

# SpiNet: A deep neural network for Schatten p-norm regularized medical image reconstruction

Aditya Rastogi and Phaneendra Kumar Yalavarthy<sup>a)</sup>

Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India

(Received 28 August 2020; revised 30 December 2020; accepted for publication 19 January 2021; published 22 March 2021)

**Purpose:** To propose a generic deep learning based medical image reconstruction model (named as SpiNet) that can enforce any Schatten p-norm regularization with  $0 < p \leq 2$ , where the  $p$  can be learnt (or fixed) based on the problem at hand.

**Methods:** Model-based deep learning architecture for solving inverse problems consists of two parts, a deep learning based denoiser and an iterative data consistency solver. The former has either L2 norm or L1 norm enforced on it, which are convex and can be easily minimized. This work proposes a method to enforce any  $p$  norm on the noise prior where  $0 < p \leq 2$ . This is achieved by using Majorization–Minimization algorithm, which upper bounds the cost function with a convex function, thus can be easily minimized. The proposed SpiNet has the capability to work for a fixed  $p$  or it can learn  $p$  based on the data. The network was tested for solving the inverse problem of reconstructing magnetic resonance (MR) images from undersampled  $k$  space data and the results were compared with a popular model-based deep learning architecture MoDL which enforces L2 norm along with other compressive sensing-based algorithms. This comparison between MoDL and proposed SpiNet was performed for undersampling rates ( $\mathbf{R}$ ) of 2 $\times$ , 4 $\times$ , 6 $\times$ , 8 $\times$ , 12 $\times$ , 16 $\times$ , and 20 $\times$ . Multiple figures of merit such as PSNR, SSIM, and NRMSE were utilized in this comparison. A two-tailed t test was performed for all undersampling rates and for all metrics for proving the superior performance of proposed SpiNet compared to MoDL. For training and testing, the same dataset that was utilized in MoDL implementation was deployed.

**Results:** The results indicate that for all undersampling rates, the proposed SpiNet shows higher PSNR and SSIM and lower NRMSE than MoDL. However, for low undersampling rates of 2 $\times$  and 4 $\times$ , there is no significant difference in performance of proposed SpiNet and MoDL in terms of PSNR and NRMSE. This can be expected as the learnt  $p$  value is close to 2 (norm enforced by MoDL). For higher undersampling rates  $\geq 6\times$ , SpiNet significantly outperforms MoDL in all metrics with improvement as high as 4 dB in PSNR and 0.5 points in SSIM.

**Conclusion:** As deep learning based medical image reconstruction methods are gaining popularity, the proposed SpiNet provides a generic framework to incorporate Schatten p-norm regularization with  $0 < p \leq 2$  with an added advantage of providing superior performance compared to its counterparts. The proposed SpiNet also has useful addition of Schatten p-norm value in regularization term being automatically chosen based on the available training data. © 2021 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.14744]

Key words: inverse problems, magnetic resonance imaging, medical image reconstruction, regularization, Schatten p-norm

## 1. INTRODUCTION

Magnetic resonance imaging (MRI) is immensely popular in clinical imaging on account of being minimally invasive, versatile in requirement of tissue contrast and capable of effective imaging of both anatomical and physiological changes in tissues (e.g., DCE MRI,<sup>1</sup> ASL,<sup>2</sup> DSC MRI<sup>3</sup> etc.). However, the mechanism of data acquisition makes MRI inherently slow compared to other three-dimensional (3D) imaging techniques like x-ray computed tomography. This problem has been addressed by developments in both hardware for data acquisition as well as faster reconstruction methods. Multicoil MRI techniques (parallel MRI),<sup>4</sup> like SENSE,<sup>5</sup> SMASH,<sup>6</sup> and GRAPPA<sup>7</sup> have been successful in reducing the number of phase encoding steps by using an array of

spatially distributed receiver coils and exploiting the redundancy present in the acquired data. However, such apparatus makes the scanner more costly and may not give the required acceleration in scan time, specially for dynamic MRI.

Hence, attempts have been made to reduce the scan time by using advanced computational algorithms in tandem with or independent to parallel MRI. The core concept behind such algorithms is the theory of compressive sensing, which allows robust recovery of signal from fewer number of measurements.<sup>8</sup> These algorithms work by constructing a loss function between acquired and computed data ( $k$  space in case of MRI) and enforcing a prior such as total variation (TV),<sup>9</sup> discrete wavelet transform (DWT),<sup>10</sup> dictionary-based transform,<sup>11</sup> and other low rank transforms.<sup>12</sup> Algorithms to solve these problems can be divided into variable splitting

algorithms,<sup>13</sup> operator splitting algorithms,<sup>14,15</sup> and nonsplitting algorithms.<sup>16,17</sup> These are typically iterative in nature and hence require high computational time and manual tuning of regularization parameters. Moreover, the priors used might not be optimal for a given application.

Recently, convolutional neural network (CNN)-based techniques have shown promising results in image reconstruction and are also computationally efficient. Moreover, the data driven approach of CNN also has a property of learning the prior tailored to the application at hand. In medical imaging, CNNs have shown promising results in image recovery/enhancement tasks like bandwidth enhancement,<sup>18,19</sup> perfusion map denoising<sup>20</sup> etc. CNNs have also been used to solve linear inverse problems using UNET architectures,<sup>21</sup> GAN,<sup>22</sup> and residual learning based architectures.<sup>23</sup> Such methods are also called discriminative learning methods.

These discussed networks do not incorporate the linear relation between measurements and required anatomical image explicitly. Also, both the linear relation and the prior are learnt implicitly. This subjects the network to find the optimal solution from a larger solution space and leads to more trainable parameters, which in turn demand more training data. Moreover, the implicitly learnt linear relationship can only provide an approximate solution. On the contrary, the traditional iterative techniques explicitly incorporate both the linear relationship and the handcrafted prior, but the prior tends to be suboptimal. Keeping in mind the advantages and limitations of both techniques, Deep Image Prior<sup>24</sup> explored the idea of integrating the merits of both techniques. The authors learnt the prior using discriminative learning and enforced the linear relationship explicitly. The latter is also called as data consistency term. The objective of the network is to learn a mapping  $f_\theta: z \rightarrow x$ , such that it minimizes the following cost function for a fixed  $z$  and  $y$ .

$$J(\theta) = \|\mathbf{A}f_\theta(z) - y\|_p^p \quad (1)$$

Here  $\mathbf{A}: x \rightarrow y$  and  $p = 1$  or  $2$ . The network can be trained for a single  $y$ , and  $z$  can be any fixed random input and  $J(\theta)$  is minimized for a fixed number of iterations. Later,  $x$  can be estimated as  $x = f_{\hat{\theta}}(z)$  where  $\hat{\theta}$  minimizes  $J(\theta)$ . In such a framework, the architecture of the network itself can work as the prior and usually architectures like encoder–decoder works better than skip connection networks like UNet and ResNet. This idea has been used for image denoising, inpainting, and super-resolution, Magnetic particle imaging (MPI)<sup>25</sup> and Deep radon transform.<sup>26</sup> One prominent limitation of this method is that it needs large number of trainable parameters to map from random signal  $z$  to required image  $x$  and the reconstruction is primarily driven by the choice of architecture.

These short comings can be addressed by forming a Tikhonov type functional of Eq. (1) which is a sum of data consistency term and prior. It was demonstrated that Deep Image Prior<sup>24</sup> can be seen as a Tikhonov type functional<sup>27</sup> for inverse problem (and vice versa<sup>25</sup>) for a continuous and differentiable  $f_\theta(\cdot)$ . This separates the prior from the data

consistency term and using operator splitting or variable splitting techniques, the prior and data consistency term can be dealt separately. The prior is implemented using a neural network and the data consistency term is solved analytically. Networks like ISTA-Net<sup>28</sup> used the operator splitting technique and networks like MoDL,<sup>29</sup> ADMM-Net<sup>30</sup> and Refs. [31,32] used variable splitting technique. Enforcing data consistency while training the network improved the accuracy of network with advantage of training samples requirement being in few hundreds. ADMM-Net and ISTA-Net used neural network to learn the transform domain in which the signal is sparse. In ADMM-Net, the prior term is solved using a piece-wise linear shrinkage function with trainable parameters to generalize a Lp norm regularization. On the contrary, ISTA-Net used L1 norm as regularization, which transforms the prior into a soft thresholding problem. Moreover, ISTA-Net enforced orthonormality on the sparsity transform. Instead of learning a transform domain, Refs. [29,31–35]. treated solving for prior term as denoising and substituted it with a neural network based denoiser. These networks iterated between these two parts for a fixed number of iterations to compute the solution. Ref. [29] showed that end to end training with shared weights of denoiser gave better results than using pretrained denoisers<sup>31,32</sup> or learning a different denoiser at each iteration. For its purpose, MoDL enforced L2 norm on the prior distribution and solved the linear relationship using conjugate gradient (CG) algorithm. In MoDL, it was also shown that CG performs better than proximal gradient method as applied in Refs. [33–35]. A limitation of MoDL was that it enforced L2 norm on noise prior, which may not be optimal for problem at hand (e.g., L2 norm is optimal for Gaussian noise and is known to promote smoothness in the reconstructed image).

To overcome this limitation and present a generic framework, this work proposes a novel model-based deep learning architecture which enforces Lp norm ( $0 < p \leq 2$ ) [also known as Schatten p-Norm] on the noise prior, which was named as SpiNet (representing Schatten p-norm regularized inversion based deep neural Network). To solve the data constraint with Schatten p-Norm, Majorization–Minimization (MM) algorithm<sup>36</sup> was deployed. To be specific, for minimization, CG was utilized and we used four MM loops. We trained and tested proposed SpiNet using the same training and testing datasets as utilized in MoDL and experimented with undersampling rates  $\mathbf{R}$  of 2×, 4×, 6×, 8×, 12×, 16×, and 20× with radial golden angle (RGA)<sup>37</sup> and variable density random sampling pattern and by adding white Gaussian noise of level 0.01. We also systematically compared proposed SpiNet results with iterative compressive sensing based recovery techniques using TV and DWT regularization. For comparison of reconstruction results, peak signal-to-noise ratio (PSNR) was utilized as a figure of merit. Another unique feature of proposed SpiNet was that the  $p$  in Schatten p-Norm regularization can be a trainable as well as fixed variable, making it very generic, and demonstrate the same by fixing the value of  $p$  to 1 and thereby enforcing L1 constraint on noise prior. Our experiments showed that for lower

undersampling rates (especially,  $2\times$  &  $4\times$ ), the performance of MoDL and proposed SpiNet are very similar. For higher undersampling rates, proposed SpiNet performs significantly better. We corroborated the last statement using two-tailed t test for unequal variance (Welsh test). In short, the novelty of this work lies in proposing a generic iterative model-based deep learning network (called it as SpiNet), which can enforce any Lp norm (Schatten p-norm) with  $0 < p \leq 2$  regularization for medical image reconstruction, where  $p$  can be a fixed or trainable parameter.

## 2. BACKGROUND

### 2.A. Notations

We used small boldface alphabets such as  $\mathbf{x}$  for vectors and capital boldface alphabets like  $\mathbf{A}$  for matrices. The symbol  $\mathbf{r}$  describes the spatial location in 3D volume in image domain, that is,  $\mathbf{r} \in \{x, y, z\}$  and  $\hat{\mathbf{r}}$  was used to denote location in 3D volume of frequency domain, that is,  $\hat{\mathbf{r}} \in \{k_x, k_y, k_z\}$ . The  $p^{th}$ -norm of any vector  $\mathbf{x}$  was denoted as  $\|\mathbf{x}\|_p$  and it was defined as  $\|\mathbf{x}\|_p = (\sum_n |x_n|^p)^{1/p}$  (Schatten p-norm), where  $x_n$  is the  $n^{th}$  element of vector  $\mathbf{x}$ . The small boldface  $\mathbf{b}(\mathbf{r}, \mathbf{c})$  was utilized to denote fully sampled  $k$  space data, where  $\mathbf{c}$  stand for coil number and small boldface  $\hat{\mathbf{b}}(\hat{\mathbf{r}}, \mathbf{c})$  was utilized to denote under sampled  $k$  space data. Superscript  $H$  was used to denote conjugate transpose of a matrix. If the matrix is real, it is equivalent to transpose of the matrix. Also symbol "\*" was used to define Hadamard product between two matrices (element wise multiplication).

### 2.B. Image reconstruction/inverse problem

Forward model is an estimation of the undersampled  $k$  space data from an anatomical image and it is illustrated in Fig. 1. The steps involved in forward modeling are described below:

1. The first step involves estimation of individual coil images from anatomical data and this can be written as

$$\hat{\mathbf{x}}(\mathbf{r}, \mathbf{c}) = \mathbf{S}(\mathbf{r}, \mathbf{c})\mathbf{x}_c(\mathbf{r}) \tag{2}$$

Here  $\mathbf{x}_c(\mathbf{r})$  is the vector  $\mathbf{x}(\mathbf{r})$  repeated  $c$  times.

2. The  $k$  space acquisition of the parallel coils can be estimated from the individual coil image using two-dimensional (2D) Fourier transform as shown in Eq. (3) using the same rationale as given in the work of Wang et al.<sup>38</sup>

$$\mathbf{b}(\hat{\mathbf{r}}, \mathbf{c}) = F \otimes F \hat{\mathbf{x}}(\mathbf{r}, \mathbf{c}) \tag{3}$$

Where  $F$  is the one-dimensional discrete Fourier transform matrix and  $\otimes$  is the Kronecker product.

3. The undersampled  $k$  space data can then be obtained via Eq. (4) using undersampling mask  $\mathbf{U}(\hat{\mathbf{r}}, \mathbf{c})$ :-

$$\hat{\mathbf{b}}(\hat{\mathbf{r}}, \mathbf{c}) = \mathbf{U}(\hat{\mathbf{r}}, \mathbf{c}) * \mathbf{b}(\hat{\mathbf{r}}, \mathbf{c}) \tag{4}$$

In our experiment, we utilized radial golden angle undersampling mask.<sup>37</sup>

The objective of inverse problem is to estimate the anatomical image  $\mathbf{x}(\mathbf{r})$  from multicoil undersampled  $k$  space data  $\hat{\mathbf{b}}(\hat{\mathbf{r}}, \mathbf{c})$ . The foundation for recovery of signals from incomplete measurements was laid by the earlier works of Candès and Tao<sup>39,40</sup> and Donoho.<sup>41</sup> For successful recovery of anatomical images from undersampled  $k$  space data, three conditions are important. First, the desired signal (anatomical image in our case) should be sparse in some domain  $\Gamma(\cdot)$ . Second, the artifacts produced by the undersampling matrix should be incoherent in  $\Gamma(\cdot)$  and finally, the reconstruction method should enforce both data consistency and sparsity constraint. The incoherence of different undersampling patterns is described in detail in Ref. [10] Anatomical image can be recovered from undersampled  $k$  space data using a cost function as shown in Eq. (5).

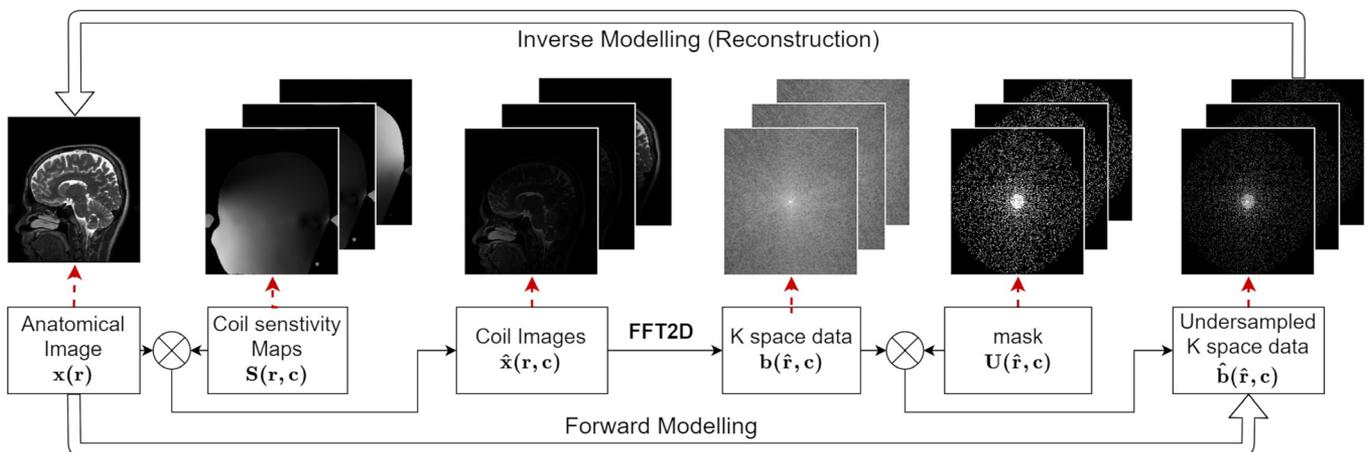


FIG. 1. Cartoon image of main steps involved in forward and inverse model (reconstruction). In forward model, undersampled  $k$  space data were estimated from the anatomical image by first constructing the coil images using coil sensitivity maps. The 2D Fourier transform of coil images gives the fully sampled  $k$  space data, which will then be multiplied with Radial Golden Angle (RGA) undersampling map as described in Ref. [37] to get the undersampled  $k$  space data or  $\hat{\mathbf{b}}(\hat{\mathbf{r}}, \mathbf{c})$ . [Color figure can be viewed at wileyonlinelibrary.com]

$$\tilde{\mathbf{x}}(\mathbf{r}) = \underset{\mathbf{x}(\mathbf{r})}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x}(\mathbf{r}) - \hat{\mathbf{b}}(\hat{\mathbf{r}}, \mathbf{c})\|_2^2 + \lambda \|\Gamma(\mathbf{x}(\mathbf{r}))\|_p^p \quad (5)$$

Here  $\mathbf{A} = \mathbf{U}(\hat{\mathbf{r}}, \mathbf{c}) * F \otimes FS(\mathbf{r}, \mathbf{c}) \mathbf{I}_c$  where  $\mathbf{I}_c$  is a matrix that produces  $c$  copies of  $\mathbf{x}(\mathbf{r})$ .

This cost function (Eq. (5)) can be solved using iterative techniques like proximal gradient,<sup>42</sup> alternating direction method of multipliers (ADMM),<sup>43</sup> iterative shrinkage–thresholding algorithm (ISTA) and its variations (TwISTA, FISTA etc),<sup>14,15</sup> split augmented Lagrangian shrinkage algorithm (SALSA)<sup>13</sup> etc. For solving Eq. (5) using these techniques, commonly used sparsifying transforms are TV<sup>9</sup> and DWT<sup>10</sup> and the sparsity constraint is enforced using  $p = 1$ , that is, L1 norm. For  $p < 1$ , the Eq. (5) is nonconvex and the methods mentioned above are not applicable. However, the nonconvex function  $\|z\|_p^p$  can be bounded by a convex function and solved using MM approach. This method is called FOCUSS and was introduced by Gorodnitsky et al.<sup>36</sup> The steps of FOCUSS are shown below. For a separable constraint  $\|z\|_p^p = \sum_{i=1}^N |z_i|^p$ , let  $y_i = |z_i|^2$ . Then  $h(y_i) = y_i^{p/2}$  is strictly concave in  $0 < p < 2$ . Utilizing Taylor expansion about a constant  $\bar{y}_i$  results in

$$h(y_i) < h'(\bar{y}_i)(y_i - \bar{y}_i) + h(\bar{y}_i) \quad (6)$$

where  $h'(\bar{y}_i) = \frac{p}{2}(\bar{y}_i^{p/2-1})$ . Hence Eq. (6) can be written as

$$h(y_i) < \frac{p}{2}(\bar{y}_i^{p/2-1})(y_i - \bar{y}_i) + h(\bar{y}_i) \quad (7)$$

As  $\bar{y}_i$  is a constant, Eq. (7) thus becomes

$$h(y_i) < \frac{p}{2}(\bar{y}_i^{p/2-1})y_i + C \quad (8)$$

For a separable constraint,  $\Gamma(\cdot)$  such that for any  $\mathbf{x}$ ,  $\Gamma(\mathbf{x}) = \gamma(x_1) \gamma(x_2) \cdots \gamma(x_M)^T$ , the cost function [Eq. (5)] will become

$$J(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2 + \lambda \sum_{i=1}^N g(x_i) \quad (9)$$

Here, the parenthesis were removed for brevity and  $g(x_i) = |\gamma(x_i)|^{2p/2} = h(y_i)$ . Substituting this in Eq. (9) and using Eq. (8) leads to

$$J(\mathbf{x}) \leq \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2 + \lambda \sum_{i=1}^N \frac{p}{2} |\gamma(\bar{x}_i)|^{2(p/2-1)} |\gamma(x_i)|^2 + C \quad (10)$$

for some constant  $\bar{\mathbf{x}}$ . By inserting Eq. (8) in (9), applying upper bound to the cost function  $J(\mathbf{x})$  becomes the majorization step. Sum of all terms consisting of only  $|\bar{x}_i|$  is replaced with  $C$  and as the constant  $C$  is not important for minimization, the solution of Eq. (10) is

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2 + \lambda' \|\mathbf{W}\Gamma(\mathbf{x})\|_2^2 \quad (11)$$

where,  $\lambda' = \lambda \frac{p}{2}$  and  $\mathbf{W} = \operatorname{diag}(|\gamma(\bar{\mathbf{x}})|^{p/2-1})$ . When  $\bar{\mathbf{x}} = \mathbf{x}$  then  $\|\mathbf{W}\Gamma(\mathbf{x})\|_2^2 = \|\Gamma(\mathbf{x})\|_p^p$  and Eq. (11) is same as Eq. (5) for a separable  $\Gamma(\cdot)$ . Solving Eq. (11) is the minimization step. The main steps to solve Eqs. (10) & (11) are given in Algorithm 1.

Algorithm 1 Majorization Minimization (MM) algorithm to estimate  $\tilde{\mathbf{x}}$

---

Require: (INPUT)  $\bar{\mathbf{x}}, \mathbf{A}, \hat{\mathbf{b}}, \Gamma(\cdot)$  and  $\maxIter$   
 $\bar{\mathbf{x}} \leftarrow \mathbf{A}^H \hat{\mathbf{b}}, j \leftarrow 0$   
 while  $j \leq \maxIter$  do  
 $\mathbf{W} = \operatorname{diag}(|\gamma(\bar{\mathbf{x}})|^{p/2-1})$   
 $\tilde{\mathbf{x}}_j = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2 + \lambda' \|\mathbf{W}\Gamma(\mathbf{x})\|_2^2$   
 $\bar{\mathbf{x}} \leftarrow \tilde{\mathbf{x}}_j$   
 $j \leftarrow j + 1$   
 end while  
 $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}}_{j-1}$

---

Recently a new perspective to solve image reconstruction and denoising problem was provided in Ref. [44] by decoupling the data consistency and prior constraint. Substituting the  $\|\Gamma(\cdot)\|_p^p$  with  $q(\cdot)$ , where the latter is the image prior, the decoupled form of Eq. (5) is

$$\tilde{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{z}\|_2^2 + \lambda q(\mathbf{z}) \quad (12)$$

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2 + \alpha \|\mathbf{x} - \tilde{\mathbf{z}}\|_2^2 \quad (13)$$

Here, Eq. (12) is equivalent to denoising  $\mathbf{x}$  with a prior  $q(\cdot)$  and Eq. (13) enforces the data consistency in vicinity of  $\tilde{\mathbf{z}}$ . If one lets  $\Gamma(\mathbf{x}) = \mathcal{N}_w(\mathbf{x})$ , where  $\mathcal{N}_w(\mathbf{x}) = \mathbf{x} - \mathbf{z}$ , and  $p = 2$ , Eq. (5) simplifies into Eq. (13). The effectiveness of noise priors as regularizers was demonstrated in Refs. [45,46]. The denoiser can be an off-the-shelf algorithm like BM3D, which was implemented for MR reconstruction in.<sup>47</sup> Researchers have also used pretrained denoiser<sup>32</sup> or have attempted end to end training along with iterative reconstruction.<sup>29,33,48</sup> Aggarwal et al.<sup>29</sup> recently proposed MoDL and showed that weight sharing among denoisers at different iterations and with end to end learning outperforms pretrained denoisers as well as end to end learning without weight sharing. Moreover, sharing weights leads to reduction in the number of parameters, they demonstrated state-of-the-art recovery with only 360 training samples.<sup>29</sup>

The formulation of MoDL is (Eq. (14))

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \hat{\mathbf{b}}\|_2^2}_{\text{Constraint}} + \lambda \underbrace{\|\mathcal{N}_w(\mathbf{x})\|_2^2}_{\text{prior}} \quad (14)$$

By substituting  $\mathcal{N}_w(\mathbf{x}) = \mathbf{x} - \mathbf{z}$ , the solution of this problem can be written as follows

$$\mathbf{z}^k = \mathcal{D}_w^k(\mathbf{x}^{k-1}) \quad (15)$$

$$\mathbf{x}^k = (\mathbf{A}^H \mathbf{A} + \lambda \mathbf{I})^{-1} (\lambda \mathbf{z}^k + \mathbf{A}^H \hat{\mathbf{b}}) \quad (16)$$

Here,  $\mathcal{D}_w^k(\cdot)$  is a CNN-based denoiser at  $k^{\text{th}}$  iteration. MoDL enforced L2-norm as prior, which may not be the best prior as reconstruction with undersampled data includes aliasing artifact as well and L2-norm is known to promote smoothness in the reconstructed image. To give a more generalized framework, in this work, we propose Schatten  $p$ -norm

constraint (prior/regularization) and corresponding cost function can be written as

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Ax} - \hat{\mathbf{b}}\|_2^2 + \lambda \|\mathcal{N}_w(\mathbf{x})\|_p^p \quad (17)$$

Similar to MoDL, the solution to Eq. (17) can be written as follows

$$\mathbf{z}^k = \mathcal{D}_w^k(\mathbf{x}^{k-1}) \quad (18)$$

$$\mathbf{x}^k = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Ax} - \hat{\mathbf{b}}\|_2^2 + \lambda \|\mathbf{x} - \mathbf{z}^k\|_p^p \quad (19)$$

Similar to above, when  $\Gamma(\mathbf{x}) = \mathcal{N}_w(\mathbf{x})$ , and  $0 < p \leq 2$ , Eq. (5) simplifies into Eq. (19). If the value of  $p = 1$ , it simplifies into traditional L1 norm regularized compressive sensing algorithm. The architecture of the network is explained in Section 3.A. The relationship between SpiNet and established MoDL as well as analytical compressive sensing reconstruction framework has been explained in detail in Appendix A.

### 3. MATERIALS AND METHODS

In this section, we will describe the implementation of proposed SpiNet and Total Variation and Discrete Wavelet Transform based iterative recovery which uses the concepts of compressive sensing (abbreviated as CS-TV). The details of architecture of MoDL are given in Ref. [29].

#### 3.A. Proposed SpiNet

The architecture of proposed SpiNet is given in Fig. 2. The network consists of two blocks, a CNN-based denoiser block (*DW*) for solving Eq. (18) and a data consistency block (*DC*) for solving Eq. (19). The latter was solved using MM algorithm as described in Eqs. (6)–(10). The Schatten p-norm regularized function  $\|\mathbf{x} - \mathbf{z}^k\|_p^p$  is a separable cost function, where  $g(x_i) = |x_i - z_i^k|^{2p/2}$ . Here,  $z^k$  is given by Eq. (18). To solve Eq. (19), we write a cost function using majorization as

$$J(\mathbf{x}) \leq \frac{1}{2} \|\mathbf{Ax} - \hat{\mathbf{b}}\|_2^2 + \lambda' \|\mathbf{W}(\mathbf{x} - \mathbf{z}^k)\|_2^2 \quad (20)$$

Here,  $\lambda' = \lambda p/2$  and  $\mathbf{W} = \operatorname{diag}(|\bar{x}_i - z_i^k|^{p/2-1})$ . Minimizing Eq. (20) w.r.t  $\mathbf{x}$  gives

$$\mathbf{x} = (\mathbf{A}^H \mathbf{A} + \lambda' \mathbf{W}^2)^{-1} (\lambda' \mathbf{W}^2 \mathbf{z}^k + \mathbf{A}^H \hat{\mathbf{b}}) \quad (21)$$

Equation (21) was minimized using CG method for  $M_1$  iterations (i.e., Minimization iterations) to estimate  $\mathbf{x}$  given  $\mathbf{W}$  and  $\mathbf{z}^k$  and was used as  $\bar{x}$  for next Majorization iteration. The Majorization step as shown in Eq. (20) has been repeated  $\hat{N}$  (i.e., number of majorization iterations) times to give the output  $\mathbf{x}^k$ , that is,  $\mathbf{x}^k = \mathbf{x}_{\hat{N}}$ . We repeat Eq. (18) and (19)  $N$  times to get the output of network  $\tilde{\mathbf{x}}$ . The denoiser block was inspired by Ref. [49] and learns a noise residue. It has been implemented using  $N_L$  layers of convolutional filters followed by batch normalization (BN) layers. First  $N_L - 1$  layers of the network have ReLU activation function after BN layer and the input to the block was added to the output of the block. In

the network, the parameters belong to a set  $\Theta$ , such that  $\Theta = \{\lambda^k, \mathcal{D}_w^k, p^k\}_{k=1}^N$  are trainable parameters, where  $N$  is the number of iterations in unrolled network and  $\mathcal{D}_w^k(\cdot)$  is the deep learning based denoiser. As the training parameters are shared among different iterations,  $\Theta = \{\lambda, \mathcal{D}_w, p\}$ . The loss function of the network has been

$$\mathcal{L}(\Theta) = \frac{1}{N_t} \sum_{i=1}^{N_t} \|\tilde{\mathbf{x}}_i - \mathbf{y}_i\|_2^2 \quad (22)$$

where  $N_t$  represents the number of images in training data and  $\{\mathbf{y}_i\}_{i=1}^{N_t}$  are the labels.

#### 3.B. Reconstruction using regularization (CS-TV)

Model-based sparse recovery method was utilized as the standard method and is described in detail in Ref. [50] for anatomical image recovery. We recover the anatomical images by incorporating two priors in Eq. (5) namely anisotropic total variation (*TV*) and two dimensional Debauchy's discrete wavelet transform (*DWT*). The modified optimization algorithm in this case will be

$$\tilde{\mathbf{x}}(\mathbf{r}) = \underset{\mathbf{x}(\mathbf{r})}{\operatorname{argmin}} \|\mathbf{Ax}(\mathbf{r}) - \hat{\mathbf{b}}(\mathbf{r}, \mathbf{c})\|_2^2 + \lambda_1 \|\psi(\mathbf{x}(\mathbf{r}))\|_1 + \lambda_2 \|\mathbf{x}(\mathbf{r})\|_{TV} \quad (23)$$

Here  $\psi(\cdot)$  represents two level Debauchy's wavelet decomposition. The L1 norm in Eq. (23) was relaxed as shown in Ref. [8] Two-dimensional DWT operator can be implemented as a matrix vector multiplication as shown in Ref. [38]. Equation (23) was solved using limited memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS)<sup>51</sup> algorithm. The regularization parameters  $\lambda_1$  and  $\lambda_2$  were chosen empirically. This method will be referred as "CS-TV" from here onwards.

## 4. EXPERIMENTAL DETAILS

#### 4.A. Dataset

The dataset utilized for training and testing of proposed SpiNet is identical to the one used in MoDL<sup>29</sup> for fair comparison. The training dataset was acquired using 3D T2 CUBE sequence with Cartesian readouts using a 12-channel head coil. The training dataset had 90 slices each from four patients and each image dimension was  $256 \times 232$  with 1 mm isotropic resolution. The dimension of training data is  $360 \times 256 \times 232 \times 12$  (slices  $\times$  rows  $\times$  column  $\times$  coils). The 90 slices selected for training had noticeable skull anatomy for training. The testing was performed on the dataset of a fifth patient and the dimensions of testing data was  $164 \times 256 \times 232 \times 12$  (slices  $\times$  rows  $\times$  column  $\times$  coils). For testing also, only those slices were selected which had skull anatomy present. The coil sensitivity maps were generated using ESPIRiT<sup>52</sup> and the undersampling maps were generated using RGA sampling. For training and testing, we retrospectively undersampled the  $k$  space data. Different randomly generated realizations of undersampling pattern were utilized for different slices, however, same pattern was used for all 12

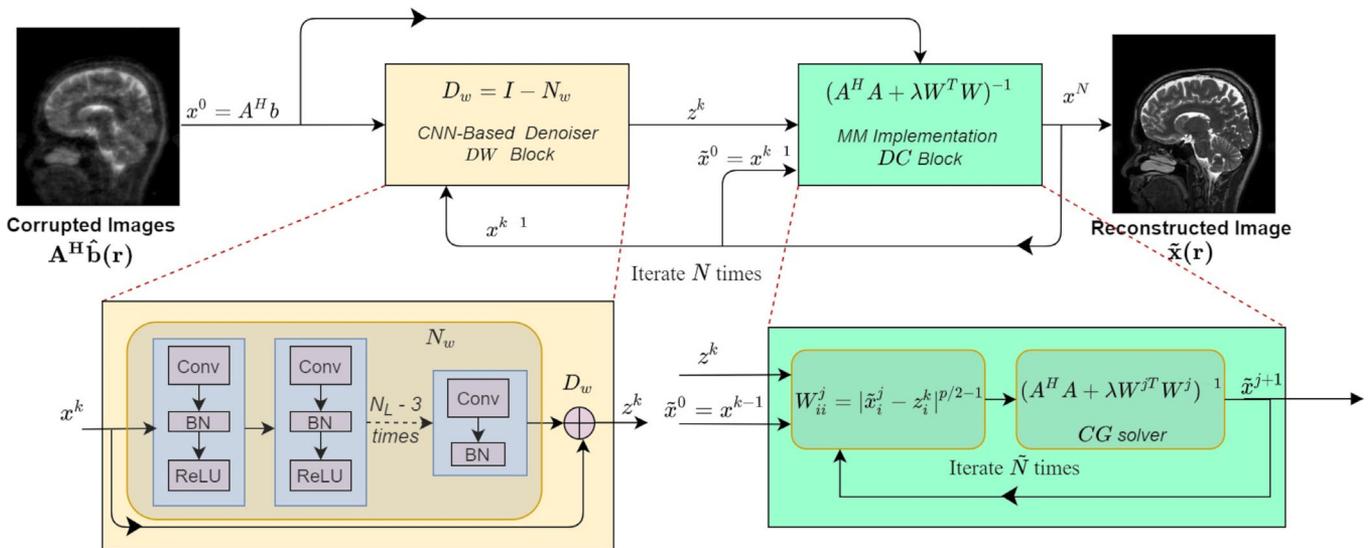


FIG. 2. Network architecture of proposed SpiNet in rolled fashion. Each iteration of the network consists of two blocks, namely denoiser block (DW) and data consistency block (DC). [Color figure can be viewed at wileyonlinelibrary.com]

coils data of same slice. For testing the sensitivity to noise, we added white Gaussian noise (WGN) with standard deviation being 0.01 (1% noise) to the undersampled  $k$  space data.

To analyze the generalizability of SpiNet on different organs or image sequence, we trained and tested SpiNet on Chest MRI data and T1-weighted dynamic contrast-enhanced (DCE) MR data of breast cancer patients. The chest MRI data consist of 100 training and 50 testing images and undersampling rate of 6x was used. More details of this dataset are given in Ref. [53]. For DCE MRI data of the breast consisted of 3000 training images from five patients and testing data of 932 images from sixth patients and undersampling rate of 20x was used. More details on this dataset can be found in Ref. [53] Respective state of the art (SOTA) methods used for chest MRI data and DCE breast cancer data are ADMM-Net<sup>53</sup> and ISTA – Net<sup>28</sup>.

### 4.B. Implementation

The training and testing dataset were retrospectively undersampled with undersampling rate  $\mathbf{R}$  of 2x, 4x, 8x, 12x, 16x and 20x using RGA undersampling scheme. The implementation details of proposed SpiNet and MoDL were given below.

#### 4.B.1. Proposed SpiNet

For training proposed SpiNet, the training dataset consisted of 360 anatomical images of brain from four patients and the testing dataset consisted of 164 images from a single patient. The dataset is described in Section 4.A in more detail. The number of iterations ( $N$ ) were ten for all undersampling rates  $\mathbf{R}$ . The number of layers ( $N_L$ ) were kept as five in all cases and the filter size and number of filters are given in Table I. The real and complex component of the  $A^H \hat{\mathbf{b}}$  were taken as separate channels as input to the first layer of neural network. White Gaussian noise (WGN) of level

TABLE I. Size of filters  $N_x \times N_y$ , number of input channels  $N_c$  and the number of filters  $N_f$  at every layer in DW block of  $k^{th}$  iteration for the proposed SpiNet as shown in Fig. 2.

Layer no.	Size of filter ( $N_x \times N_y \times N_c \times N_f$ )
1	$3 \times 3 \times 2 \times 64$
2 - 4	$3 \times 3 \times 64 \times 64$
5	$3 \times 3 \times 64 \times 2$

These hyperparameters were same  $\forall k$  as weights were shared.

0.01 was added to training and testing data and RGA undersampling mask was utilized. All trainable parameters were shared among the ten iterations. For ease of convergence, we followed the training strategy of MoDL and trained the network initially for one iteration, that is,  $N = 1$  for 100 epochs and then used the learnt weights for initializing a network with ten iterations, which was again trained for 100 epochs. ADAM<sup>30</sup> optimizer was deployed for optimization as this scheme maintains two learning rates corresponding to each parameter. These learning rates were estimated from the first and second moments of gradients. The learning rate was  $10^{-3}$  and other parameters for ADAM were  $\beta_1 = 0.9, \beta_2 = 0.999$  and  $\epsilon = 10^{-8}$ . The number of majorization iterations  $\hat{N}$  was kept as four and number of CG iterations  $M_1$  was also kept as four. The value of  $p$  was initialized as 0.9 for training. Other initializations ranging of  $p \in [0.7, 1.2]$  were also tested, but no major difference in or learnt  $p$  value was noticed. The total number of trainable parameters were 113,922.

#### 4.B.2. MoDL

For training MoDL network, the training and testing dataset were same as proposed SpiNet. The number of iterations  $N$ , number of layers  $N_L$ , filter size and number of filters were

kept same as SpiNet. For Solving Eq. (16), CG method was used till the residue was  $< 10^{-5}$ . The codes for MoDL implementation were borrowed from <https://github.com/hkaggarwal/modl> and more details of MoDL implementation are present in Ref. [29].

#### 4.C. Figure of merit

For comparing reconstruction results using methods discussed in this work, PSNR, normalized root mean square error (NRMSE), and structural similarity index metric (SSIM)<sup>55</sup> were utilized as choice of metric. PSNR (in dB) =  $20\log_{10}\frac{\max}{\text{mse}}$ , where mse is the mean square error between reconstructed image and target image. NRMSE between estimate  $\hat{x}$  and target  $x$  is given by  $\text{NRMSE} = \hat{x}$  and target  $x$  is given by  $\text{NRMSE} = \frac{\text{rmse}(x,\hat{x})}{\max\{x\}-\min\{x\}}$ . The PSNR  $\in (0, \infty)$  (measured in dB), is the measure of signal compare to noise present in the reconstructed image, and higher value of PSNR represents better reconstruction. SSIM metric  $\in (0, 1]$  and it takes value 1 when both images are same. The exponents for the luminance, contrast, and structural terms were set to 1 and  $\sigma$  for Gaussian weighing function was set to 1.5. NRMSE  $\in [0, \infty)$  and the value is 0 when both images are exactly same.

#### 4.D. Statistical test

We performed two-tailed t test for unequal variance (also known as Welch test) for different undersampling rates. The null hypothesis was that MoDL and proposed SpiNet have same mean performance in terms of a given metric. Welch test was conducted for averaged PSNR, NRMSE, and SSIM for  $\mathbf{R} = 2\times, 4\times, 6\times, 8\times, 12\times, 16\times, \text{ and } 20\times$  using the test dataset of 164 images. Significance level  $\alpha = 0.05$  was selected for testing. The results are shown in Table II for PSNR, Table III for NRMSE and Table IV for SSIM, where a  $P < 0.05$  means that there is significant difference in the performance of these two methods. Here  $P$  denotes the probability that the difference in the performance of these two methods happened by chance.

TABLE II. Results of two-tailed Welch test performed on averaged PSNR values (in dB) of MoDL and proposed SpiNet for different undersampling rates ( $\mathbf{R}$ ).

R	MoDL (PSNR)	SpiNet (PSNR)	Better	Significant	$P$ -value
2 $\times$	42.41	42.53	SpiNet	No	0.34
4 $\times$	40.83	40.97	SpiNet	No	0.26
6 $\times$	39.01	39.96	SpiNet	Yes	$1.41e^{-11}$
8 $\times$	38.59	38.84	SpiNet	Yes	0.03
12 $\times$	36.44	37.31	SpiNet	Yes	$7.40e^{-08}$
16 $\times$	34.49	36.25	SpiNet	Yes	$2.61e^{-29}$
20 $\times$	32.09	35.39	SpiNet	Yes	$8.34e^{-25}$

The last column denotes the  $P$  value for the statistical test, when its value is less than 0.05, then we dismiss the null hypothesis that both methods have same average PSNR.

TABLE III. Results of two-tailed Welch test performed on NRMSE values (in %) of MoDL and proposed SpiNet for different undersampling rates ( $\mathbf{R}$ ).

R	MoDL (%)	SpiNet (%)	Better	Significant	$P$ -value
2 $\times$	0.77	0.76	SpiNet	No	0.74
4 $\times$	0.93	0.92	SpiNet	No	0.70
6 $\times$	1.11	1.04	SpiNet	Yes	$1.55e^{-4}$
8 $\times$	1.24	1.15	SpiNet	Yes	$3.82e^{-3}$
12 $\times$	1.70	1.44	SpiNet	Yes	$1.26e^{-15}$
16 $\times$	1.93	1.65	SpiNet	Yes	$4.07e^{-14}$
20 $\times$	2.61	1.84	SpiNet	Yes	$3.34e^{-44}$

The last column denotes the  $P$  value for the statistical test, when its value is less than 0.05, then we dismiss the null hypothesis that both methods have same average NRMSE.

TABLE IV. Results of two-tailed Welch test performed on SSIM values of MoDL and proposed SpiNet for different undersampling rates ( $\mathbf{R}$ ).

R	MoDL	SpiNet	Better	Significant	$P$ -value
2 $\times$	0.98	0.99	SpiNet	Yes	$4.96e^{-22}$
4 $\times$	0.97	0.98	SpiNet	Yes	$6.30e^{-27}$
6 $\times$	0.90	0.98	SpiNet	Yes	$7.56e^{-85}$
8 $\times$	0.96	0.97	SpiNet	Yes	$5.10e^{-31}$
12 $\times$	0.93	0.96	SpiNet	Yes	$2.50e^{-38}$
16 $\times$	0.90	0.95	SpiNet	Yes	$1.09e^{-74}$
20 $\times$	0.89	0.94	SpiNet	Yes	$5.63e^{-54}$

The last column denotes the  $P$  value for the statistical test, when its value is less than 0.05, then we dismiss the null hypothesis that both methods have same average SSIM.

#### 4.D.1. Computational implementation

All computations were carried out on a Linux workstation with Intel i9 processor, 2.10 GHz clock speed, having 128 GB RAM and two Quadro RTX 8000 GPUs with 48 GB memory each. Preprocessing and postprocessing steps were performed in MATLAB and both neural networks were implemented in Tensorflow v1.10. The proposed SpiNet is available as open-source at <https://github.com/adityarastogi2k12/SpiNet>.

## 5. RESULTS

### 5.A. Results of specific image

We compare the performance of MoDL, proposed SpiNet, and CS-TV on a single test image in this subsection. The iteration wise comparison between them has been presented in Section 5.B, where proposed SpiNet was shown to reconstruct high quality image in lesser number of iterations.

The performance of MoDL and proposed SpiNet was compared in Fig. 3 for undersampling rates  $\mathbf{R}$  of 4 $\times$ , 8 $\times$ , 12 $\times$ , 16 $\times$ , and 20 $\times$ . Figure 3(i) shows the results in the whole brain region and the PSNR values are given in parenthesis. Figure 3(ii) shows the performance of MoDL and proposed

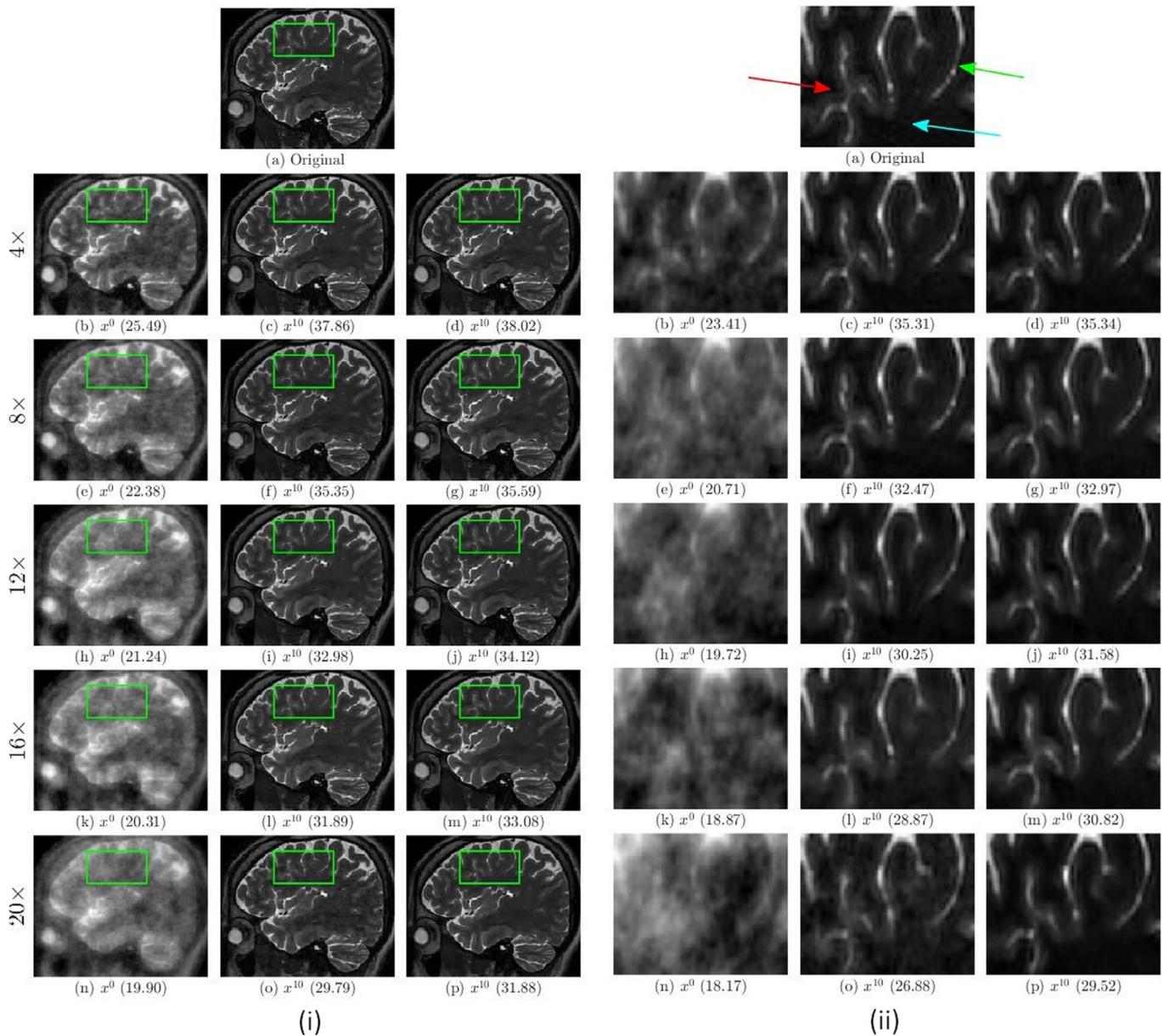


FIG. 3. Comparison of performance of MoDL and proposed SpiNet in the (i) whole brain region and (ii) in region of interest (zoomed version of green bounding box in (i)). In both subfigures (i) and (ii), the top row represents the original/target image from fully sampled  $k$  space. The undersampling rates ( $\mathbf{R}$ ) were mentioned on the left edge of (i). The first column from second row onwards represents the aliased image estimated from zero filled  $k$  space and is input to both MoDL and proposed SpiNet. The second column of (i) and (ii) represents output of MoDL after  $10^{th}$  iteration and the third column of (i) and (ii) represents output of  $10^{th}$  iteration of proposed SpiNet. The PSNR values are shown in parenthesis. It can be seen that for lower  $\mathbf{R}$  the performance of MoDL and proposed SpiNet is similar but as  $\mathbf{R}$  increases the gap between performance of proposed SpiNet and MoDL increases. Such regions are marked by red, green, and cyan arrows in original of (ii). The significance of these arrows is explained in Section 5.A. Readers are advised to see the image in digital copy in full brightness of the screen. [Color figure can be viewed at wileyonlinelibrary.com]

SpiNet in region bounded by green box in Fig. 3(i). From both subfigures, we can observe that for low undersampling rate (namely  $4\times$ ) the performance of MoDL and proposed SpiNet is very close to each other and the difference is mainly due to the statistical nature of white Gaussian noise of level 0.01 added to the undersampled  $k$  space. However, as the undersampling rate increases, the difference in performance between both networks increases in which proposed SpiNet performs better than MoDL in the whole image as well as the bounded region. For  $16\times$  and  $20\times$  undersampling, there is

performance difference of  $\approx 2$  dB between MoDL and proposed SpiNet in both whole image and bounded region. Also, the cloud like artifacts in the cortical region of brain (dark/hypointense region marked by cyan colored arrow) are more visible in image estimated by MoDL compared to proposed SpiNet result. Moreover, hyperintensive regions marked by red and green arrows are better reconstructed by SpiNet as compared to MoDL. T2-weighted MR images show hyperintensive signals for structures containing high water content like CSF, edema, or blood. The arrows marked by red and

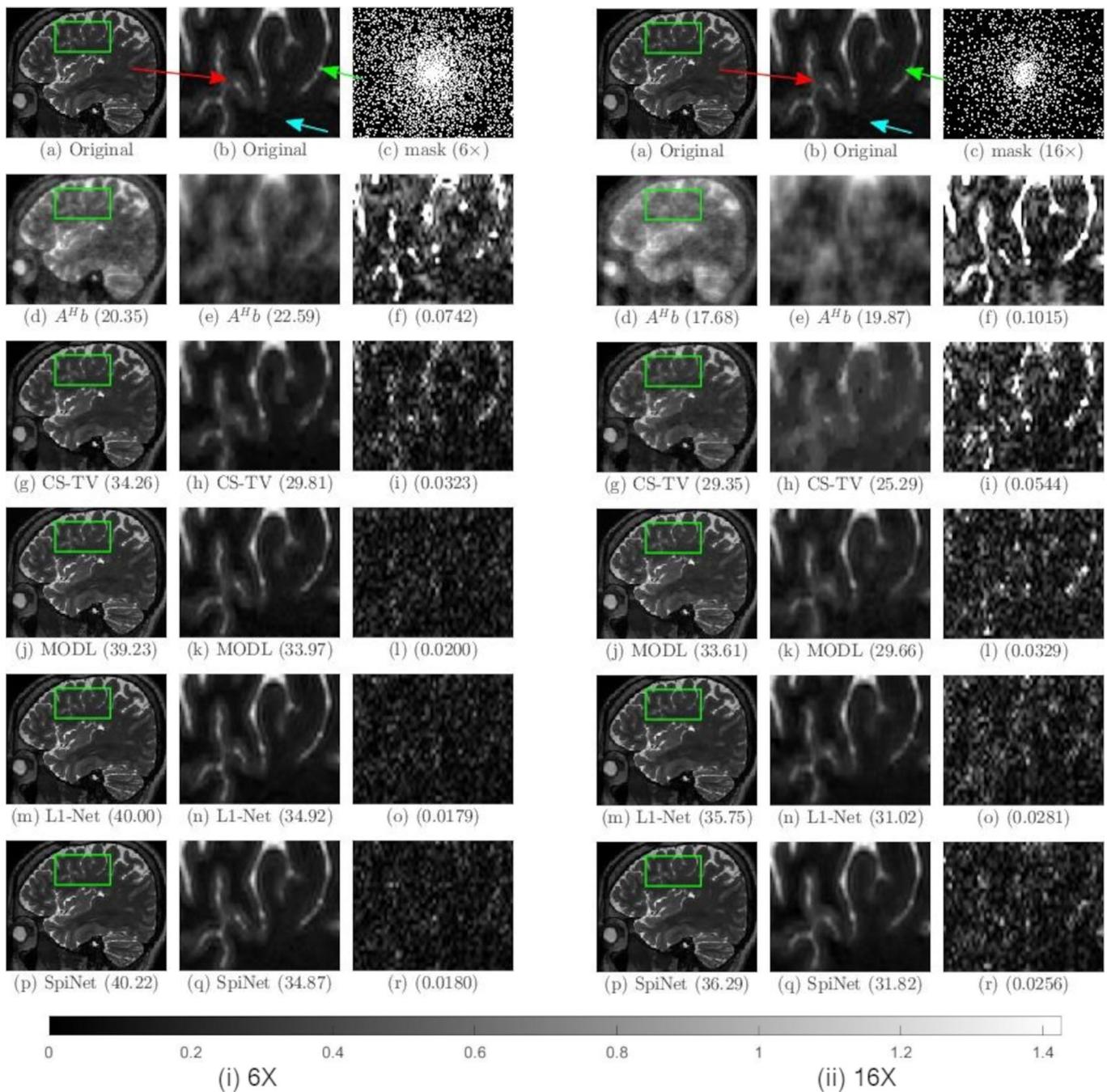


FIG. 4. Comparison of performance of all discussed methods in this work, CS-TV in third row, MoDL in fourth row, L1-Net results in fifth row, and proposed SpiNet results in sixth row, respectively, for (i) 6x and (ii) 16x undersampling rates using variable-density Cartesian random sampling pattern given in (c). The second row in both (i) and (ii) represents the image estimated from zero filled  $k$  space, which is also the input to the MoDL, L1-Net and proposed SpiNet. The L1-Net was implemented using SpiNet in which the value of  $p$  was kept constant and equal to 1. The third column in both (i) and (ii) from second row onward represents four times the difference between the estimated image and the original image in ROI (green bounding boxed image). The parenthesis values in first two columns of (i) and (ii) represents the PSNR and third column numbers represent the root mean square error. Areas where MoDL is not able to properly reconstruct the features or has artifacts are marked by red, green, and cyan arrows in (b) of (i) and (ii). The significance of these arrows is explained in Section 5.A. Readers are advised to see the image in digital copy in full brightness of the screen. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

green color marks the fissures and sulci which contains CSF. Pathologies like CSF leak can cause distortion of these structures. The shape of sulci is also correlated with aging and cognitive decline.<sup>56,57</sup> Moreover, pathological processes, such as demyelination or inflammation, often increases water content in tissues, which increases signal strength in T2-weighted

MR images. The cyan arrow represents lipids, which appear dark in white matter. Cloud like artifacts in that region can be mistaken with white matter diseases which shows as bright regions in white matter. Additionally, such artifacts can effect postprocessing on MR images like segmentation/classification (artifact can act as adversarial noise), quantitative

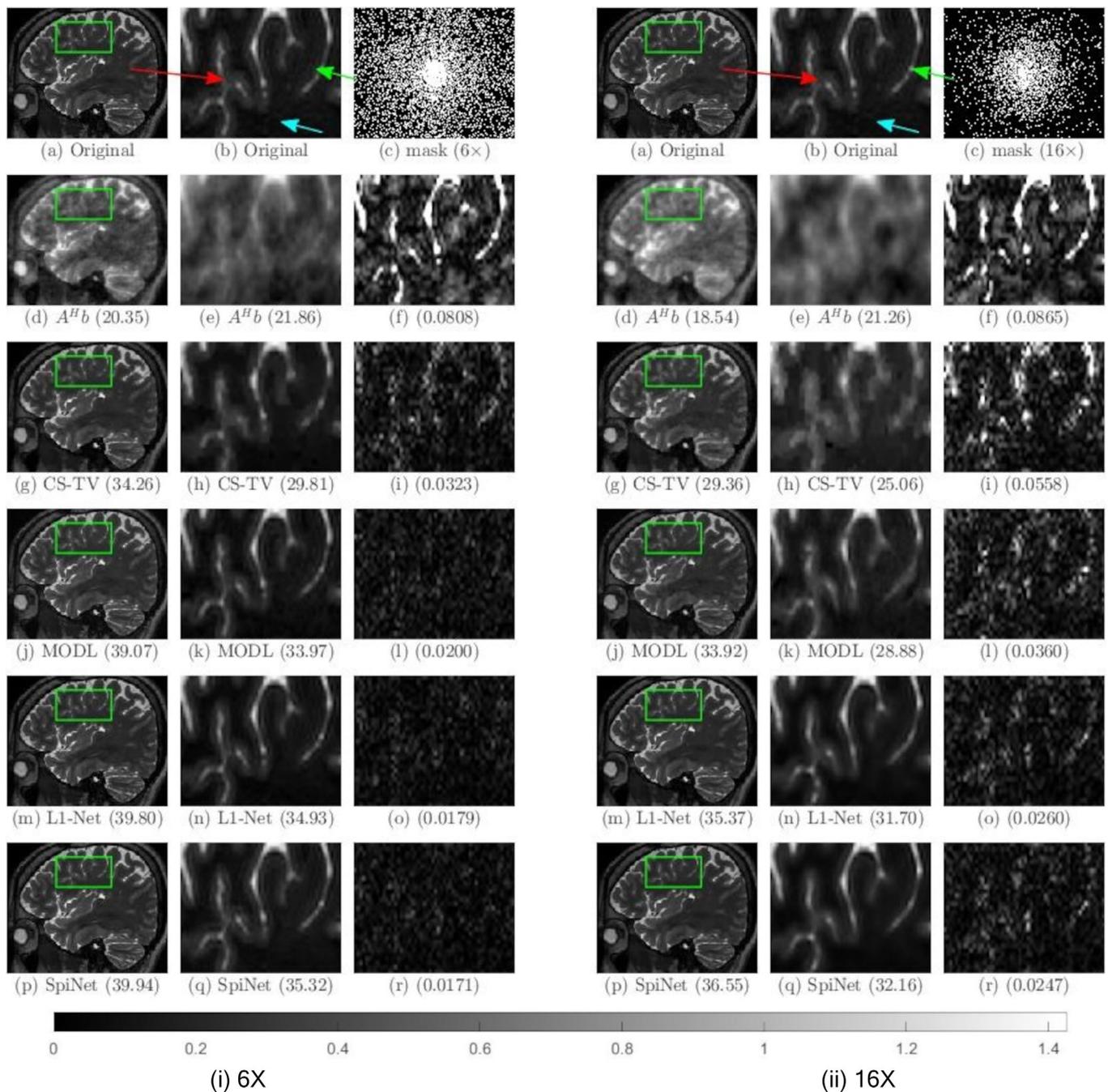


FIG. 5. Comparison of performance of all discussed methods in this work, CS-TV in third row, MoDL in fourth row, L1-Net results in fifth row, and proposed SpiNet results in sixth row, respectively, for (i) 6x and (ii) 16x undersampling rates using radial golden angle random sampling pattern given in (c). The second row in both (i) and (ii) represents the image estimated from zero filled  $k$  space which is also the input to the MoDL, L1-Net and proposed SpiNet. The L1-Net was implemented using SpiNet in which the value of  $p$  was kept constant and equal to 1. The third column in both (i) and (ii) from second row onward represents four times the difference between the estimated image and the original image in ROI (green bounding boxed image). The parenthesis values in first two columns of (i) and (ii) represents the PSNR and third column numbers represent the root mean square error. Areas where MoDL is not able to properly reconstruct the features or has artifacts are marked by red, green and cyan arrows in (b) of (i) and (ii). The significance of these arrows are explained in Section 5.A. Readers are advised to see the image in digital copy in full brightness of the screen. [Color figure can be viewed at wileyonlinelibrary.com]

modeling of tissue parameters in case of diffusion tensor imaging, dynamic contrast-enhanced imaging etc.

In the original implementation of MoDL, variable-density Cartesian random sampling pattern was used as shown in Fig. 4. To check versatility of proposed SpiNet, we compared the performance of MoDL and SpiNet for variable-density

Cartesian random sampling pattern and the results are shown in Fig. 4(i) for 6x and in Fig. 4(ii) for 16x undersampling. For comparison, L1-Net results are also shown, that was implemented exactly as SpiNet with enforced Schatten  $p$ -norm having constant value equal to 1. We also implemented CS-TV algorithm as described in Section 2.B to estimate high

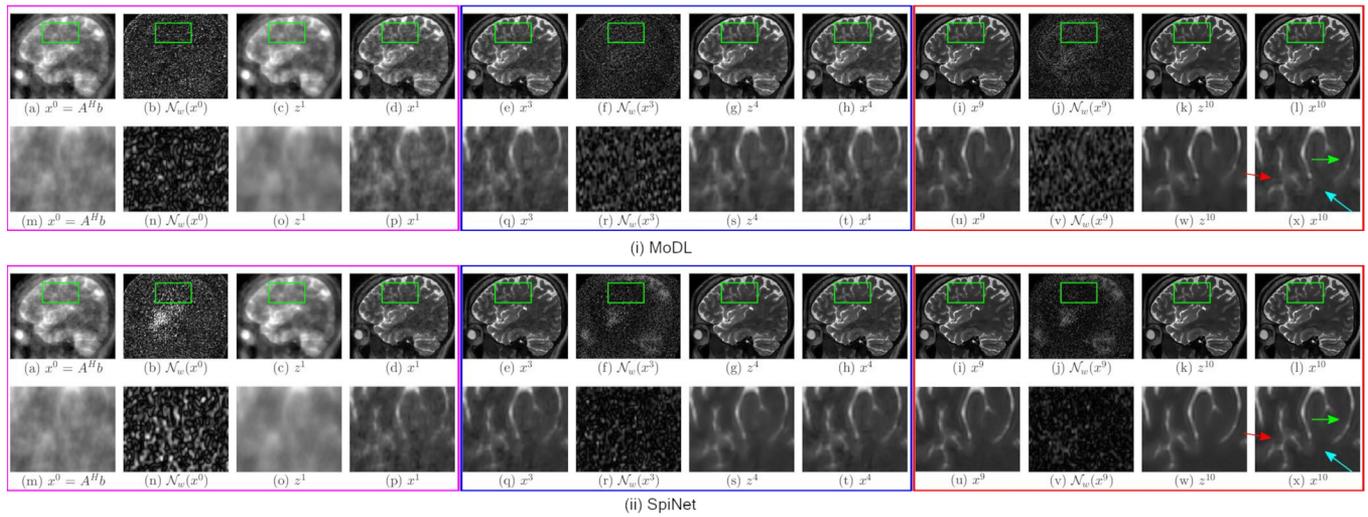


FIG. 6. Performance of (i) MoDL and (ii) proposed SpiNet across different iterations of corresponding networks in recovering the image from  $16\times$  undersampled  $k$  space data. In both figures (a) represents the input to the network, (b) represents the noise estimate  $\mathcal{N}_w(x^0)$  in the first iteration, (c) represents the denoised image  $z^1$  after first iteration, which is the output of denoiser block  $DW$ , and (d) represents the output of data consistency block ( $DC$ ) in first iteration. Images (e)-(h) and (i)-(l) represent the same for  $4^{th}$  and  $10^{th}$  iteration, respectively, and in the same order. The bottom rows of both subfigures represent zoomed version of corresponding images above them bounded by the green box. In both (i) and (ii) the magenta box shows the input, noise estimate, denoised output, and  $DC$  block output for  $1^{st}$  iteration and the blue and red boxes show the same variables for  $4^{th}$  and  $10^{th}$  iteration, respectively, and in the same order. The magnitude of noise images has been multiplied by 10 for clarity. From image (t) in (i) and (ii), it can be seen that SpiNet is able to achieve better quality of reconstruction than MoDL after  $4^{th}$  iteration. From (x) in (i) and (ii) it can also be seen that SpiNet has sharper edges at the "x" shaped structure (marked by red colored arrow). Readers are advised to see the image in digital copy in full brightness of the screen. [Color figure can be viewed at wileyonlinelibrary.com]

resolution image. The first column shows the performance of above-mentioned methods in terms of PSNR (in parenthesis). The second column shows the zoomed view of area bounded by green box in first column. Here  $A^Hb$  represents the image estimated from zero filled  $k$  space. The first image in third column represents the undersampling mask and rest images represent the error in estimated image for the bounded area. These error images were multiplied with four for ease of viewing. The values in parenthesis represent RMSE (root mean square error) with respect to fully sampled image. From the results, it can be observed that in terms of both PSNR and RMSE, we get considerable improvement over  $A^Hb$  by all methods. However, deep learning based methods perform better than CS-TV algorithm. It can also be seen that for  $\mathbf{R} = 6\times$ , L1-Net and SpiNet perform better than MoDL by  $\approx 0.7$  dB in the whole image and  $\approx 1$  dB in the bounded region. Moreover, the performance of SpiNet and L1-Net is similar and the small difference can be attributed to the random noise. At higher undersampling rate, CS-TV has visible staircase artifacts in cortical region. Moreover, the performance of all deep learning based methods is considerably better than CS-TV. For  $\mathbf{R} = 16\times$ , L1-Net and SpiNet perform better than MoDL by  $\approx 1.5$  and  $\approx 2.5$  dB, respectively, in the whole image and by  $\approx 3$  and  $\approx 3.5$  dB, respectively, in the bounded region.

Figure 5 represents the results of RGA undersampling pattern with  $\mathbf{R} = 6\times$  and  $16\times$ . The undersampling pattern resembles variable-density pattern in figure because the image is adjusted to match the aspect ratio of nearby images. It can be seen that for  $\mathbf{R} = 6\times$ , L1-Net and SpiNet perform better than MoDL by  $\approx 0.7$  dB in the whole image and  $\approx 1$  dB in the

bounded region. Additionally the performance of L1-Net and SpiNet is comparable to each other. For  $\mathbf{R} = 16\times$ , the performance of CS-TV shows the same trend as in Fig. 4(ii). Similarly, L1-Net and proposed SpiNet perform better than MoDL by a margin of  $\approx 2.5$  dB in the whole image and  $\approx 2$  dB in the bounded region. The performance of proposed SpiNet is better than L1-Net by a margin of  $\approx 0.7$  dB. From error images, it can also be observed that the error in L1-Net and SpiNet resembles noise, the error in MoDL estimate has higher concentration in the CSF region (area shown by red and green arrows) and thus has higher structural loss at some regions compared to other regions.

### 5.B. Comparison with MoDL as a function of iterations

Here, we compare the performance of MoDL and proposed SpiNet at different stages of reconstruction for a single image. Figure 6 shows estimation of high resolution MR image  $x^{10}$  (output of network after ten iterations) from aliased image  $x^0$  (input to the network) estimated from zero filled undersampled  $k$  space with undersampling rate  $\mathbf{R} = 16\times$  for MoDL in Fig. 6(i) and SpiNet in Fig. 6(ii). Both subfigures show the input  $x^{k-1}$ , noise estimate  $\mathcal{N}_w(\cdot)$ , denoised version  $z^k$  (i.e., output of  $DW$  block) and output of  $DC$  block  $x^k$  for  $k = 1, 4 \& 10$  (bounded in magenta, blue, and red boxes, respectively). The second row in both subfigures shows the zoomed in region that is bounded by the green box in the corresponding images above them. The noise estimates are multiplied by 10 for visibility. From the output of  $DC$  block of  $4^{th}$  iteration (i.e., image (t)) in Figs. 6(i) and 6(ii) it can be seen

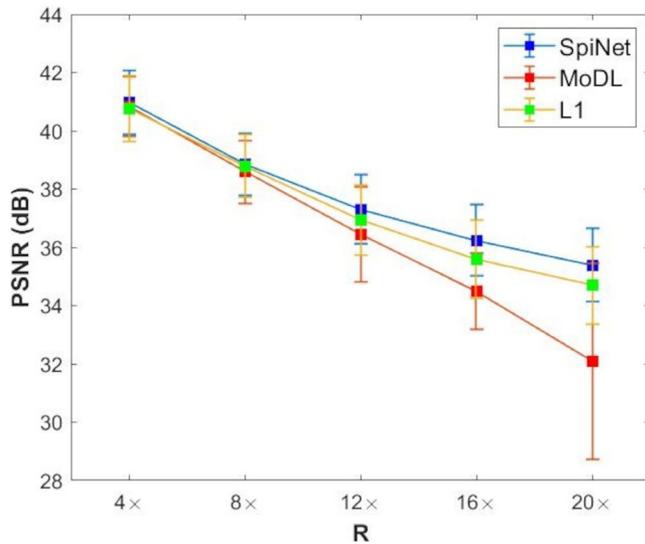


FIG. 7. Comparison of the performance of MoDL (red), L1-Net (green) and proposed SpiNet (blue) with respect to averaged PSNR (in dB) on the test dataset of 164 images as a function of undersampling ( $R$ ) rates. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

that output of proposed SpiNet has less cloud like artifact in the cortical region than MoDL. Even the output of  $DC$  block of  $1^{st}$  iteration (i.e., image (p)) for SpiNet has sharper edges and better contrast than corresponding output of MoDL. From this, it is clear that proposed SpiNet recovers high quality image in lesser number of iterations.

### 5.C. Statistical analysis

Figure 7 shows the performance of MoDL (red), L1-Net (green), and proposed SpiNet (blue) for  $4\times$ ,  $8\times$ ,  $12\times$ ,  $16\times$ , and  $20\times$  undersampling in terms of averaged PSNR based on 164 samples of test dataset. Similar trend to that of a single test image was followed. As the undersampling rate  $R$  increases, as expected, the proposed SpiNet starts to outperform MoDL. Moreover, for higher undersampling rate, the error bar (standard deviation) of MoDL is wider than proposed SpiNet, which also implies that proposed SpiNet can be considered as a more reliable method for reconstruction. The performance of L1-Net is also better than MoDL at higher  $R$ , but performs inferior to the proposed SpiNet.

We also performed a two-tailed t test with unequal variance (Welch test) using the 164 images of test dataset. The details of the test are given in Section 4.D. We take averaged PSNR as metric for the test and the significance threshold is 0.05. The results of the test are given in Table II as function of undersampling rate ( $R$ ), where the difference in performance is significant if  $P < 0.05$ . From the table, it can be observed that for lower undersampling rates of  $2\times$  and  $4\times$ , the performance of MoDL and proposed SpiNet is comparable. However, for higher undersampling rates the difference in performance of two methods is significant. This table shows that for higher undersampling rate, the performance of proposed SpiNet is significantly better than MoDL. The same conclusion can be reached from Table III, where t test

TABLE V. Training and testing time for MoDL and proposed SpiNet in seconds (rounded to nearest integer).

Method	MoD	Proposed SpiNet
Training Time (in sec)	273	474
Testing Time (in sec)	35	61

performed using NRMSE as the metric. However, when the test was conducted using SSIM as metric, SpiNet performs significantly better than MoDL even at low undersampling rates as well and is the results are shown in Table IV. However, SSIM as an evaluation metric should only be used in tandem with PSNR or NRMSE as these models were not trained using SSIM as loss function. Therefore they do not guarantee high recovery of SSIM and may not be consistent to show a trend. One such observation is for  $R = 6\times$  in case of MoDL, where there is a drop in average SSIM value.

### 5.D. Runtime

Table V compares the training and testing time for MoDL and proposed SpiNet in seconds. The training time is for 1 epoch of 360 training samples and the testing time is for the test dataset of 164 images. From the table, it is clear that the MoDL is  $\approx 1.5\times$  faster than SpiNet in training and  $\approx 2\times$  faster in testing.

The reason for longer training and testing time for SpiNet are the multiple function calls and loop iterations because of  $MM$  algorithm. Moreover, the operations performed in proposed SpiNet are higher than MoDL. The optimization equation of SpiNet requires a) additional calculation of  $\mathbf{W}^2$  ( $\approx O(3m)$  operations) at every majorization step (i.e.,  $4 \times N$  times) and b) matrix operations using  $\mathbf{W}^2$  ( $\approx O(4m^2)$ ) at every majorization step (i.e.,  $4 \times N$  times) for  $N = 10$ . In case of brain dataset,  $m = 256 \times 232$ . These computations are not required in case of MoDL. The above-mentioned two factors contribute to longer training and testing time in case of the proposed SpiNet. However, the proposed SpiNet is still advantageous over MoDL as it provides better recovery in terms of PSNR and can reconstruct same quality images in lesser iterations as shown in Fig. 6 and Table VI. From Table VI, it is also apparent that for  $R \geq 6\times$ , the reconstruction of SpiNet trained and tested for  $N = 5$  outperforms MoDL trained and tested for  $N = 10$ .

### 5.E. Learnt $p$ -values

The enforced Schatten  $p$ -norm value, that is,  $p$  of SpiNet was a trainable parameter. The learnt value of  $p$  for undersampling rates of ( $R$ ) of  $2\times$ ,  $4\times$ ,  $6\times$ ,  $8\times$ ,  $12\times$ ,  $16\times$ , and  $20\times$  are shown in Table VII. From the table it can be seen that for smaller  $R$  (lesser undersampling rates) the  $p$  values are close to 2, that is, the constraint norm is closer to L2 norm. However, as we increase the undersampling rate, the value of  $p$  decreases. A possible reason for this is that for lower  $R$ , the noise prior  $\mathcal{N}_w(\cdot)$  resembles more to Gaussian noise (which

TABLE VI. This table compares the PSNR (in dB) for MoDL and SpiNet when trained and tested for  $N = 5$  and  $N = 10$  for different undersampling rates ( $\mathbf{R}$ ).

R	MODL		SpiNet	
	$N = 10$ (PSNR)	$N = 5$ (PSNR)	$N = 10$ (PSNR)	$N = 5$ (PSNR)
2×	42.41	42.29	42.53	42.51
4×	40.83	40.52	40.97	40.77
6×	39.01	38.85	39.96	39.61
8×	38.59	38.08	38.84	38.62
12×	36.44	36.2	37.31	37.06
16×	34.49	33.86	36.25	35.78
20×	32.09	31.47	35.39	35.17

It can be seen that the results for SpiNet for  $N = 5$  are superior to the results of MoDL for  $N = 10$ .

is the noise added to the  $k$  space). However, as we increase the undersampling rate, the distribution of noise is not best estimated with L2 norm as the distribution deviated from Gaussian. Moreover, the unit ball of L $p$  norm where  $p \approx 1.4$  resembles the unit ball of elastic-net,<sup>58</sup> which is shown to give better performance than both L2 and L1 norm. For SpiNet trained with variable density undersampling pattern of  $\mathbf{R} = 16\times$ , learnt  $p$  value = 1.40 and for RGA undersampling pattern for  $\mathbf{R} = 6\times$ , the learnt  $p$  value is 1.46. However, the learnt  $p$  value for breast data for RGA pattern of  $\mathbf{R} = 20\times$  is 1.22 and for chest MRI data (mentioned in Section 4.D) for RGA pattern of  $\mathbf{R} = 6\times$  is 1.13 as shown in Table X and Section 6. This shows that the learnt  $p$  value depends less on the undersampling pattern and more on the data type.

## 5.F. Fivefold Cross validations

We performed a fivefold cross-validation study on proposed SpiNet for 16× undersampling rate in which the data of four of five patients were taken as training data and one patient data were used for testing, leading to five models. The results in terms of averaged PSNR (in dB) of the test data are shown in Table VIII. It was found that the performance of proposed SpiNet is similar in all cases, which suggests that the network is robust and able to generalize across the data.

TABLE VII. This table shows the learnt  $p$  value for proposed SpiNet for different undersampling rates ( $\mathbf{R}$ ).

R	PSNR (dB)	$p$
2×	42.53	1.81
4×	40.97	1.69
6×	39.96	1.42
8×	38.85	1.40
12×	37.31	1.39
16×	36.25	1.38
20×	35.39	1.36

TABLE VIII. This table shows results (in terms of averaged PSNR (in dB)) of test data (164 images) for fivefold cross validation performed for 16× undersampling rate.

Test patient data	PSNR (dB)
Patient 1	35.52
Patient 2	35.51
Patient 3	35.62
Patient 4	35.72
Patient 5	35.55

TABLE IX. Results of ablation study for undersampling rate of 12× for deciding the ratio between number of Majorization iterations to number of Minimization iteration ( $MM$  ratio).

Network	MM Ratio	$p$	PSNR (dB)
Lp	2/8	1.5	37.09
Lp	3/6	1.45	37.14
Lp	4/5	1.42	37.31
Lp	5/5	1.37	37.35
Lp	5/4	1.43	37.40
Lp	4/4	1.39	37.34

From here, it can be seen that majorization iterations  $\geq 4$  give similar results in terms of PSNR of reconstructed image.

## 5.G. Ablation study

To decide the number of majorization iterations ( $\hat{N}$ ) and number of minimization iterations ( $M_1$ ), we performed an ablation study for undersampling rate of 12× and the results for the same are shown in Table IX. From Table IX, it can be inferred that for  $\hat{N} \geq 4$ , the proposed SpiNet performs similar in terms of PSNR and the learnt  $p$  - values are also similar. However as the total number of iterations ( $\hat{N} \times M_1$ ) increases, both computational time and memory requirement increases. Hence we chose  $\hat{N} = 4$  and  $M_1 = 4$  which has the minimum number of total iterations among options where  $\hat{N} \geq 4$ .

## 6. GENERALIZABILITY

To test the generalization ability of the proposed SpiNet, as discussed in Section 4.A the chest MRI as well as DCE-MR data of the breast were utilized. The chest MRI data and the undersampling masks were same as used in implementation of ADMM-Net.<sup>53</sup> For experiments on contrast enhance breast MRI data, the training and testing techniques are same as used in.<sup>54</sup> Chest MR results are compared with ADMM-Net and the results for the breast data is compared with ISTA – Net<sup>+</sup>,<sup>28</sup> and these results are shown in Table X. It can be seen that the proposed SpiNet outperforms both state of the art networks for the respective datasets by a margin  $\approx 1.5$  dB and assert that the proposed SpiNet has the capability to work across various datasets.

Table XI shows the performance of SpiNet on chest and breast dataset without and with retraining the network in fourth and fifth columns respectively. It can be seen from the

TABLE X. This table shows performance for SpiNet for Chest and Breast dataset and compares the results with their state of the art methods.

Data	R	Training images (number)	SOTA (PSNR)	SpiNet tuned (PSNR)	p
Chest	6×	100	ADMM-Net/37.17 ± 4.1	38.6 ± 3.6	1.13
Breast	20×	3000	ISTA-Net <sup>+</sup> /30.19 ± 1.8	31.7 ± 1.1	1.22

TABLE XI. This table shows the performance of SpiNet when trained on brain data and tested on chest/breast data (column fourth) and when fine tuned for the respective datasets (column fifth).

Testing data	R	Input (PSNR)	Not tuned (PSNR)	Tuned (PSNR)
Chest	6×	23.97 ± 3.8	30.8 ± 4.2	38.6 ± 3.6
Breast	20×	18.36 ± 1.3	23.2 ± 2.1	31.7 ± 1.1

Second column shows the PSNR (in dB) of the aliased images generated from the undersampled  $k$  space data.

table that retraining leads to better reconstruction. An important observation is that the PSNR of SpiNet without retraining is still higher than that of aliased images. This is expected as introducing a data consistency term should give results at least as good or better than the input. This is one of the silent features of model-based DL architectures, which has been shown here.

## 7. LIMITATIONS

There are some inherent limitations for the proposed SpiNet, which can make its application domain narrow. First being that Majorization–Minimization algorithm that was utilized is only applicable to separable priors, which can be written in form  $\|\Gamma(\mathbf{x})\|_p^p = \sum_i |\gamma(x_i)|^p$ . This might not be true for all inverse problems and a reformulation of Eq. (5) will be required to make it more generic. Second, theoretically the convergence of the network is not known and hence its application for complex nonlinear inverse problems may be limited. Finally, SpiNet like all deep learning models focuses on decreasing the cost function over whole training set, hence it can be invariant to minor abnormalities, which can be of clinical significance, but do not contribute much to training loss (Osteophyte, early stages of heterotopic ossification). Hence, the validation of this network for dataset with more pathologies may be needed to claim generality of the network.

## 8. CONCLUSION

In this work, we proposed a novel architecture that can enforce regularization of any norm  $p$  (Schatten  $p$ -norm) in unrolled deep learning based scheme for solving an inverse problem. We demonstrated the idea for reconstruction of MR images from undersampled  $k$  space data and compared proposed SpiNet results with MoDL, which utilizes  $p = 2$ . We illustrated that the proposed SpiNet has the capability to learn the  $p$  value as well as work for a fixed  $p$  value (as shown with  $p = 1$ ). For different undersampling rates ( $\mathbf{R}$ ), keeping all

hyper-parameters same, it was demonstrated that for lower  $\mathbf{R}$ , MoDL and SpiNet performance is similar. For higher  $\mathbf{R}$  ( $\geq 6\times$ ), the proposed SpiNet performs significantly better than MoDL. We have also provided detailed explanation of our observations and have highlighted the limitations of our approach, will try to address these limitations in future work. The trained network is available for enthusiastic users as open source at <https://github.com/adityarastogi2k12/SpiNet>.

## CONFLICTS OF INTEREST

There is no conflict of interest declared by Authors.

## ACKNOWLEDGMENT

This work was supported by the Science & Engineering Research Board (SERB) core research grant (No. CRG/2018/000672) as well as WIPRO GE-CDS Collaborative Laboratory of Artificial Intelligence in Healthcare and Medical Imaging (AIHMI).

## DATA AVAILABILITY

The dataset containing T2 weighted images of brain is available at <https://github.com/hkaggarwal/modl>. Breast DCE MRI dataset<sup>59,60</sup> is available at The Cancer Imaging Archive (TCIA)<sup>61</sup> for researchers who meet the criteria and gain approvals to access the research database. Chest MRI data used for additional experiments are available at <https://github.com/yangyan92/Deep-ADMM-Net>. The authors do not own any of these datasets. The trained SpiNet network is made available for enthusiastic users as open source at <https://github.com/adityarastogi2k12/SpiNet>.

## APPENDIX A

### A.1. MoDL being special case of SpiNet

The cost function for SpiNet is given in Eq. (17). The minimization is achieved by variable splitting technique as shown in Eqs. (18) and (19), for prior and data consistency terms, respectively. The cost function for Eq. (19) can be upper bounded using Majorization–Minimization algorithm as shown in Eq. (20). Here,  $\lambda' = \lambda p/2$  and  $\mathbf{W} = \text{diag}(|\bar{x}_i - z_i^k|^{p/2-1})$ . Minimizing Eq. (20) w.r.t  $\mathbf{x}$  leads to Eq. (21).

For  $p = 2$ ,  $\mathbf{W} = \mathbf{I}$ , and  $\lambda' = \lambda$ , simplifies Eq. (21) to (16), which is the update equation in the MoDL. Hence, for  $p = 2$ , SpiNet simplifies into MoDL. Hence in Fig. 2, the DC block of SpiNet simplifies to that of MoDL,<sup>29</sup> making MoDL as special case of SpiNet.

### A.2. SpiNet as a special case of compressive sensing based reconstruction framework

Compressive sensing framework for solving the inverse problem involves minimizing cost function given in Eq. (5).

The norm enforced on the prior is mostly  $p = 1$  as it is convex and enforces sparsity. In case of SpiNet  $\Gamma(\mathbf{x}) = \mathcal{N}_w(\mathbf{x})$  which is the noise present  $\mathbf{x}$ . Moreover,  $\mathcal{N}_w(\mathbf{x}) = \mathbf{x} - \mathcal{D}_w(\mathbf{x}) = \mathbf{x} - \mathbf{z}$  is also sparse for the optimal solution. It was shown in Ref.44 that a prior regularized optimization function can be split into two equations where one contains the prior term and other the data consistency term as shown in Eqs. (18) and (19), respectively. Solving Eq. (18) is analogous to denoising. In analytical compressive sensing frameworks, a handcrafted prior gets utilized, which is not data dependent. Hence in traditional compressive sensing framework the denoiser is not learnt based on the training data. Therefore, in Fig. 2, the DW block is not CNN based and can be fixed to be a denoiser of choice. Hence SpiNet becomes a special case of conventional compressive sensing framework, when the denoiser is trainable/data-driven.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: yalavarthy@iisc.ac.in

## REFERENCES

- Gribbestad IS, Gjesdal KI, Gunnar N, Steinar L, Hjelstuen MHB, Alan J. An introduction to dynamic contrast-enhanced MRI in oncology in Dynamic contrast-enhanced magnetic resonance imaging in oncology: Springer; 2005:1–22.
- Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*. 2015;73:102–116.
- Rempp KA, Brix G, Wenz F, Becker CR, Gückel F, Lorenz WJ. Quantification of regional cerebral blood flow and volume with dynamic susceptibility contrast-enhanced MR imaging. *Radiology*. 1994;193:637–641.
- Larkman DJ, Nunes RG. Parallel magnetic resonance imaging. *Phys Med Biol*. 2007;52:R15.
- Pruessmann KP, Markus W, Scheidegger MB, Peter B. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42:952–962.
- Sodickson DK, McKenzie CA, Ohliger MA, Yeh EN, Price MD. Recent advances in image reconstruction, coil sensitivity calibration, and coil array design for SMASH and generalized parallel MRI. *Magn Reson Mater Phys Biol Med*. 2001;13:158–163.
- Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47:1202–1210.
- Michael L, David D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182–1195.
- Rudin LI, Stanley O, Emad F. Nonlinear total variation based noise removal algorithms. *Physica D*. 1992;60:259–268.
- Michael L, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. *IEEE Signal Process Mag*. 2008;25:72–82.
- Goud LS, Mathews J. A blind compressive sensing frame work for accelerated dynamic MRI, in 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI); IEEE 2012:1060–1063.
- Lingala SG, Hu Y, DiBella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. *IEEE Trans Med Imag*. 2011;30:1042–1054.
- Afonso MV, Bioucas-Dias JM, Figueiredo MAT. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans Image Process*. 2010;19:2345–2356.
- Bioucas-Dias JM, Figueiredo MA. Two-step iterative shrinkage/thresholding algorithm for linear inverse problems. *IEEE Trans Image Process*. 2007;16:2980–2991.
- Amir B, Marc T. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci*. 2009;2:183–202.
- Seung-Jean K, Kwangmoo K, Michael L, Stephen B, Dimitry G. An interior-point method for large-scale L1 regularized least squares. *IEEE J Selected Top Sign Process*. 2007;1:606–617.
- Figueiredo MAT, Nowak RD, Wright SJ. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J Select Top Sign Process*. 2007;1:586–597.
- Gutta S, Kadimesetty VS, Kalva SK, et al. Deep neural networkbased bandwidth enhancement of photoacoustic data. *J Biomed Opt*. 2017;22:116001.
- Awasthi N, Jain G, Kalva SK, Pramanik M, Yalavarthy PK. Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography. *IEEE Trans Ultrason Ferroelectr Frequ Contr*. 2020;67:2660–2673.
- Kadimesetty VS, Sreedevi G, Sriram G, Yalavarthy PK. Convolutional neural network-based robust denoising of low-dose computed tomography perfusion maps. *IEEE Trans Radiat Plasma Med Sci*. 2018;3:137–152.
- Hwan JK, McCann MT, Emmanuel F, Michael U. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process*. 2017;26:4509–4522.
- Guang Y. Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. 2017;37:1310–1321.
- Taejoon EO, Yohan J, Taeseong K, Jinseong J, Ho-Joon L, Dosik H. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magn Reson Med*. 2018;80:2188–2201.
- Dmitry U, Andrea V. Lempitsky V. Deep image prior in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:9446–9454.
- Dittmer S, Kluth T, Thorstein RHM, Maass P. Deep image prior for 3D magnetic particle imaging: A quantitative comparison of regularization techniques on Open MPI dataset arXiv e-prints. 2020:arXiv-2007.
- Ji He, Jianhua MA. Radon inversion via deep learning in Medical Imaging 2019: Physics of Medical Imaging; International Society for Optics and Photonics. 2019;10948:1094810.
- Dittmer S, Kluth T, Maass P, Baguer DO. Regularization by architecture: A deep prior approach for inverse problems. *J Math Imaging Vision*. 2020;62:456–470.
- Jian Z, Bernard G, ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2018:1828–1837.
- Aggarwal HK, Mani MP, Jacob M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2018;38:394–405.
- Kingma DP, Jimmy BA, Adam: A method for stochastic optimization arXiv preprint arXiv:1412.6980. 2014.
- Kai Z, Zuo WG, Shuhang ZL. Learning deep CNN denoiser prior for image restoration in Proceedings of the IEEE conference on computer vision and pattern recognition; 2017:3929–3938.
- Lei Z, Wangmeng Z. Image restoration: From sparse and low-rank priors to deep priors [lecture notes]. *IEEE Signal Process Mag*. 2017;34:172–179.
- Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055–3071.
- Patrick P, Max W. Recurrent inference machines for solving inverse problems arXiv preprint arXiv:1706.04008. 2017.
- Morteza M, Hatem M, Vardan P, Shreyas V, David D, John P, eds. Recurrent generative adversarial networks for proximal learning and automated compressive image recovery arXiv preprint arXiv:1711.10046. 2017.
- Gorodnitsky IF, Rao BD. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Trans Signal Process*. 1997;45:600–616.
- Feng LI, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn Reson Med*. 2014;72:707–717.
- Huiyuan W, Vieira J. 2-D wavelet transforms in the form of matrices and application in compressed sensing in 2010 8th World Congress on Intelligent Control and Automation: IEEE; 2010:35–39.

39. Candes EJ, Terence T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans Inf Theory*. 2006;52:5406-5425.
40. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory*. 2006;52:489-509.
41. Donoho DL. Compressed sensing. *IEEE Trans Inf Theory*. 2006;52:1289-1306.
42. Tyrrell RR. Monotone operators and the proximal point algorithm. *SIAM J Contr Opt*. 1976;14:877-898.
43. Stephen B, Neal P, Eric C. Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc 2011.
44. Venkatakrishnan SV, Bouman CA, Brendt W. Plug-and-play priors for model based reconstruction, in 2013 IEEE Global Conference on Signal and Information Processing: IEEE; 2013:945-948.
45. Yaniv R, Michael E, Peyman M. The little engine that could: regularization by denoising (RED). *SIAM J Imaging Sci*. 2017;10:1804-1844.
46. Siavash AB, Matthias Z. Image restoration using autoencoding priors. in Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications; SCITEPRESS 2018;5:33-44.
47. Eksioğlu EM. Decoupled algorithm for MRI reconstruction using nonlocal block matching model BM3D-MRI. *J Math Imaging Vision*. 2016;56:430-440.
48. Schlemper JO, Caballero J, Hajnal JV, et al. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imag*. 2017;37:491-503.
49. Kai Z, Wangmeng Z, Yunjin C, Deyu M, Lei Z. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans Image Process*. 2017;26:3142-3155.
50. Yi G, Goud LS, Yinghua Z, Marc LR, Nayak KS. Direct estimation of tracer-kinetic parameter maps from highly undersampled brain dynamic contrast enhanced MRI. *Magn Reson Med*. 2017;78:1566-1578.
51. Liu DC, Jorge N. On the limited memory BFGS method for large scale optimization. *Math Program*. 1989;45:503-528.
52. Uecker M, Lai P, Murphy MJ, et al. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med*. 2014;71:990-1001.
53. Sun J, Li H, Xu Z, et al. Deep ADMM-Net for compressive sensing MRI. in Advances in neural information processing systems; 2016:10-18.
54. Aditya R, Yalavarthy PK. Comparison of iterative parametric and in direct deep learning-based reconstruction methods in highly under-sampled DCE-MR Imaging of the breast. *Med Phys*. 2020;47:4838-4861.
55. Zhou W, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
56. Liu T, Wen W, Zhu W, et al. The effects of age and sex on cortical sulci in the elderly. *NeuroImage*. 2010;51:19-27.
57. Liu T, Wen W, Zhu W, et al. The relationship between cortical sulcal variability and cognitive performance in the elderly. *NeuroImage*. 2011;56:865-873.
58. Trevor H, Robert T, Martin W. *Statistical Learning with Sparsity: the Lasso and Generalizations*. Boca Raton: CRC Press; 2015.
59. Huang W, Li X, Chen Y, et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. *Trans Oncol*. 2014;7:153-166.
60. Huang W, Li X, Chen Y, et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. *Trans Oncol*. 2014;7:153-166.
61. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057.