Low-Rank Adaptation for Efficient Contrast Specialization of 3D MRI Diffusion Models

Blake E. Dewey[1,2][0000−0003−4554−5058], Samuel W. Remedios[2][0000−0001−8634−8128], Muhammad F. A. Chaudhary[2][0000−0002−0807−4560], Aaron Carass[2][0000−0003−4939−5085], and Jerry L. Prince[2][0000−0002−6553−0876]

[1] Department of Neurology, Johns Hopkins University, Baltimore, MD, USA
[2] Image Analysis and Communications Laboratory, Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
blake.dewey@jhu.edu

Abstract. Diffusion models in medical imaging have emerged as important priors suited for numerous downstream analysis tasks. Training a separate 3D diffusion model from scratch for each MRI contrast is computationally expensive and difficult to scale. Pretrained models enable data- and time-efficient adaptation, but conventional fine-tuning requires maintaining a complete set of model weights per target contrast. We study Low-Rank Adaptation (LoRA) as a parameter-efficient mechanism for contrast specialization of pretrained 3D MRI diffusion models without modifying the backbone weights. Adapting a pretrained T_1-weighted model to T_2-weighted MRI, we analyze model behavior under different adapter ranks and training hyperparameters, evaluating anatomical distribution alignment with WASABI and perceptual similarity with FID. LoRA achieves anatomical alignment comparable to full fine-tuning (WASABI 2.10 vs. 2.29) while updating only 9.6% of model parameters, with modest but still competitive perceptual quality (FID 38.41 vs. 34.37). The gap between WASABI and FID indicates that low-rank updates recover global anatomical alignment more readily than fine-grained perceptual texture. The resulting adapters are lightweight and interchangeable: a rank-16 adapter requires 15.04 MB compared with 157.34 MB for a full fine-tuned checkpoint. The source code and model weights are made available at: https://github.com/piksl-research/t2w-lora-diffusion.

Keywords: generative models · magnetic resonance imaging · parameter-efficient adaptation.

1 Introduction

Diffusion models, particularly denoising diffusion probabilistic models (DDPMs) [10], have demonstrated strong performance in 3D medical image generation [6,30]. In magnetic resonance imaging (MRI), DDPMs have emerged as general-purpose

generative priors. They can support contrast synthesis, data augmentation for supervised learning, reconstruction and super-resolution in inverse problems, lesion inpainting, domain harmonization across sites and protocols, and uncertainty-aware anatomical analysis [6, 17, 21–24, 30]. However, training such models from scratch requires large-scale datasets and substantial computational resources, making it impractical to train separate models for each imaging contrast.

Pretrained diffusion models offer a pathway toward efficient cross-contrast adaptation. Conventional full fine-tuning reduces time and data requirements relative to training from scratch [18] but updates and stores the entire set of model weights for each target contrast, resulting in limited scalability for multi-contrast deployment. Parameter-efficient fine-tuning (PEFT) methods address this by updating only a subset of parameters while preserving the pretrained weights. Low-Rank Adaptation (LoRA) introduces structured low-rank updates to convolutional layers, enabling contrast-specific specialization without modifying the base architecture [11]. The target deployment scenario is maintaining one large pretrained 3D generative backbone with lightweight contrast-specific adapters for distribution-level generative tasks such as simulation, augmentation, harmonization studies, and foundation-model reuse.

Recent work has explored structured parameterizations for 3D convolutional kernels in 3D diffusion models [16]. These approaches primarily focused on increasing representational flexibility of parameter-efficient updates and were evaluated on preprocessed T_1-weighted imaging across datasets. However, the effectiveness of LoRA for contrast adaptation in 3D diffusion models has not been systematically characterized. In particular, the relationship between adapter rank, training hyperparameters, anatomical distribution alignment, and perceptual image quality is still unknown in 3D generative settings.

In this work, we investigate parameter-efficient contrast specialization in 3D diffusion models of MRI. Starting from a large-scale pretrained 3D T_1-weighted model, we adapt to a smaller T_2-weighted MRI dataset using low-rank updates and compare against full fine-tuning and training from scratch. Our contributions are:

1. **Reusable anatomical priors for cross-contrast adaptation.** Comparing zero-shot T_1 generation, T_2 training from scratch, full fine-tuning, and LoRA shows that the pretrained 3D diffusion backbone contains anatomical structure that transfers across contrast.
2. **Low-dimensional contrast specialization.** LoRA recovers anatomical distribution alignment comparable to full fine-tuning while updating fewer than 10% of parameters on a multi-dataset cohort of 4,981 T_2-weighted volumes.
3. **Joint perceptual and anatomical evaluation.** We assess adaptation quality using both FID [9] and WASABI [13], examining the relationship between perceptual realism and anatomical distribution alignment.
4. **Efficiency and modularity analysis.** We quantify checkpoint size across ranks and motivate a multi-adapter strategy for MRI contrasts.

2 Methods

2.1 Flow-Matching DDPMs

Models in this work use flow-matching DDPMs [7] as the basis for generation. Let $x_0 \sim p(x_0)$ denote a 3D MRI sample and $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denote a noise sample of the same dimensions. In the forward diffusion process, each state x_t ($t \in [0, 1]$) is expressed as a linear combination between data and noise:

$$x_t = \alpha_t x_0 + \beta_t \epsilon, \tag{1}$$

where α_t and β_t are determined by the chosen noise schedule (making $x_1 = \epsilon$) and $\alpha_t^2 + \beta_t^2 = 1$ for all t to preserve variance across the entire diffusion trajectory.

The reverse process is defined as a reparameterization:

$$x_s = \alpha_s \hat{x}_0 + \beta_s \hat{\epsilon}, \tag{2}$$

where denoising proceeds by first estimating $\hat{x}_0$ and $\hat{\epsilon}$ from x_t, then reparameterizing to an earlier timepoint x_s ($s < t$). In flow-matching DDPMs, a deep neural network predicts a flow matching field: $\hat{\mu} = \hat{\epsilon} - \hat{x}_0 = f_\theta(x_t, t)$. This yields estimates of:

$$\hat{x}_0 = \left(x_t - \beta_t f_\theta(x_t, t)\right) / \left(\alpha_t + \beta_t\right), \tag{3}$$

$$\hat{\epsilon} = \left(x_t + \alpha_t f_\theta(x_t, t)\right) / \left(\alpha_t + \beta_t\right). \tag{4}$$

2.2 Low-Rank Adaptation (LoRA)

Let $W \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}}}$ denote the weights and $b \in \mathbb{R}^{C_{\text{out}}}$ denote the bias of a linear layer in a pretrained model applied to input vector $x \in \mathbb{R}^{C_{\text{in}}}$. In full fine-tuning, all parameters of W are updated. PEFT methods instead model a structured update:

$$Y = (W + \Delta W)x + b, \tag{5}$$

where X is the input feature and ΔW has significantly fewer trainable parameters than W.

In LoRA, ΔW is decomposed into two low-rank matrices, $A \in \mathbb{R}^{C_{\text{out}} \times r}$ and $B \in \mathbb{R}^{r \times C_{\text{in}}}$, where $r \ll \min\left(C_{\text{in}}, C_{\text{out}}\right)$ and the rank r controls the capacity of the adaptation. This allows a final form of:

$$Y = \left(W + \frac{\alpha}{r} AB\right) x + b, \tag{6}$$

where the total parameters to update are $r\left(C_{\text{in}} + C_{\text{out}}\right) \ll C_{\text{in}} C_{\text{out}}$ and α is a scaling hyperparameter.

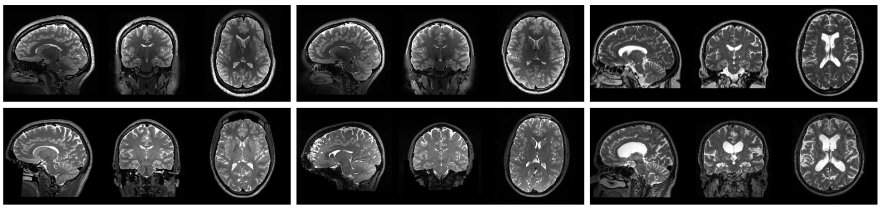

Fig. 1. Triplanar (sagittal, coronal, and axial) views of 6 volumes in the training set.

2.3 Adapting 3D Convolution Layers

Let $W \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times k_d \times k_h \times k_w}$ denote the weights of a 3D convolutional layer in a pretrained diffusion model applied to input tensor $X \in \mathbb{R}^{C_{\text{in}} \times k_d \times k_h \times k_w}$. ΔW is parameterized using a low-rank factorization in channel space. The kernel update is reshaped as $\Delta W_{\text{flat}} \in \mathbb{R}^{C_{\text{out}} \times (C_{\text{in}} k_d k_h k_w)}$, which is decomposed as $A \in \mathbb{R}^{r \times (C_{\text{in}} k_d k_h k_w)}$ and $B \in \mathbb{R}^{C_{\text{out}} \times r}$, and then reshaped to match W:

$$Y = \left(W + \frac{\alpha}{r} \text{reshape}(AB) \right) * X + b, \tag{7}$$

where $*$ denotes 3D convolution.

2.4 Evaluation Criteria

We evaluate adaptation quality using complementary metrics: Fréchet Inception Distance (FID) [9] to measure perceptual similarity, and WASABI (Wasserstein Anatomical Similarity of Brain Indices) [13] to measure anatomical distribution alignment.

FID was computed using `pytorch-fid` [26]. `pytorch-fid` uses a pretrained InceptionV3 model to extract features, resulting in a 2048-length vector for each image. As InceptionV3 [28] operates in 2D, 32 slices spaced 4 mm apart were extracted from the center of the volume in each slice orientation (axial, sagittal, and coronal) for a total of 96 slices per volume. This slice-wise FID is a pragmatic proxy for perceptual similarity and does not directly measure volumetric coherence.

WASABI computes Earth Mover's Distance [25] between distributions of regional brain volumes computed from two different samples, using squared Euclidean distance as the distance metric. Fifty-two volumes (averaged over left and right) are computed using SynthSegV2 [3] with the robust segmentation option [4]. WASABI is bootstrapped one thousand times over random subsets (80% of total) to provide statistics on the metrics. As WASABI aggregates regional volumes, it assesses global morphometric plausibility to complement FID results.

Table 1. Distribution of subjects (S) and image volumes (V) by dataset for T_2-weighted experiments.

Dataset	Subjects (S)	Volumes (V)	Train S	Train V	Test S	Test V
HBN [1]	1155	1155	1048	1048	107	107
HCP1200 [29]	1113	1113	1010	1010	103	103
IXI [12]	552	552	501	501	51	51
MPILEMON [2]	225	225	204	204	21	21
NIMHHRVD [20]	247	251	224	227	23	24
OASIS3 [15]	1021	1638	927	1488	94	150
NARRATIVES [19]	46	47	–	–	46	47
Total	4359	4981	3914	4478	445	503

3 Experiments and Results

3.1 Pretrained Model and Adaptation Dataset

Pretrained Model. We begin with a previously trained 3D U-Net-based flow-matching diffusion model pretrained on 72k high-resolution, T_1-weighted brain MRI volumes from 38 online datasets [23]. This pretrained model achieves an FID of 27.2 on withheld subjects from the training datasets and 39.9 on two completely withheld datasets. The pretrained model achieves a WASABI score of 4.36 ± 0.569 against withheld test data (compared to the lower limit of 0.48 ± 0.065 computed between subsets of the test data).

Contrast Adaptation Dataset. The dataset used for contrast adaptation consists of 4,981 high-resolution, T_2-weighted brain MRI volumes from 7 online datasets, with 4,478 volumes used for training (Table 1 and Figure 1). A minimal preprocessing pipeline was applied to these images: resample and reorientation to 1mm axial, padding/cropping to 192x224x192mm, linear min-max normalization and quantization to 16-bit integer values. Approximately 10% of the volumes from each dataset were withheld as testing data with no subjects appearing in both training and testing data.

3.2 Model Implementation

We built all models using a MONAI [5] 3D U-Net with five resolution levels (16–256 channels, 2x per level), attention at the bottleneck and final layer, and two ResNet blocks per level. Training used the Adam optimizer [14] with a batch size of 24 (across 8 NVIDIA H200 GPUs) and random spatial augmentation (3D rotations up to $\pm 5°$ and 3D translations up to ± 5 mm). The pretrained T_1-weighted model was trained for 150 epochs with exponential moving average (EMA) updates applied to each batch. LoRA adapters were applied to all 3D convolutional and linear layers. A matrices were initialized with uniform Kaiming [8] and B matrices were initialized to zero. Bias vectors and normalization layers were also updated during training ($\sim$35k parameters). We additionally trained (i) a model from

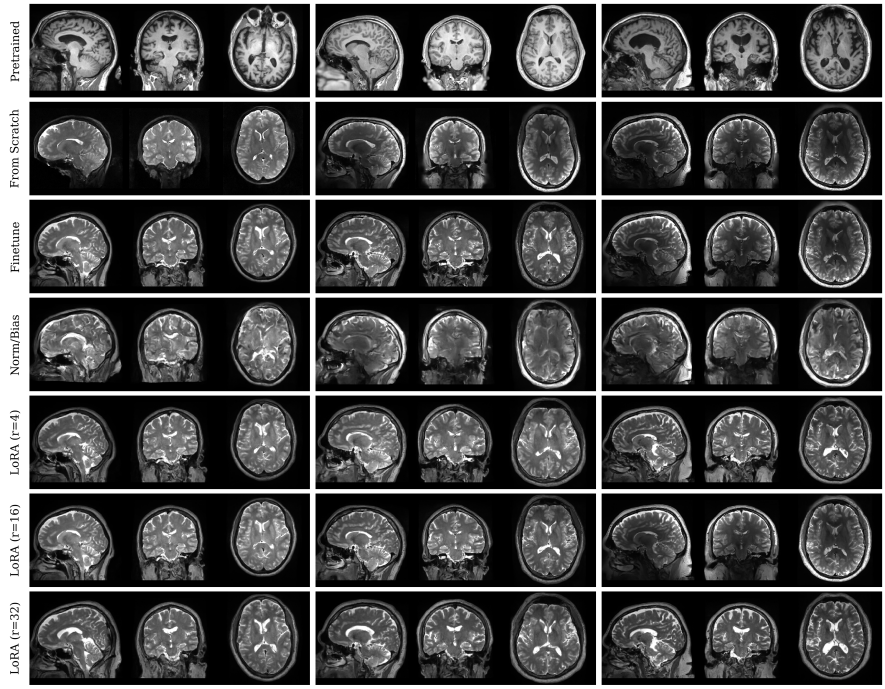

Fig. 2. Triplanar view (sagittal, coronal, and axial) views of 3 synthetic volumes generated by each of our tested models in Table 2. Columns are generated using the same random seed.

scratch on T_2-weighted data, (ii) a fully fine-tuned model initialized from the T_1 checkpoint, and (iii) a minimal baseline updating only normalization and bias parameters. Inference was performed using the denoising diffusion implicit model (DDIM) framework [27] with 64 inference steps.

3.3 Baseline Performance

Table 2 summarizes anatomical alignment (WASABI) and perceptual similarity (FID). The pretrained T_1 model exhibits severe mismatch on T_2 data (WASABI $= 31.29 \pm 0.89$; FID $= 77.65$), so the source model is not usable in a zero-shot setting, as expected. Training from scratch on T_2 improves WASABI to 5.63 ± 0.26, but remains worse than adaptation-based strategies despite extended training (FID $48.56 \rightarrow 42.59$ from 200 to 600 epochs). Full fine-tuning reaches WASABI $= 2.29 \pm 0.17$ and FID $= 34.37$ within 200 epochs. Together, these baselines indicate that the pretrained T_1 backbone contains reusable anatomical structure, as adaptation is substantially better than relearning the T_2 distribution from scratch. Figure 2 shows qualitative results for baselines in rows 1–3.

Table 2. Anatomical and perceptual performance for contrast adaptation from T_1-weighted to T_2-weighted. WASABI scores are mean score ± standard deviation. All LoRA models trained with learning rate of 5×10^{-4}. Best results per column are shown in bold and second-best are underlined.

Method	WASABI ↓	FID ↓
T_1 Pretrained	31.29 ± 0.89	77.65
From Scratch	5.63 ± 0.26	42.59
Full Fine-Tuning	<u>2.29 ± 0.17</u>	**34.37**
Norm/Bias Only	25.36 ± 1.25	94.73
LoRA ($r = 4$)	2.58 ± 0.15	46.44
LoRA ($r = 16$)	**2.10 ± 0.14**	<u>38.41</u>
LoRA ($r = 32$)	2.77 ± 0.25	42.76

3.4 LoRA Performance

We conducted sweeps over rank ($r \in \{4, 16, 32\}$), learning rate (10^{-4}, 5×10^{-4}, 10^{-3}), and training duration (up to 300 epochs) to determine stable optimization regimes. LoRA required a higher learning rate than the 10^{-5} for full fine-tuning due to the zero initialization of the B matrices. As shown in Fig. 3, low learning rates (10^{-4}) produced reduced anatomical and contrast diversity (e.g., similar contrasts with larger ventricles across the samples, WASABI > 12 and FID increase ~ 7 across ranks). Conversely, high learning rates (10^{-3}) introduced instability and degraded perceptual quality (e.g., FID = 43.9 for $r = 16$). A learning rate of 5×10^{-4} yielded stable convergence and consistently strong anatomical alignment across ranks.

Training duration further differentiated adaptation strategies. Full fine-tuning converged rapidly, with minimal improvement beyond 200 epochs and slight degradation in FID at later epochs, suggesting mild overfitting. LoRA exhibited similar behavior: for $r = 16$, FID increased from 38.4 at 200 epochs to 40.6 at 300 epochs despite stable anatomical alignment.

Updating only normalization layers and bias terms (no LoRA adapters; $r = 0$) failed to recover either anatomical or perceptual fidelity (WASABI = 25.36, FID = 94.7), indicating that specialization is not explained by simple feature rescaling and requires spatially structured updates to the denoising model.

Using optimal hyperparameters (5×10^{-4}, 200 epochs), LoRA with $r = 16$ achieved anatomical alignment comparable to full fine-tuning (WASABI = 2.10 ± 0.14 vs. 2.29 ± 0.17). The rank sweep was non-monotonic: $r = 4$ remained competitive but degraded both metrics, while $r = 32$ worsened relative to $r = 16$ (WASABI 2.77 vs. 2.10; FID 42.76 vs. 38.41). Thus, adapter capacity is not a monotonic quality knob in this 3D diffusion setting.

Perceptual similarity followed a different trend from anatomical alignment. Full fine-tuning achieved the lowest FID (34.37), while LoRA $r = 16$ approached but did not match it (38.41). This anatomy–texture dissociation is the central tradeoff: low-rank updates recover global morphometric distribution alignment

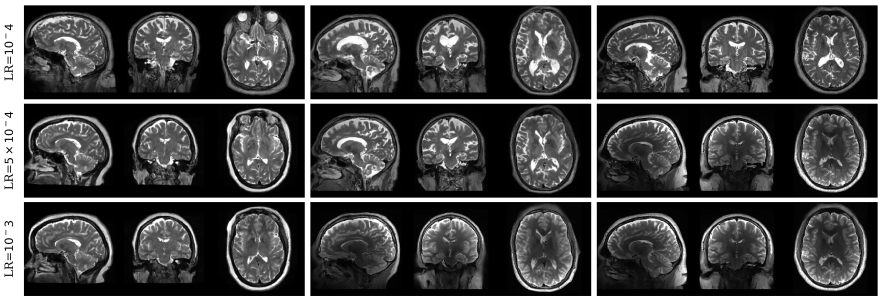

Fig. 3. Triplanar views (sagittal, coronal, and axial) of 3 synthetic volumes generated by a LoRA model (r=16) changing the learning rate used in training. Columns are generated using the same random seed.

Table 3. Trainable parameter count and checkpoint size for different adaptation strategies. Relative size is reported with respect to full fine-tuning.

Method	**# Params**	**Size (MB)**	**Size (%)**
From Scratch	41,245,217	157.34	100.0
Full Fine-Tuning	41,245,217	157.34	100.0
Norm/Bias Only	35,521	0.14	0.09
LoRA ($r = 4$)	1,012,161	3.86	2.45
LoRA ($r = 16$)	3,943,569	15.04	9.56
LoRA ($r = 32$)	7,852,609	29.96	19.04

more readily than fine-grained T_2 appearance. Qualitatively (Fig. 2), LoRA samples show plausible global anatomy and contrast structure while retaining modest texture differences relative to unconstrained fine-tuning.

3.5 Efficiency and Multi-Adapter Deployment

The training-time efficiency gain comes from adapting a pretrained model rather than training from scratch: on 8 NVIDIA H200 GPUs, pretraining required 1 hour 12 minutes per epoch, compared with 4 minutes 31 seconds for contrast adaptation, due to a smaller training set. LoRA does not reduce peak training memory or per-epoch compute relative to full fine-tuning because both require the same forward and backward passes. Its advantage here is storage, parameter, and deployment modularity (Table 3). For k target contrasts, full fine-tuning stores $157.34 \cdot k$ MB, whereas a shared backbone with rank-16 adapters stores $157.34 + 15.04 \cdot k$ MB. This reaches a storage advantage after two target contrasts and uses 277.66 MB for eight contrasts, compared with 1,258.72 MB for eight full fine-tuned checkpoints.

4 Conclusion

We demonstrated that pretrained 3D diffusion models contain reusable anatomical structure for cross-contrast specialization: adaptation from a T_1 prior outperformed training a T_2 model from scratch. LoRA recovered anatomical distribution alignment comparable to full fine-tuning with 9.6% of model parameters, but retained a FID gap, indicating that global anatomy and fine-grained texture respond differently to constrained adaptation. The non-monotonic rank sweep further shows that larger adapters are not automatically better for 3D diffusion specialization. These findings position LoRA as both a practical modular-deployment mechanism and a probe of what transfers in pretrained 3D MRI generative models. Future work should characterize low-data scaling, sequential multi-contrast adaptation, source-contrast retention, and downstream task utility.

Acknowledgments. This work received support from DoD CDMRP HT9425-24-1-0785 (Dewey), CDMRP W81XWH2010912 (Prince), and NIH R01EB036013 (Prince). Data used in this paper were sourced from open datasets; due to space constraints, please visit the citations of the appropriate data descriptor manuscripts for relevant information.

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

References

1. Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N. et al.: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data **4**(1), 170181 (Dec 2017)
2. Babayan, A., Erbey, M., Kumral, D., Reinelt, J.D., Reiter, A.M.F., Röbbig, J., Schaare, H.L., Uhlig, M. et al.: A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. Scientific Data **6**(1), 180308 (Feb 2019)
3. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (May 2023)
4. Billot, B., Magdamo, C., Cheng, Y., Arnold, S.E., Das, S., Iglesias, J.E.: Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. Proceedings of the National Academy of Sciences **120**(9), e2216399120 (Feb 2023)
5. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A. et al.: MONAI: An open-source framework for deep learning in healthcare (Nov 2022), arXiv.2211.02701
6. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional Diffusion Models for Semantic 3D Brain MRI Synthesis. IEEE Journal of Biomedical and Health Informatics **28**(7), 4084–4093 (Jul 2024)
7. Gao, R., Hoogeboom, E., Heek, J., De Bortoli, V., P., M.K., Salimans, T.: Diffusion Meets Flow Matching, https://diffusionflow.github.io/

8. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1026–1034. Santiago, Chile (Dec 2015)
9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6629–6640. NIPS'17 (Dec 2017)
10. Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models (Dec 2020), arXiv.2006.11239
11. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models. In: International Conference on Learning Representations (2022)
12. IXI Dataset – Brain Development, https://brain-development.org/ixi-dataset/
13. Jafrasteh, B., Peng, W., Wan, C., Luo, Y., Adeli, E., Zhao, Q.: WASABI: A Metric for Evaluating Morphometric Plausibility of Synthetic Brain MRIs. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 684–694 (2026)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
15. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J. et al.: OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease (Dec 2019)
16. Li, B., Chang, Z., Liang, T., Li, C., Tanaka, T., Aoki, S., Zhao, Q., Sun, Z.: Parameter-Efficient Fine-Tuning of 3D DDPM for MRI Image Generation Using Tensor Networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. 15963, pp. 382–392 (2026)
17. Mallardi, G., Calefato, F., Lanubile, F., Logroscino, G., Tafuri, B.: Diffusion Models for Neuroimaging Data Augmentation: Assessing Realism and Clinical Relevance. Journal of Medical Systems **49**(1), 161 (Nov 2025)
18. Moon, T., Choi, M., Lee, G., Ha, J.W., Lee, J.: Fine-tuning diffusion models with limited data. In: NeurIPS 2022 Workshop on Score-Based Methods (2022)
19. Nastase, S.A., Liu, Y.F., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N., Chen, J., Honey, C.J. et al.: The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. Scientific Data **8**(1), 250 (Sep 2021)
20. Nugent, A.C., Thomas, A.G., Mahoney, M., Gibbons, A., Smith, J.T., Charles, A.J., Shaw, J.S., Stout, J.D. et al.: The NIMH intramural healthy volunteer dataset: A comprehensive MEG, MRI, and behavioral resource. Scientific Data **9**(1), 518 (Aug 2022)
21. Pollak, C., Kügler, D., Bauer, T., Rüber, T., Reuter, M.: FastSurfer-LIT: Lesion inpainting tool for whole-brain MRI segmentation with tumors, cavities, and abnormalities. Imaging Neuroscience **3**, imag_a_00446 (Jan 2025)
22. Qraitem, M., Saenko, K., Plummer, B.A.: From Fake to Real: Pretraining on Balanced Synthetic Images to Prevent Spurious Correlations in Image Recognition. In: Computer Vision – ECCV 2024. pp. 230–246 (2025)
23. Remedios, S.W., Carass, A., Prince, J.L., Dewey, B.E.: Diffusion-Driven Generation of Minimally Preprocessed Brain MRI (Oct 2025), arXiv.2510.26834
24. Remedios, S.W., Wei, S., Carass, A., Dewey, B.E., Prince, J.L.: Cycle-Consistent Zero-Shot Through-Plane Super-Resolution for Anisotropic Head MRI. In: Information Processing in Medical Imaging. pp. 249–264 (2025)

25. Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In: Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). pp. 59–66 (1998)
26. Seitzer, M.: pytorch-fid: FID Score for PyTorch (Feb 2026), https://github.com/mseitzer/pytorch-fid
27. Song, J., Meng, C., Ermon, S.: Denoising Diffusion Implicit Models (Oct 2022), arXiv.2010.02502
28. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2818–2826. Las Vegas, NV, USA (Jun 2016)
29. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L. et al.: The Human Connectome Project: A data acquisition perspective. NeuroImage **62**(4), 2222–2231 (Oct 2012)
30. Zhao, C., Guo, P., Yang, D., Tang, Y., He, Y., Simon, B., Belue, M., Harmon, S. et al.: MAISI-v2: Accelerated 3D High-Resolution Medical Image Synthesis with Rectified Flow and Region-specific Contrastive Loss (Aug 2025), arXiv.2508.05772