

MRI RECOVERY WITH A SELF-CALIBRATED DENOISER

Sizhuo Liu*, Philip Schniter†, and Rizwan Ahmad*

*Department of Biomedical Engineering, Ohio State University, Columbus OH, 43210, USA

†Department of Electrical and Computer Engineering, Ohio State University, Columbus OH, 43210, USA

ABSTRACT

Plug-and-play (PnP) methods that employ application-specific denoisers have been proposed to solve inverse problems, including MRI reconstruction. However, training application-specific denoisers is not feasible for many applications due to the lack of training data. In this work, we propose a PnP-inspired recovery method that does not require data beyond the single, incomplete set of measurements. The proposed self-supervised method, called recovery with a self-calibrated denoiser (ReSiDe), trains the denoiser from the patches of the image being recovered. The denoiser training and a call to the denoising subroutine are performed in each iteration of a PnP algorithm, leading to a progressive refinement of the reconstructed image. For validation, we compare ReSiDe with a compressed sensing-based method and a PnP method with BM3D denoising using single-coil MRI brain data.

Index Terms— plug-and-play, unsupervised learning, self-supervised learning, MRI reconstruction, denoising.

1. INTRODUCTION

Magnetic resonance imaging (MRI) suffers from slow data acquisition, which can lead to longer scan times and compromised spatial or temporal resolution. The data acquisition for MRI occurs in the spatial Fourier domain, called k-space, and can be accelerated by prospective undersampling. Compressed sensing (CS) has been extensively applied to facilitate image recovery from highly undersampled k-space data. All major MRI vendors now offer CS-inspired reconstruction methods on their commercial scanners. More recently, deep learning (DL)-based reconstruction methods have been shown to outperform CS in terms of image quality. Typically, these methods rely on supervised learning to train a convolutional neural network (CNN) that recovers images from undersampled k-space data or aliased images [1]. Other supervised learning techniques are inspired by variational optimization methods where an iterative algorithm is unrolled and iterates between data-consistency enforcement and CNN application, which provides regularization [2].

This work was funded by the National Institute of Health grants R01HL135489 and R01EB029957.

Despite the improvements offered by DL-based methods, their extension to applications where training data are scarce remains challenging. For example, collecting fully sampled k-space data for dynamic applications with time-varying physiology is challenging. In addition, the test data may have significantly different image contrast, resolution, or anatomical features compared to the training database. This mismatch between the training and testing data can lead to performance degradation. Plug-and-play (PnP) methods, which solve an inverse problem by calling a denoising subroutine in each iteration, can circumvent this problem. PnP methods do not require training data when an off-the-shelf denoiser, e.g., BM3D [3], is used. However, more recently, it has been demonstrated that the performance of PnP can be improved by employing an application-specific denoiser, typically trained using DL [4]. These denoisers do not require fully sampled k-space data but instead can be trained on images or image patches. However, these high-quality images or image patches may still not be available for some applications, especially dynamic applications.

Recently, several methods have been proposed that utilize only undersampled k-space data for training [5, 6]. For example, using the Noise2Noise approach, Lehtinen et al. demonstrated that a denoiser or a network that performs image recovery can be trained from input and the target images that are both corrupted with k-space undersampling or noise [7]; however, acquiring the data twice for the same imaging slice is not always practical. A broader challenge with these methods is that they still require training data, and databases of undersampled MRI measurements are not readily available.

More recently, there has been an increased interest in developing self-supervised methods that require no training data at all [8, 9]. For example, Yaman et al. have proposed a self-supervised learning method that partitions the undersampled k-space into multiple subsets and trains a network that infers one subset from the other [9]. This method, however, is specific to MRI reconstruction and its extension to other inverse problems is not evident.

In this work, we propose and validate a self-supervised method, called recovery with a self-calibrated denoiser (ReSiDe), for MRI reconstruction that requires no data beyond the single set of undersampled measurements. The method is broadly applicable to other inverse problems.

2. METHOD

2.1. Compressed sensing and plug-and-play

We consider the problem of recovering an N -pixel image $\mathbf{x} \in \mathbb{C}^N$ from noisy measurements $\mathbf{y} \in \mathbb{C}^M$ collected with a known forward operator $\mathbf{A} \in \mathbb{C}^{M \times N}$, i.e., $\mathbf{y} = \mathbf{Ax} + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{C}^M$ is additive white Gaussian noise with variance τ^2 . For acceleration rate $R \triangleq N/M > 1$, \mathbf{x} cannot be uniquely inferred from \mathbf{y} . A typical CS-based formulation to infer \mathbf{x} from \mathbf{y} takes the form,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\tau^2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \phi(\mathbf{x}) \right\}, \quad (1)$$

where $\phi(\mathbf{x})$ a sparsity-inducing penalty. Typically, $\phi(\mathbf{x}) = \lambda \|\Psi \mathbf{x}\|_1$, where Ψ is the wavelet transform, or $\phi(\mathbf{x}) = \lambda \text{TV}(\mathbf{x})$, where $\text{TV}(\cdot)$ represents total variation. The tuning parameter, $\lambda > 0$, controls the regularization strength. Primal-dual splitting (PDS) [10] is a popular algorithm to solve Eq. 1 and is given in Algorithm 1, with $\mathbf{f}(\mathbf{u}) = \text{prox}_{\nu\phi}(\mathbf{u})$, which is the proximal operator defined as

$$\text{prox}_{\nu\phi}(\mathbf{u}) \triangleq \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + \frac{1}{2\nu} \|\mathbf{x} - \mathbf{u}\|_2^2 \right\}. \quad (2)$$

The proximal operator in Eq. 2 can be interpreted as the maximum a posteriori estimation of \mathbf{x} from the noisy measurement $\mathbf{u} = \mathbf{x} + \mathbf{w}$, where \mathbf{w} is additive white Gaussian noise with variance ν , and \mathbf{x} has the probabilistic prior $p(\mathbf{x}) \propto \exp(-\phi(\mathbf{x}))$. Besides PDS, the proximal operator also appears in other algorithms commonly used to solve Eq. 1, including alternating directions method of multipliers (ADMM). With this denoising viewpoint, Bouman et al. proposed replacing the proximal operator with a call to a powerful denoiser such as BM3D [11]. The approach has been called “plug-and-play” because essentially any denoiser $\mathbf{f}(\cdot)$ can be “plugged into” ADMM, PDS, or other algorithms used to solve Eq. 1.

Algorithm 1 PnP Algorithm

Require: $\nu > 0, \tau, \mathbf{A}, \mathbf{y}, f$

- 1: $\gamma = \frac{\nu}{\tau^2} \|\mathbf{A}\|_2^2, \mathbf{x}_0 = \mathbf{A}^\text{H} \mathbf{y}, \mathbf{z}_0 = \mathbf{Ax}_0 - \mathbf{y}$
- 2: **for** $t = 1, 2, 3, \dots$ **do**
- 3: $\mathbf{u}_t = \mathbf{x}_{t-1} - \frac{\nu}{\tau^2} \mathbf{A}^\text{H} \mathbf{z}_{t-1}$
- 4: $\mathbf{x}_t = \mathbf{f}(\mathbf{u}_t)$
- 5: $\mathbf{v}_t = 2\mathbf{x}_t - \mathbf{x}_{t-1}$
- 6: $\mathbf{z}_t = \frac{\gamma}{1+\gamma} \mathbf{z}_{t-1} + \frac{1}{1+\gamma} (\mathbf{Av}_t - \mathbf{y})$
- 7: **end for**
- 8: **return** $\hat{\mathbf{x}} \leftarrow \mathbf{x}_t$

Instead of using generic denoisers (e.g., BM3D or BM4D), it is possible to train application-specific denoisers based on DL and use them in PnP. It has been shown that application-specific denoisers can outperform generic denoisers [12].

Also, most generic denoisers are designed only for real-valued 2D images, with limited options available for higher-dimensional or complex-valued images. In contrast, DL denoisers can be trained and implemented for images of any dimension or size.

2.2. Recovery with a self-calibrated denoiser (ReSiDe)

Training an application-specific denoiser, $\mathbf{f}(\cdot)$, in Algorithm 1 relies on supervised learning, which requires access to clean images. Once trained, the denoiser acts as an implicit prior and can facilitate image recovery when $R > 1$. However, clean and high-quality images may not be available for many applications. In response, we propose a PnP-inspired method that iteratively trains the denoiser from the image being recovered. The resulting method, called ReSiDe, is implemented using a modified PDS algorithm (Algorithm 2).

Algorithm 2 ReSiDe Algorithm

Require: $\nu > 0, \tau, \mathbf{A}, \mathbf{y}, f$

- 1: $\gamma = \frac{\nu}{\tau^2} \|\mathbf{A}\|_2^2, \mathbf{x}_0 = \mathbf{A}^\text{H} \mathbf{y}, \mathbf{z}_0 = \mathbf{Ax}_0 - \mathbf{y}$
- 2: **for** $t = 1, 2, 3, \dots$ **do**
- 3: $\tilde{\mathbf{x}}_{t-1} = \mathbf{x}_{t-1} + \mathcal{N}(\mathbf{0}, \sigma_t^2 \mathbf{I})$
- 4: $\theta_t = \text{argmin}_{\theta} \sum_{i=1}^P \mathcal{L}(f(\mathcal{I}[\tilde{\mathbf{x}}_{t-1}]_i; \theta), \mathcal{I}[\mathbf{x}_{t-1}]_i)$
- 5: $\mathbf{u}_t = \mathbf{x}_{t-1} - \frac{\nu}{\tau^2} \mathbf{A}^\text{H} \mathbf{z}_{t-1}$
- 6: $\mathbf{x}_t = f(\mathbf{u}_t; \theta_t)$
- 7: $\mathbf{v}_t = 2\mathbf{x}_t - \mathbf{x}_{t-1}$
- 8: $\mathbf{z}_t = \frac{\gamma}{1+\gamma} \mathbf{z}_{t-1} + \frac{1}{1+\gamma} (\mathbf{Av}_t - \mathbf{y})$
- 9: **end for**
- 10: **return** $\hat{\mathbf{x}} \leftarrow \mathbf{x}_t$

In Algorithm 2, the current estimate, \mathbf{x}_{t-1} , is contaminated with complex-valued, zero-mean white Gaussian noise of variance σ_t^2 to generate a noisy image, $\tilde{\mathbf{x}}_{t-1}$ (Line 3). Then, we train a network by feeding it $P \geq 1$ pairs of noisy-clean patches as input-output (Line 4). The noisy patches are extracted from $\tilde{\mathbf{x}}_{t-1}$, while the clean patches are extracted from \mathbf{x}_{t-1} . The locations of the patches within \mathbf{x}_{t-1} or $\tilde{\mathbf{x}}_{t-1}$ are selected randomly. The patch extraction is performed by the operator \mathcal{I} , with $\mathcal{I}[\cdot]_i$ representing the i^{th} patch. The network is parameterized by θ , with θ_t representing the network parameters at the t^{th} iteration. In each iteration, the network is trained de novo from a random initialization of the network parameters. Finally, we use the trained network to denoise \mathbf{u}_t (Line 6). Here, the operator $\mathcal{L}(\cdot)$ defines the loss function used in training. We chose $\mathcal{L}(f(\mathcal{I}[\tilde{\mathbf{x}}_{t-1}]_i; \theta), \mathcal{I}[\mathbf{x}_{t-1}]_i) = \|f(\mathcal{I}[\tilde{\mathbf{x}}_{t-1}]_i; \theta) - \mathcal{I}[\mathbf{x}_{t-1}]_i\|_2^2$.

3. EXPERIMENT AND RESULTS

3.1. Experimental settings

For validation, two brain datasets are used, i.e., a 320×320 T1-weighted and a 384×384 T2-weighted brain MRI. The

multi-coil k-space data were downloaded from fastMRI (<https://fastmri.org/>) and converted to complex-valued single-coil images using ESPIRiT-based sensitivity maps [13]. The k-space corresponding to the single-coil image was undersampled at $R = 1.8$ using two different Cartesian sampling masks: variable density pseudo-random sampling with undersampling in only one direction (M1), shown in Figures 1 and 3, and random sampling with undersampling in both directions (M2), shown in Figures 2 and 4. For both sampling patterns, 32 central lines were fully sampled.

The denoiser architecture is similar to one we have previously reported for supervised training [14]. In summary, the network architecture is inspired by DnCNN [12], with five convolutional layers, four ReLu layers in between, and a skip connection. The first four convolutional layers have $64 \times 3 \times 3$ kernels, and the last convolutional layer has two 3×3 kernels. We split real and imaginary parts into two channels.

In each iteration, we selected $P = 144$ patches of size 64×64 . We decreased the noise standard deviation, σ_t , after every ten iterations such that the training signal-to-noise ratio, $\text{SNR}_t = 20 \log_{10}(\|\mathbf{x}_t\|_2 / (\sqrt{2N}\sigma_t))$, increased from 10 dB to 40 dB in the increments of 5 dB after every ten iterations, i.e., $\text{SNR}_t = 10 + 5 \lfloor (t - 1)/10 \rfloor$, where $\lfloor k \rfloor$ rounds k to the nearest integer less than or equal to k . The training was performed on a workstation equipped with a single NVIDIA GPU (GeForce RTX 2080 Ti). We set the minibatch size to 16 and used the Adam optimizer with a learning rate of 1×10^{-3} . The number of epochs was fixed at 100, and the algorithm was executed for 70 iterations.

To compare ReSiDe with other unsupervised methods, we also recovered the images using a CS method, called L1-wavelet, and PnP-BM3D. The L1-wavelet reconstruction was performed using the BART toolbox (<https://mrirecon.github.io/bart/>), while PnP-BM3D was performed using Algorithm 1, with $f(\cdot)$ representing a call to the BM3D denoiser, which was applied separately to real and imaginary parts of the image. BM3D also takes noise variance as an input; this parameter was manually tuned and left fixed for all experiments. The performance evaluation was based on NMSE defined as $20 \log_{10}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2)$.

3.2. Results

Figures 1–4 show the reconstructions for T1 and T2 images from two different sampling patterns, M1 and M2. The first row includes the true image from fully sampled k-space, zero-filled reconstruction, and images from L1-wavelet, PnP-BM3D, and ReSiDe reconstructions. The second row shows the sampling pattern along with the corresponding error maps. Table 1 summarizes NMSE values from different reconstruction methods. Overall, ReSiDe outperforms L1-wavelet and PnP-BM3D in all four cases, with an average advantage of 3.06 dB over L1-wavelet and 0.91 dB over PnP-BM3D. All

three reconstruction methods perform better for the sampling pattern that is less structured (M2). Figure 5 shows the convergence curve of ReSiDe under three different SNR_t settings, i.e., a fixed SNR_t of 10 dB, a fixed SNR_t of 25 dB, and SNR_t that progressively increases over iterations.

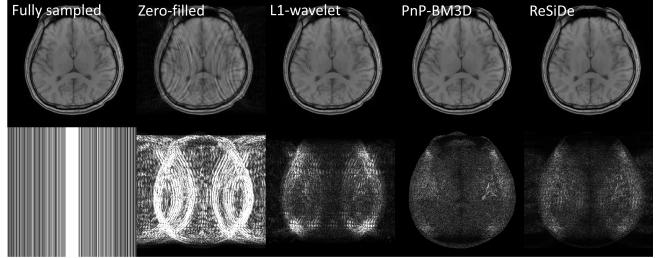


Fig. 1: T1-weighted brain image reconstruction using the structured sampling (M1). The second row shows the absolute error map after 1.5-fold amplification.

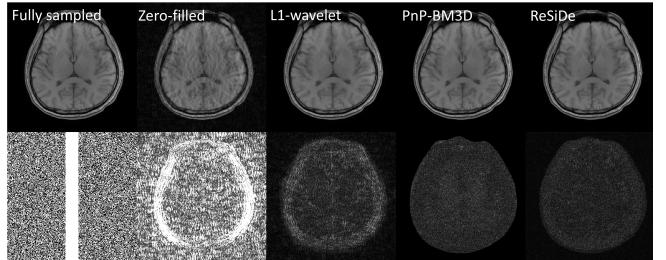


Fig. 2: Same as Figure 1 but with the random sampling (M2).

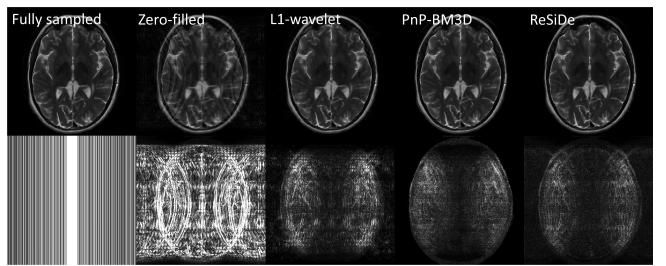


Fig. 3: T2-weighted brain image reconstruction using the structured sampling (M1). The second row shows the absolute error map after 1.5-fold amplification.

4. DISCUSSION

MRI reconstruction methods have evolved rapidly in the last two decades, with the DL-based methods taking over CS-based methods in the latest, ongoing efforts. Most DL methods used for MRI recovery require access to a large corpus of high-quality training data in the form of raw k-space data or images. There are several MRI applications, e.g., contrast-based dynamic applications, where obtaining such training

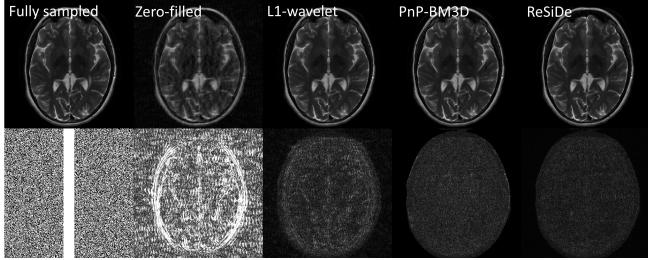


Fig. 4: Same as Figure 3 but with the random sampling (M2).

Contrast	Sampling	Zero-filled	L1-wavelet	PnP-BM3D	ReSiDe
T1	M1	-13.50	-24.31	-27.37	-27.59
T1	M2	-12.97	-26.97	-28.82	-29.96
T2	M1	-11.86	-21.67	-23.60	-24.08
T2	M2	-11.52	-24.04	-25.81	-27.59

Table 1: NMSE (dB) for the T1 and T2-weighted brain MRI data undersampled with M1 and M2 sampling patterns.

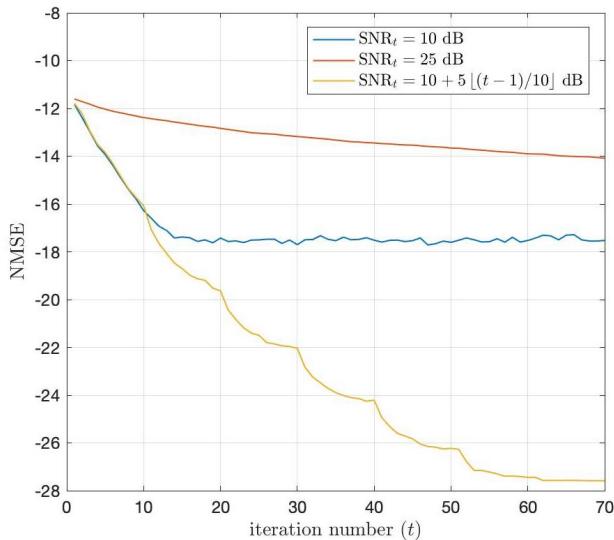


Fig. 5: Change in NMSE of ReSiDe with the iteration number. The results are shown for the T2-weighted image with random sampling pattern (M2) and three different SNR_t settings. Similar behavior was observed for other three combinations of sampling and image contrast.

data is challenging or infeasible. To facilitate such applications, we propose a self-calibrated scheme called ReSiDe. In contrast to the traditional PnP approach, where an off-the-shelf or an application-specific trained denoiser is used, ReSiDe iteratively trains the denoiser from the image being recovered, obviating the need for training data while enabling image-specific training of the denoiser. For validation, ReSiDe is applied to single-coil MRI data with two different contrasts and two different sampling patterns.

We observed ReSiDe to converge for a wide range of σ_t values. However, the speed of convergence and the final

NMSE were greatly impacted by σ_t . Selecting a fixed, low training SNR_t (e.g., 10 dB) improved the convergence speed but led to lower final NMSE and generated overly smooth images. In contrast, selecting a fixed, high training SNR_t (e.g., 25 dB) improved the final NMSE but offered a very slow convergence. To improve both the convergence speed and the final NMSE, we adopted a scheme where the starting SNR_t of 10 dB was increased by 5 dB after every ten iterations (Figure 5).

A major limitation of ReSiDe is the reconstruction time, with each iteration taking approximately 3-4 minutes on a workstation with a single GPU. ReSiDe shares this limitation with other training-free DL methods such as deep image prior [15]. The computation speed of ReSiDe can be potentially improved by training the denoiser after every k^{th} iteration, with $k > 1$, simplifying the denoiser architecture, retaining network weights from one iteration to the next, and transitioning σ_t based on the relative change in the image, i.e., $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2 / \|\mathbf{x}_t\|_2$. We also believe that the performance of ReSiDe can be further improved by optimizing the denoiser architecture and employing a different loss function. The architecture used in this work is identical to the one we previously used to train an application-specific denoiser from a large number of training patches. Optimizing the denoiser architecture and evaluating various loss functions as well as application to multi-coil MRI will be considered in future studies. Also, we will target dynamic applications, including cardiac MRI, and compare ReSiDe to a wider range of self-supervised methods.

In this work, we have utilized a specific method to train the denoiser, i.e., we first add noise to \mathbf{x}_{t-1} and then train the denoiser to remove this noise [16, 17]. Several other “blind denoising” methods, e.g., Stein’s Unbiased Risk Estimator (SURE)-based methods [18, 19], Noise2Void [20], Noise2Self [21], and Noise2Inverse [22], have been proposed that can also train a network (denoiser) from a single noisy image. In general, these methods employ either SURE or cross-validation to train the network. In the case of latter, the network is typically trained by estimating one subset of pixels from another subset. We believe many of these self-calibrated denoisers, including Noise2Void and Noise2Self, can also be used within ReSiDe and may offer faster convergence or lower NMSE.

5. CONCLUSION

A self-calibrated method is proposed for MRI reconstruction. The proposed method, called ReSiDe, outperforms other training-free methods, namely L1-wavelet and PnP-BM3D, in terms of image quality. Although ReSiDe is described and applied to single-coil MRI data, it is potentially applicable to multi-coil MRI and other inverse problems.

6. REFERENCES

- [1] Guang Yang et al., “DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.
- [2] Kerstin Hammernik et al., “Learning a variational network for reconstruction of accelerated MRI data,” *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [3] Aram Danielyan et al., “BM3D frames and variational image deblurring,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1715–1728, 2011.
- [4] Sizhuo Liu et al., “Free-breathing cardiovascular MRI using a plug-and-play method with learned denoiser,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1748–1751.
- [5] Jonathan Tamir et al., “Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data,” in *Proc. Intl. Soc. Mag. Reson. Med*, 2019, vol. 27, p. 0660.
- [6] Elizabeth Cole et al., “Unsupervised MRI reconstruction with generative adversarial networks,” *arXiv preprint arXiv:2008.13065*, 2020.
- [7] Jaakko Lehtinen et al., “Noise2Noise: Learning image restoration without clean data,” *arXiv preprint arXiv:1803.04189*, 2018.
- [8] Kyong Jin et al., “Time-dependent deep image prior for dynamic MRI,” *arXiv e-prints*, pp. arXiv–1910, 2019.
- [9] Burhaneddin Yaman et al., “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.
- [10] Shunsuke Ono, “Primal-dual plug-and-play image restoration,” *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1108–1112, 2017.
- [11] Venkatakrishnan et al., “Plug-and-play priors for model based reconstruction,” in *2013 IEEE GlobalSIP*. IEEE, 2013, pp. 945–948.
- [12] Kai Zhang et al., “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [13] Martin Uecker et al., “ESPIRiT—An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA,” *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014.
- [14] Rizwan Ahmad et al., “Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery,” *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 105–116, 2020.
- [15] Dmitry Ulyanov et al., “Deep image prior,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [16] Magauiya Zhussip et al., “Extending Stein’s unbiased risk estimator to train deep denoisers with correlated pairs of noisy images,” *Advances in neural information processing systems*, vol. 32, pp. 1465–1475, 2019.
- [17] Jun Xu et al., “Noisy-as-clean: learning self-supervised denoising from corrupted image,” *IEEE Transactions on Image Processing*, vol. 29, pp. 9316–9329, 2020.
- [18] Christopher A Metzler et al., “Unsupervised learning with Stein’s unbiased risk estimator,” *arXiv preprint arXiv:1805.10531*, 2018.
- [19] Hemant K Aggarwal et al., “Ensure: Ensemble Stein’s unbiased risk estimator for unsupervised learning,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1160–1164.
- [20] Alexander Krull et al., “Noise2Void-learning denoising from single noisy images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2129–2137.
- [21] Joshua Batson et al., “Noise2Self: Blind denoising by self-supervision,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 524–533.
- [22] Allard Hendriksen et al., “Noise2Inverse: Self-supervised deep convolutional denoising for tomography,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1320–1335, 2020.