Contrast-Reversal Augmentation for Adult-to-Neonatal MR Image Reconstruction: Pretraining and Fine-Tuning Strategies

Stephen Moore[1,2,3], Lara Leijser[3,4,5], Richard Frayne[1,2,3,6], and Roberto Souza[3,7]

[1] Biomedical Engineering, University of Calgary, Calgary, AB, Canada
stephen.moore@ucalgary.ca
[2] Seaman Family MR Research Centre, Foothills Medical Centre, Calgary, AB, Canada
[3] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada
[4] Pediatrics, Division of Neonatology, University of Calgary, Calgary, AB, Canada
[5] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada
[6] Radiology and Clinical Neuroscience, University of Calgary, Calgary, AB, Canada
[7] Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada

Abstract. We investigated the effect of incorporating contrast-informed augmentations into pretraining data on the adult-to-neonatal adaptation behavior of deep learning-based magnetic resonance (MR) image reconstruction models. Models were pretrained on unaugmented adult data, contrast-informed augmented adult data, or a mixture of both, then fine-tuned with varying amounts of neonatal data. Evaluation on held-out neonatal and adult test sets assessed neonatal adaptation and adult-domain retention. Models pretrained on a mix of unaugmented and augmented adult data produced the best zero-shot generalization to neonatal data. Models pretrained on exclusively augmented adult data reduced adult-domain performance and did not consistently improve zero-shot neonatal performance. All pretrained models performed significantly better on neonatal reconstruction after fine-tuning with neonatal data and generally exceeded neonatal-only training in low-data settings. These results suggest that contrast-informed augmentation can improve neonatal MR image reconstruction when incorporated into a mixed pretraining distribution rather than used as a full replacement for real adult data.

Keywords: MR Reconstruction · Domain Adaptation · Infant Imaging.

1 Introduction

Magnetic resonance (MR) imaging produces high-quality soft-tissue images without ionizing radiation, but has a long acquisition time [11]. Acceleration techniques reduce this time by undersampling k-space [11]. Advanced methods, including parallel imaging (PI) [11, 5, 16] and compressed sensing (CS) [12], reconstruct these data and have been clinically deployed for more than two decades.

Deep learning (DL)-based reconstruction methods have emerged as alternatives to these conventional methods and perform well on in-domain data, but can fail under domain shift [10]. A domain shift may arise from factors that include variations in anatomy, acquisition protocols, and scanner hardware [9, 10]. Because clinical imaging is heterogeneous, robustness under domain shift is essential for clinical deployment.

Neonatal brain MR is informative both clinically and in research [3, 2]. It is relevant to DL-based MR reconstruction for two primary reasons: it represents a specific domain shift from adult data characterized by differences in anatomy, tissue contrast, signal-to-noise characteristics, and a higher prevalence of motion artifacts [3, 1]; and large high-quality neonatal datasets are scarce, limiting the feasibility of training DL models directly on neonatal data. This combination of domain shift and data scarcity motivates approaches that improve the generalizability of adult-trained DL-based MR image reconstruction models. Neonatal MR reconstruction, therefore, provides a clinically relevant test case for studying how reconstruction models transfer when target-domain data are scarce and distinct from comparatively abundant adult training data.

Here, we investigate how contrast-informed data augmentation affects zero-shot generalization and few-shot transfer of adult-trained DL-based MR reconstruction models to neonatal data. We specifically test whether neonatal-informed augmentation is most beneficial as a replacement for adult data or as a supplement to adult data. We then evaluate how these pretraining distributions affect neonatal fine-tuning performance and adult-domain retention.

2 Background

Unrolled optimization network architectures, such as the end-to-end variational network (E2E-VarNet), are common in DL-based MR reconstruction [19]. These models learn data-driven priors and perform well on data from distributions matching their training data distribution [6], but are sensitive to domain shift: differences in anatomy, acquisition parameters, and image statistics can degrade performance [17, 9]. Consequently, domain adaptation and generalization are the focus of a significant body of recent DL-based MR reconstruction research [9, 4]. Common strategies to maximize performance under domain shift include data augmentation and target-domain fine-tuning [9, 4]. These approaches are prevalent in adjacent tasks [14, 18, 7, 4], but in adult-to-neonatal DL-based MR reconstruction, to our knowledge, no studies have investigated their joint application.

3 Methods

3.1 Experimental Overview

We investigated the effect of neonatal contrast-informed augmentation on the generalizability of DL-based MR reconstruction models from adult to neonatal

data and on the transfer of these models after fine-tuning with small amounts of neonatal data. Models were trained on exclusively unaugmented adult data, an equal proportion of unaugmented and augmented adult data, or exclusively augmented adult data. Models were evaluated on held-out neonatal and adult data before and after fine-tuning with one, three, five, and ten neonatal volumes to measure zero-shot generalization, few-shot transfer, and adult-domain retention. Separate models were also trained on the neonatal data subsets used for fine-tuning to measure the performance of models trained directly on neonatal data. Our code is available at: https://github.com/moorestephen/MICCAI-RIME-2026.

3.2 Data

The adult dataset was sourced from the publicly available NYU fastMRI dataset and consisted of raw, fully sampled multi-coil **k**-space data from $n = 80$ T2-weighted brain MR volumes [13]. The fastMRI brain dataset includes variable in-plane matrix sizes and resolutions [13]. Because the dataset consists of standard 2D protocols (Multi-slice Spin Echo and Turbo Spin Echo), the voxels are anisotropic with a typical slice thickness of 5.0 mm. The neonatal dataset was collected as part of the P3 Cohort study and the P3 Brain Health study and consisted of raw, fully sampled multi-coil **k**-space data from $n = 32$ T2-weighted volumes (axial T2 FSE 3 mm) [15]. All MR acquisitions were obtained around term-equivalent age (40-44 weeks postmenstrual age). The zero-padded data had a matrix size of 512×512. Zero-padding was removed during preprocessing, yielding acquired **k**-space matrices of 312×312 for model training and evaluation. The P3 Cohort and P3 Brain Health studies were approved by the University of Calgary's Conjoint Health Research Ethics Board (REB20-1635 and REB21-1446, respectively). Written informed consent was obtained from parents/legal guardians. Data were used in accordance with the approved protocols.

3.3 Model Training

The models were trained using one of three datasets derived from the adult data. The Unaug-Only, Mixed, and Aug-Only models were trained on exclusively unaugmented adult data, an equal proportion of unaugmented and augmented adult data, and exclusively augmented adult data, respectively. The neonatal dataset was split into 20 held-out test volumes, 10 candidate fine-tuning volumes, and 2 validation volumes; fine-tuning subsets of 1, 3, 5, and 10 volumes were nested subsets of the 10 candidate fine-tuning volumes. Neonatal models were trained exclusively on the neonatal data reserved for fine-tuning as controls. The adult dataset was split into 50 volumes for training, 5 volumes for validation, and 25 volumes for held-out testing.

All models shared the same E2E-VarNet architecture [19], loss function (structural similarity index measure (SSIM) [21]), optimizer (Adam [8]), and training hyperparameters to isolate the effect of training data distributions and fine-tuning protocols. During model training, the initial learning rate was 1×10^{-3}

for 40 epochs, after which it was reduced to 1×10^{-4}. The models were fine-tuned using a learning rate of 1×10^{-4} for 40 epochs and 1×10^{-5} for 10 epochs. Separate experiments investigated retrospective Cartesian undersampling with $R = 4$ and $R = 8$.

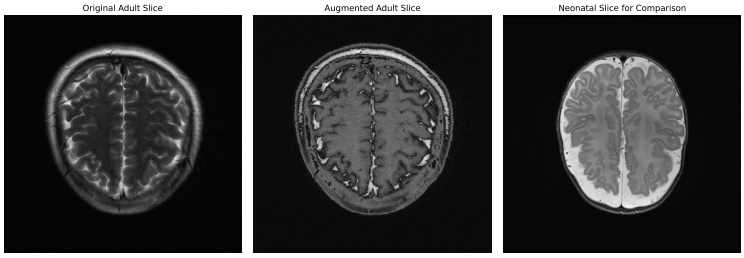

Fig. 1. Representative example of contrast-informed augmentation applied to adult MR data. The original adult slice, augmented adult slice, and a reference neonatal slice are included from left to right.

3.4 Augmentation Protocol

The contrast-informed augmentation was applied in image space. Multi-coil data y_c were inverse Fourier transformed and coil-combined using ESPIRiT-generated sensitivity maps S_c [20], yielding a complex image with magnitude $m(\mathbf{r})$ and phase $\phi(\mathbf{r})$. k-Means clustering ($k = 4$, fixed random seed) was applied to the magnitude image; after ordering clusters by mean intensity, the lowest-intensity cluster was excluded and the remaining clusters assigned to approximate WM, GM, and CSF. Denoting the mean magnitude intensities of the WM and GM clusters by μ_{WM} and μ_{GM}, respectively, and their voxel sets by Ω_{WM} and Ω_{GM}, the contrast-informed transformation was then defined as $m'(\mathbf{r}) = \mu_{\mathrm{WM}} + \mu_{\mathrm{GM}} - m(\mathbf{r})\ \forall\ \mathbf{r} \in \Omega_{\mathrm{WM}}$, $0.8\,[\mu_{\mathrm{WM}} + \mu_{\mathrm{GM}} - m(\mathbf{r})]\ \forall\ \mathbf{r} \in \Omega_{\mathrm{GM}}$, and $m(\mathbf{r})$ otherwise. The transformed magnitude was then Gaussian-smoothed ($\sigma = 0.5$), and additive Gaussian noise was applied ($\epsilon \sim \mathcal{N}(0, 0.02^2)$), both within the sensitivity-map support. The original phase was retained, and the augmented image was re-expanded to individual coils using the original sensitivity maps and Fourier transformed to generate augmented k-space. The same sensitivity maps were reused because the augmentation did not modify the assumed coil geometry or spatial sensitivity profiles.

3.5 Evaluation Metrics and Statistical Analysis

Reconstruction quality was quantified using SSIM and peak signal-to-noise ratio (PSNR). Because SSIM was also used as the reconstruction loss, PSNR was included as a complementary pixel-error-based metric. Metrics were computed

by slice and averaged by volume. We performed Friedman tests to assess overall differences among models within each pretraining and fine-tuning regime and, when significant, *post-hoc* pairwise Bonferroni-corrected Wilcoxon signed-rank tests ($\alpha = 0.05$).

4 Results

4.1 Augmented Data Improved Zero-Shot Transfer Only When Combined with Unaugmented Adult Data

The models pretrained on adult-derived data distributions (Unaug-Only, Mixed, and Aug-Only) were evaluated on held-out neonatal test data to assess their zero-shot transfer ability. The Mixed models produced reconstructions with significantly higher mean SSIM and PSNR values than the Unaug-Only and Aug-Only models at both acceleration factors (Table 1). Unaug-Only produced reconstructions of $R = 4$ undersampled data with significantly higher mean SSIM and PSNR values than Aug-Only. On $R = 8$ undersampled data, Aug-Only produced reconstructions with a significantly higher mean SSIM but a significantly lower mean PSNR than Unaug-Only.

	R4		R8	
Model	SSIM	PSNR	SSIM	PSNR
Unaug-Only	0.862 ± 0.020	32.66 ± 0.80	0.763 ± 0.026	29.32 ± 0.57
Mixed	$\mathbf{0.896 \pm 0.014}$	$\mathbf{33.60 \pm 0.66}$	$\mathbf{0.815 \pm 0.022}$	$\mathbf{29.69 \pm 0.59}$
Aug-Only	0.842 ± 0.020	32.28 ± 0.78	0.789 ± 0.023	29.23 ± 0.59

Table 1. Neonatal reconstruction performance, reported as mean ± standard deviation across test volumes. Bold values indicate the highest mean performance for each acceleration factor and metric.

4.2 Fine-tuning Improved Neonatal Reconstruction Performance

Under $R = 4$ undersampling, the Mixed model, fine-tuned with one neonatal volume, had mean SSIM and PSNR values of 0.964 ± 0.010 and 39.35 ± 1.10, respectively, and significantly outperformed the neonatal-only control model trained on one neonatal volume (SSIM = 0.950 ± 0.009 and PSNR = 35.89 ± 0.67). Pairwise comparison of the models fine-tuned on 10 neonatal data volumes revealed a significant difference in mean SSIM only between Unaug-Only and Aug-Only, and between Unaug-Only and Mixed. Performance gains between fine-tuning subsets were modest compared to the gain from no fine-tuning to fine-tuning: the mean SSIM values of the fine-tuned models, irrespective of the distribution of pretraining data and the number of volumes of neonatal data used during fine-tuning, ranged from 0.957 to 0.970 (Figure 2).

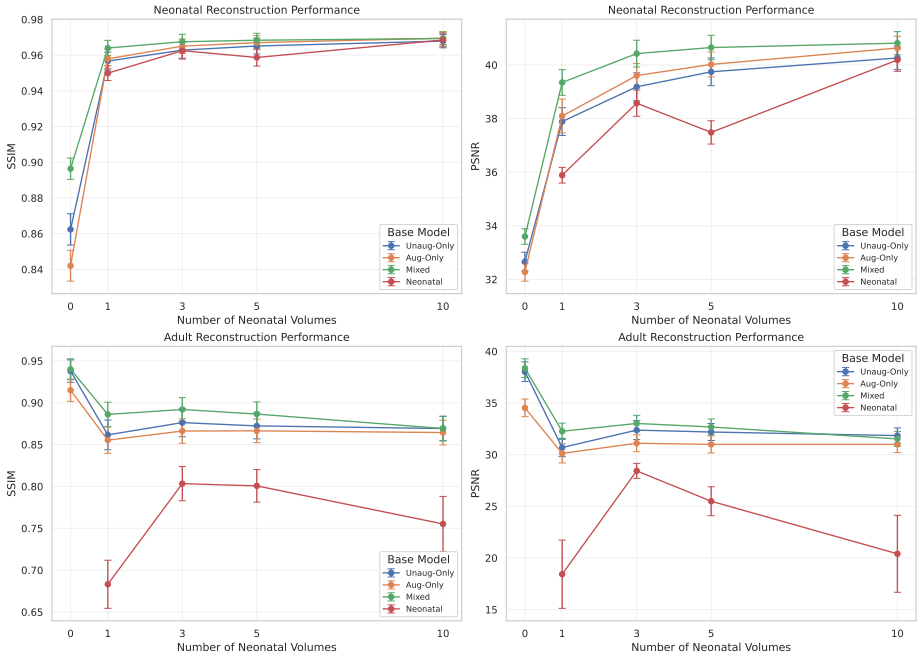

Fig. 2. Neonatal reconstruction performance as function of the number of neonatal fine-tuning or training volumes. Panels display $R = 4$ undersampling experiments. Mean SSIM (left column) and PSNR (right column) are shown for neonatal (top row) and adult (bottom row) data. Compares the Unaug-Only, Mixed, Aug-Only, and Neonatal models. For the adult-trained models, the x-axis indicates the number of neonatal volumes used for fine-tuning; for the neonatal-only model, it indicates the number of neonatal volumes used for training.

Under $R = 8$ undersampling, all fine-tuned pretrained models produced reconstructions with significantly higher mean SSIM and PSNR values than those produced by the Neonatal models. The best-performing models for each pretraining distribution were those fine-tuned with 10 neonatal volumes; of these, the fine-tuned Mixed model had the highest mean SSIM (0.948 ± 0.012, versus 0.945 ± 0.011 and 0.946 ± 0.012 for Aug-Only and Unaug-Only, respectively) and PSNR (36.02 ± 0.76 versus 35.73 ± 0.78 and 35.54 ± 0.67 for Aug-Only and Unaug-Only, respectively). Of these, only the Aug-Only and Unaug-Only fine-tuned models had no significantly different mean SSIM values, and only the Aug-Only and Mixed fine-tuned models had no significantly different mean PSNR values. As at $R = 4$, the largest gains followed the first neonatal fine-tuning volume.

4.3 Adult-Domain Retention After Neonatal Fine-Tuning

Before neonatal fine-tuning, Unaug-Only and Mixed produced reconstructions with mean SSIM and PSNR values not significantly different under both $R = 4$

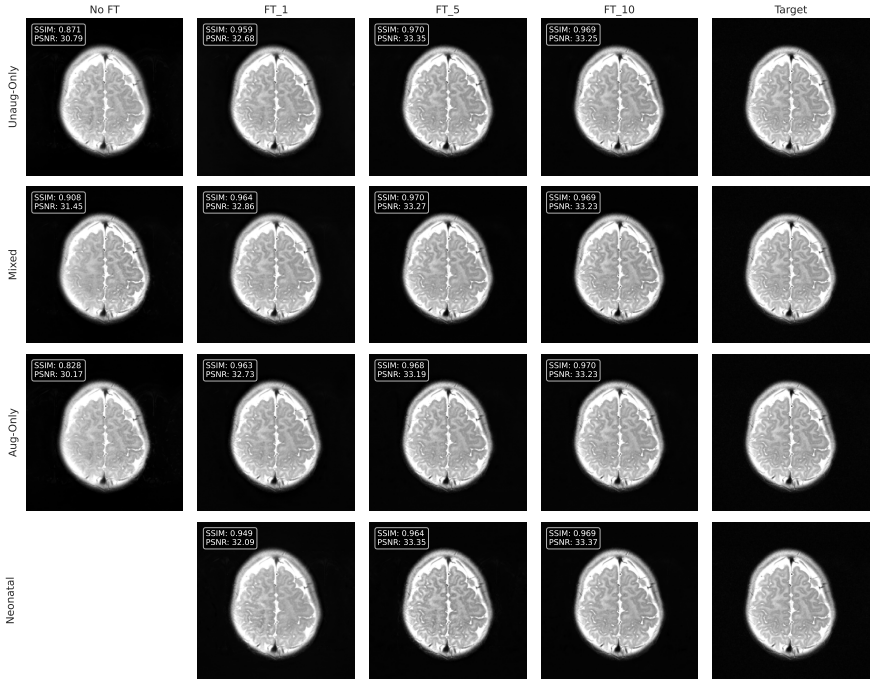

Fig. 3. Representative neonatal reconstructions of $R = 4$ undersampled data from adult-trained, fine-tuned, and neonatal-control models. Rows show reconstructions from models trained initially on unaugmented adult data (Unaug-Only), mixed unaugmented and augmented adult data (Mixed), augmented adult data only (Aug-Only), or neonatal data only (Neonatal). For the adult-trained models, columns show the base model before neonatal fine-tuning (No FT) and after fine-tuning with 1, 5, or 10 neonatal volumes; the 3-volume fine-tuning condition is omitted for visual clarity. The final column shows the fully sampled target image. Reconstructions are cropped to the central 256×256 pixels for display. SSIM and PSNR are displayed for each reconstruction and were calculated from the uncropped images. Similar trends were observed under $R = 8$ undersampling; representative reconstructions were omitted due to space constraints.

and $R = 8$ undersampling. The Aug-Only model produced reconstructions with significantly lower SSIM and PSNR values compared to both other models. For Unaug-Only and Mixed, neonatal fine-tuning reduced adult-domain performance, with the largest relative decrease between the pretrained baseline models and those fine-tuned on one neonatal volume (Figure 2). The Aug-Only and fine-tuned Aug-Only models performed similarly on the held-out adult test set.

5 Discussion

We designed and implemented a neonatal contrast-informed augmentation protocol and investigated how its incorporation into adult pretraining data affected zero-shot neonatal generalization, few-shot neonatal adaptation, and adult-domain retention. Mixed pretraining produced the best zero-shot neonatal reconstruction, while augmented-only pretraining did not consistently improve zero-shot performance and reduced adult-domain retention. This suggests that neonatal-informed augmentation was useful, but only when it remained paired with unaugmented adult data. Practically, these findings suggest that target-informed augmentation should be treated as a supplement to real source-domain data rather than as a replacement distribution, particularly when source-domain retention remains important.

The zero-shot generalization results suggest that the augmentation was beneficial when it expanded the distribution of adult-derived training data, but not when it replaced all unaugmented adult data. Augmented-only training sought to introduce neonatal-like contrast to the training set, but may have sacrificed realistic acquisition and reconstruction statistics. However, the observed performance gains may be due to the inherent variability introduced by augmentation; further experiments investigating the performance gains arising from contrast-informed augmentations compared to more general transformations (rotations, reflections, *etc.*) would help determine whether the improved generalizability is related specifically to the target domain-informed nature of the augmentations.

Neonatal fine-tuning significantly improved neonatal reconstruction performance across all pretraining distributions, and the fine-tuned models outperformed the Neonatal control models for each number of neonatal volumes. The largest gain occurred from no fine-tuning (*i.e.*, pretraining baselines) to fine-tuning with one neonatal volume, suggesting that even minimal target-domain supervision can strongly calibrate adult-pretrained reconstruction models to neonatal data. This suggests that adult pretraining is valuable when neonatal data are scarce, especially considering the poor performance of the Neonatal-Only models on adult test data.

This study has several limitations. First, the neonatal dataset was relatively small, collected at a single site, and limited to one acquisition protocol. The adult and neonatal acquisitions also differed in through-plane resolution (typically 5 mm versus 3 mm, respectively), which was not explicitly addressed in these slice-wise experiments. Second, the contrast-informed augmentation protocol was intentionally simple and approximated relatively few aspects of the adult-to-neonatal domain shift; specifically, it modified tissue contrast but retained adult anatomy, including skull and extracranial structures. The augmentation strategy was also not compared with established domain-adaptation approaches, limiting conclusions regarding its relative effectiveness. Third, all experiments used retrospectively undersampled data. Future work should evaluate prospectively accelerated acquisitions and include a more extensive clinical image-quality assessment to determine whether the observed metric improvements translate to meaningful improvements in neonatal MR reconstruction.

6 Conclusion

This study investigated the effect of adult pretraining data distributions on the generalizability of adult-trained DL-based MR reconstruction models to neonatal data. Models trained on both unaugmented and contrast-informed augmented adult data generalized better to neonatal data than either type independently, and models trained on exclusively augmented adult data did not consistently perform better on neonatal test data and performed worse on held-out adult data. After neonatal fine-tuning, adult-pretrained models performed significantly better compared to their baselines and generally outperformed neonatal-only models trained with the same amount of neonatal data, especially under higher acceleration. These findings suggest that contrast-informed augmentation is most useful when it expands, rather than replaces, the adult pretraining distribution.

Acknowledgments. This work was supported by research funding provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Alberta Innovates LevMax Health Programs, the Digital Research Alliance of Canada compute infrastructure, and funding from the Alberta Children's Hospital Foundation, Calgary Health Foundation, and the Canada First Research Excellence Fund (One Child Every Child).

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

References

1. Antonov, N.K., et al.: Feed and Wrap MRI Technique in Infants. Clinical Pediatrics **56**(12), 1095–1103 (2017)
2. Dubois, J., et al.: The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. Neuroscience **276**, 48–71 (2014)
3. Dubois, J., et al.: MRI of the Neonatal Brain: A Review of Methodological Challenges and Neuroscientific Advances. Journal of Magnetic Resonance Imaging **53**(5), 1318–1343 (2021)
4. Fabian, Z., Heckel, R., Soltanolkotabi, M.: Data augmentation for deep learning based accelerated MRI reconstruction with limited data. In: Proceedings of the 38th International Conference on Machine Learning. pp. 3057–3067. PMLR (2021)
5. Griswold, M.A., et al.: 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)
6. Heckel, R., et al.: Deep learning for accelerated and robust MRI reconstruction. Magnetic Resonance Materials in Physics, Biology and Medicine **37**(3), 335–368 (2024)
7. Jiang, J., et al.: Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets. Medical Physics **46**(10), 4392–4404 (2019)
8. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization (2017)
9. Knoll, F., et al.: Assessment of the generalization of learned image reconstruction and the potential for transfer learning. Magnetic Resonance in Medicine **81**(1), 116–128 (2019)

10. Knoll, F., et al.: Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues. IEEE Signal Processing Magazine **37**(1), 128–140 (2020)
11. Larkman, D.J., Nunes, R.G.: Parallel magnetic resonance imaging. Physics in Medicine & Biology **52**(7), R15 (2007)
12. Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed Sensing MRI. IEEE Signal Processing Magazine **25**(2), 72–82 (2008)
13. Muckley, M.J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M., Arberet, S., Nickel, D., Ramzi, Z., Ciuciu, P., Starck, J.L., Teuwen, J., Karkalousos, D., Zhang, C., Sriram, A., Huang, Z., Yakubova, N., Lui, Y.W., Knoll, F.: Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. IEEE Transactions on Medical Imaging **40**(9), 2306–2317 (Sep 2021). https://doi.org/10.1109/TMI.2021.3075856, https://ieeexplore.ieee.org/abstract/document/9420272
14. Omidi, A., et al.: Improving skull-stripping for infant MRI via weakly supervised domain adaptation using adversarial learning. Computers in Biology and Medicine **196**, 110903 (2025)
15. Pekarsky, C., Skiffington, J., Leijser, L.M., Slater, D., Metcalfe, A.: Social Media Recruitment Strategies to Recruit Pregnant Women Into a Longitudinal Observational Cohort Study: Usability Study. Journal of Medical Internet Research **24**(12), e40298 (Dec 2022). https://doi.org/10.2196/40298, https://www.jmir.org/2022/12/e40298
16. Pruessmann, K.P., et al.: 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)
17. Richiardi, J., et al.: Chapter 6 - Domain shift, domain adaptation, and generalization: A focus on MRI. In: Trustworthy AI in Medical Imaging, pp. 127–151. The MICCAI Society book Series, Academic Press (2025)
18. Sanford, T.H., et al.: Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model. American Journal of Roentgenology **215**(6), 1403–1410 (2020)
19. Sriram, A., et al.: End-to-End Variational Networks for Accelerated MRI Reconstruction. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 64–73. Springer International Publishing (2020)
20. Uecker, M., et al.: ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. Magnetic Resonance in Medicine **71**(3), 990–1001 (2014)
21. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)