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

Abstract. Low-field MRI reduces the infrastructure burden of brain imaging but suffers from low signal-to-noise ratio and long acquisitions. We benchmark U-Net, CascadeNet, MoDL, DUN-DD, and E2E-VarNet on real OSI2 ONE low-field data using nine-subject Leave-One-Subject-Out (LOSO) validation at $R = 2$. E2E-VarNet achieved the highest mean SSIM (0.8212 ± 0.0139) and PSNR (30.06 ± 0.75 dB), closely followed by CascadeNet (0.8179 ± 0.0141; 30.03 ± 0.75 dB). Relative to U-Net (0.5054 ± 0.0696; 22.56 ± 1.93 dB), these correspond to gains of 0.3158 SSIM and 7.50 dB for E2E-VarNet and 0.3125 SSIM and 7.47 dB for CascadeNet. CascadeNet uses 98 K parameters versus 1.48 M for E2E-VarNet. Zero-shot tests from $R = 2$ to $R = 12$ show progressive degradation for all models, with E2E-VarNet and CascadeNet retaining the strongest high-acceleration performance.

Keywords: Low-field MRI · K-space reconstruction · Physics-informed deep learning · Unrolled networks · Data consistency · OSI2 ONE dataset · Cross-validation

1 Introduction

Magnetic resonance imaging (MRI) is the standard for soft-tissue diagnosis, but conventional 1.5–3 T scanners require expensive infrastructure that limits access in low-resource and rural settings [4]. Low-field systems below 0.1 T reduce this burden through compact permanent-magnet designs, but their low signal-to-noise ratio (SNR) and long fully sampled acquisitions make acceleration difficult. Undersampling k-space shortens acquisition at the cost of aliasing and potential loss of diagnostic structure.

Physics-informed unrolled networks alternate learned refinement with data consistency (DC), providing a physical constraint that can reduce anatomically implausible reconstruction. Most reconstruction work, however, targets high-field, multi-channel MRI. The single-channel OSI2 ONE system has much lower SNR, Rician noise, and no parallel-imaging coil sensitivities, making direct transfer of high-field models inappropriate.

Contributions. We (i) resolve a scanner-specific metadata issue in which slice information is stored in the non-standard `kspace_encode_step_2` field; (ii) benchmark five reconstruction paradigms on real 50 mT data with rigorous nine-fold LOSO validation; and (iii) evaluate four scanner-grounded undersampling families and a zero-shot $R = 2$–12 robustness sweep.

2 Related Work

MRI reconstruction estimates x from undersampled measurements $y = F_{\Omega}x + \eta$. Zero filling introduces aliasing, while compressed sensing exploits sparsity through regularised optimisation [13]. U-Net models [16] offer faster reconstruction but do not explicitly enforce k-space fidelity [6]. Physics-informed and unrolled methods incorporate the acquisition model through DC. CascadeNet [17] applies hard DC after CNN cascades; MoDL [1] alternates a shared denoiser with conjugate-gradient DC; E2E-VarNet [18] uses gradient-based DC; and DUN-DD [9] jointly processes k-space and image-domain representations. We compare these approaches on real OSI2 ONE data.

3 Methods

3.1 Dataset and K-Space Assembly

Experiments use the OSI2 ONE dataset acquired on a 47 mT Halbach-based system [2]: 427 ISMRMRD HDF5 files across six sub-datasets, of which 249 fully sampled files were retained. The acquisitions use a single-channel receiver coil and 3D Turbo Spin Echo readout [7]. The 38 slices are encoded through `kspace_encode_step_2`, rather than the conventional slice field [10]; treating the files as standard volumes therefore scrambles anatomy. Each file is assembled into a $38 \times 136 \times 150$ array (slice $\times$ phase encoding $\times$ readout), incomplete phase-encoding coverage is zero-padded, and magnitude is normalised by the 99th percentile (Fig. 1).

3.2 Leave-One-Subject-Out Cross-Validation

Nine subjects (9033, 9070, 9074, 9092, 9101, 9110, 9133, 9139, and 9147) are evaluated with nine-fold LOSO validation. In each fold, one subject is completely withheld and the remaining eight provide training data; each subject is held out exactly once. This prevents correlated slices from the same subject appearing in both training and evaluation.

Four acquisition-grounded mask families are implemented: variable-density random with a fully sampled central 12.5% at $R = 2$; circular windowing; partial

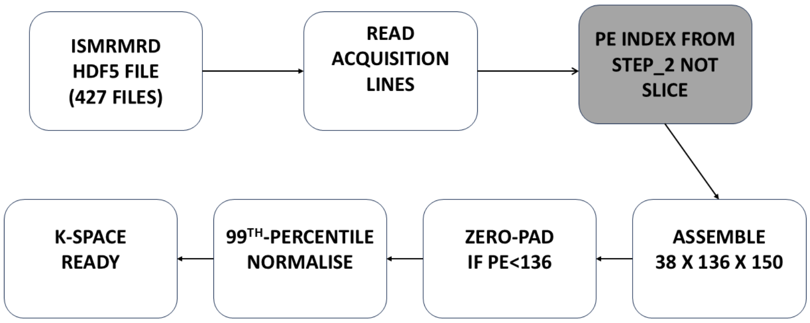

Fig. 1. K-space assembly pipeline for the OSI2 ONE ISMRMRD format. The non-standard slice encoding (`kspace_encode_step_2`) is corrected before reconstruction.

Table 1. Held-out subject for each LOSO fold. Training uses the other eight subjects.

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

Fourier [14]; and locally low-rank (LLR) Poisson-disk sampling [19]. The complete fold definition is retained in the supplementary material; the main text preserves the subject-wise protocol and all evaluation conditions.

3.3 Model Architectures

We evaluate U-Net [16] (image-domain, no DC), CascadeNet [17] (k-space, hard DC), MoDL [1] (k-space, conjugate-gradient exact DC), DUN-DD [9] (dual-domain, soft DC), and E2E-VarNet [18] (k-space, gradient-based DC). Their configurations are summarised in Table 2. U-Net uses image-domain encoder–decoder regression; CascadeNet uses five CNN/DC cascades; MoDL uses eight conjugate-gradient unrolled iterations; DUN-DD uses five dual-domain iterations with $\lambda_{dc} = 0.5$; and E2E-VarNet uses eight gradient-based DC iterations.

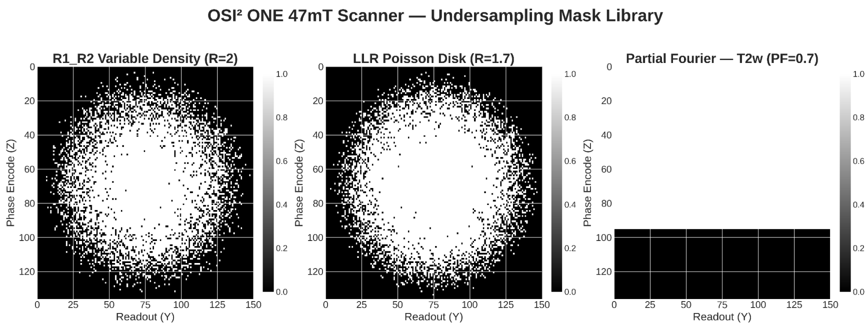

Fig. 2. Undersampling mask library used in the benchmark. The masks are grounded in OSI2 ONE acquisition protocols.

Table 2. Architecture comparison. Params = trainable parameters; DC = data consistency.

Model	**Params**	**Iter.**	**DC**	**Domain**
U-Net	1.93 M	1	None	Image
CascadeNet	98 K	5	Hard	K-space
MoDL	114 K	8	CG exact	K-space
DUN-DD	77 K	5	Soft	Dual
E2E-VarNet	1.48 M	8	Gradient	K-space

3.4 Loss Function and Training

Training minimises the hybrid objective

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

where $\hat{K}$ is reconstructed k-space, K^* the fully sampled target, and $\mathcal{F}^{-1}$ the inverse Fourier transform. The Huber term promotes robust frequency-domain fidelity and the image-domain L_1 term preserves structure and edges. Models are trained for up to 40 epochs with Adam [11], initial learning rate 10^{-3}, cosine annealing [12], gradient clipping at norm ≤ 1.0, batch size 8, and dual NVIDIA T4 GPUs. No explicit image-space augmentation is used; the variable-density mask is regenerated independently for each sample at each epoch, providing changing sampling realisations.

4 Experiments and Results

4.1 Evaluation Metrics

Metrics are computed on normalised magnitude images after inverse Fourier transformation of predicted k-space. SSIM [20] is the primary structural metric and PSNR is the complementary pixel-fidelity metric. Mean and standard deviation are calculated across the nine held-out subjects.

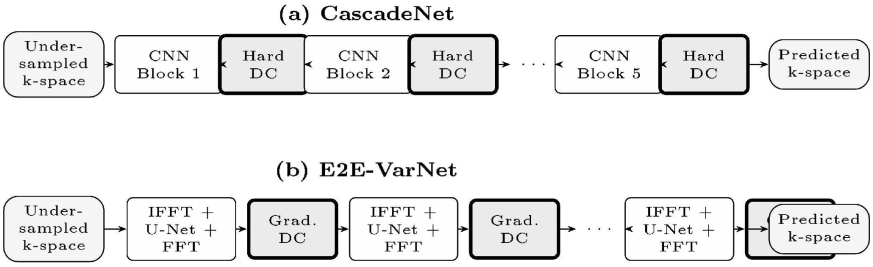

Fig. 3. Architecture of the two leading models. Thick borders denote data-consistency blocks.

Table 3. Nine-fold LOSO performance at $R = 2$. Differences are relative to U-Net.

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

4.2 LOSO Cross-Validation Comparison at $R = 2$

Table 3 reports performance at the training acceleration factor. E2E-VarNet achieves the highest mean SSIM (0.8212 ± 0.0139) and PSNR (30.06 ± 0.75 dB). CascadeNet is nearly identical (0.8179 ± 0.0141; 30.03 ± 0.75 dB) while using only 98 K parameters versus 1.48 M for E2E-VarNet. DUN-DD also substantially improves over U-Net; MoDL has the largest fold-to-fold variability.

4.3 Visual Reconstruction and Error Map Analysis

Figure 4 shows a representative mid-brain slice at $R = 2$. Physics-informed models preserve anatomical structure more convincingly than U-Net; CascadeNet and E2E-VarNet show the strongest agreement with the reference, DUN-DD preserves major structures, and MoDL is comparatively smooth. These are representative-slice observations, not substitutes for the nine-fold aggregate statistics.

4.4 Zero-Shot Generalization and Reliability Sweep

Models trained strictly at $R = 2$ are evaluated without retraining at $R = 4, 6, 8, 10, 12$. Table 4 reports the new aggregate PSNR and SSIM results. The representative $R = 2$ slice-27 values are included separately because the accompanying visual sweep uses the clearest mid-brain slice. Across acceleration, both metrics decline for every model; E2E-VarNet and CascadeNet retain the

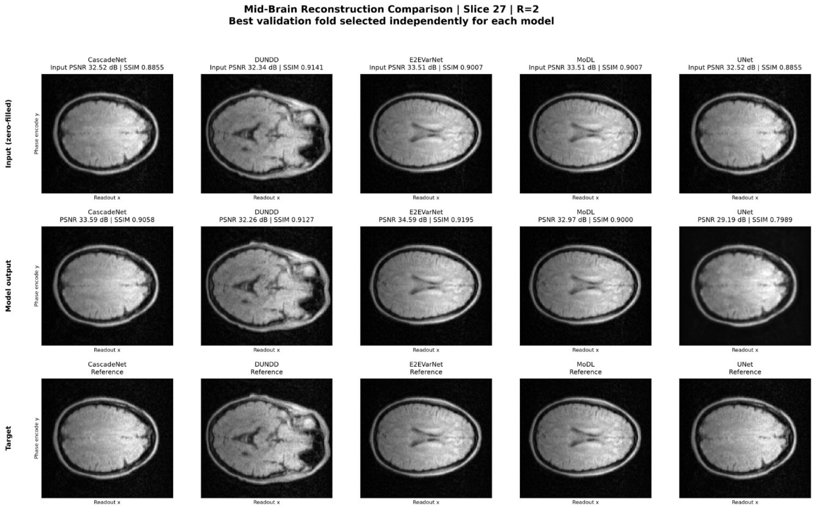

Fig. 4. Representative mid-brain reconstruction comparison at $R = 2$ for slice 27. Input, model output, and target/reference images are shown for CascadeNet, DUN-DD, E2E-VarNet, MoDL, and U-Net.

strongest values at the most aggressive factors, while U-Net starts substantially lower and converges toward the other models at high acceleration.

Table 4. Zero-shot acceleration sweep. Models are trained only at $R = 2$. Aggregate columns report the new results; $R = 2$ (slice 27) reports the representative slice values used for visual inspection. Mean is the mean of the six aggregate acceleration values $R = 2$–12.

			PSNR (dB)				
Model	$R = 2$	$R = 2$ (slice 27)	$R = 4$	$R = 6$	$R = 8$	$R = 10$	$R = 12$
CascadeNet	30.98	33.59	26.23	24.41	23.38	22.75	22.27
DUN-DD	29.56	32.26	24.16	22.13	21.18	20.66	20.32
E2E-VarNet	30.92	34.59	26.11	24.23	23.18	22.49	21.99
MoDL	29.40	32.97	24.22	22.35	21.42	20.93	20.58
U-Net	26.56	29.19	24.21	22.87	22.14	21.73	21.42
			SSIM				
Model	$R = 2$	$R = 2$ (slice 27)	$R = 4$	$R = 6$	$R = 8$	$R = 10$	$R = 12$
CascadeNet	0.8303	0.9058	0.6228	0.5384	0.4927	0.4631	0.4417
DUN-DD	0.8280	0.9127	0.5908	0.4970	0.4493	0.4204	0.3999
E2E-VarNet	0.8331	0.9195	0.6261	0.5404	0.4947	0.4650	0.4438
MoDL	0.8070	0.9000	0.5687	0.4845	0.4429	0.4180	0.4014
U-Net	0.6139	0.7989	0.5399	0.4932	0.4662	0.4484	0.4361

Aggregate means across $R = 2$–12: CascadeNet 25.00 dB / 0.5648; DUN-DD 23.00 dB / 0.5309; E2E-VarNet 24.82 dB / 0.5672; MoDL 23.15 dB / 0.5204; U-Net 23.16 dB / 0.4996.

Figure 6 provides a representative CascadeNet visual sweep. The new slice-27 examples show that image quality decreases as acceleration increases, with progressively reduced fine detail and contrast.

Visual inspection of the representative slice-27 sweep shows model-specific failure patterns. E2E-VarNet and CascadeNet remain visually usable through $R = 4$, while important anatomical features become blurred from $R = 6$ onward. DUN-DD maintains anatomical detail through $R = 6$ but exhibits a linear

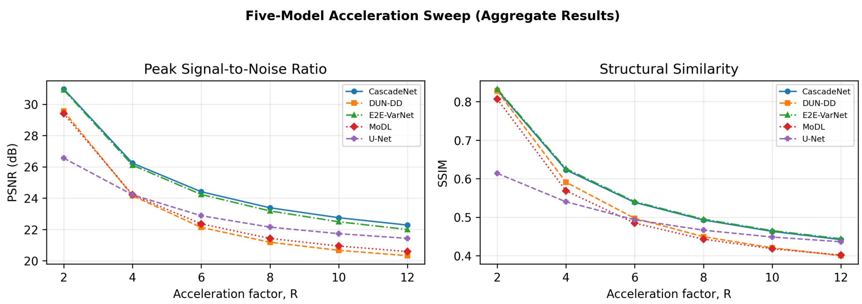

Fig. 5. Aggregate acceleration sweep for the five reconstruction models. PSNR and SSIM are plotted as acceleration increases from $R = 2$ to $R = 12$ using the new reported results.

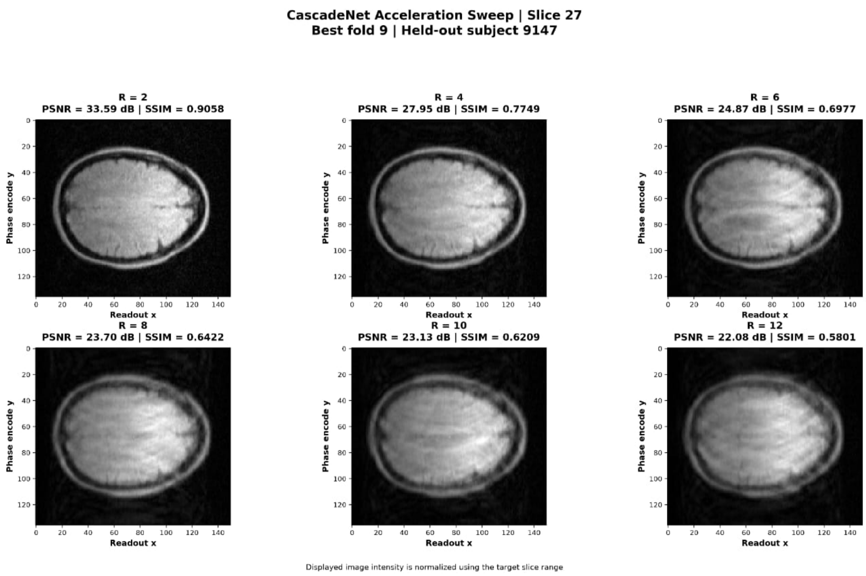

Fig. 6. CascadeNet reconstruction quality across acceleration factors for slice 27 in the best validation fold (held-out subject 9147). Each panel reports PSNR and SSIM for the reconstructed brain image.

artefact in the centre. MoDL maintains detail through $R = 4$; an artefact then emerges and worsens through $R = 12$ as the images become blurrier. U-Net begins losing anatomical detail at $R = 4$ and becomes progressively blurrier through $R = 12$.

5 Discussion

5.1 Data Consistency as the Critical Robustness Prior

Every unrolled architecture substantially outperforms U-Net in the subject-wise benchmark. E2E-VarNet provides the largest gain (+0.3158 SSIM, +7.50 dB PSNR), followed closely by CascadeNet (+0.3125 SSIM, +7.47 dB). The result is consistent with DC acting as a physical anchor: measured k-space constrains learned updates and reduces sensitivity to unseen anatomy and acquisition variation.

5.2 CascadeNet Efficiency and Safety

CascadeNet is especially attractive for resource-constrained reconstruction. Its mean SSIM and PSNR are nearly indistinguishable from E2E-VarNet while its 98 K parameters are about 15× fewer than E2E-VarNet's 1.48 M. Hard DC prevents the learned component from overwriting acquired k-space samples, providing a direct physical constraint.

5.3 Dual-Domain, MoDL, and Limitations

DUN-DD reaches SSIM 0.8009 and PSNR 28.72 dB, with stable SSIM variability (SD 0.0177). MoDL reaches 0.7396 SSIM and 27.22 dB PSNR but has the largest variability (SSIM SD 0.0844; PSNR SD 2.15 dB). The dataset contains only nine volunteer scans with fully sampled data, so performance on pathological cases and larger populations is unknown. Training uses retrospective undersampling of fully sampled acquisitions; available real R1/R2 acquisitions were obtained in separate sessions with inconsistent contrasts, preventing clean paired real-world evaluation. Larger datasets and prospectively paired acquisitions are needed.

6 Impact in Resource-Constrained Settings

Physics-informed reconstruction can improve low-field MRI quality while accommodating subject-wise distribution shift. The acceleration sweep shows progressive degradation across all models, with E2E-VarNet and CascadeNet retaining the strongest aggregate performance at high acceleration. U-Net starts substantially lower at $R = 2$ and becomes closer to the other models at higher acceleration, although its visual detail degrades from $R = 4$ onward. CascadeNet also achieves near-leading quality with roughly 15× fewer parameters than E2E-VarNet. This efficiency is relevant to portable scanners and resource-limited reconstruction hardware, although clinical deployment requires prospective validation.

7 Conclusion

We present a reproducible benchmark of five reconstruction paradigms on the OSI2 ONE 47 mT dataset. Scanner-specific k-space assembly and nine-fold LOSO validation address key data-engineering and generalisation issues. At $R = 2$, E2E-VarNet achieves the highest mean SSIM and PSNR, while CascadeNet provides nearly identical quality with about $15\times$ fewer parameters. Physics-informed models outperform U-Net at the training acceleration, while the zero-shot sweep shows that robustness depends on acceleration factor: E2E-VarNet and CascadeNet retain the strongest high-acceleration results, whereas DUN-DD and MoDL decline below U-Net at the highest factors. The sweep is retrospective and synthetic, not a clinically validated acceleration limit; prospective paired acquisitions and broader populations are required before deployment claims.

Acknowledgments. The authors thank the instructors of the SPARK Academy 2025 summer school on deep learning in medical imaging for background knowledge that informed this work: Zhang Dong, Confidence Raymond, Aondona Moses Iorumbur, Udunna C. Anazodo, and Sanni Henry Ananyi, and thank Linshan Liu (MiND Lab, Montreal Neurological Institute, McGill University) for administrative support.

The authors acknowledge SPARK Academy support from the Digital Research Alliance of Canada, the University of Washington Azure GenAI for Science Hub (eScience Institute and Microsoft; PI: Mehmet Kurt), the Lacuna Fund for Health and Equity (PI: Udunna Anazodo, 0508-S-001), the Radiological Society of North America, the R&E Foundation Derek Harwood-Nash International Education Scholar Grant, McGill HBHL, and NSERC (PI: Udunna Anazodo, DGECR-2022-00136), and the McGill University Doctoral Internship Program.

Disclosure of Interests. The authors declare no competing interests.

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. Cooley, C., et al.: A portable brain MRI scanner for underserved settings. Nat. Biomed. Eng. **5**, 229–237 (2021)
4. Geethanath, S., Vaughan, J.: Accessible magnetic resonance imaging: A review. J. Magn. Reson. Imaging **57**, 1–15 (2023)
5. Hammernik, K., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. **79**(6), 3055–3071 (2018)
6. Heckel, R., et al.: Deep learning for MRI reconstruction: A review. arXiv preprint (2024)

7. Hennig, J., Nauerth, A., Friedburg, H.: RARE imaging: A fast imaging method for clinical MR. Magn. Reson. Med. **3**(6), 823–833 (1986)
8. Iliçak, E., et al.: Physics-informed deep unrolled network for portable MR image reconstruction. arXiv preprint (2025)
9. 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)
10. Inati, S., et al.: ISMRM raw data format: A proposed standard for MRI raw datasets. Magn. Reson. Med. **77**(1), 411–421 (2017)
11. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: ICLR 2015 (2015)
12. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: ICLR 2017 (2017)
13. Lustig, M., Donoho, D., Pauly, J.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**(6), 1182–1195 (2007)
14. McGibney, G., et al.: Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. Magn. Reson. Med. **30**(1), 51–59 (1993)
15. O'Reilly, T., et al.: In vivo 3D brain and extremity MRI at 50 mt using a permanent magnet Halbach array. Magn. Reson. Med. **85**, 495–505 (2021)
16. 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
17. 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)
18. 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
19. Trzasko, J., Manduca, A.: Local versus global low-rank promotion in dynamic MRI series reconstruction. In: Proceedings ISMRM. vol. 19, p. 4371 (2011)
20. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)