

ROBUST SELF-GUIDED DEEP IMAGE PRIOR

Evan Bell¹, Shijun Liang², Qing Qu³, Saiprasad Ravishankar^{2,4}

¹Dept. of Mathematics, Michigan State University, East Lansing, MI.

²Dept. of Biomedical Engineering, Michigan State University, East Lansing, MI.

³Dept. of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI.

⁴Dept. of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI.

ABSTRACT

In this work, we study the deep image prior (DIP) for reconstruction problems in magnetic resonance imaging (MRI). DIP has become a popular approach for image reconstruction, where it recovers the clear image by fitting an overparameterized convolutional neural network (CNN) to the corrupted/undersampled measurements. To improve the performance of DIP, recent work shows that using a reference image as an input often leads to improved reconstruction results compared to vanilla DIP with random input. However, obtaining the reference input image often requires supervision and hence is difficult in practice. In this work, we propose a self-guided reconstruction scheme that uses no training data other than the set of undersampled measurements to simultaneously estimate the network weights and input (reference). We introduce a new regularization that aids the joint estimation by requiring the CNN to act as a powerful denoiser. The proposed self-guided method gives significantly improved image reconstructions for MRI with limited measurements compared to the conventional DIP and the reference-guided method while eliminating the need for any additional data.

Index Terms— Magnetic resonance imaging, compressed sensing, machine learning, deep learning, Deep image prior.

1. INTRODUCTION

MRI [1] is a non-invasive modality that is widely used to image for both human anatomical structures and functions. However, the MRI data acquisition process is slow, and traditional image reconstruction methods require many measurements to produce high-quality images. This limits efficiency and image quality in applications including dynamic imaging and may lead to improper clinical diagnosis. Consequently, developing reconstruction methods that can produce clinically useful images from highly undersampled measurements is desirable.

Recently, compressed sensing (CS) [2], has enabled the reconstruction of MR images from undersampled data, leading to reduced scan times. CS relies on a prior assumption about the structure of MR images, and the reconstruction framework incorporates a regularization term based on this assumption. Two common choices for this prior in classical CS-MRI are sparsity in the wavelet domain [3] and total variation of the reconstructed image [4]. While classical CS assumes known sparsity bases of signals or incoherence, methods using learned image models for reconstruction have proved more effective, such as those involving (patch-based) synthesis dictionary learning [5, 6]. Recent advances in transform learning [7, 8]

also provide an efficient alternative framework for sparse modeling in MRI. Other recent approaches have used explicitly (with supervision) learned regularizers [9] that also have the potential to yield improved image quality.

Deep learning has received a great deal of attention in the medical imaging field over the past few years, as it has shown superior performance in denoising and imaging inverse problems such as MRI reconstruction compared to several conventional methods. For example, end-to-end trained CNNs have demonstrated success in image reconstruction. One popular network choice is the U-Net architecture [10, 11]. A wide range of other architectures have also been used successfully for MRI reconstruction, including transformers [12] and generative adversarial networks [13] (cf. review in [8]). Importantly, hybrid-domain methods [14] that enforce data consistency (i.e., the reconstruction should be consistent with the measured data) during both training and reconstruction, show enhanced stability and performance. However, while these deep learning models achieve state-of-the-art performance in important imaging tasks, they often require extensive training on a large data set.

Numerous studies, beginning with the deep image prior [15], have shown that even in the absence of training data, the architecture of a CNN may serve as a sufficiently robust prior to permit network learning and image reconstruction from only undersampled measurements. The inductive bias of untrained CNNs allows them to provide denoising, inpainting, super-resolution, and even compressed sensing [16] without any training data set. A recent work [17] proposes incorporating some prior information into DIP-based reconstruction by using a carefully selected reference image as network input during training. This reference-guided approach significantly improves the reconstruction quality and stability, while eliminating the need for fully supervised training. However, this method still relies on the availability of a suitable reference image, which may not always be the case. It is also unclear from [17] how such a reference can be suitably selected based on only undersampled measurements of an unknown test image.

Inspired by the ability of reference-based guidance to improve the performance of DIP reconstruction, *we consider the setting where absolutely no reference or training data is available. We propose a self-guided DIP method, which eliminates the need for a separate reference image and gives much better MRI reconstruction quality than the reference-guided method as well as other schemes.* We also draw inspiration in our approach from recent advances such as randomized smoothing [18].

2. DEEP IMAGE PRIOR BASED MRI RECONSTRUCTION

2.1. Multi-coil MRI Reconstruction

To ensure accurate reconstructions, ill-posed inverse problems such as MR image reconstruction from undersampled k-space (Fourier

E. Bell (*belleva1@msu.edu*) and S. Liang (*liangs16@msu.edu*) are corresponding authors (equal contributors). The authors thank Dr. Rongrong Wang for feedback on the work. This work was supported in part by the National Science Foundation (NSF) grants CCF-2212065 and CCF-2212066.

domain) measurements generally require additional priors or models on the underlying image. Typically, this is achieved by incorporating explicit or implicit regularizers during image reconstruction. The regularizer imposes additional constraints on the desired type of solution resulting in a more stable eventual solution. For multi-coil MRI reconstruction of an image $\mathbf{x} \in \mathbb{C}^q$, the optimization problem can be mathematically formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \sum_{c=1}^{N_c} \|\mathbf{A}_c \mathbf{x} - \mathbf{y}_c\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad (1)$$

where $\mathbf{y}_c \in \mathbb{C}^p$, $c = 1, \dots, N_c$, represent the acquired k-space measurements from N_c coils. We write the forward system operator as $\mathbf{A}_c = \mathbf{M}\mathbf{F}\mathbf{S}_c$, where $\mathbf{M} \in \{0, 1\}^{p \times q}$ is a masking operator that captures the pattern for sampling data in k-space, $\mathbf{F} \in \mathbb{C}^{q \times q}$ is the Fourier transform operator, and $\mathbf{S}_c \in \mathbb{C}^{q \times q}$ is the c th coil-sensitivity map (a diagonal matrix). Here, the explicit regularizer $\mathcal{R}(\cdot)$ is used to restrict the solutions to the space of desirable images.

Choices for the regularizer in MRI reconstruction can vary from the ℓ_1 penalty on wavelet coefficients or a total variation penalty to patch-based sparsity in learned dictionaries or, as in our method, the proximity of the reconstructed image to a CNN-denoised image.

2.2. Deep Image Prior for MRI Reconstruction

MRI reconstruction using DIP is typically formulated as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{c=1}^{N_c} \|\mathbf{A}_c \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}) - \mathbf{y}_c\|_2^2, \quad \hat{\mathbf{x}} = \mathbf{f}_{\hat{\boldsymbol{\theta}}}(\mathbf{z}), \quad (2)$$

where $\mathbf{f}$ is a CNN with parameters $\boldsymbol{\theta}$ and $\mathbf{z}$ is a typically fixed network input that is randomly chosen (e.g., a random Gaussian vector or tensor). We will refer to this formulation as “vanilla DIP” in this work.

2.3. Overfitting and Spectral Bias in Deep Image Prior

Since DIP relies on training a network using noisy or incomplete data, the corresponding corruptions will eventually appear in the network output if it is trained until loss function convergence. This phenomenon not only affects the performance of DIP in image denoising, which has been well-studied, but also in inverse problems such as MRI reconstruction, where the measurement operator has a non-trivial null space. In Fig. 1, we can observe that the MRI reconstruction quality peaks quickly and then gradually decreases as training continues. This demonstrates the need for an early stopping criterion when using vanilla DIP to solve inverse problems.

Another phenomenon observed when using DIP is *spectral bias*, where the network learns low frequency image content more quickly and more accurately than high frequency content [19]. Spectral bias may limit the performance of DIP, since the network may not learn relevant high frequency content before overfitting. However, it may also help explain the effectiveness of DIP in some settings, because noise generally contains more high frequency content than natural images.

2.3.1. Understanding Spectral Bias and Overfitting for DIP MRI

To understand the spectral bias present in vanilla DIP MRI reconstruction, we use a frequency band metric to investigate the difference between reconstructed frequencies and those of the ground truth. We analyze the multi-coil k-space of the output image $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z})$ at each network update iteration and compare it to that of the target

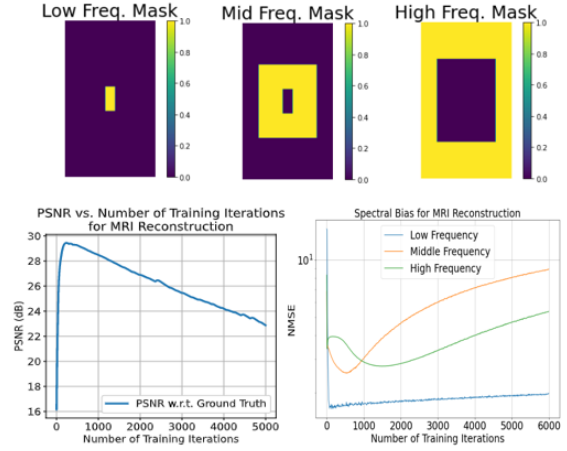


Fig. 1: Top row: the three masks used to compute the frequency band-based metric. Bottom row: reconstruction PSNR plot on the left illustrates the overfitting issue that occurs during MRI reconstruction. Spectral bias also affects the performance of DIP for MRI reconstruction (right plot), as different frequency bands are reconstructed at different rates.

image’s k-space $\mathbf{y}_c^*$ to understand the convergence dynamics of different frequency components (see Fig. 1). To do this, we compute the following metric for low, middle, and high frequency bands:

$$\text{NMSE} := \frac{\sum_{c=1}^{N_c} \|\mathbf{M}_{\text{freq}} \mathbf{A}_c \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}) - \mathbf{M}_{\text{freq}} \mathbf{y}_c^*\|_2^2}{\sum_{c=1}^{N_c} \|\mathbf{M}_{\text{freq}} \mathbf{y}_c^*\|_2^2} \quad (3)$$

where $\mathbf{M}_{\text{freq}}$ is the frequency band mask.

2.4. Reference-Guided DIP

The reference-guided DIP formulation was proposed in [17] as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{c=1}^{N_c} \|\mathbf{A}_c \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}) - \mathbf{y}_c\|_2^2, \quad \hat{\mathbf{x}} = \mathbf{f}_{\hat{\boldsymbol{\theta}}}(\mathbf{z}).$$

This formulation is identical to the problem in (2), except that the input to the network is no longer fixed random noise, but is instead a reference image that is very similar to the one being reconstructed. The input to the network introduces some additional structural information, and we can consider the network as essentially performing image refinement or style transfer rather than image generation from scratch. This method is quite reasonable in cases where a data set of structurally similar images is available (the FastMRI data set is a good example), and there is a systematic way to choose the network input image from the data set based on only undersampled k-space observations in the test scan.

In [17], the input image seems to be chosen ad hoc by hand. As a more realistic modification of this method, we use an approach similar to the recent LOND-N-MRI [20] method to search for the reference image (using a distance metric such as euclidean distance or other metric) that is most similar to an estimated test reconstruction from undersampled data.

2.5. Self-Guided DIP

To circumvent the need for a prior chosen reference to guide DIP, we introduce the following method which adaptively estimates such

a reference, which we call **self-guided DIP**:

$$\hat{\theta}, \hat{z} = \arg \min_{\theta, z} \underbrace{\sum_{c=1}^{N_c} \|A_c \mathbb{E}_{\eta} [f_{\theta}(z + \eta)] - y_c\|_2^2}_{\text{data consistency}} + \underbrace{\alpha \|\mathbb{E}_{\eta} [f_{\theta}(z + \eta)] - z\|_2^2}_{\text{denoiser regularization}} \quad (4)$$

$$\hat{x} = \mathbb{E}_{\eta \sim P_{\eta}} [f_{\hat{\theta}}(\hat{z} + \eta)] \quad (5)$$

In this optimization, z is no longer a reference image, but is instead initialized to a zero-filled (for missing k-space) least squares reconstruction, and η is random noise drawn from some distribution P_{η} (either uniform or Gaussian in our experiments). The first term in the optimization enforces data consistency, while the second term enforces that f should act as a denoiser. We also note that in our scheme, z is optimized, in contrast to both vanilla and reference-guided DIP. Hence we call this method “self-guided,” because at each iteration the network’s “reference” is updated.

Another interesting aspect of this scheme is that the best performance is achieved when the magnitude of η is quite large. The addition of this noise to the network input at each iteration is similar to training techniques used to promote adversarial robustness [21]. Additionally, computing both the loss and final reconstruction as an expectation bears resemblance to recent work on the use of randomized smoothing to train robust models.

In Fig. 2, we demonstrate the importance of the second term in the optimization. Without this term, z is not able to be updated appropriately, which leads to unstable training and poor performance. This demonstrates the effectiveness of using denoising performance as a regularizer.

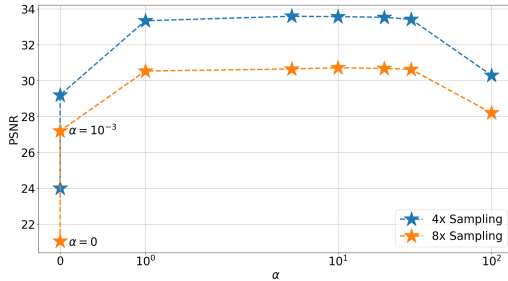


Fig. 2: Self-guided deep image prior: effect of regularization.

2.6. Data Correction

To further increase the quality of the network output, we apply a simple data correction post-processing operation after the optimization is complete. This is similar to the reference-guided DIP paper [17] and is given by

$$y_{c_{\text{new}}} = \overline{\mathbf{M}} \mathbf{F} \mathbf{S}_c(\hat{x}) + \mathbf{M}^T y_c, \quad \hat{x}_{\text{corrected}} = \sum_{c=1}^{N_c} \mathbf{S}_c^H \mathcal{F}^H y_{c_{\text{new}}}, \quad (6)$$

where $\overline{\mathbf{M}}$ samples only the frequencies not in the mask, $\mathbf{M}$, setting values for other frequencies to zero. Assuming the sensitivity maps are appropriately normalized, this step is akin to solving the least squares data-fitting problem with the originally missing frequencies compensated using the information from $\mathbf{M} \mathbf{F} \mathbf{S}_c(\hat{x})$.

Algorithm 1 Self-Guided Algorithm

Require: Initial reconstructed image z^0 , random Gaussian noise or uniform noise η , k-space sampling mask $\mathbf{M}$, regularization parameter α , parameters of optimizer for θ and z (we used Adam optimizer), number of optimization iterations T .

- 1: Initialize reconstruction network parameters θ with randomly initialized weights. Set $z = z^0$.
- 2: **for** iteration < maximal number T **do**
- 3: Generate k inputs $z + \eta$ for k realizations of η .
- 4: Compute the loss and calculate gradients.
- 5: Update θ using network optimizer.
- 6: Update z using input optimizer.
- 7: **end for**
- 8: Compute $x_{\text{out}} = \mathbb{E}_{\eta} [f_{\theta}(z + \eta)]$.
- 9: Apply the data correction operation to x_{out} .
- 10: **return** learned net. parameters θ . output reconstruction x_{out} .

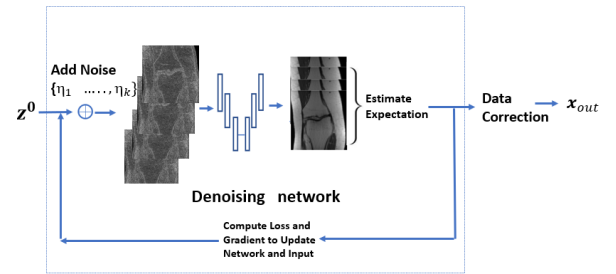


Fig. 3: Flow chart of the proposed self-guided DIP algorithm.

3. EXPERIMENTS AND RESULTS

Data set. We tested the different methods on the multi-coil FastMRI knee data set [22, 23]. 15 separate images (all from different sequences) were used as test data. The coil sensitivity maps for all cases were obtained using the BART toolbox [24] and all the coil sensitivity maps were estimated from under-sampled (center of k-space) data to create a realistic testing setup.

Training setup. We provide a comparison of five reconstruction methods: vanilla DIP, reference-guided DIP, self-guided DIP, compressed sensing with wavelet regularization, and a CNN trained in an end-to-end supervised manner on a set of 3000 images. For compressed sensing, we used the SigPy package¹, and the regularization parameter was tuned and set as $\lambda = 10^{-6}$. During training, the network weights were initialized randomly (normally distributed). For all of the deep network methods, the network architecture used was a deep U-Net ($\sim 3 \times 10^8$ parameters). The network parameters were optimized using Adam with a learning rate of 3×10^{-4} .

For the self-guided method, we observed that the noise η can be drawn from different distributions such as the normal or uniform distribution with essentially identical performance. For our experiments we draw η from $U(0, m)$, where m is $\frac{1}{2}$ of the maximum value of the magnitude of z . In this case, z is also optimized using Adam with a learning rate of 1×10^{-1} . At each iteration, we estimated the expectation inside the loss function using 4 realizations of η . For all three unsupervised methods besides compressed sensing, the data correction outlined in Section 2.6 was applied. Since the supervised U-Net is a pure image-domain network, there was no

¹<https://github.com/mikgroup/sigpy>

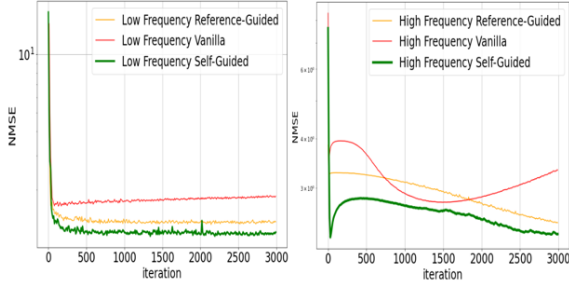


Fig. 4: Error in the low and high frequencies of the reconstruction, with different methods plotted over the iterations at 4x undersampling. In general, self-guided DIP demonstrates better convergence in both frequency bands when compared to the other methods.

post-processing performed.

Evaluation. We tested each of the methods for performing reconstructions at 4x acceleration (25.0% sampling) and 8x acceleration (12.5% sampling). Variable density 1-D random Cartesian (phase-encode) undersampling was performed in all cases. We quantified the reconstruction quality of the different methods using the peak signal-to-noise ratio (PSNR) in decibels (dB). We also computed the frequency band metric to study the spectral bias and overfitting in each method.

Ax	Vanilla DIP	Reference- Guided	CS Recon	Self-Guided DIP	Supervised U-Net
4x	30.2	33.17	29.3	33.59	33.15
8x	28.75	30.23	27.8	30.72	30.27

Table 1: Average reconstruction PSNR values (in dB) for 15 images at 4x and 8x undersampling or acceleration (Ax).

3.1. Results

Table 1 provides a comparison of the average PSNR values for reconstruction over the testing set at both 4x and 8x acceleration. The proposed self-guided DIP outperforms the vanilla DIP, the reference-guided DIP, compressed sensing, and even the supervised model trained on many images. Visual comparisons provided in Figs. 5 and 6 for 8x and 4x undersampling also show the benefits offered by self-guided DIP.

We also conducted experiments to understand the reconstruction of different frequencies across the three DIP-based methods. To do this, we used the same frequency band metric introduced previously. We computed this metric over 5 images for the 4x acceleration, and the average result is shown in Fig. 4. We observe that the self-guided method shows reduced spectral bias and compared to the others (high frequencies are reconstructed sooner and more accurately), and also shows less overfitting in both frequency bands considered, especially compared to vanilla DIP.

4. CONCLUSIONS

We proposed a novel self-guided deep MRI reconstruction technique that iteratively optimizes the network input while also training a reconstruction model that is robust to large random perturbations of this input. This was achieved by introducing a new regularization term that encourages the reconstructor to act as a denoiser. Empirically, we demonstrated promising results on the multi-coil FastMRI

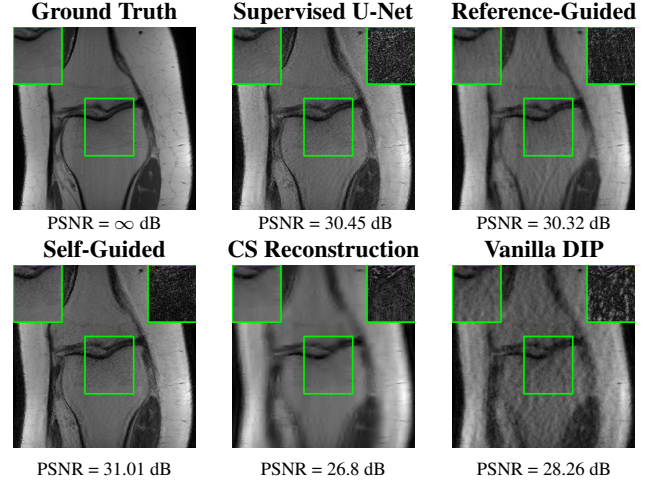


Fig. 5: Comparison of reconstructions of a knee image using the proposed self-guided method at 8x acceleration versus supervised learning, vanilla DIP, compressed sensing, and reference-guided reconstruction. The inset panel on the top left in each image corresponds to a region of interest (indicated by the green bounding box in the image) in the image, while the inset panel on the top right depicts the corresponding error map.

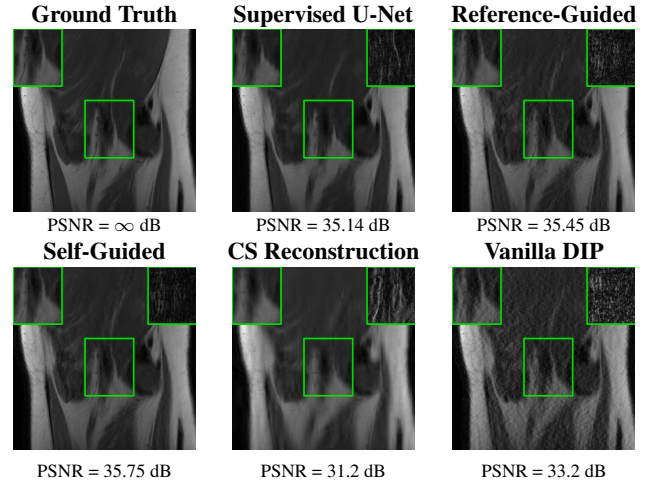


Fig. 6: Same comparisons/setup as Fig. 5, but for 4x acceleration.

data set. Notably, however, this approach does not involve any pre-training, and can thus readily handle changes in the measurement data. Moreover, this self-guided method showed better performance than the same model trained globally in a supervised manner on a large data set (with lengthy training times). This shows that highly adaptive learning approaches may have the potential to outperform traditional data-driven learning approaches in image reconstruction. In the future, we hope to carry out more theoretical analyses to better understand the performance of both vanilla and self-guided DIP for MRI reconstruction. In particular, neural tangent kernel analysis is a technique that has recently led to progress in this direction, and we believe the analysis could be extended to these cases. We also plan to study whether similar self-guided schemes could improve the performance of DIP for other tasks, such as denoising or super-resolution.

5. REFERENCES

- [1] J. A. Fessler, "Model-Based Image Reconstruction for MRI," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 81–89, 2010.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] M. Kivanc Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.
- [4] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [5] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.
- [6] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.
- [7] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1072–1086, 2012.
- [8] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2020.
- [9] A. Ghosh, M. T. McCann, and S. Ravishankar, "Bilevel learning of 11 regularizers with closed-form gradients (blorc)," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1491–1495.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.
- [11] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Im. Proc.*, vol. 26, no. 9, pp. 4509–22, Sept. 2017.
- [12] C. Feng, Y. Yan, H. Fu, L. Chen, and Y. Xu, "Task Transformer Network for Joint MRI Reconstruction and Super-Resolution," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2021.
- [13] K. Lei, M. Mardani, J. M. Pauly, and S. Vasanawala, "Wasserstein gans for mr imaging: From paired to unpaired training," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 105–115, Jan 2021.
- [14] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, 2019.
- [15] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [16] M. Z. Darestani and R. Heckel, "Accelerated MRI with untrained neural networks," *arXiv preprint arXiv:2007.02471*, 2021.
- [17] D. Zhao, F. Zhao, and Y. Gan, "Reference-driven compressed sensing mr image reconstruction using deep convolutional neural networks without pre-training," *Sensors*, vol. 20, no. 1, 2020.
- [18] J. Cohen, E. Rosenfeld, and Z. Kolter, "Certified adversarial robustness via randomized smoothing," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1310–1320.
- [19] Z. Shi, P. Mettes, S. Maji, and C. G. M. Snoek, "On measuring and controlling the spectral bias of the deep image prior," *arXiv preprint arXiv:2107.01125*, 2021.
- [20] S. Liang, A. Sreevatsa, A. Lahiri, and S. Ravishankar, "LOND-N-MRI: Adaptive Local Neighborhood-Based Networks for MR Image Reconstruction from Undersampled Data," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 2022, pp. 1–4.
- [21] H. Zhang, Y. Yu, J. Jiao, E. P. Xing, L. E. Ghaoui, and M. I. Jordan, "Theoretically principled trade-off between robustness and accuracy," *International Conference on Machine Learning*, 2019.
- [22] J. Zbontar et al, "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI," 2019, arXiv preprint arXiv:1811.08839.
- [23] F. Knoll et al, "fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning," *Radiology: Artificial Intelligence*, vol. 2, no. 1, pp. e190007, 2020.
- [24] M. Uecker, "mrrecon/bart: version 0.4.03," Apr. 2018.