according to typical settings mentioned by the authors.
For 1D undersampling, Fig. 6(a)–(f) show that, when the acceleration factor is low (AF = 4), except JSENSE has obvious artifacts, other compared methods provide the images with nice artifacts suppression while JDSI achieves the best performance visually and quantitatively. However, when the acquisition is highly accelerated (AF = 8), all compared methods own dramatic degradation and yield results exhibiting obvious artifacts, whereas JDSI has the smallest reconstruction error and outperforms them by about 1.39∼10.53 dB in PSNR (Fig. 6(g)–(l)).

For 2D undersampling, Fig. 7(a)–(e) show that JDSI provides the best reconstructions, while compared methods have edge artifacts and details blurring. Consistent quantitative results of more datasets are summarized in Table IV.

These results demonstrate that, for calibration-based reconstruction, JDSI can achieve the best ability of the artifacts suppression and details preservation, under different tested undersampling patterns and acceleration factors.

2) Calibrationless Reconstruction: In general, the need to obtain ACS signals prolongs the overall data acquisition time and sometimes increases the susceptibility to subject motion [14], [15]. Therefore, calibrationless reconstruction has become a research hotspot due to it does not rely on ACS signals (i.e.,...
TABLE IV
RLNE ($\times 10^2$)/PSNR (dB)/SSIM ($\times 10^2$) OF BRAIN RECONSTRUCTIONS UNDER CALIBRATION-BASED UNDERSAMPLING PATTERNS [MEAN$\pm$STD]

| Dataset | Method | RLNE | PSNR | SSIM |
|---------|--------|------|------|------|
| T2 weighted | JSENSE | 12.62$\pm$1.58 | 30.97$\pm$1.14 | 90.25$\pm$1.92 |
| | pFISTA-SENSE | 6.53$\pm$1.48 | 36.93$\pm$1.39 | 96.28$\pm$1.27 |
| | MoDL | 6.52$\pm$1.46 | 36.82$\pm$1.42 | 96.29$\pm$0.97 |
| | Joint-ICNet | 7.20$\pm$1.47 | 35.93$\pm$1.37 | 95.93$\pm$1.46 |
| | JDSI | 6.07$\pm$1.80 | 37.56$\pm$1.86 | 96.39$\pm$1.71 |
| | JSENSE | 9.54$\pm$1.52 | 31.39$\pm$0.98 | 89.40$\pm$2.11 |
| T2 FLAIR | JSENSE | 36.85$\pm$4.45 | 21.66$\pm$1.51 | 73.39$\pm$4.36 |
| | pFISTA-SENSE | 18.07$\pm$4.13 | 27.96$\pm$1.79 | 87.55$\pm$2.42 |
| | MoDL | 15.92$\pm$1.35 | 28.92$\pm$1.36 | 90.54$\pm$1.27 |
| | Joint-ICNet | 16.49$\pm$2.01 | 28.64$\pm$1.82 | 89.35$\pm$2.41 |
| | JDSI | 13.52$\pm$2.25 | 30.42$\pm$1.82 | 91.43$\pm$2.19 |

TABLE V
RLNE ($\times 10^2$)/PSNR (dB)/SSIM ($\times 10^2$) OF BRAIN RECONSTRUCTIONS UNDER CALIBRATIONLESS UNDERSAMPLING PATTERNS [MEAN$\pm$STD]

| Dataset | Method | RLNE | PSNR | SSIM |
|---------|--------|------|------|------|
| T2 FLAIR | P-LORAKS | 9.22$\pm$1.64 | 32.08$\pm$2.24 | 91.38$\pm$1.45 |
| | SAKE | 9.02$\pm$1.90 | 32.33$\pm$2.17 | 91.40$\pm$2.23 |
| | DOTA | 5.02$\pm$0.91 | 37.39$\pm$2.53 | 96.72$\pm$0.91 |
| | HDSLR | 4.95$\pm$0.90 | 37.51$\pm$2.07 | 96.12$\pm$1.08 |
| | JDSI | 4.33$\pm$0.86 | 38.69$\pm$2.16 | 97.15$\pm$0.83 |

Note: The means and standard deviations are computed over all test data. The lowest RLNE, highest PSNR and SSIM values are bold faced.

D. Robustness to Patients
In general, the training datasets are mainly composed of healthy volunteers because they are relatively easy to conduct large-scale acquisition. However, the data of patients are often different from that of the healthy subjects, mainly reflected in the distribution and intensity of pathological tissues. These differences between the training and target data may lead to the performance drop of deep learning methods. Thus, exploring the robustness of the proposed method to patients is essential for clinical applications.

Here, we used a brain dataset of healthy subjects to train all deep learning methods, and then to reconstruct a brain image with white matter lesion. The patient data were obtained according to the information provided in fastMRI [40] which were annotated as non-specific white matter lesion. Fig. 9(a) and (f) show that the proposed JDSI recovered the pathological tissues (Marked with yellow arrows) most like the fully sampled image. Whereas other compared methods do not recover the lesion well, resulting in over-smooth (Fig. 9(b)), structure loss (Fig. 9(c)), intensity loss (Fig. 9(d)), and residual artifacts (Fig. 9(e)).

The observation demonstrates that the proposed JDSI owns some robustness to patients.

E. Robustness to Different ACS Lines
The ACS is crucial for accelerated MRI reconstruction, once the number of ACS lines is limited, the pre-estimated sensitivity maps become inaccurate, leading to the degradation of reconstruction [12], [14]. Thus, refining sensitivity maps during the reconstruction may be important.

Here, to validate the robustness of our joint network to different ACS lines, we compared the proposed JDSI with other two networks which use pre-estimated sensitivity maps from JSENSE [12] and ESPiRiT [6], respectively. The network architectures of these two are consistent with JDSI, except for removing the sensitivity-update-related modules.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Fig. 8. Calibrationless reconstruction results of the T2 FLAIR brain dataset using different methods. (a) (Or (g)) is the fully sampled image and the corresponding undersampling pattern. (b)–(f) (Or (h)–(l)) are reconstructed images and the corresponding error maps, from top to bottom. Note: RLNE/PSNR(dB)/SSIM are listed for each reconstruction.

Fig. 9. Reconstruction results of the T2 FLAIR brain dataset with lesion using different methods. (a) Is the fully sampled image, magnified details, and the corresponding undersampling pattern. (b)–(f) Are reconstructed images, magnified details, and the corresponding error maps, from top to bottom. Note: The white matter lesions are marked with yellow arrows. RLNE/PSNR(dB)/SSIM are listed for each reconstruction.

Fig. 10 shows that, for the tested number of ACS lines, the proposed JDSI consistently outperforms other two methods in terms of three evaluation criteria. Moreover, with a decrease of ACS lines (<16 lines), the accuracy of two pre-estimated sensitivity maps decreases, resulting in significant evaluation criteria degradations. By updating the sensitivity maps in the network, these criteria of the proposed JDSI are only slightly reduced when the ACS lines are reduced. Thus, updating the sensitivity maps provides robustness to ACS lines.

Representative images in Fig. 11(a)–(d) show that, when ACS lines are enough (ACS = 24), all three methods can yield good sensitivity estimations and suppress artifacts well. The sensitivity map of JDSI is closest to ground truth due to sensitivity refinement. However, when the ACS lines are limited (Fig. 11(e)–(h)), JSENSE and ESPRIT provide inferior sensitivity maps, resulting in obvious artifacts in reconstructed images, whereas our JDSI provides consistently superior sensitivity estimation with nice image artifacts suppression.
Fig. 10. Quantitative comparison of reconstructions using different methods with different number of ACS lines of a brain dataset. (a)–(c) Are the means of RLNE, PSNR, SSIM of the reconstructions, respectively. Note: The 1D Cartesian undersampling pattern with $AF = 4$ is used. The difference between each mask is the number of ACS lines. The means are computed over all test slices.

Fig. 11. Representative reconstructed brain images with and without enough ACS lines. (a) (Or (e)) is the fully sampled image, the corresponding undersampling pattern, and the ground truth sensitivity map of $14^{th}$ coil. (b)–(d) (Or (f)–(h)) are reconstruction errors and estimated sensitivity maps, from top to bottom. Note: The 1D Cartesian undersampling patterns are used. RLNE/PSNR(dB)/SSIM are listed for each reconstruction.

It demonstrates the importance of faithful sensitivity estimation for high-quality image reconstructions. Our joint network overcomes the limitation of conventional sensitivity estimation strategies under limited ACS lines, providing high-quality and robust reconstructions.

F. Reader Study

Since objective criteria (i.e., RLNE, PSNR, and SSIM) might not reflect image quality in terms of diagnostically important features, we further invited three radiologists (with 12, 29, and 30 years of clinical experience) to independently evaluate the reconstruction images. They were blind to the reconstruction methods. To avoid rater bias, the order of reconstructed images was presented randomly for each slice. Three clinical-concerned criteria including signal-to-noise ratio, artifacts, and overall image quality were used (Table VI). The reader study was conducted through our cloud platform [21], [41], [42] at https://csrc.xmu.edu.cn/CloudBrain.html.

The images used for the reader study were T2 FLAIR images reconstructed under the 1D Cartesian undersampling pattern with $AF = 6$. They included 10 healthy subjects and 10 patients. The patient data were obtained according to the information provided in fastMRI+ [40] which were annotated as non-specific white matter lesion. To reduce the burden of radiologists’ scoring, each subject contained about central three slices which have sufficient tissue information.

Table VII and Fig. 12 show that the proposed JDSI obtains highest mean scores. It is the only one which all three criteria are slightly over 4, indicating the reconstructed images are suitable for diagnosis. These results are consistent with the superiority of JDSI on the above-mentioned objective criteria. The scores

| Score | Signal-to-noise ratio | Artifacts | Overall image quality |
|-------|-----------------------|-----------|----------------------|
| 5     | Excellent             | None      | Excellent            |
| 4     | Good                  | Minimal   | Good                 |
| 3     | Fair                  | Acceptable| Adequate             |
| 2     | Low                   | Significant| Poor                |
| 1     | Poor                  | Massive   | Non-diagnostic       |

Note: The score of each criterion has a range from 0 to 5 with precision of 0.1.

Fig. 12. Representative images for the reader study. (a)–(d) Are reconstructed images from a healthy subject. (e)–(h) Are reconstructed images from a patient. Note: Each image lists the mean scores of three radiologists in the signal-to-noise ratio/artifacts/overall image quality.

| Method   | Signal-to-noise ratio | Artifacts | Overall image quality |
|----------|-----------------------|-----------|----------------------|
| MoDL     | 3.59±0.30             | 3.65±0.44 | 3.64±0.29            |
| Joint-ICNet | 3.72±0.22             | 3.59±0.27 | 3.70±0.20            |
| HDSL    | 3.51±0.15             | 3.52±0.15 | 3.53±0.15            |
| JDSI     | 4.01±0.19             | 4.01±0.18 | 4.02±0.18            |

Note: Images from 10 healthy subjects and 10 patients are involved in the reader study. The means and standard deviations are computed over all images, respectively. The highest scores are bold faced.
of other compared methods mean their images are good for diagnosis. Besides, the differences between JDSI and other compared methods are statistically significant according to all P-values of the Wilcoxon signed rank test < 0.001.

In summary, the image quality improvements obtained by the proposed JDSI is significant and the overall quality steps into the excellent level for clinical diagnosis.

G. Comparison With the State-of-the-Art E2E-VarNet

Here, we further compared another state-of-the-art deep learning methods, called E2E-VarNet [31]. It employs a U-Net as the sensitivity estimator at the beginning of the network to avoid time-consuming ESPRIT pre-estimation [6], and the end-to-end training manner makes it superior to some methods that rely on pre-estimated sensitivity maps. However, E2E-VarNet only uses the ACS region for sensitivity estimation and loses high-frequency information, while ignores the importance of updating sensitivity maps during the reconstruction process. These problems may degrade final reconstructions, especially when the AF is high.

Differently, our proposed JDSI uses the whole k-space rather than only the ACS region to capture more high-frequency information, while integrates a customized sensitivity estimation module to update sensitivity maps and make them further approximate the ground truth sensitivity maps. Thus, JDSI can simultaneously achieve more faithful sensitivity estimation and high-quality image reconstruction.

Fig. 13(a)–(c) show that, when the AF is low, both two methods can yield good sensitivity estimations and image reconstructions. Besides, JDSI achieves better reconstruction both visually and quantitatively, since its sensitivity map is much closer to ground truth. However, for a higher AF in Fig. 13(d)–(f), E2E-VarNet provides inferior sensitivity map, resulting in image artifacts and the lesion structure loss, whereas our JDSI provides consistently superior sensitivity estimation with nice image artifacts suppression and lesion details preservation.

V. DISCUSSIONS

In this work, we first clearly demonstrate the key role of a faithful coil sensitivity estimation with high-frequency information in MRI reconstruction. Following this idea, we then elaborately design network modules that estimate sensitivity maps using all measured k-space and explicitly constrain these maps in the loss function to obtain better sensitivity maps. Integrating them into a deep image reconstruction network, we successfully observe the mutual promotion of sensitivity estimation and image reconstruction, which gradually leads to a final high-quality image. More importantly, the clear design motivation and network behavior also make our further optimization traceable.

An impressive advantage of the proposed JDSI is its superior performance in both healthy subjects and patients when imaging acceleration is high and/or autocalibration signals are limited. In particular, JDSI can clearly reconstruct the margin and contrast of lesions. This is conductive to better monitoring of subtle changes in lesion structures during follow-up to determine whether the condition is improving (lesion absorption) or deteriorating (lesion expansion). Besides, its reliability is also supported by three experienced radiologists from a clinical diagnostic perspective.

In the future, we anticipate that the following aspects are worth exploring: 1) Test its usability of other anatomies (e.g., knee, spine, and cardiac) and attempt to utilize emerging networks [43], [44] for further optimization or adaptation. 2) Extend the idea of deep coil sensitivity estimation to other MRI applications that rely on accurate coil correlation modeling and require high resolution and acceleration, such as dynamic and quantitative MRI.

VI. CONCLUSION

In this work, we present a Joint Deep Sensitivity estimation and Image reconstruction network (JDSI) for accelerated magnetic resonance imaging (MRI). Our network simultaneously achieves the faithful estimation of sensitivity maps and the reconstruction of high-quality images.

Extensive results on in vivo datasets demonstrate that, for both calibration-based and calibrationless reconstruction, the proposed JDSI provides improved and more robust reconstruction performance than state-of-the-art methods. Besides, the reliability of our JDSI in the diagnostic perspective is validated by three experienced radiologists even under highly accelerated scanning.

In summary, a faithful deep sensitivity estimation is expected to be beneficial for clinical MRI applications, which need high imaging accelerations and/or have limited acquisition of autocalibration signals to improve patient comfort and reduce motion-induced artifacts.

ACKNOWLEDGMENT

The authors thank Yirong Zhou and Jiayu Li for supporting the reader study on the Cloud Brain Imaging platform; Xinlin Zhang and Dan Ruan for the helpful discussions; Drs. Michael Lustig,
Mathews Jacob, Dosik Hwang, Justin P. Haldar, and Anuroop Sriram for sharing their codes online; the editors and reviewers for the valuable comments.

REFERENCES

[1] G. H. Mukesh, A. O’Shea, and R. Weissleder, “Advances in clinical MRI technology,” Sci. Transl. Med., vol. 11, no. 523, 2019, Art. no. eaba2591.

[2] M. Lustig, D. Donoho, and J. M. Pauly, “SparSE: MRI: The application of compressed sensing for rapid MR imaging,” Magn. Reson. Med., vol. 58, no. 6, pp. 1182–1195, 2007.

[3] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: Sensitivity encoding for fast MRI,” Magn. Reson. Med., vol. 42, no. 5, pp. 952–962, 1999.

[4] M. A. Griswold et al., “Generalized autocalibrating partially parallel acquisitions (GRAPPA),” Magn. Reson. Med., vol. 47, no. 6, pp. 1202–1210, 2002.

[5] Z.-P. Liang, “Spatiotemporal imaging with partially separable functions,” in Proc. IEEE 4th Int. Symp. Biomed. Imag., 2007, pp. 988–991.

[6] M. Uecker et al., “ESPIRIT—An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA,” Magn. Reson. Med., vol. 71, no. 3, pp. 990–1001, 2014.

[7] S. Ravishankar and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” IEEE Trans. Med. Imag., vol. 30, no. 5, pp. 1028–1041, May 2011.

[8] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, “A fast wavelet-based reconstruction method for magnetic resonance imaging,” IEEE Trans. Med. Imag., vol. 30, no. 9, pp. 1649–1660, Sep. 2011.

[9] 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,” Med. Image Anal., vol. 18, no. 6, pp. 843–856, 2014.

[10] Y. Liu, Z. Zhan, J.-F. Cai, D. Guo, Z. Chen, and X. Qu, “Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging,” IEEE Trans. Med. Imag., vol. 35, no. 9, pp. 2130–2140, Sep. 2016.

[11] X. Zhang et al., “A guaranteed convergence analysis for the projected fast iterative soft-thresholding algorithm in parallel MRI,” Med. Image Anal., vol. 69, 2021, Art. no. 101987.

[12] L. Ying and J. Sheng, “Joint image reconstruction and sensitivity estimation in SENSE (JSENSE),” Magn. Reson. Med., vol. 57, no. 6, pp. 1196–1202, 2007.

[13] M. Lustig and J. M. Pauly, “SPIRiT: Iterative self-consistent parallel imaging reconstruction based on double-domain generative adversarial network,” IEEE J. Biomed. Health Inform., vol. 26, no. 9, pp. 4371–4377, Sep. 2022.

[14] X. Wang, B. P. Sutton, F. Lam, and Z.-P. Liang, “DeepSENSE: Learning coil sensitivity functions for SENSE reconstruction using deep learning,” Magn. Reson. Med., vol. 87, no. 4, pp. 1894–1902, 2022.

[15] A. Srivastava et al., “End-to-end variational networks for accelerated MRI reconstruction,” in Proc. Med. Image Comput. Comput. Assist. Interv., 2020, pp. 64–73.

[16] 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., 2021, pp. 5266–5275.

[17] M. Arvinte, S. Vishwanath, A. H. Tewfik, and J. I. Tamir, “J-sense deep: Accelerated MRI reconstruction via unrolled alternating optimization,” in Proc. Med. Image Comput. Comput. Assist. Interv., 2021, pp. 350–360.

[18] J. Zhang and B. Ghahem, “ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1828–1837.

[19] Z. Wang et al., “A sparse model-inspired deep thresholding network for exponential signal reconstruction-application in fast biological spectroscopy,” IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 10, pp. 7578–7592, Oct. 2023.

[20] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. Med. Image Comput. Comput. Assist. Interv., 2015, pp. 234–241.

[21] J. Zbontar et al., “fastMRI: An open dataset and benchmarks for accelerated MRI,” 2019, arXiv: 1811.08839.

[22] S. Kiryu et al., “Clinical impact of deep learning reconstruction in MRI,” Radiographics, vol. 43, no. 6, 2023, Art. no. e220133.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[24] R. Zhao et al., “FastMRI++, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data,” Sci. Data, vol. 9, no. 1, pp. 152–157, 2022.

[25] Y. Zhou et al., “XCloud-pFISTA: A medical intelligence cloud for accelerated MRI,” in Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc., 2021, pp. 3290–3292.

[26] Y. Zhou et al., “CloudBrain-ReconAI: An online platform for MRI reconstruction and image quality evaluation,” 2022, arXiv: 2212.01878.

[27] F. Shamshad et al., “Transformers in medical imaging: A survey,” Med. Image Anal., vol. 88, 2023, Art. no. 102802.

[28] H. Chang and J. C. Ye, “Score-based diffusion models for accelerated MRI,” Med. Image Anal., vol. 80, 2022, Art. no. 102479.