This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material

Physics-Informed Deep Learning for Accelerated Open-Source Low-Field MRI Reconstruction

Tolough Nelson Aondongu[1], Muhammad Inuwa Muhammad[2], Tella Adetayo[2], Victor Ejike Onyedim[2], Joseph John[3], Ayatullah Hanif Showunmi[2], Zhang Dong[6], Confidence Raymond[4,5], Udunna C. Anazodo[4,5], Aondona Moses Iorumbur[2,5], and Sanni Henry Ananyi[2]

[1] University of Ilorin, Kwara State, Nigeria
[2] Federal University of Technology, Minna, Niger State, Nigeria
[3] Federal University of Lafia, Nasarawa State, Nigeria
[4] Montreal Neurological Institute, McGill University, Montreal, Canada
[5] Medical Artificial Intelligence Lab (MAI Lab), Lagos, Nigeria
[6] Department of Electrical & Computer Engineering, University of British Columbia, Vancouver, BC, Canada

Purpose and Scope

This supplementary material expands the experimental description of the main manuscript and is intended to make the OSI2 ONE reconstruction benchmark easier to reproduce. It does not introduce alternative numerical results. All dataset definitions, preprocessing choices, model configurations, training settings, evaluation conditions, and reported values below are reconstructed from the authoritative main manuscript. Where the manuscript does not specify an implementation detail, this supplement does not invent a value.

1 Dataset and Data Engineering

1.1 Source Data

The experiments use the open-source OSI2 ONE dataset acquired on a 47 mT Halbach-based low-field MRI system [2]. The dataset contains 427 ISMRMRD HDF5 files across six sub-datasets. The main experiment retains the 249 files identified as fully sampled; partially sampled files are excluded from the supervised reconstruction benchmark. The acquisitions use a single-channel receiver coil and 3D Turbo Spin Echo readout [3].

Nine unique human subjects form the LOSO benchmark: 9033, 9070, 9074, 9092, 9101, 9110, 9133, 9139, and 9147. The subject-level split, rather than a random slice split, is essential because slices from one person share anatomy and acquisition characteristics.

1.2 Scanner-Specific K-Space Assembly

The raw ISMRMRD metadata requires special handling. Slice information is stored in the non-standard `kspace_encode_step_2` field rather than the conventional slice field [5]. A generic volume reader can therefore assign slices incorrectly and produce geometrically scrambled anatomy.

Each retained file is reordered into a $38 \times 136 \times 150$ k-space array (slice $\times$ phase encoding $\times$ readout). Files with fewer than 136 phase-encoding lines are zero-padded. Each assembled volume is then normalised by its 99th-percentile magnitude before undersampling and model processing.

1.3 Preprocessing Integrity Checklist

A faithful reproduction should verify the following sequence before training:

1. Read the ISMRMRD acquisition dimensions and metadata fields.
2. Use `kspace_encode_step_2` as the slice index when assembling the volume.
3. Reorder measurements to $38 \times 136 \times 150$.
4. Zero-pad incomplete phase-encoding coverage to 136 lines.
5. Inspect inverse-Fourier-transformed volumes for anatomically coherent slice order.
6. Apply the same 99th-percentile magnitude normalisation to the resulting volumes.

The slice-order check is particularly important: reconstruction learning cannot correct a systematic anatomical ordering error introduced before the model receives its input.

2 LOSO Cross-Validation

The benchmark uses nine folds. In fold i, all data belonging to one subject are withheld from training and used only for evaluation; the remaining eight subjects form the training pool. Every subject is therefore held out exactly once.

The primary $R = 2$ benchmark uses the variable-density random mask. Cross-validation means and standard deviations are calculated across the nine held-out subjects/folds, not across randomly partitioned slices.

3 Undersampling Masks

Four undersampling families are implemented to represent acquisition patterns associated with OSI2 ONE:

Variable density. The principal $R = 2$ training/evaluation mask retains a fully sampled central region corresponding to 12.5% of k-space and samples the remaining locations with a spatially varying probability favouring low spatial frequencies. The mask is regenerated independently for each training sample at every epoch.

Table 1. Complete LOSO fold definition.

Fold	Held-out	Training subjects
1	9033	9070, 9074, 9092, 9101, 9110, 9133, 9139, 9147
2	9070	9033, 9074, 9092, 9101, 9110, 9133, 9139, 9147
3	9074	9033, 9070, 9092, 9101, 9110, 9133, 9139, 9147
4	9092	9033, 9070, 9074, 9101, 9110, 9133, 9139, 9147
5	9101	9033, 9070, 9074, 9092, 9110, 9133, 9139, 9147
6	9110	9033, 9070, 9074, 9092, 9101, 9133, 9139, 9147
7	9133	9033, 9070, 9074, 9092, 9101, 9110, 9139, 9147
8	9139	9033, 9070, 9074, 9092, 9101, 9110, 9133, 9147
9	9147	9033, 9070, 9074, 9092, 9101, 9110, 9133, 9139

Table 2. Model configurations used in the benchmark.

Model	Parameters	Iterations/Cascades	DC	Domain
U-Net [9]	1.93 M	1	None	Image
CascadeNet [10]	98 K	5	Hard	K-space
MoDL [1]	114 K	8	CG exact	K-space
DUN-DD [4]	77 K	5	Soft ($\lambda_{dc} = 0.5$)	Dual
E2E-VarNet [11]	1.48 M	8	Gradient	K-space

Circular window. This reproduces the circular phase-encoding coverage of the native acquisition.

Partial Fourier. This represents asymmetric k-space coverage in which one side of phase encoding is incompletely sampled [8].

LLR Poisson-disk. This provides variable-density, incoherent sampling for the locally low-rank setting [12].

The zero-shot acceleration protocol is defined in Section 6. Models trained with the $R = 2$ protocol are evaluated at higher acceleration factors without retraining.

4 Model Configurations

4.1 U-Net

The U-Net receives an inverse-Fourier-transformed magnitude representation and performs image-domain encoder–decoder regression. No explicit k-space data-consistency step is applied.

4.2 CascadeNet

CascadeNet contains five CNN refinement cascades. After each learned refinement, hard data consistency restores measured k-space samples. Thus acquired measurements cannot be overwritten by the learned component.

Table 3. Common training configuration reconstructed from the main manuscript.

Setting	Configuration
Optimiser	Adam [6]
Initial learning rate	10^{-3}
Schedule	Cosine annealing [7]
Maximum epochs	40
Batch size	8
Gradient clipping	Global norm ≤ 1.0
Image-space augmentation	None
Mask generation	Independent variable-density mask per sample/epoch

4.3 MoDL

MoDL uses eight unrolled iterations. A shared CNN denoiser is alternated with a conjugate-gradient exact data-consistency solution, providing an explicit model-based reconstruction operation while sharing learned parameters across iterations.

4.4 DUN-DD

DUN-DD uses five iterations with parallel image- and k-space-domain processing. The representations are fused using a residual attention U-Net and soft data consistency with $\lambda_{dc} = 0.5$.

4.5 E2E-VarNet

E2E-VarNet uses eight gradient-based data-consistency iterations. A refinement U-Net predicts a k-space correction that is incorporated through the variational reconstruction scheme.

5 Training and Optimisation

All models use the same broad optimisation protocol so that the architectural comparison remains controlled. All model training for this study was performed on Compute Canada.

The optimisation objective is

$$L(\hat{K}, K^*) = 0.7L_{\text{Huber}}(\hat{K}, K^*) + 0.3L_1\left(\mathcal{F}^{-1}\hat{K}, \mathcal{F}^{-1}K^*\right). \tag{1}$$

The k-space Huber term provides robust frequency-domain fidelity, while the image-domain L_1 term promotes structural and edge preservation. The main manuscript does not state the numerical Huber transition parameter; therefore no value is asserted here. A reproduction should use the same value as the released training implementation.

Table 4. Authoritative nine-fold LOSO results at $R = 2$.

Model	SSIM	SD	ΔSSIM	PSNR	SD	ΔPSNR
U-Net	0.5054	0.0696	–	22.56	1.93	–
DUN-DD	0.8009	0.0177	+0.2955	28.72	1.00	+6.16
CascadeNet	0.8179	0.0141	+0.3125	30.03	0.75	+7.47
E2E-VarNet	0.8212	0.0139	+0.3158	30.06	0.75	+7.50
MoDL	0.7396	0.0844	+0.2342	27.22	2.15	+4.66

The variable-density mask is regenerated independently for every training sample at every epoch. Consequently, the same fully sampled source slice can be presented under multiple sampling realisations, reducing the opportunity to memorise one aliasing pattern.

6 Evaluation

Metrics are computed on normalised magnitude images after inverse Fourier transformation of reconstructed k-space. SSIM [13] is the primary structural metric; PSNR is reported in decibels as a complementary measure of pixel fidelity. Mean and standard deviation are calculated across the nine held-out subjects.

6.1 Primary $R = 2$ Results

E2E-VarNet is the best aggregate model by both reported metrics. CascadeNet is close to E2E-VarNet but uses approximately 15 times fewer parameters. DUN-DD provides a large improvement over U-Net, while MoDL shows the largest subject-wise variability.

Table 5. New zero-shot acceleration results. Models are trained only at $R = 2$.

	PSNR (dB)						
Model	$R = 2$	$R = 4$	$R = 6$	$R = 8$	$R = 10$	$R = 12$	**Mean**
CascadeNet	30.98	26.23	24.41	23.38	22.75	22.27	25.00
DUN-DD	29.56	24.16	22.13	21.18	20.66	20.32	23.00
E2E-VarNet	30.92	26.11	24.23	23.18	22.49	21.99	24.82
MoDL	29.40	24.22	22.35	21.42	20.93	20.58	23.15
U-Net	26.56	24.21	22.87	22.14	21.73	21.42	23.16
	SSIM						
Model	$R = 2$	$R = 4$	$R = 6$	$R = 8$	$R = 10$	$R = 12$	**Mean**
CascadeNet	0.8303	0.6228	0.5384	0.4927	0.4631	0.4417	0.5648
DUN-DD	0.8280	0.5908	0.4970	0.4493	0.4204	0.3999	0.5309
E2E-VarNet	0.8331	0.6261	0.5404	0.4947	0.4650	0.4438	0.5672
MoDL	0.8070	0.5687	0.4845	0.4429	0.4180	0.4014	0.5204
U-Net	0.6139	0.5399	0.4932	0.4662	0.4484	0.4361	0.4996

7 Zero-Shot Acceleration Sweep: Quantitative and Visual Analysis

Models are trained only at $R = 2$ and then evaluated without retraining at $R = 4, 6, 8, 10$, and 12. The sweep therefore measures robustness to a distribution shift in acceleration rather than performance after separate training at each acceleration factor.

Across the aggregate results, both PSNR and SSIM decline as acceleration increases. E2E-VarNet has the highest mean PSNR/SSIM among the five models across the sweep, with CascadeNet close behind. The visual behaviour in the representative slice-27 figures provides additional model-specific context. E2E-VarNet remains visually usable through $R = 4$, whereas important anatomical features become blurred from $R = 6$ onward. CascadeNet follows the same pattern. DUN-DD maintains anatomical detail through $R = 6$, although a linear artefact is visible in the centre. MoDL maintains anatomical detail through $R = 4$; an artefact then initiates and worsens through $R = 12$ as the images become progressively blurrier. U-Net begins losing anatomical detail at $R = 4$, with progressively blurrier reconstructions through $R = 12$. These visual observations are representative-slice findings and do not replace the aggregate metrics in Table 5. The sweep is a retrospective synthetic undersampling experiment and does not establish equivalent performance on prospectively acquired high-acceleration scans. It should therefore not be treated as a clinically validated operating limit; larger populations, pathological cases, prospective paired acquisitions, and scanner-specific validation are required before clinical deployment claims.

7.1 Representative Slice-27 Acceleration Figures

The following figures are grouped exclusively in this section so that the acceleration-dependent anatomical changes can be inspected directly. They are representative

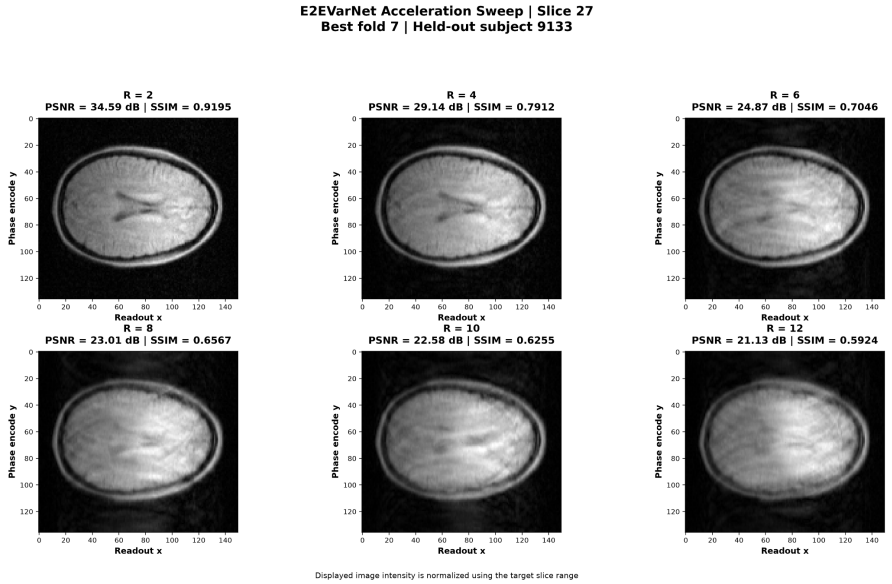

Fig. 1. E2E-VarNet acceleration sweep for slice 27. The representative images remain visually usable through $R = 4$; from $R = 6$ onward, important anatomical features become increasingly blurred.

slice-27 examples from the reported best-fold visualisations and should be interpreted together with the aggregate results above.

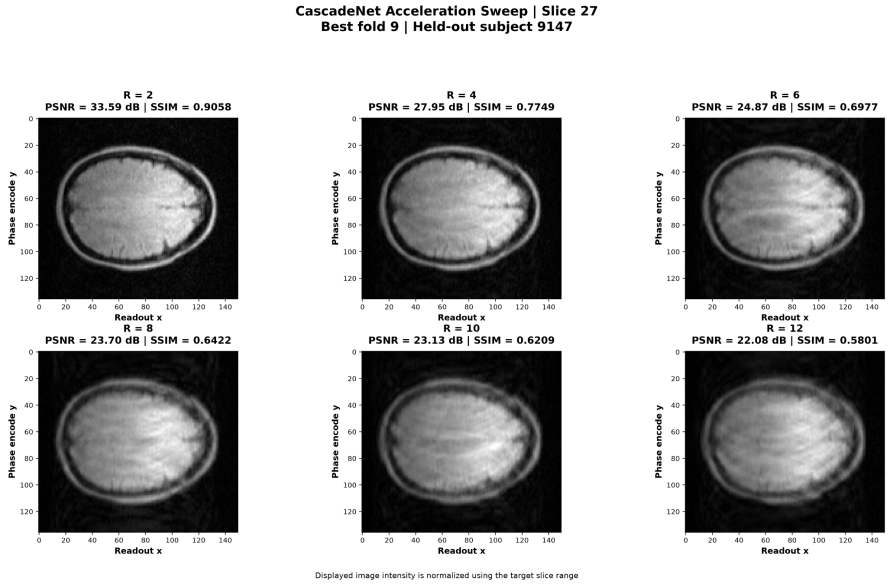

Fig. 2. CascadeNet acceleration sweep for slice 27. The representative images remain visually usable through $R = 4$; from $R = 6$ onward, important anatomical features become increasingly blurred.

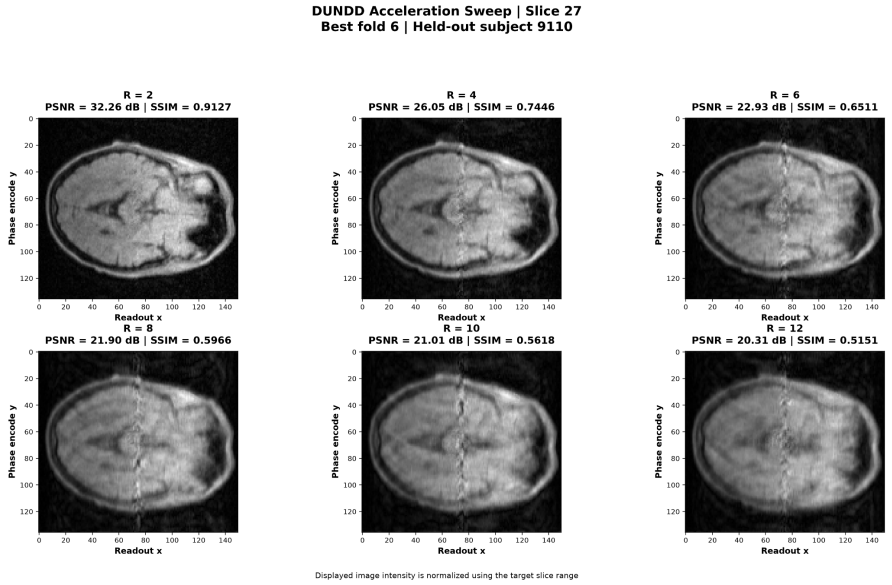

Fig. 3. DUN-DD acceleration sweep for slice 27. Anatomical detail is maintained through $R = 6$, with a linear artefact visible in the centre of the reconstruction.

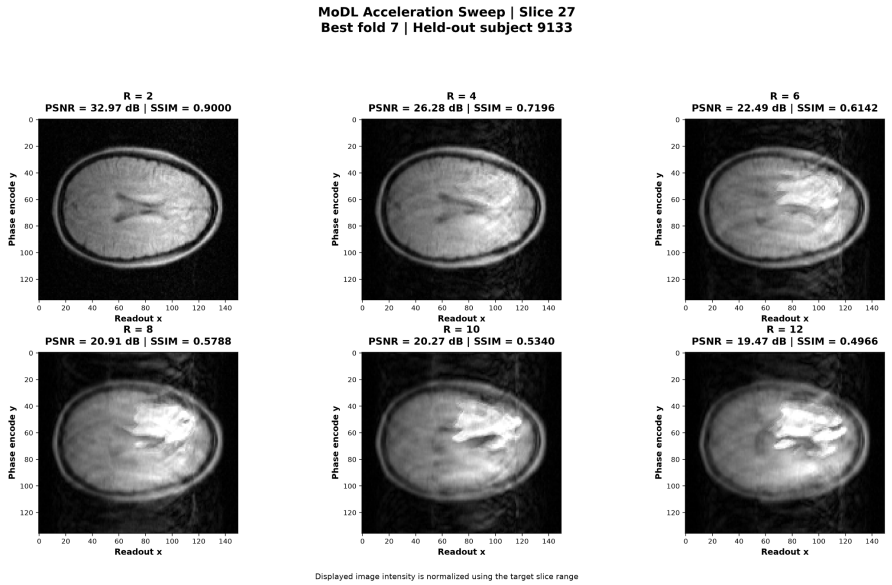

Fig. 4. MoDL acceleration sweep for slice 27. Anatomical detail is maintained through $R = 4$; an artefact begins to appear and worsens through $R = 12$, alongside increasing blur.

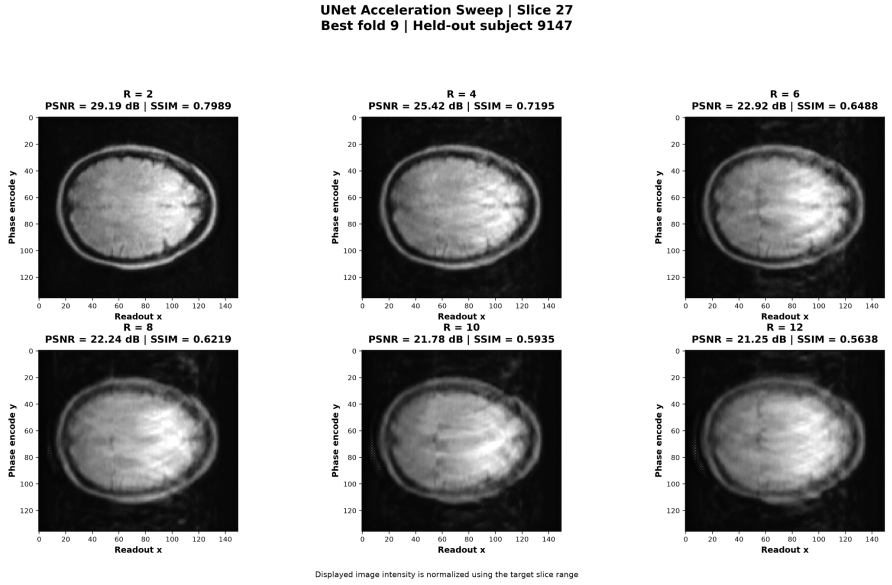

Fig. 5. U-Net acceleration sweep for slice 27. Anatomical detail begins to be lost at $R = 4$ and becomes progressively more blurred through $R = 12$.

8 Step-by-Step Reproduction Protocol

The following procedure translates the main manuscript into an implementation-oriented checklist without introducing unreported hyperparameters.

1. Obtain the OSI2 ONE ISMRMRD HDF5 dataset and identify the 249 fully sampled files used by the study.
2. For every retained file, parse acquisition metadata and explicitly read slice position from `kspace_encode_step_2`.
3. Reorder acquisition lines into a $38 \times 136 \times 150$ k-space array; zero-pad phase encoding when fewer than 136 lines are present.
4. Inverse-transform representative volumes and verify anatomically coherent slice ordering before generating learning examples.
5. Apply 99th-percentile magnitude normalisation consistently.
6. Construct the $R = 2$ variable-density random mask with a fully sampled central 12.5% region. Regenerate the mask independently for each training sample at each epoch.
7. For each LOSO fold, select one of the nine subjects as the held-out subject and use only the other eight subjects for training.
8. Instantiate U-Net, CascadeNet, MoDL, DUN-DD, and E2E-VarNet with the configurations in Table 2.
9. Train each model for up to 40 epochs using Adam, initial learning rate 10^{-3}, cosine annealing, batch size 8, and gradient clipping at global norm 1.0. All training was performed on Compute Canada.
10. Optimise the stated 0.7/0.3 Huber–image-L_1 hybrid loss.
11. Evaluate reconstructed normalised magnitude images using SSIM and PSNR after inverse Fourier transformation.
12. Aggregate the primary $R = 2$ results across all nine held-out subjects and report mean and standard deviation.
13. For the zero-shot experiment, retain the trained $R = 2$ models and evaluate them at $R = 4, 6, 8, 10$, and 12 without retraining.
14. Compare every acceleration factor with the zero-filled reference and preserve the distinction between aggregate LOSO results and best-validation-fold visual/curve examples.

9 Reproducibility and Interpretation Notes

The main results should be interpreted as a subject-wise benchmark on real low-field source data with retrospective synthetic undersampling.

The manuscript further notes that available real R1/R2 acquisitions were obtained in separate sessions with inconsistent contrasts, preventing a clean paired comparison against fully sampled acquisitions.

10 Additional Material Retained Outside the Main Text

The following details are expanded here to preserve reproducibility while keeping the main manuscript concise:

- the complete nine-fold subject mapping;
- the scanner-specific k-space assembly sequence and integrity checks;
- detailed descriptions of all four undersampling mask families;
- per-architecture descriptions of data-consistency mechanisms;
- the common training configuration and hybrid loss;
- the authoritative numerical tables for the $R = 2$ benchmark and zero-shot sweep; and
- an implementation-oriented reproduction sequence and interpretation limits.

Summary

The supplementary material is derived from the main manuscript rather than the earlier supplementary version and preserves the authoritative numerical results and reproduction details. The acceleration-sweep results and their visual interpretation are presented together in Section 7.

References

1. Aggarwal, H., Mani, M., Jacob, M.: MoDL: Model-based deep learning architecture for inverse problems. IEEE Trans. Med. Imaging **38**(2), 394–405 (2019)
2. van den Broek, R., Lena, B., Dong, Y., Ilicak, E., Imre, B., Najac, C., Kolbitsch, K., Staring, M., Webb, A.: Low-field MRI dataset with OSI2 ONE. Zenodo (2026). https://doi.org/10.5281/zenodo.19661402
3. Hennig, J., Nauerth, A., Friedburg, H.: RARE imaging: A fast imaging method for clinical MR. Magn. Reson. Med. **3**(6), 823–833 (1986)
4. Iliçak, E., et al.: Physics-guided dual-domain network with attention-based fusion for portable MRI reconstruction. In: Proceedings of the IEEE ISBI 2026 (2026)
5. Inati, S., et al.: ISMRM raw data format: A proposed standard for MRI raw datasets. Magn. Reson. Med. **77**(1), 411–421 (2017)
6. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: ICLR 2015 (2015)
7. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: ICLR 2017 (2017)
8. McGibney, G., et al.: Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. Magn. Reson. Med. **30**(1), 51–59 (1993)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Navab, N., et al. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Schlemper, J., et al.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **37**(2), 491–503 (2018)

11. Sriram, A., et al.: End-to-end variational networks for accelerated MRI reconstruction. In: Martel, A., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 64–73. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_7
12. Trzasko, J., Manduca, A.: Local versus global low-rank promotion in dynamic MRI series reconstruction. In: Proceedings ISMRM. vol. 19, p. 4371 (2011)
13. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)