Anatomy- and Site-Guided 3D Patch Diffusion for Robust MRI Harmonization

Barnabé Hache[1], Vincent Roca[2], Grégory Kuchcinski[3], Dorian Manouvriez[2], and Renaud Lopes[1]

[1] Inserm U1172, LilNCog – Lille Neuroscience & Cognition, Lille, France
barnabe.hache@inserm.fr
[2] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41 – UAR 2014 – PLBS, Lille, France
[3] CHU Lille, Department of Neuroradiology, Lille, France

Abstract. Pooling multi-site T1-weighted MRI increases power but introduces scanner-driven variability that can bias analyses. We propose a 3D patch-based conditional diffusion framework for MRI harmonization that preserves fine anatomical details while aligning image contrast across sites. Our method constructs site-invariant anatomical maps using gradient and high-frequency features, which guide patch-wise diffusion synthesis together with learnable site embeddings. Processing overlapping high-resolution patches avoids latent-decoder blurring and reduces memory requirements, enabling faithful preservation of fine anatomical detail. To further enhance anatomical fidelity, we employ manifold-constrained classifier-free guidance, stabilizing diffusion sampling while maintaining subject-specific structures. We evaluate our framework on a multi-site traveling-subject dataset and a clinically acquired pathological cohort, comparing against state-of-the-art harmonization methods. Results show that our approach effectively reduces site-related intensity variability, preserves anatomical and tissue-specific information, and maintains clinically relevant brain–cognition associations. These findings demonstrate that patch-based, anatomy-guided diffusion harmonization provides a robust, high-fidelity solution for multi-site MRI, with clear practical benefits for clinical datasets and downstream analyses. Code and model weights are available at https://github.com/barna-hache/PatchLevelDiffusionMRIHarmonization.

Keywords: Harmonization · MRI · Diffusion Model

1 Introduction

Combining structural magnetic resonance imaging (MRI) across sites increases statistical power but introduces non-biological scanners and site effects that bias downstream analyses and reduce reproducibility [13,1]. Classical harmonization methods such as ComBat [10] effectively correct site effects at the feature level (e.g., cortical thickness) but operate on derived measurements, limiting their suitability when image-level consistency is required.

Image-level harmonization methods aim to mitigate scanner-related variability using learning-based solutions. Most approaches rely on U-Net or GAN-based domain translation frameworks, including DeepHarmony [5], HACA3 [26], STGAN [16], and IGUANe [19]. Many incorporate structural–contrast disentanglement, for example via gamma perturbations to isolate anatomical representations [2]. However, insufficient structural constraints may lead to poor generalization or unintended anatomical alterations [4].

Diffusion-based harmonization methods [25,15,18] leverage the high generative fidelity of diffusion models [6] to align contrast while preserving anatomy. Most adopt latent diffusion to reduce computational cost, but compression–decoding operations may blur fine structural details. Full-resolution volumetric diffusion models, in contrast, remain memory-intensive and difficult to scale to clinical datasets [25].

Importantly, harmonization methods must remain robust under real-world clinical variability. While many approaches are validated on healthy research cohorts, clinical datasets exhibit greater heterogeneity due to pathology and acquisition variability, making anatomical faithfulness critical.

To address these challenges, we propose a patch-based, anatomy-guided 3D diffusion framework for multi-site T1-weighted brain MRI harmonization. Processing overlapping patches at native resolution preserves fine anatomical detail while reducing memory requirements. The model disentangles site-specific contrast from structural information via conditioning on a site-invariant anatomical prior and a learnable site embedding, enabling controlled target-site harmonization without anatomical displacement. Compared to latent diffusion, this strategy avoids decoder-induced blurring, and, relative to GAN-based translation, diffusion sampling provides a more stable generative process [6,25].

Our contributions are : (i) a 3D patch-level conditional diffusion pipeline that harmonizes MRI volumes to an arbitrary target site while preserving anatomical structure; (ii) an anatomy-conditioning strategy with randomized gamma perturbation to prevent leakage of site-specific intensity cues into the structural prior; and (iii) an extensive evaluation on the SRPBS traveling-subject dataset, comparing our method to leading baselines using quantitative metrics that assess image intensity distributions, anatomical segmentation consistency, and site-related variance via mixed-effects modeling, complemented by assessments on a clinically acquired pathological cohort to evaluate the preservation of biologically relevant markers.

2 Method

Overview. We propose a 3D patch-based framework to harmonize T1-weighted brain MRI volumes from arbitrary source sites into the contrast space of a chosen target site while preserving subject-specific anatomy. The pipeline (Fig. 1) comprises: (i) a site-invariant anatomical representation, (ii) paired 3D patch extraction with coarse anatomical context, (iii) conditional diffusion-based patch synthesis guided by anatomy and a learnable site embedding, and (iv) patch-wise

inference with overlap and Hann-weighted fusion to reconstruct a full volume. At test time, unseen source sites can be harmonized to a trained target site as anatomy is provided by site-agnostic maps, whereas target appearance is supplied via a learned site embedding derived from the site label.

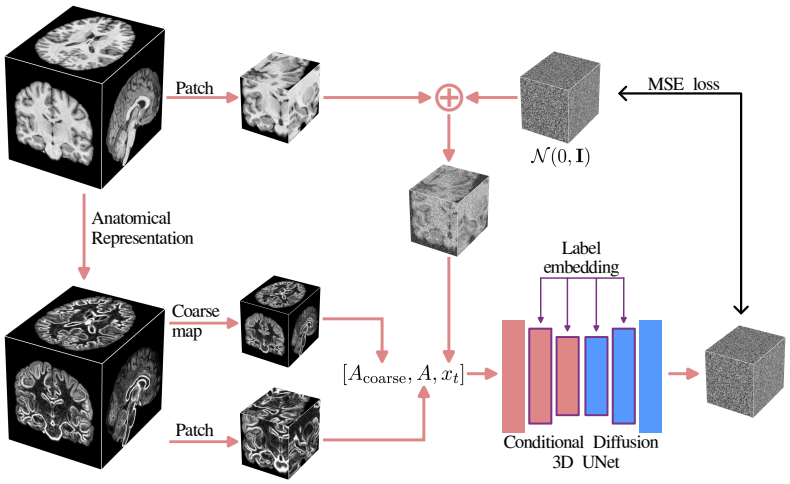

Fig. 1. Overview of the training pipeline. Anatomical representations are computed and paired ($80 \times 96 \times 80$) patches with coarse context are used for diffusion synthesis. Site labels are embedded and injected via cross-attention.

Anatomical representation. To obtain a site-invariant anatomical prior we suppress contrast cues by first applying a global gamma perturbation,

$$I^{\gamma} = \text{sign}(I)\,|I|^{\exp(\gamma_{\text{val}})}, \quad \gamma \sim \mathcal{U}(\gamma_{\min}, \gamma_{\max}),$$

with $(\gamma_{\min}, \gamma_{\max}) = (-0.5, 0.5)$ during training and $\gamma = 0$ at inference. From I^{γ} we compute gradient magnitude and high-frequency content

$$G = \|\nabla I^{\gamma}\|_2, \qquad H = \left|I^{\gamma} - \mathcal{G}_{\sigma_{\text{hf}}}(I^{\gamma})\right|,$$

and form the anatomical map $A = \alpha G + (1-\alpha)H$, which is 1^{st}–99^{th} percentile-normalized to $[-1, 1]$. Noisy image patches x_t patches of size $80 \times 96 \times 80$ (i.e. 1/8 of the volume) are sampled inside a brain bounding box. The corresponding anatomical map patch A is then extracted at identical coordinates and a coarse context A_{coarse} is obtained by trilinear downsampling the whole anatomical map to the same patch size.

Style embedding. Each acquisition site label is associated with a learnable embedding $S_k \in \mathbb{R}^{512}$ that is injected via cross-attention layers. The embedder is trained jointly with the diffusion model. At inference, harmonization is obtained by replacing the source embedding with the target embedding S_{target}.

Diffusion model. We adopt the denoising diffusion probabilistic model framework of [11] and explicitly condition the reverse denoising on two complementary signals: an anatomical map A that encodes subject-specific morphology, and a site-style embedding S that encodes scanner-dependent contrast. The forward noising process remains standard: $q(x_t \mid x_{t-1}) = \mathcal{N}\big(x_t; \sqrt{1-\beta_t}\, x_{t-1},\, \beta_t \mathbf{I}\big)$, with closed-form marginal

$$q(x_t \mid x_0) = \mathcal{N}\big(x_t; \sqrt{\bar{\alpha}_t}\, x_0,\, (1-\bar{\alpha}_t)\mathbf{I}\big), \quad \bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s).$$

The denoiser predict the noise conditioned on anatomy and style:

$$\epsilon_\theta(x_t, t, c), \qquad c = (A, A_{\text{coarse}}, S).$$

This design lets the model recover subject-specific structure from A while using S to modulate site contrast during denoising. We train the network with the noise-prediction objective $\mathcal{L}_{\text{MSE}} = \mathbb{E}_{x_0,t,\epsilon}\big[\|\epsilon - \epsilon_\theta(x_t, t, c)\|_2^2\big]$, where $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon$, with x_0 being the brain MRI patch, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

Classifier-Free Guidance++. We adopt the manifold-constrained classifier-free guidance method (CFG++) [3] during inference. As in standard CFG [12], we compute unconditional and style-conditioned noise predictions with anatomical inputs,

$$\epsilon_{\text{uncond}} = \epsilon_\theta(x_t, t; A, A_{\text{coarse}}, S = \varnothing), \qquad \epsilon_{\text{cond}} = \epsilon_\theta(x_t, t; A, A_{\text{coarse}}, S = S_{\text{target}}).$$

CFG++ forms the guided update in the usual way, $\hat{\epsilon} = \epsilon_{\text{uncond}} + \lambda\big(\epsilon_{\text{cond}} - \epsilon_{\text{uncond}}\big)$, but crucially treats the unconditional prediction as a manifold anchor. The anchored estimate $\hat{\epsilon}$ is used only to obtain a denoised image estimate $\hat{x}_0$,

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\hat{\epsilon}}{\sqrt{\bar{\alpha}_t}},$$

while the preceding noisy state is reconstructed using the unconditional noise term to remain on the learned noisy-data manifold:

$$x_{t-1} = \sqrt{\alpha_{t-1}}\,\hat{x}_0 + \sqrt{1-\alpha_{t-1}}\,\epsilon_{\text{uncond},t}.$$

In contrast, classic CFG effectively substitutes $\hat{\epsilon}$ directly into the sampling step (i.e., uses the guided prediction in place of the unconditional noise when computing x_{t-1}), which can push samples off the model manifold, require large guidance scales, and cause instability or poor invertibility. By anchoring on ϵ_{uncond}

and applying the conditional component as a tangent perturbation, CFG++ preserves anatomical fidelity, improves stability at moderate guidance scales, reduces artefacts and mode collapse, and restores invertibility useful for editing and harmonization [3]. During training, the style conditioning S_{target} is dropped with probability 0.15 to enable unconditional prediction.

Inference and reconstruction. Sampling uses a deterministic DDIM scheduler (50 steps) [21]. Patches are generated on a dense grid with $> 50\%$ overlap; each patch $\hat{x}_i$ is multiplied by a separable 3D Hann window w_i and fused by weighted averaging:

$$\hat{X}(\mathbf{v}) = \frac{\sum_i w_i(\mathbf{v})\, \hat{x}_i(\mathbf{v})}{\sum_i w_i(\mathbf{v})}.$$

This reduces edge artefacts and enforces spatial consistency. Final volumes are z-score normalized.

Implementation details. We implement a 3D conditional U-Net with residual blocks (GroupNorm, SiLU), sinusoidal timestep embeddings processed through a 2-layer MLP (dim 512), and both self- and cross-attention on flattened 3D tokens. The network has four encoder–decoder levels, attention at the two middle resolutions, and skip connections. Cross-attention uses 512-dim conditioning vectors with 64-dim heads. Training used Adam (learning rate 10^{-4}), batch size 5, for 2058 epochs on a single 48 GB GPU. On-the-fly data augmentation (random affine, elastic, and left–right flips) was applied via TorchIO. Inputs are concatenated as $[x_t, A, A_{\text{coarse}}]$ and the model outputs a single-channel noise prediction; the site embedding S is injected at multiple resolutions via cross-attention. Inference uses DDIM with 50 timesteps and the 'AIBL' site label (see training data.) as style embedding. Measured timings on our setup is 9 minutes 54 seconds per volume (preprocessing and harmonization, averaged over 5 volumes).

3 Experiments and Results

Datasets.

Training data. We assembled a training set from seven public collections of healthy adults: SALD [24], IXI, OASIS-3 [14], NMorphCH, AIBL [7], HCP Young Adult, and ICBM. From each acquisition site we randomly sampled 100 T1w volumes, yielding 1,000 images across 10 sites.

Traveling-subject data. Evaluation includes the SRPBS traveling-subject cohort [22]: 9 healthy males (age 24–32) scanned at 11 sites, comprising 97 T1w scans acquired on 10 different scanners.

Clinical data and derived cohorts. We curated a clinical database of all patients referred to the Memory Center of Lille (France) between 2011 and 2023. For each MRI–report pairs separated by ≤ 365 days, we collected age, sex, Mini-Mental State Examination score (MMSE) and the clinical diagnosis. Diagnostic labels were manually curated and binarized into Alzheimer's disease

(AD, NINCDS/ADRDA criteria [17]) and Alzheimer's disease–related dementias. From this pool we derived site-balanced cohort constructed by selecting five AD patients per site matched to the cohort mean age (16 sites, 80 volumes).

Preprocessing. All T1-w images were processed to reduce trivial acquisition variability: skull-stripping (HD-BET), bias-field correction (ANTs), affine registration to 1mm MNI space (FSL), cropping to $160 \times 192 \times 160$ voxels, and median-based intensity normalization. The observed native resolution ranges motivated resampling to a common 1mm isotropic grid.

Competing methods. We compare our method to state-of-the-art harmonization models with public code and pretrained weights (STGAN, HACA3, IGUANe) and a non-harmonized baseline (T1w-preprocessed images). Post harmonization skull stripping is applied where needed to ensure brain evaluation.

Ablation Study 1. CFG and CFG++ Across Guidance Strengths. During our diffusion model inference, we evaluate classical CFG (guidance strength ω) and manifold-constrained CFG++ (λ) against competing methods on SRPBS volumes. Structural Similarity Index (SSIM) is computed pairwise per subject using the 1^{st}–99^{th} percentile intensity range; Wasserstein Distance (WD) is computed between site-wise mean histograms (256 bins); Pearson Correlation Coefficient (PCC) is computed voxel-wise on flattened volumes. The results show (Table. 1) that both diffusion models, CFG and CFG++, achieve strong harmonization across SRPBS sites, with CFG++ consistently outperforming classical CFG. In particular, CFG++ with $\lambda = 0.6$ achieves a low WD while maintaining high SSIM and PCC, demonstrating accurate structural preservation, robust intensity alignment, and an improved overall balance between intensity matching and structural fidelity compared to baseline methods. Accordingly, this setting is used in all subsequent experiments, highlighting the effectiveness of manifold-constrained guidance for multi-site harmonization. Sample visualization in Fig. 2 (a) shows that Diffusion harmonized MRI is closer to the target.

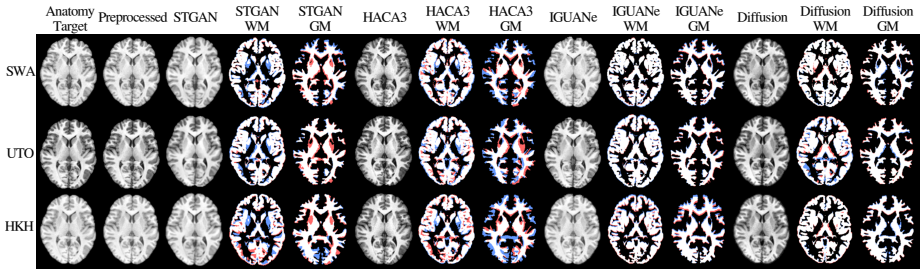

Fig. 2. (a) Central axial slices of the harmonized images. (b) SPM tissues segmentation. White and gray matter maps were compared voxel-wise with the Anatomy Target segmentation: white indicates agreement, blue under-segmentation, and red over-segmentation.

Table 1. Multi-site harmonization performance on the SRPBS dataset. The table reports voxel-level metrics for diffusion-based models (CFG, CFG++) across guidance strengths, and baseline harmonization methods. For each metric, the best method is in **bold** while the second best is in ***bold italic***

Methods		SSIM ↑	WD ↓	PCC ↑
CFG	$\omega = 0.5$	0.868 ± 0.060	0.017 ± 0.008	0.976 ± 0.014
	$\omega = 1$	0.867 ± 0.060	0.016 ± 0.008	0.976 ± 0.014
	$\omega = 2$	0.867 ± 0.060	0.018 ± 0.009	0.976 ± 0.013
CFG++	$\lambda = 0.4$	0.868 ± 0.059	0.012 ± 0.006	0.977 ± 0.014
	$\lambda = 0.6$	***0.869±0.059***	***0.010±0.006***	***0.978±0.014***
	$\lambda = 0.8$	0.866 ± 0.060	0.012 ± 0.007	0.976 ± 0.015
Baselines	Preprocessed	0.847 ± 0.094	0.036 ± 0.023	0.970 ± 0.032
	STGAN	$\mathbf{0.876 \pm 0.042}$	0.020 ± 0.011	$\mathbf{0.980 \pm 0.008}$
	HACA3	0.848 ± 0.071	$\mathbf{0.007 \pm 0.003}$	0.969 ± 0.020
	IGUANE	0.847 ± 0.093	0.033 ± 0.018	0.971 ± 0.032

Ablation Study 2. Site-invariance of the anatomical conditioning map. We assessed whether the proposed anatomical map removes site-specific information while preserving anatomy using a quantitative ablation on our training volumes. For each SRPBS site, we computed normalized intensity histograms of all scans (100 bins, 0.5–99.5% percentile clipping), both for raw images and anatomical maps. Inter-site discrepancies were measured using the squared Maximum Mean Discrepancy (MMD) with an RBF kernel ($K(x, y) = \exp(-\gamma\|x - y\|^2)$), computed for every pair of acquisition sites rather than matched subjects. Thus, each value corresponds to a site-pair comparison, and the "per-site" summary is obtained by averaging MMD values over all site pairs involving that site.

Pairwise MMD values were compared using a paired two-sided t-test between raw and anatomical-map representations on identical site pairs, showing a significant reduction for anatomical maps ($t = 4.32$, $p = 8.68 \times 10^{-5}$). Averaging MMD per site and performing the same paired test confirmed the effect at the site level ($t = 4.56$, $p = 0.0014$), indicating that anatomical maps reduce site variability consistently rather than being driven by a few site pairs.

Task 1. Linear mixed-effects modeling on brain tissue. Gray matter (GM) and white matter (WM) probability maps were obtained using SPM25 [23] on SRPBS subjects, and tissue ratios were computed as the tissue volume divided by total intracranial volume (i.e., GM + WM + CSF). For each harmonization method, we fitted a linear mixed-effects model with *site* as a fixed effect and *subject* as a random intercept:

$$\texttt{tissue\_ratio} \sim \texttt{site} + (1 \mid \texttt{subject}).$$

Marginal R^2 quantifies variance explained by site alone, whereas conditional R^2 reflects variance explained by both site and subject, capturing preserved inter-subject (biological) variability. Effective harmonization is characterized by low marginal R^2 and high conditional R^2.

Across both GM and WM (Table 2), the diffusion model is the most consistent method, jointly reducing marginal R^2 while increasing conditional R^2 relative

to preprocessed data. Figure 2(b) further shows that diffusion yields fewer tissue segmentation errors. STGAN and HACA3 achieve strong performance in one tissue but not consistently across both, indicating less stable behavior across tissue types.

Table 2. Marginal and conditional R^2 values for gray and white matter across harmonization methods. For both R^2, the best method is in **bold** while the second best is in ***bold italic***

	Preprocessed	STGAN	HACA3	IGUANe	Diffusion
			Gray Matter		
R^2 marginal ↓	0.729	**0.359**	0.687	0.650	***0.507***
R^2 conditional ↑	***0.885***	0.771	0.797	0.853	**0.909**
			White Matter		
R^2 marginal ↓	0.599	0.360	**0.127**	0.471	***0.214***
R^2 conditional ↑	0.909	0.734	**0.965**	0.916	***0.942***

Task 2. Preservation of brain volume and cognition associations. We evaluated whether harmonization preserves clinically meaningful structure–cognition links by computing partial correlations (Pearson r, controlling for age and sex) between FreeSurfer regional volumes [9] and MMSE scores. Analysis focused on predefined memory-related ROIs (hippocampus, entorhinal and parahippocampal cortices, temporal and medial regions) [20,8]. Method-level performance is summarized as the mean absolute partial r across ROIs.

As shown in Fig. 3, the diffusion method achieves the highest global mean absolute r (0.135), with improved correlations relative to preprocessed data in 6 of 8 ROIs, compared to 3/8 for STGAN, 5/8 for HACA3, and 3/8 for IGUANe. This indicates that diffusion harmonization not only preserves structure–MMSE associations but does so consistently across memory-related regions.

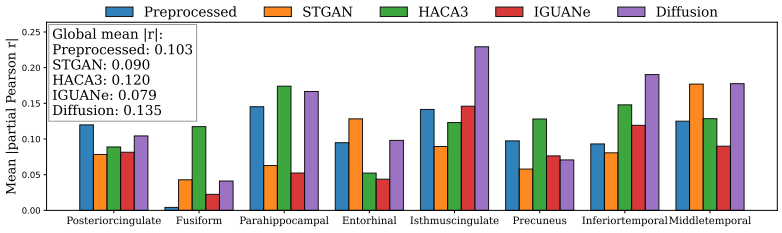

Fig. 3. Mean absolute partial correlation |r| between normalized regional brain volumes and MMSE across harmonization methods. Bars show regional |r| for each ROI; global mean |r| per method is indicated in the right-top corner.

4 Conclusion

We presented a 3D patch-based conditional diffusion framework for multi-site T1-weighted MRI harmonization, leveraging site-invariant anatomical maps to preserve fine structural details while aligning scanner-specific contrast. The patch-wise design reduces memory requirements and avoids decoder-induced blurring, while anatomical conditioning ensures faithful preservation of subject-specific anatomy. Extensive experiments on traveling-subject and clinical cohorts demonstrate that our method provides robust harmonization, effectively reducing site-related variability without compromising biologically relevant signals, highlighting its practical utility for clinical datasets.

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

References

1. Abbasi, S., Lan, H., Choupan, J., Sheikh-Bahaei, N., Pandey, G., Varghese, B.: Deep learning for the harmonization of structural mri scans: a survey. BioMedical Engineering OnLine **23**(1) (Aug 2024). https://doi.org/10.1186/s12938-024-01280-6
2. Cackowski, S., Barbier, E.L., Dojat, M., Christen, T.: Imunity: A generalizable vae-gan solution for multicenter mr image harmonization. Medical Image Analysis **88**, 102799 (Aug 2023). https://doi.org/10.1016/j.media.2023.102799
3. Chung, H., Kim, J., Park, G.Y., Nam, H., Ye, J.C.: Cfg++: Manifold-constrained classifier free guidance for diffusion models (2024). https://doi.org/10.48550/ARXIV.2406.08070
4. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation p. 529–536 (2018). https://doi.org/10.1007/978-3-030-00928-1_60
5. Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., Sotirchos, E.S., Saidha, S., Oh, J., Pham, D.L., Calabresi, P.A., van Zijl, P.C., Prince, J.L.: Deepharmony: A deep learning approach to contrast harmonization across scanner changes. Magnetic Resonance Imaging **64**, 160–170 (Dec 2019). https://doi.org/10.1016/j.mri.2019.05.041
6. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis (2021). https://doi.org/10.48550/ARXIV.2105.05233
7. Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., Ames, D.: The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. International Psychogeriatrics **21**(4), 672–687 (Aug 2009). https://doi.org/10.1017/s1041610209009405
8. Ferrarini, L., Palm, W.M., Olofsen, H., van der Landen, R., Jan Blauw, G., Westendorp, R.G., Bollen, E.L., Middelkoop, H.A., Reiber, J.H., van Buchem, M.A., Admiraal-Behloul, F.: Mmse scores correlate with local ventricular enlargement in the spectrum from cognitively normal to alzheimer disease. NeuroImage **39**(4), 1832–1838 (Feb 2008). https://doi.org/10.1016/j.neuroimage.2007.11.003

9. Fischl, B.: Freesurfer. NeuroImage **62**(2), 774–781 (Aug 2012). https://doi.org/10.1016/j.neuroimage.2012.01.021
10. Fortin, J.P., Parker, D., Tunç, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T.: Harmonization of multi-site diffusion tensor imaging data. NeuroImage **161**, 149–170 (Nov 2017). https://doi.org/10.1016/j.neuroimage.2017.08.047
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020). https://doi.org/10.48550/ARXIV.2006.11239
12. Ho, J., Salimans, T.: Classifier-free diffusion guidance (2022). https://doi.org/10.48550/ARXIV.2207.12598
13. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics **8**(1), 118–127 (Apr 2006). https://doi.org/10.1093/biostatistics/kxj037
14. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., Raichle, M.E., Cruchaga, C., Marcus, D.: Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease (Dec 2019). https://doi.org/10.1101/2019.12.13.19014902
15. Lan, H., Varghese, B.A., Sheikh-Bahaei, N., Sepehrband, F., Toga, A.W., Choupan, J.: Diffusion based multi-domain neuroimaging harmonization method with preservation of anatomical details. NeuroImage **316**, 121297 (Aug 2025). https://doi.org/10.1016/j.neuroimage.2025.121297
16. Liu, M., Zhu, A.H., Maiti, P., Thomopoulos, S.I., Gadewar, S., Chai, Y., Kim, H., Jahanshad, N.: Style transfer generative adversarial networks to harmonize multisite <scp>mri</scp> to a single reference image to avoid overcorrection. Human Brain Mapping **44**(14), 4875–4892 (Jul 2023). https://doi.org/10.1002/hbm.26422
17. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. Neurology **34**(7), 939–939 (Jul 1984). https://doi.org/10.1212/wnl.34.7.939
18. Pinaya, W.H.L., Tudosiu, P.D., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain Imaging Generation with Latent Diffusion Models, p. 117–126. Springer Nature Switzerland (2022). https://doi.org/10.1007/978-3-031-18576-2_12
19. Roca, V., Kuchcinski, G., Pruvo, J.P., Manouvriez, D., Lopes, R.: Iguane: A 3d generalizable cyclegan for multicenter harmonization of brain mr images. Medical Image Analysis **99**, 103388 (Jan 2025). https://doi.org/10.1016/j.media.2024.103388
20. Sawada, Y., Satoh, T., Saba, H., Kato, Y., Kuwada, T., Shima, S., Murakami, K., Sasaki, M., Abe, Y., Harano, K.: Assessment of localized brain regions correlated with mmse using vbm analysis of structural mri in a japanese sample. NeuroImage: Reports **5**(2), 100264 (Jun 2025). https://doi.org/10.1016/j.ynirp.2025.100264
21. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models (2020). https://doi.org/10.48550/ARXIV.2010.02502
22. Tanaka, S.C., Yamashita, A., Yahata, e.a.: A multi-site, multi-disorder resting-state magnetic resonance image database. Scientific Data **8**(1) (Aug 2021). https://doi.org/10.1038/s41597-021-01004-8
23. Tierney, T.M., Alexander, N.A., Ashburner, e.a.: Spm 25: open source neuroimaging analysis software. Journal of Open Source Software **10**(110), 8103 (Jun 2025). https://doi.org/10.21105/joss.08103

24. Wei, D., Zhuang, K., Ai, L., Chen, Q., Yang, W., Liu, W., Wang, K., Sun, J., Qiu, J.: Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset. Scientific Data **5**(1) (Jul 2018). https://doi.org/10.1038/sdata.2018.134
25. Wu, M., Yu, M., Jing, S., Yap, P.T., Zhang, Z., Liu, M.: Unpaired volumetric harmonization of brain mri with conditional latent diffusion. Medical Image Analysis **107**, 103849 (Jan 2026). https://doi.org/10.1016/j.media.2025.103849
26. Zuo, L., Liu, Y., Xue, Y., Dewey, B.E., Remedios, S.W., Hays, S.P., Bilgel, M., Mowry, E.M., Newsome, S.D., Calabresi, P.A., Resnick, S.M., Prince, J.L., Carass, A.: Haca3: A unified approach for multi-site mr image harmonization. Computerized Medical Imaging and Graphics **109**, 102285 (Oct 2023). https://doi.org/10.1016/j.compmedimag.2023.102285