ECHO: Estimated Composite Hybrid Observation for Zero-Shot Self-Supervised MRI Reconstruction

Wenlei Shang[1], Wenjian Liu[1], Peng Hu[1,2], and Zijian Zhou[1]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
zhouzj@shanghaitech.edu.cn

Abstract. Zero-shot self-supervised learning (ZS-SSL) for MRI reconstruction optimizes network parameters directly on the acquired undersampled measurements, eliminating the need for large training datasets. This paradigm mitigates domain shift issues commonly observed in database trained models when encountering unseen sampling trajectories, acceleration rates, contrasts, or anatomies, while offering improved interpretability relative to other zero-shot approaches. Nonetheless, current ZS-SSL methods face three major limitations: (1) data-splitting strategies introduce training-inference mismatches in effective acceleration rates, leading to suboptimal reconstruction; (2) over-parameterized networks incur substantial computational overhead and increase overfitting risk; and (3) the inherently limited supervision further constrains achievable performance. We propose ECHO (Estimated Composite Hybrid Observation), another ZS-SSL MRI reconstruction framework designed to address these challenges. Instead of partitioning acquired k-space data, ECHO constructs training inputs from intermediate reconstructions that are statistically consistent with inference-time observations, thereby eliminating acceleration mismatch. Moreover, ECHO employs an under-parameterized, physics-informed architecture tailored to the zero-shot regime, significantly reducing computational complexity while maintaining reconstruction accuracy. Extensive experiments demonstrate that ECHO consistently improves reconstruction quality and achieves substantial training speedup. Our code is publicly available at https://github.com/ShanghaiTech-Hu-Lab/ECHO

Keywords: MRI reconstruction · Zero-shot self-supervised learning · Self-supervised learning

1 Introduction

Magnetic Resonance Imaging (MRI) reconstruction from undersampled k-space data is a classical inverse problem governed by well-defined physical models.

Accelerated MRI aims to recover high-quality images from incomplete measurements to reduce scan time while preserving diagnostic fidelity. However, undersampling renders the problem highly ill-posed, motivating decades of methodological development, including parallel imaging and compressed sensing techniques [9, 16, 8, 20, 15]. More recently, deep learning, particularly physics-guided unrolled networks, has substantially advanced reconstruction performance by embedding the MRI forward model into iterative optimization frameworks [28, 22, 1]. Despite their effectiveness, these data-driven methods typically require large training datasets and often exhibit limited generalization to unseen sampling patterns, acceleration factors, or contrasts.

Zero-shot MRI reconstruction addresses this limitation by optimizing networks directly on the acquired undersampled measurements. Representative approaches include Deep Image Prior (DIP) [23, 13], Implicit Neural Representations (INR) [12, 19], and Zero-Shot Self-Supervised Learning (ZS-SSL) [25, 24]. Among these, ZS-SSL is particularly appealing due to its explicit integration of MRI physics through model-based data consistency combined with learned regularization. However, existing ZS-SSL methods inherit the data-splitting strategy and large network architectures from SSDU (Self-Supervised Learning via Data Undersampling) [26]. Data splitting introduces a mismatch between training and inference distributions, especially in effective acceleration rates, thereby leading to suboptimal reconstruction. Additionally, over-parameterized networks are ill-suited to the extremely limited data available in zero-shot settings, increasing computational cost and overfitting risk.

To overcome these limitations, we propose ECHO (Estimated Composite Hybrid Observation) for ZS-SSL MRI reconstruction. By constructing inputs statistically consistent with inference-time observations, ECHO eliminates data-splitting and training-inference mismatch, achieving high reconstruction performance. We also demonstrate that an under-parameterized physics-based network is particularly suitable for zero-shot reconstruction, and systematically analyze the impact of the number of randomized secondary masks on reconstruction performance.

2 Preliminaries & Related Work

2.1 Physics-based MRI Reconstruction

In accelerated MRI, the forward operator is typically ill-conditioned due to undersampling, necessitating regularization for stable image recovery. The reconstruction problem is commonly formulated as

$$x^* = \arg\min_x \frac{1}{2}\|A_\Omega S x - y_\Omega\|_2^2 + \lambda R(x), \tag{1}$$

where y_Ω is the acquired undersampled k-space measurements, S denotes coil sensitivity maps, and $A_\Omega = M_\Omega F$ is the encoding operator with Fourier transform F and undersampling mask M_Ω. The regularizer $R(x)$ introduces prior information, and λ balances data consistency and regularization.

Physics-guided deep learning methods unroll iterative optimization algorithms such as ADMM [4] and ISTA [2] into trainable networks. These methods achieve strong interpretability and competitive performance with relatively shallow convolutional architectures [22] [28].

2.2 Zero-shot MRI Reconstruction

Among zero-shot MRI reconstruction methods, DIP-based methods [23, 13] reconstruct images by mapping fixed random noise to the target image through a convolutional network, with the network architecture acting as an implicit prior. Self-guided DIP [13] further improves performance by replacing the fixed noise with trainable inputs and jointly optimizing both inputs and network parameters. INR methods model images as continuous coordinate-based functions, with architectures such as sinusoidal representation networks (SIREN) [21] and Instant Neural Graphics Primitives (Instant-NGP) [19] that improve high-frequency modeling and convergence through periodic activations and learnable encodings.

ZS-SSL [25] is commonly built on SSDU [26]. It partitions the acquired undersampled k-space into three subsets: one used as the network input, one as the supervision label, and a third for self-validation to establish a stopping criterion and prevent overfitting. Because training relies only on a subset of the acquired measurements, whereas inference uses all acquired k-space, a mismatch arises between training and inference, especially in the effective acceleration rate. Moreover, the self-validation strategy further reduces the amount of data available for training, which may degrade reconstruction performance. SPICER [10] was proposed to mitigate this discrepancy by acquiring multiple undersampled measurements using masks drawn from the same distribution. However, such multi-acquisition strategy may be impractical in clinical settings due to extended scan time and increased sensitivity to motion.

3 Method

To avoid data-splitting and extended scanning time, ECHO tends to use intermediate reconstruction and secondary undersampling for model training, relaxing the requirement of differently undersampled k-spaces. Therefore, similar to a Siamese network [5], ECHO consists of an estimator network and a target network, sharing the same architecture and weights.

Given an acquired undersampled k-space, the estimator produces an estimate of the fully sampled k-space. Then an explicit data-consistency (DC) projection is applied, yielding a composite hybrid observation that combines estimated and acquired k-space measurements. This estimation step solely refines the missing k-space entries and is performed without gradient backpropagation. Therefore, in the worst case, the hybrid observation reduces to the noisy zero-filled image used in Noise2Recon [6], guaranteeing stable training from scratch.

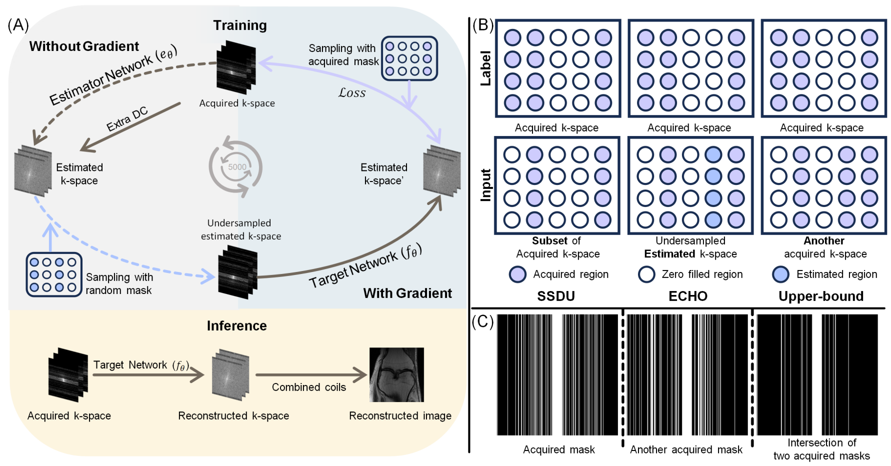

Fig. 1. (A) Overview of the training scheme of the proposed ECHO method. (B) Schematic illustration of the input–label pairs used in SSDU, ECHO, and the upper-bound experiments. (C) Examples of undersampling masks.

The hybrid observation is then re-undersampled using a different mask independently drawn from the same distribution as the acquisition mask with the same acceleration rate, before fed into the target network. The target network is optimized by minimizing the loss between its output and the corresponding acquired k-space sample. After backpropagation, the estimator network copies the updated parameters, enabling the next estimation step.

During training, ECHO alternates between the estimator network (for k-space refinement) and target network (for parameter optimization). Once training is finished, the target network takes the acquired k-space as input and directly outputs the full k-space, which is used to produce the final image after Fourier transform and coil combination. The general framework and data flow of ECHO are illustrated in Fig. 1(A).

3.1 Intuitions on ECHO's Behavior

We hypothesize that ECHO behaves similarly to alternating optimization (AO) algorithms [3]. Prior self-supervised works [25, 18, 11] have shown that using the acquired measurements as training targets is the only way to obtain an unbiased estimate of supervised learning, and that multiple undersampled k-spaces of a fully sampled image constitute the ideal inputs for training to reduce the training-inference gap [10]. For zero-shot, the only acquired measurement is directly used as the target. However, differently undersampled k-spaces derived from a fully sampled image are unavailable. ECHO introduces an estimated fully sampled image $\hat{x}$ as an explicit auxiliary variable, leading to the joint optimization problem $\min_{\hat{x},\theta} \|A_\Omega f_\theta(\hat{x}_{\hat{\Omega}}) - y_\Omega\|_2^2$, where f_θ denotes the reconstruction network, and $\hat{\Omega}$ the randomized secondary mask with the same distribution as Ω. This formulation naturally admits an alternating optimization strategy in which $\hat{x}$ and θ are updated in turn. While direct update of $\hat{x}$ is rather complicated, we leverage the prior [10] that suitable network inputs should correspond to undersampled k-spaces of a fully sampled image and approximate it as $\hat{x}^{(t)} = f_{\theta^{(t-1)}}(y_\Omega)$. With $\hat{x}^{(t)}$ fixed, the update of θ reduces to standard network training via stochastic gradient descent, with gradients stopped through $\hat{x}^{(t)}$.

In practice, the AO mechanism transitions ECHO from the SSDU setting to the upper-bound setting, where different real-acquired k-spaces are indeed available, as shown in Fig. 1 (B) and (C). Compared with splitting the acquired k-space and contrasting two zero-filled images in SSDU, ECHO iteratively refines the zero-filled image toward a fully sampled image, thereby reducing the distribution gap between training and inference inputs with respect to both undersampling rates and patterns.

3.2 Implementation Details

The base network consists of 4 convolutional layers, each followed by a PReLU activation except for the final layer, with 32 hidden channels. The depth and width of the base network are selected to yield a parameter count comparable to that of ISTA-Net$^+$ [28], thereby limiting overfitting risk in the zero-shot setting. The network directly operates on multi-coil images for both input and output, avoiding explicit coil combination and expansion before and after the data-consistency module, which improves computational efficiency. This design further allows the data-consistency step to admit a closed-form solution. The reconstruction backbone is unrolled for 5 iterations, and the model is trained for 5,000 steps. These hyperparameters were selected based on validation in one subject with 10 slices.

We evaluated our method on the fastMRI knee and brain datasets [27]. Twenty subjects were randomly selected and each subject contained 10 slices. Gaussian-distributed undersampling masks were used. The auto-calibration region contained 24 or 16 center k-space lines, for acceleration factors of $4\times$ and $6\times$, respectively. Reconstruction performance was evaluated using SSIM and PSNR against the ground truth images.

4 Comparison to Existing Zero-shot Approaches

Performance of ECHO was compared with Total Variation (TV), vanilla DIP, self-guided DIP, SIREN, and Instant-NGP. Quantitative results are summarized in Table 1, where ECHO consistently achieves the highest SSIM and PSNR across all settings. Representative reconstructions are shown in Fig. 2 (A), demonstrating that ECHO achieves improved reconstruction than competing methods.

Because fastMRI has fully sampled k-space, to verify the effectiveness of using the hybrid k-space for reconstruction, we compared different ZS-SSL training

Table 1. The reconstruction performance on the fastMRI knee and brain.

	Knee				Brain			
	4×		6×		4×		6×	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
TV	0.8793	33.36	0.8396	31.84	0.9338	33.89	0.9063	31.07
Vanilla DIP	0.8503	32.45	0.8193	31.12	0.9236	33.65	0.9079	32.16
Self-Guided DIP	0.8774	33.60	0.8299	31.71	0.9177	34.20	0.8889	32.27
SIREN	0.8894	34.10	0.8517	32.20	0.9375	34.50	0.9125	32.22
Instant-NGP	0.8915	34.32	0.8588	32.44	0.9281	34.30	0.9094	32.09
ECHO	**0.9107**	**35.65**	**0.8744**	**33.04**	**0.9420**	**35.64**	**0.9219**	**33.27**

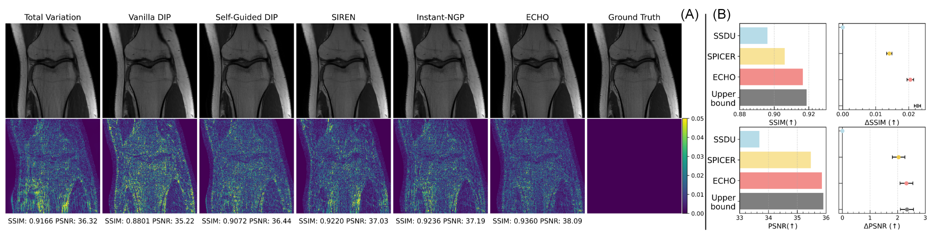

Fig. 2. (A) Reconstructed 4 × undersampled knee images with SSIM and PSNR values. ECHO achieves the highest image quality. (B) ZS-SSL performance with different input constructions; mean gains over SSDU with 95 % confidence intervals indicate statistical significance.

paradigms, including SSDU, SPICER, and the upper-bound setting by constructing different k-space inputs. While SSDU uses subsets of the acquired k-space, SPICER uses two fixed differently undersampled k-spaces, and the upper-bound setting uses real-acquired k-spaces with randomized masks. Another undersampled k-space is used as the label to avoid autoregressive.

As shown in Fig. 2 (B), both ECHO and the upper-bound setting outperform SPICER, indicating that using different masks drawn from the same distribution instead of using only the single acquired mask is not the performance bottleneck. Moreover, the small performance gap between ECHO and the upper-bound suggests that the hybrid k-space can effectively substitute for real-acquired k-space during training.

5 Ablation Studies and Mechanistic Analysis

To gain a deeper understanding of ECHO and the factors contributing to its performance, a series of ablation studies were conducted on the fastMRI knee data.

5.1 Training Dynamics of the Hybrid K-space

As shown in previous sections, ECHO achieved the highest reconstruction performance with a small gap compared to the upper-bound setting. We hypothesize that this behavior arises because ECHO progressively refines the network inputs towards the ideal training input, namely the undersampled fully sampled image [10]. To validate this hypothesis, the evolution of the hybrid k-space during training is analyzed. As shown in Fig. 3 (A), the hybrid k-space initially resembles noisy zero-filled reconstructions with strong aliasing but progressively approaches the fully sampled reference. Consistently, Fig. 3 (B) visualizes the feature distributions of the hybrid k-space using DreamSim embeddings [7, 14] and t-SNE [17]. The features of the hybrid k-space progressively move toward those of the fully sampled reference. Meanwhile, after applying undersampling masks, the features of the undersampled hybrid k-space cluster around those of the inference k-space input, demonstrating that the training–inference distributions are matched. In addition, the under-parameterized physics-based network of ECHO effectively avoided overfitting according to Fig. 3 (C), where SSIM and PSNR remain stable across 200 slices throughout training.

5.2 Impact of Randomized Masks

In ZS-SSL MRI reconstruction, the measurements available are fixed, and data diversity is mainly determined by the randomized undersampling masks used during training. We therefore hypothesize that the number of randomized masks plays a critical role, complementary to the mask distribution. To verify this, we varied the number of random masks used for composite hybrid k-space undersampling, and compared Gaussian masks with uniform masks of the same acceleration. Each experiment is repeated three times for stability. As shown in Fig. 3 (D), reconstruction performance improves with more masks used and gradually saturates; and Gaussian masks consistently outperform uniform masks across all settings.

5.3 Robustness to Network Architectures

Selecting network hyper-parameters can be challenging for most deep learning methods. Although larger models often improve performance in supervised learning, they may not have the benefits of ZS-SSL and can increase the risk of overfitting because of the only one available training data. In contrast, physics-based networks can achieve strong performance with significantly fewer parameters. We therefore hypothesize that ECHO is less sensitive to the reconstruction backbone, as its performance is mainly constrained by input construction rather than model capacity. We evaluated models with 3, 4, 5 unrolled layers, hidden channel sizes of 32 and 64, and used ISTA-Net$^+$ as a reference. As shown in Fig. 3 (E), though careful hyperparameter tuning may yield minor performance gains, ECHO maintains stable SSIM and PSNR across different model capacities, verifying our hypothesis.

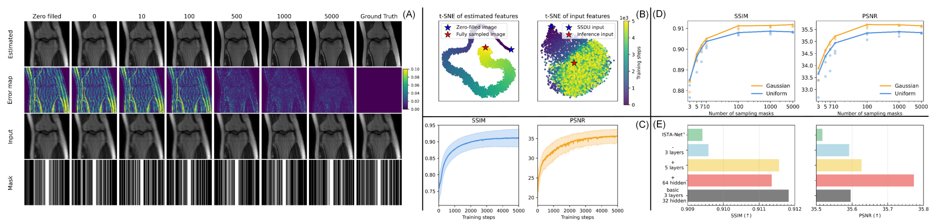

Fig. 3. (A) Estimated images, corresponding error maps, undersampled estimated images used as network inputs, and the associated undersampling masks at different training steps. (B) t-SNE visualization of feature embeddings, computed using DreamSim, for the estimated images and network inputs. (C) Average SSIM and PSNR over 200 slices throughout the entire training process. (D) Influence of the number and type of secondary undersampling masks on reconstruction performance. (E) Influence of network hyperparameter on reconstruction performance.

6 Conclusion

In conclusion, we introduced ECHO, a different ZS-SSL framework for MRI reconstruction that avoids data splitting. By constructing hybrid k-spaces from the intermediate reconstruction of a single undersampled acquisition, input-label pairs with sufficient supervision effect can be obtained for network training. The constructed inputs also eliminate training-inferenece mismatch that is typically encountered in self-supervised methods, leading to superior reconstruction performance of ECHO.

Acknowledgments. This study was funded in part by National Natural Science Foundation of China (Grant/Award Number: 82302295); Beijing Natural Science Foundation (Grant/Award Number: 1S24094).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE transactions on medical imaging **38**(2), 394–405 (2018)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences **2**(1), 183–202 (2009)
3. Bezdek, J.C., Hathaway, R.J.: Convergence of alternating optimization. Neural, Parallel & Scientific Computations **11**(4), 351–368 (2003)
4. Chan, S.H., Wang, X., Elgendy, O.A.: Plug-and-play admm for image restoration: Fixed-point convergence and applications. IEEE Transactions on Computational Imaging **3**(1), 84–98 (2016)

5. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15750–15758 (2021)
6. Desai, A.D., Ozturkler, B.M., Sandino, C.M., Boutin, R., Willis, M., Vasanawala, S., Hargreaves, B.A., Re, C., Pauly, J.M., Chaudhari, A.S.: Noise2recon: enabling snr-robust mri reconstruction with semi-supervised and self-supervised learning. Magnetic Resonance in Medicine **90**(5), 2052–2070 (2023)
7. Fu, S., Tamir, N., Sundaram, S., Chai, L., Zhang, R., Dekel, T., Isola, P.: Dreamsim: Learning new dimensions of human visual similarity using synthetic data. arXiv preprint arXiv:2306.09344 (2023)
8. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (grappa). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **47**(6), 1202–1210 (2002)
9. Hollingsworth, K.G.: Reducing acquisition time in clinical mri by data undersampling and compressed sensing reconstruction. Physics in Medicine & Biology **60**(21), R297 (2015)
10. Hu, Y., Gan, W., Ying, C., Wang, T., Eldeniz, C., Liu, J., Chen, Y., An, H., Kamilov, U.S.: Spicer: Self-supervised learning for mri with automatic coil sensitivity estimation and reconstruction. Magnetic resonance in medicine **92**(3), 1048–1063 (2024)
11. Klug, T., Atik, D., Heckel, R.: Analyzing the sample complexity of self-supervised image reconstruction methods. Advances in Neural Information Processing Systems **36**, 65869–65893 (2023)
12. Kunz, J.F., Ruschke, S., Heckel, R.: Implicit neural networks with fourier-feature inputs for free-breathing cardiac mri reconstruction. IEEE Transactions on Computational Imaging (2024)
13. Liang, S., Bell, E., Qu, Q., Wang, R., Ravishankar, S.: Analysis of deep image prior and exploiting self-guidance for image reconstruction. IEEE Transactions on Computational Imaging (2025)
14. Lin, K., Krainovic, A., Wang, K., Heckel, R.: Improving deep learning for accelerated mri with data filtering. arXiv preprint arXiv:2508.13822 (2025)
15. Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed sensing mri. IEEE signal processing magazine **25**(2), 72–82 (2008)
16. Lustig, M., Pauly, J.M.: Spirit: iterative self-consistent parallel imaging reconstruction from arbitrary k-space. Magnetic resonance in medicine **64**(2), 457–471 (2010)
17. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
18. Millard, C., Chiew, M.: A theoretical framework for self-supervised mr image reconstruction using sub-sampling via variable density noisier2noise. IEEE transactions on computational imaging **9**, 707–720 (2023)
19. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG) **41**(4), 1–15 (2022)
20. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: sensitivity encoding for fast mri. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **42**(5), 952–962 (1999)
21. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)

22. Sun, J., Li, H., Xu, Z., et al.: Deep admm-net for compressive sensing mri. Advances in neural information processing systems **29** (2016)
23. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018)
24. Velasquez, A., Jha, S., Alkhouri, I.R.: On the dataless training of neural networks. arXiv preprint arXiv:2510.25962 (2025)
25. Yaman, B., Hosseini, S.A.H., Akçakaya, M.: Zero-shot self-supervised learning for mri reconstruction. arXiv preprint arXiv:2102.07737 (2021)
26. Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M.: Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magnetic resonance in medicine **84**(6), 3172–3191 (2020)
27. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839 (2018)
28. Zhang, J., Ghanem, B.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1828–1837 (2018)