Evaluation of Supervision in Deep Learning networks for accelerated-MRI reconstruction

Iason Skylitsis
Department of Biomedical Engineering & Physics
Amsterdam UMC
Amsterdam, The Netherlands
iason.sky@gmail.com

Dimitrios Karkalousos
Department of Biomedical Engineering & Physics
Amsterdam UMC
Amsterdam, The Netherlands
d.karkalousos@amsterdamumc.nl

Matthan W. A. Caan
Department of Biomedical Engineering & Physics
Amsterdam UMC
Amsterdam, The Netherlands
m.w.a.caan@amsterdamumc.nl

***Abstract*—Supervised methods typically require fully sampled data for training, which can be difficult and time-consuming to acquire. To address this limitation, self-supervised methods have been proposed in the literature as an alternative approach. In this study, we investigate the efficiency and reconstruction quality of different supervised and self-supervised methods for Magnetic Resonance Imaging (MRI) reconstruction. Furthermore, we investigate the efficacy of using a pre-trained model for inference, as well as the individual contribution of each method's components to the final results. We extensively compare a pre-trained Cascades of Independently Recurrent Inference Machines (CIRIM) to a range of other methods, including self-supervised methods such as Noise2Recon and Self-Supervision via Data Undersampling (SSDU), supervised methods like U-Net and ResNet, and traditional methods such as Compressed Sensing and Proximal Gradient. We evaluate the performance of these methods in terms of reconstruction quality, using metrics such as SSIM, PSNR, MSE, and NMSE. Our results indicate that while the models trained on the given dataset (Stanford 3D FSE Knees) show slightly higher performance, the pre-trained CIRIM model performs comparably and yields high-quality reconstructions. Our findings suggest that a network like CIRIM, pre-trained on high signal-to-noise (SNR) acquisitions, can generalize well to new datasets and provide high-quality reconstructions.**

***Index Terms*—self-supervised, supervised, MRI reconstruction**

I. Introduction

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that produces high-resolution images of the organs and the tissues of the body. The main limiting factor of MRI is the long scanning times, which in a clinical setting are of utmost importance. To reduce the prolonged scanning times several techniques were introduced such as parallel imaging (PI) [20], sensitivity encoding (SENSE) [16], and compressed sensing (CS) [15].

With the rise of Machine Learning (ML), several new methods of MRI reconstruction emerged to accelerate the process even further. ML methods can be further classified into three main categories: Supervised, Unsupervised, and Self-Supervised. In the supervised learning paradigm a retrospectively undersampled image is used for training, while a fully sampled image is used to compute the loss between the predicted and the ground truth output. Some of the most popular supervised learning approaches that are also compared in this study are the U-Net [18], the ResNet [10], and the Cascades of Independently Recurrent Inference Machines (CIRIM) [12].

U-Net is a convolutional neural network (CNN) introduced by Ronnenberger et al. [19] in 2015 for biomedical image segmentation and was later used for reconstruction purposes. It consists of an encoder pathway for capturing contextual information and a decoder pathway for upsampling feature maps to reconstruct the output image. Skip connections enable information flow between the encoder and decoder at multiple spatial resolutions. Although such a network architecture can perform well, its performance is still limited due to operating only in image space without any MR physics knowledge incorporated. Nevertheless, U-Net has shown promise in improving MRI image quality, reducing artifacts, and accelerating acquisition times. Its effectiveness in MRI reconstruction makes it a widely used approach with potential applications in clinical workflows for shorter scan times and increased efficiency.

ResNet (Residual Neural Network) is a deep learning network introduced by He et al. [10] in 2016, that has been applied to MRI reconstruction to accelerate the process of image reconstruction from undersampled k-space data. ResNet-based models leverage residual connections to effectively capture the complex relationships between the undersampled k-space data and fully sampled image data, leading to improved reconstruction accuracy and reduced artifacts compared to traditional methods. These models typically consist of multiple residual blocks containing convolutional layers and are trained using paired undersampled k-space data and fully sampled image data.

The CIRIM (Cascades of Independently Recurrent Inference Machines) model, introduced by Karakalousos et al. [12] in 2022, represents a novel approach to MRI reconstruction. The CIRIM model leverages the power of recurrent neural networks (RNNs) and cascades to achieve improved de-aliasing and denoising while maintaining stable gradient calculations. The key innovation of CIRIM lies in the use of independently recurrent neural networks (IndRNNs) as recurrent units, which efficiently capture long-range dependencies in MRI data. The cascades in CIRIM allow for deep but balanced RNNs, enabling effective optimization and leveraging the prior information of an MR image. The CIRIM model

performs unrolled optimization using gradient descent with data consistency (DC) enforced implicitly. In recent studies, CIRIM has demonstrated superior performance in denoising and generalization capabilities.

Other popular supervised learning approaches include the Recurrent Inference Machines (RIM) [14, 17], the Variational Network (VN) [9], the Learned Primal-Dual reconstruction technique (LPDNet) [2], the Model-Based Deep learning technique (MoDL) [3], and the KIKI-net [7].

Although supervised learning approaches achieve high-quality reconstructions, they require fully sampled reference data for training. However, acquiring fully-sampled data is time-consuming and not always possible to collect. To address this issue several unsupervised and self-supervised methods have emerged. Some of the most popular self-supervised methods are Self-Supervision via Data Undersampling (SSDU) [23] and Noise2Recon [6].

The SSDU leverages data undersampling to train a physics-guided MRI reconstruction neural network without fully sampled data. The approach involves partitioning available measurements into two disjoint sets, one for data consistency (DC) and the other for calculating the training loss. This approach shares similarities with the widely used concept of cross-validation in machine learning, which partitions available data into two sets for training and validation. However, the key difference is that in the SSDU approach partitioning is performed for each slice in the dataset, while cross-validation partitions the entire dataset only once.

On the other hand, the Noise2Recon model enables joint MRI reconstruction and denoising using both labeled and unlabeled scans in semi-supervised and self-supervised settings. The undersampled scans that have a ground truth reference are utilized in a supervised manner, while the undersampled scans without ground truth are augmented with complex Gaussian noise and a loss is calculated between the clean and noisy reconstruction.

Several methods have been proposed for self-supervised MRI reconstruction. Acar et al. [1] successfully applied the SSDU method to dynamic MRI reconstruction. Furthermore. Hu et al. proposed Self-Supervised Learning with a Parallel Network Training Framework [11], extending SSDU using a parallel self-supervised framework for improved accelerated Cartesian MRI reconstruction. Yaman et al. introduced Zero-Shot Self-Supervised Learning [24], where they extended the two disjoint sets introduced in SSDU, to three with the third one being used for early stopping. On the other hand, HQS-Net [22] uses a Variable Splitting with Quadratic Penalty (VSQP) based iterative network. Finally, Cui et al. introduced Self-Supervised Learning on Score-Based Models for MRI Reconstruction (Self-Score) [5], using score-based diffusion models for MRI reconstruction.

Unsupervised learning approaches have also been proposed in the literature such as Unsupervised MRI Reconstruction with Generative Adversarial Networks [4], where Cole et al. introduced GANs to iteratively reconstruct MRI in the absence of fully-sampled data. Zero-Shot Learned Adversarial Transformers (SLATER) [13] uses a deep adversarial network with cross-attention transformers to map noise and latent variables onto coil-combined MR images. Unsupervised Deep Basis Pursuit [21] leverages the basis pursuit optimization method, by using a deep neural network prior instead of l_1-regularization.

In this study, we aim to compare and contrast various MRI reconstruction methods, including both supervised and self-supervised techniques. We provide a comprehensive description of each method, including its theoretical foundations, key algorithmic steps, and any modifications employed in the context of our study. We systematically evaluate and compare these methods in order to better understand their strengths and limitations, and identify the contribution of supervision in MRI reconstruction. To set the stage, we provide a brief primer on MRI, introducing the forward model of MRI acquisition and the inverse problem of MRI reconstruction.

II. Methods

A. Accelerated MRI acquisition

The acquisition of MRI data takes place in the frequency domain, commonly known as k-space. When a filtering mask denoted as P is applied to the acquired k-space frequencies, a undersampled k-space denoted as y is obtained, which can be converted to image space x using an inverse Fourier transform. Conversely, a Fourier transform is used to convert from image space to k-space.

The forward model of accelerated MRI acquisition mathematically describes how the measured data are obtained from an underlying reference image. It can be formulated as:

$$y = PFS_i^H x + n_i, \quad i = 1, \ldots, c \tag{1}$$

where y represents the acquired k-space data after applying the filtering mask P, F denotes the Fourier transform, S_i represents the sensitivity encoding operator for the i-th of c receiver coils, H represents the Hermitian complex conjugate, x represents the underlying reference image in the image space, and n_i represents the measured noise for the i-th coil. We can introduce the encoding operator E as:

$$E^i = PFS_i \tag{2}$$

So the forward model can be formulated as:

$$y = E^i x + n_i, \quad i = 1, \ldots, c \tag{3}$$

B. Accelerated MRI Reconstruction

The forward model allows us to obtain the undersampled k-space data, given the true image x, subsampling mask P, and sensitivity maps S_i. However, the objective is to approximate the true image from the undersampled k-space, a task commonly referred to as the inverse problem of MRI reconstruction. In compressed sensing (CS) this is addressed as a constrained optimization problem using the ℓ_1 norm. Another way to approach this is by using a maximum a posteriori (MAP) estimator and minimize the log-likelihood

derived from the forward model. This is solved dynamically over multiple timesteps in the RIM model. On the other hand, U-Net and ResNet work in the image space, without any prior knowledge of the underlying physics, optimizing the weights of the network using gradient descent to achieve a better approximation of the ground truth image. These approaches represent different strategies for tackling the inverse problem in MRI reconstruction, ranging from physics-based optimization to data-driven learning, each with its unique strengths and limitations.

C. SSDU

To enable physics-guided deep learning MRI reconstruction in the absence of fully sampled data, the SSDU method divides the acquired undersampled data indices, denoted by Ω, into two separate and non-overlapping sets, Θ and Λ, such that $\Omega = \Theta \cup \Lambda$. The k-space locations corresponding to Θ are utilized in the Data Consistency (DC) units during training. The DC term, which in SSDU's case is the Proximal Gradient method, is responsible for keeping the prediction space close to the latent space of the input. On the other hand, the k-space points corresponding to Λ are utilized to define the loss function. The loss function that the network tries to minimize is defined as follows:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(y^i_{\Lambda}, E^i_{\Lambda}(f(y^i_{\Theta}, E^i_{\Theta}; \theta))). \tag{4}$$

To elaborate further, the reconstruction network generates an output image $f(y^i_{\Theta}, E^i_{\Theta}; \theta)$ by utilizing only the k-space indices specified by Θ using the DC block. This output image is then transformed back into k-space using the encoding operator E^i_{Λ}, which is defined by the k-space indices in Λ. The acquired k-space data at these locations with the corresponding reconstructed k-space values, the training loss is computed in the k-space domain. The network parameters are then updated based on this training loss.

Equation 4 is solved iteratively in the DC block of the unrolled network using the conjugate gradient method, while a convolutional neural network (CNN) based on the ResNet architecture described in II-G is used for generating the reconstructions.

Inspired by the original paper, we compute masks using Gaussian random selection with a ratio $\rho = 0.4$. The ratio ρ is defined as the proportion of the set of indices used for the loss function (Λ) to the overall set of acquired undersampled data indices (Ω), i.e. $\rho = |\Lambda|/|\Omega|$.

D. Noise2Recon

The fundamental principle of Noise2Recon's consistency framework enables the use of both fully-sampled and undersampled scans concurrently, enhancing reconstruction in label-limited scenarios while increasing robustness to noise. In addition, Noise2Recon is not limited to any specific model and can be extended to unsupervised settings where fully-sampled references are not available.

The training process of Noise2Recon incorporates two distinct pathways for data flow. In the supervised training pathway, fully-sampled scans undergo retrospective undersampling, followed by reconstruction using the model f_θ, and optimization with respect to the target. Conversely, undersampled scans are augmented with masked, zero-mean complex Gaussian noise and also pass through the same model f_θ, resulting in noisy reconstructions. In the supervised pathway, the loss $\mathcal{L}sup$ is computed conventionally between the reconstruction and the target. In the unsupervised pathway, a consistency loss denoted as $\mathcal{L}cons$ is employed, which measures the discrepancy between the clean and noisy reconstructions, thus enforcing noise invariance in the model. The total loss is a weighted sum of the supervised and consistency losses, represented as $\mathcal{L}total = \mathcal{L}sup + \lambda\mathcal{L}_{cons}$, where λ is a hyperparameter that determines the balance between the two loss components.

The U-Net architecture, as described in II-G, served as the backbone model for the Noise2Recon method.

E. Dataset

We conducted experiments on the publicly available dataset of fully-sampled 3D fast-spin echo (FSE) multi-coil knee scans obtained from http://mridata.org [8]. The dataset comprised 19 subjects, which were partitioned into 14 subjects (4480 slices) for training, 2 subjects (640 slices) for validation, and 3 subjects for testing (960 slices). Fully-sampled data were retrospectively undersampled using a 2D Poisson Disc undersampling mask.

F. Evaluation Metrics

In our study, we conducted a comprehensive evaluation of the reconstructed normalized magnitude images obtained from the complex-valued estimations x_τ in comparison to the reference image x, utilizing well-established evaluation metrics. These metrics included the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Normalized Mean Square Error (NMSE). These metrics were calculated and reported per volume, providing quantitative measures of the accuracy, quality, and fidelity of the reconstructed images, allowing us to thoroughly assess the performance and effectiveness of the examined models.

G. Experiments

One of the main things we wanted to try in our experiments was how well a pre-trained model can generalize in other data distributions. To that end, an in-house 3T T1 3D Brains dataset was used, consisting of 11 subjects scanned on a 3.0 T Phillips scanner in Amsterdam UMC. The data were visually checked to ascertain that they were not affected by motion artifacts. After scanning, raw data were exported and stored for offline reconstruction experiments. The training set was composed of ten subjects (approximately 3000 slices) and the validation set of one subject (approximately 300 slices). We trained a U-Net, a ResNet and a CIRIM model with that dataset using Gaussian

2D mask with 12x acceleration factor, denoted with (SPT) in the results.

The remaining methods were trained on the training set of the Stanford Knee 3D FSE dataset using a Poisson 2D Disc mask with 12x acceleration factor. More specifically, we trained a Noise2Recon with three supervised subjects (N2R 3SUBJ), a Noise2Recon with one supervised subject (N2R 1SUBJ), a fully self-supervised Noise2Recon (N2R FULLSS), an SSDU with five residual blocks (SSDU 5RB), an SSDU with one residual block (SSDU 1RB), a U-Net and a ResNet. The supervised methods are denoted with (S) while the semi-supervised or self-supervised methods are denoted with (SS) in the results.

All methods were evaluated on the test set of the Stanford Knee 3D FSE dataset, and additional reconstructions were generated using Compressed Sensing (CS), Proximal Gradient, and Zero-Filled SENSE. Our selection of methods was not arbitrary. We chose the ResNet because it serves as the regularizer for the SSDU method, and the Proximal Gradient is used in the DC block. Furthermore, we selected the U-Net because it is the backbone architecture used in the Noise2Recon approach.

The configuration of hyperparameters for each method was considered carefully and was the same for all experiments. More specifically, the 2D U-Net configuration utilized in this study included 4 pooling layers, with the first convolution in the model having 32 output channels. Each resolution of the U-Net consisted of a convolutional block with two 3×3 convolutions, followed by instance normalization and a leaky Rectified Linear Unit (ReLU) activation function with a slope of α=-0.2. The total number of trainable parameters in the model was 7.76 million.

The ResNet-based CNN, which was used in this study, consisted of an input layer, an output layer, and 5 residual blocks with skip connections that promote information flow during network training. Each residual block consisted of two convolutional layers, with the first layer followed by a rectified linear unit, and the second layer followed by a constant multiplication layer with a factor of C = 0.1. All layers have a kernel size of 3×3 and 64 channels.

Finally, the hyperparameters of the CIRIM model utilized in this study were 5 cascades, 128 filters, and 8 time-steps, further showcasing the versatility and adaptability of the CIRIM model for MRI reconstruction tasks. It is worth noting that for all methods the l_1 loss was computed and all models were trained upon convergence, due to the limited time-frame of this study.

III. Results

Table I presents an analysis of the overall performance of all methods on the test set from the Stanford Knee 3D FSE dataset. The best-performing models were all trained on the same data distrubution, the Stanford Knee 3D FSE dataset. The U-Net model trained in a supervised manner was found to be the best-performing in terms of SSIM. The Noise2Recon, a semi-supervised approach trained on three supervised subjects, was the second-best performing model, followed by the ResNet, which was also trained in a supervised manner. Furthermore, the highest performance in terms of pre-trained models was achieved by the CIRIM method. It is worth noting that the models trained on the given dataset showed higher performance but only by a small margin over the pre-trained CIRIM.

The reconstructions of an axial slice from the Stanford Knee 3D FSE dataset are depicted in Figure 1. Upon visual inspection, the reconstructions generated by the U-Net and ResNet models appear very similar to the ground truth. However, the reconstructions produced by the Noise2Recon and CIRIM models appear to have lower SNR, particularly in the case of Noise2Recon. On the other hand, the SSDU and Proximal Gradient models produce visually similar reconstructions. While the reconstruction from the Compressed Sensing (CS) method is comparable to that of ResNet, it appears more blurred. In our experiments, we observed that the Fully Unsupervised Noise2Recon and the pre-trained ResNet were unable to resolve the background noise, resulting in more noisy reconstructions. Additionally, the pre-trained U-Net showed poor generalization capabilities and produced distorted images. The supplementary material includes example reconstructions of two additional subjects for reference.

IV. Discussion

In this study, we systematically evaluated and compared a variety of MRI reconstruction methods to further understand and highlight the contribution of supervision in MRI reconstruction. The main self-supervised methods investigated were the SSDU and Noise2Recon. As was mentioned in II-C, the SSDU method uses a ResNet as a regularizer and proximal gradient as a DC method. To better comprehend the individual contributions of each component to the overall performance, we examined each part of the method separately. More specifically, we trained and compared an SSDU with one residual block and one with five residual blocks to investigate the contribution of the network complexity. Also, a supervised ResNet was trained in a supervised manner as a baseline. Furthermore, we created reconstructions only using the proximal gradient method and compared the respective results.

Similarly, in the Noise2Recon approach a U-Net is used as the backbone architecture. We wanted our comparisons to be extensive so we trained a fully self-supervised version of the method, one with only one supervised subject and one with three supervised subjects, inspired by the original paper. Furthermore, we trained the U-Net in a supervised manner to use it as a baseline.

Moreover, we wanted to investigate how well a pre-trained network can generalize and how its results compare with the ones derived from the self-supervised methods. Our main hypothesis was that a well-trained pre-trained model with strong prior knowledge can generalize well enough and achieve comparable or even better results than the self-supervised approaches. To that end, we chose an in-house dataset of

Table I

This table presents the SSIM, PSNR, MSE, and NMSE scores for all evaluated methods on the Stanford Knee 3D FSE dataset, with performance reported for 12 times acceleration using Poisson 2D undersampling. The method names are listed in the first column and the best-performing values highlighted in bold. The methods are sorted based on their SSIM scores, and the (↑) and (↓) symbols indicate whether higher or lower values are better, respectively. Methods denoted with (SS) are self-supervised or semi-supervised, (S) indicates supervised methods, and (SPT) denotes a supervised pre-trained model trained on 3T T1 3D Brains with Gaussian 2D 12x.

Method	SSIM↑	PSNR↑	MSE↓	NMSE↓
UNET (S)	**0.853 ± 0.049**	28.41 ± 4.26	0.002 ± 0.002	0.251 ± 0.232
N2R 3SUBJ (SS)	0.846 ± 0.037	29.03 ± 3.52	**0.001 ± 0.001**	0.212 ± 0.156
RESNET (S)	0.838 ± 0.045	27.02 ± 4.56	0.002 ± 0.003	0.350 ± 0.360
CIRIM (SPT)	0.822 ± 0.048	28.25 ± 3.39	0.002 ± 0.001	0.253 ± 0.172
N2R 1SUBJ (SS)	0.814 ± 0.045	**29.15 ± 3.42**	**0.001 ± 0.001**	**0.207 ± 0.153**
CS	0.793 ± 0.051	26.15 ± 4.07	0.003 ± 0.002	0.418 ± 0.350
SSDU 5RB (SS)	0.792 ± 0.130	24.58 ± 8.10	0.004 ± 0.006	0.717 ± 0.912
SSDU 1RB (SS)	0.791 ± 0.139	24.72 ± 8.61	0.004 ± 0.006	0.710 ± 0.940
Proximal Gradient	0.662 ± 0.026	20.45 ± 1.84	0.009 ± 0.004	1.482 ± 0.482
N2R FULLSS (SS)	0.629 ± 0.026	22.34 ± 1.99	0.006 ± 0.003	0.960 ± 0.323
Zero-Filled SENSE	0.603 ± 0.041	18.00 ± 0.56	0.016 ± 0.002	2.590 ± 0.435
UNET (SPT)	0.574 ± 0.036	25.90 ± 0.82	0.003 ± 0.000	0.419 ± 0.021
RESNET(SPT)	0.449 ± 0.008	23.42 ± 2.63	0.005 ± 0.003	0.758 ± 0.401

high signal-to-noise ratio (SNR) acquisitions of 3T T1 3D Brains to pre-train three different networks. More specifically we trained a CIRIM, a ResNet, and a U-Net. The reason we chose CIRIM is that it incorporates physics knowledge and has demonstrated generalization capabilities. Furthermore, we pre-trained a ResNet and a U-Net in the same dataset, to investigate how well they could generalize given the fact that they do not incorporate any MR physics knowledge.

The results of our study indicate that the CIRIM model, which was trained on the in-house 3T T1 3D brains dataset, could achieve comparable results to the top-performing methods that were trained on the dataset used for evaluation in either a fully supervised or a semi-supervised manner, which still required a part of supervision (e.g. N2R 3SUBJ). Moreover, our findings revealed that SSDU with five residual blocks performed similarly to that with one block, highlighting that the DC block's optimization method, proximal gradient, is more crucial than network complexity. Notably, Proximal Gradient's performance is only slightly lower than SSDU, supporting the idea that ResNet's contribution as a regularizer is minimal for SSDU.

It is worth noting that due to time constraints, our experiments were stopped upon convergence rather than full convergence. However, we believe that this does not impact the validity of our claims, and our findings provide valuable insights into the effectiveness of various MRI reconstruction methods.

V. Conclusion

Overall, our findings suggest that a well-trained physics-based pre-trained model with strong prior knowledge, such as CIRIM, can generalize well to an unseen dataset. In contrast, the U-Net and ResNet pre-trained models, which operate without incorporating any MR physics knowledge, did not manage to generalize that well to the unseen dataset. Furthermore, the performance of the pre-trained CIRIM indicates that high SNR acquisitions are more important than structure in building strong prior knowledge. This is supported by the fact that even though the CIRIM was pre-trained on a brains dataset using a Gaussian 2D instead of a Poisson 2D mask, it managed to generalize well to the knees dataset.

References

[1] M. Acar, T. Çukur, and İ. Öksüz. Self-supervised dynamic mri reconstruction. In *Machine Learning for Medical Image Reconstruction: 4th International Workshop, MLMIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4*, pages 35–44. Springer, 2021.

[2] J. Adler and O. Oktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6): 1322–1332, June 2018. doi: 10.1109/tmi.2018.2799231. URL https://doi.org/10.1109/tmi.2018.2799231.

[3] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2): 394–405, Feb. 2019. doi: 10.1109/tmi.2018.2865356. URL https://doi.org/10.1109/tmi.2018.2865356.

[4] E. K. Cole, J. M. Pauly, S. S. Vasanawala, and F. Ong. Unsupervised mri reconstruction with generative adversarial networks, 2020.

[5] Z.-X. Cui, C. Cao, S. Liu, Q. Zhu, J. Cheng, H. Wang, Y. Zhu, and D. Liang. Self-score: Self-supervised learning on score-based models for mri reconstruction, 2022.

[6] A. D. Desai, B. M. Ozturkler, C. M. Sandino, R. Boutin, M. Willis, S. Vasanawala, B. A. Hargreaves, C. M. Ré, J. M. Pauly, and A. S. Chaudhari. Noise2recon: Enabling joint mri reconstruction and denoising with semi-supervised and self-supervised learning, 2022.

[7] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, and D. Hwang. scpKIKI/scp -net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80(5):

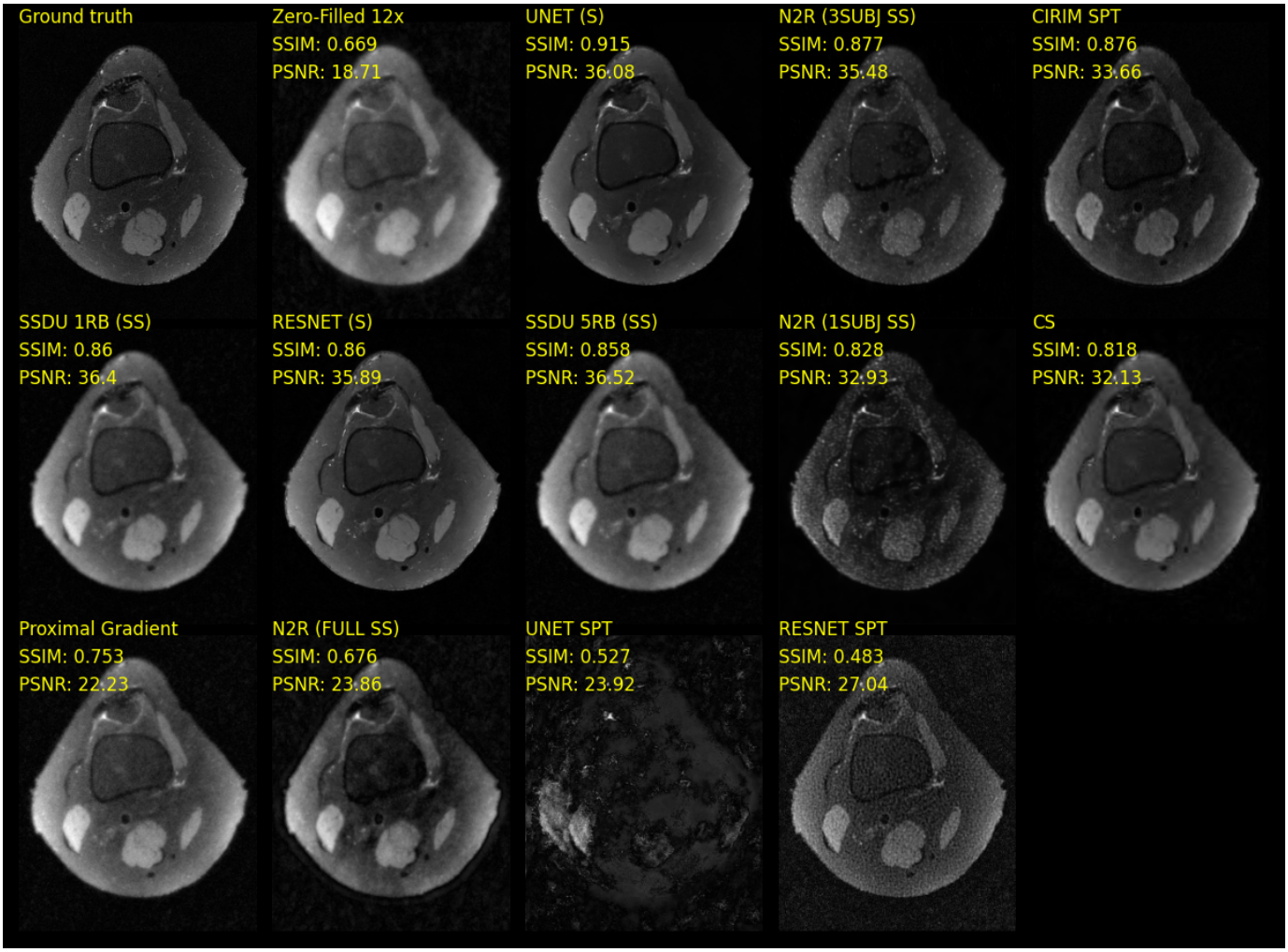

Fig. 1. Visual comparison of a reconstructed axial slice from the test set of the Stanford Knee 3D FSE dataset, using all evaluated methods. The top row displays the ground truth and zero-filled reconstruction, while the remaining reconstructions are arranged in descending order based on SSIM. Each reconstructed image is accompanied by corresponding SSIM and PSNR metrics.

2188–2201, Apr. 2018. doi: 10.1002/mrm.27201. URL https://doi.org/10.1002/mrm.27201.

[8] K. Epperson, A. M. Sawyer, M. Lustig, M. Alley, M. Uecker, P. Virtue, P. Lai, and S. Vasanawala. Creation of fully sampled mr data repository for compressed sensing of the knee. In *2013 Meeting Proceedings. Section for Magnetic Resonance Technologists*, 2013.

[9] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6): 3055–3071, Nov. 2017. doi: 10.1002/mrm.26977. URL https://doi.org/10.1002/mrm.26977.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] C. Hu, C. Li, H. Wang, Q. Liu, H. Zheng, and S. Wang. Self-supervised learning for MRI reconstruction with a parallel network training framework. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 382–391. Springer International Publishing, 2021. doi: 10.1007/978-3-030-87231-1_37. URL https://doi.org/10.1007/978-3-030-87231-1_37.

[12] D. Karkalousos, S. Noteboom, H. E. Hulst, F. M. Vos, and M. W. A. Caan. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. *Physics in Medicine &amp Biology*, 67(12): 124001, June 2022. doi: 10.1088/1361-6560/ac6cc2. URL https://doi.org/10.1088/1361-6560/ac6cc2.

[13] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Çukur. Unsupervised mri reconstruction via zero-shot learned adversarial transformers, 2022.

[14] K. Lønning, P. Putzky, J.-J. Sonke, L. Reneman, M. W. Caan, and M. Welling. Recurrent inference machines for

reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, Apr. 2019. doi: 10.1016/j.media.2019.01.005. URL https://doi.org/10.1016/j.media.2019.01.005.

[15] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, Mar. 2008. doi: 10.1109/msp.2007.914728. URL https://doi.org/10.1109/msp.2007.914728.

[16] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, Nov. 1999. doi: 10.1002/(sici)1522-2594(199911)42:5⟨952::aid-mrm16⟩3.0.co;2-s. URL https://doi.org/10.1002/(sici)1522-2594(199911)42:5⟨952::aid-mrm16⟩3.0.co;2-s.

[17] P. Putzky and M. Welling. Recurrent inference machines for solving inverse problems, 2017.

[18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://doi.org/10.1007/978-3-319-24574-4_28.

[19] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[20] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, Oct. 1997. doi: 10.1002/mrm.1910380414. URL https://doi.org/10.1002/mrm.1910380414.

[21] J. I. Tamir, S. X. Yu, and M. Lustig. Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data, 2020.

[22] A. Q. Wang, A. V. Dalca, and M. R. Sabuncu. Neural network-based reconstruction in compressed sensing mri without fully-sampled training data. In *Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 3*, pages 27–37. Springer, 2020.

[23] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, July 2020. doi: 10.1002/mrm.28378. URL https://doi.org/10.1002/mrm.28378.

[24] B. Yaman, S. A. H. Hosseini, and M. Akçakaya. Zero-shot self-supervised learning for mri reconstruction, 2021.

VI. Supplementary Material

The reconstructions of two distinct axial slices from the Stanford Knee 3D FSE dataset are depicted in Figures S1 and S2. Figure S1 displays the reconstructions generated by all the evaluated models. It can be observed that the reconstructions generated by the U-Net and ResNet appear very similar to the ground truth, achieving a high SSIM score of 0.916 and 0.909 respectively. The reconstructions produced from the Noise2Recon approach with three supervised subjects and the pre-trained CIRIM preserve the overall structure of the image, but appear noisier. Notably, the Compressed Sensing approach outperformed all of the fully self-supervised methods, including Noise2Recon with one supervised subject, which produced a noisier reconstruction. The reconstructions produced from SSDU and Proximal Gradient are visually similar, preserving the overall structure but having a high contrast that resembles the zero-filled reconstruction. Furthermore, the reconstructions produced by the SSDU with one and five residual blocks are almost identical, supporting our hypothesis that the complexity of the network has minimal contribution to the quality of the produced reconstructions. The fully self-supervised Noise2Recon and the pre-trained ResNet managed to preserve most of the overall structure but failed to resolve background noise, resulting in noisy reconstructions. Lastly, the pre-trained U-Net model showed poor generalization capabilities and produced distorted images.

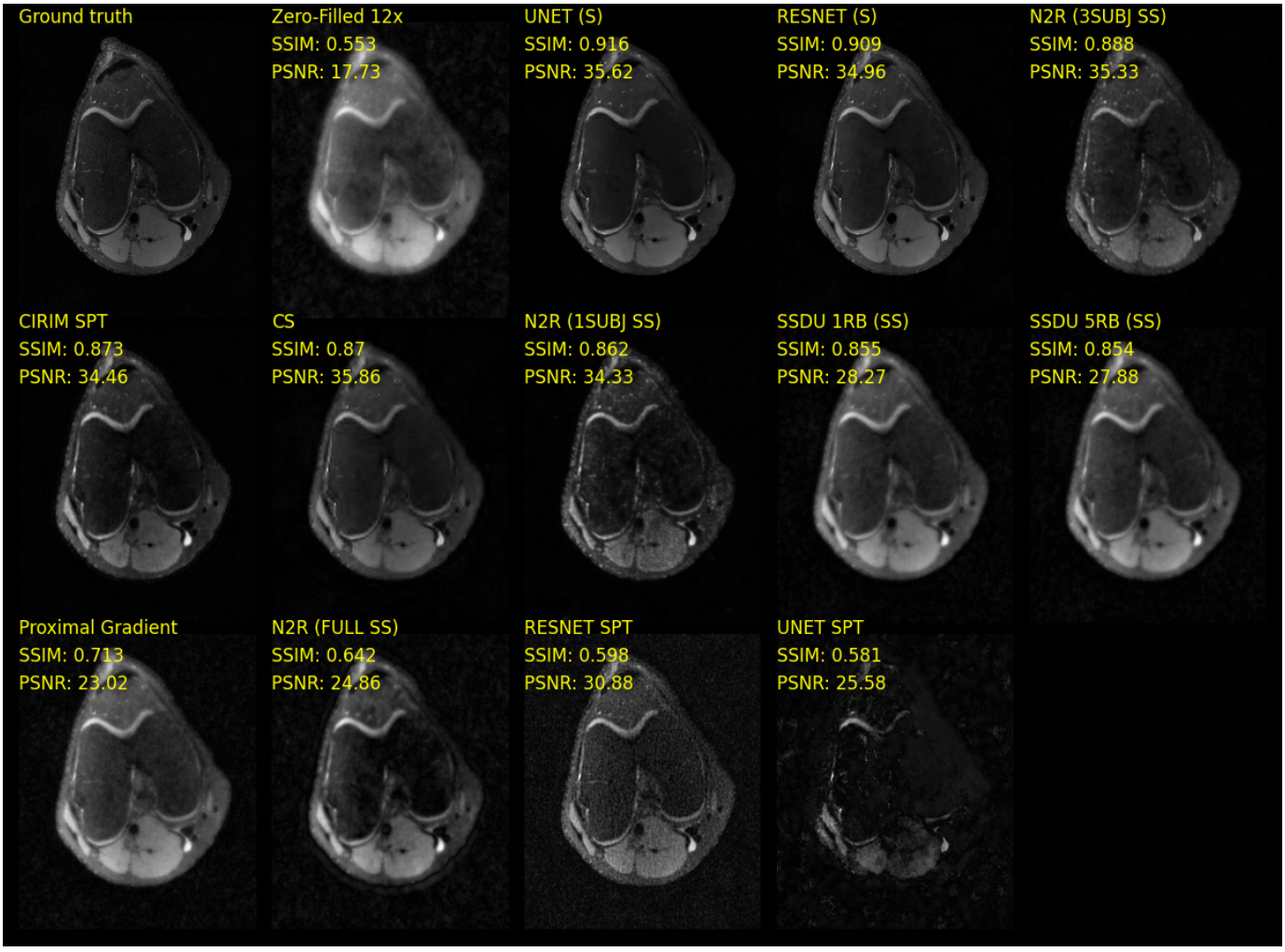

Fig. S1. Visual comparison of a reconstructed axial slice from the test set of the Stanford Knee 3D FSE dataset, using all evaluated methods. The top row displays the ground truth and zero-filled reconstruction, while the remaining reconstructions are arranged in descending order based on SSIM. Each reconstructed image is accompanied by corresponding SSIM and PSNR metrics.

In Figure S2, all evaluated models are presented, with the pre-trained CIRIM model achieving the highest SSIM reconstruction and demonstrating its excellent generalization capabilities. The two supervised methods (U-Net and ResNet) produced visually similar reconstructions, with U-Net achieving a higher PSNR value. Compressed Sensing once again outperformed all of the fully self-supervised methods, including Noise2Recon with one supervised subject, which produced a noisier reconstruction. Moreover, the Proximal Gradient method achieved better results than the SSDU in terms of both SSIM and PSNR, producing

a sharper reconstruction. Similarly, the fully self-supervised Noise2Recon achieved higher SSIM and PSNR than the other self-supervised methods. In this example, the reconstructions produced by the SSDU are very similar to the Zero-Filled reconstructions in terms of SSIM and PSNR. Finally, the pre-trained U-Net model once again demonstrated poor generalization capabilities and produced distorted images, but achieved a higher SSIM score than the ResNet, which is noisy but has preserved the overall structure of the image.

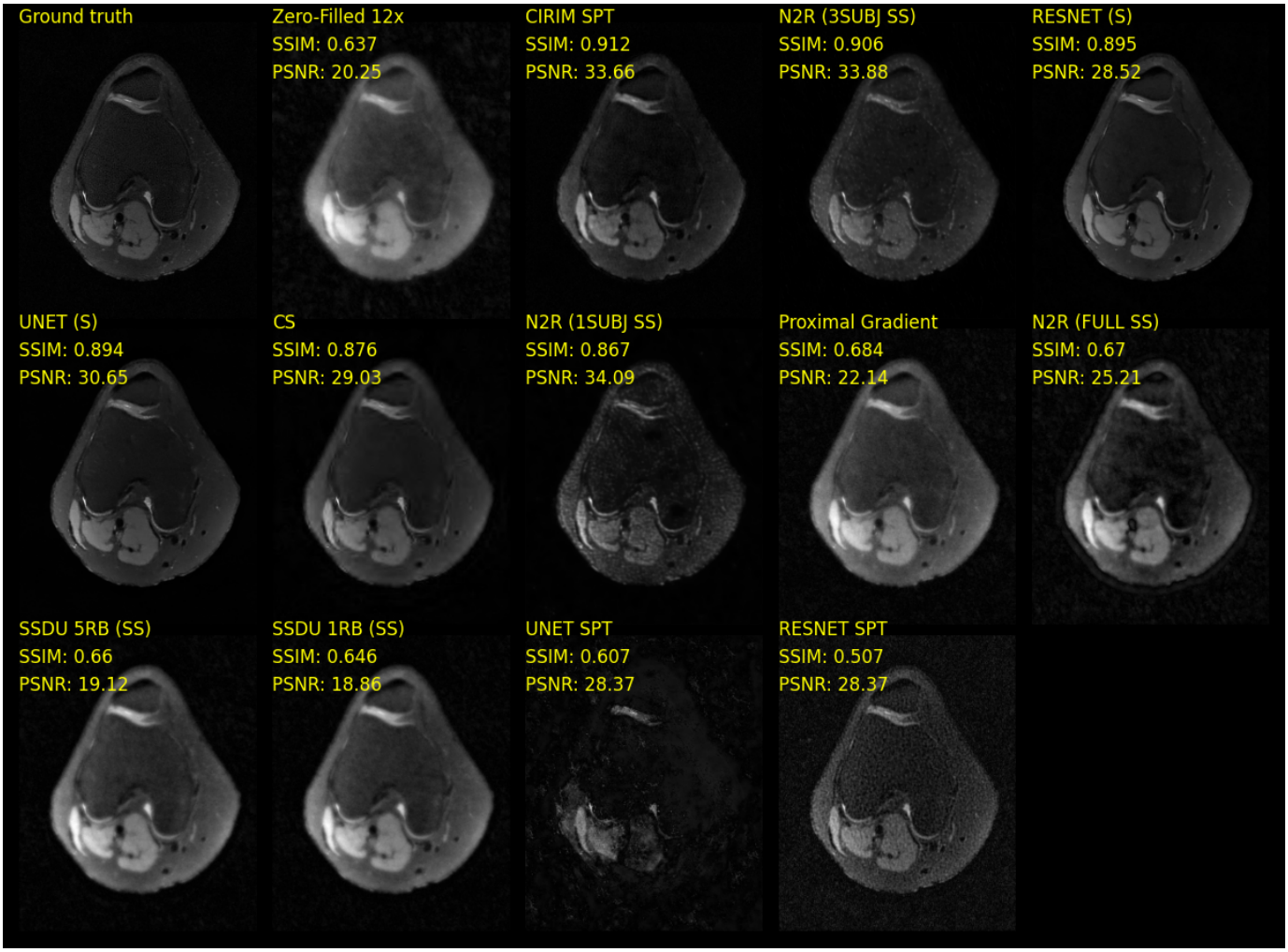

Fig. S2. Visual comparison of a reconstructed axial slice from the test set of the Stanford Knee 3D FSE dataset, using all evaluated methods. The top row displays the ground truth and zero-filled reconstruction, while the remaining reconstructions are arranged in descending order based on SSIM. Each reconstructed image is accompanied by corresponding SSIM and PSNR metrics.