


Patch-level Anatomy- and Style-Guided 3D Diffusion for Multi-site T1 MRI Harmonization

Barnabé Hache¹ 

BARNABE.HACHE@INSERM.FR

Vincent Roca²

VINCENT.ROCA@CHU-LILLE.FR

Grégory Kuchcinski^{1,2,3}

GREGORY.KUCHCINSKI@CHU-LILLE.FR

Dorian Manouvriez²

DORIAN2.MANOUVRIEZ@CHU-LILLE.FR

Renaud Lopes⁴

RENAUD.LOPES@CHU-LILLE.FR

¹ Univ. Lille, Inserm, CHU Lille, U1172 - LilNCog - Lille Neuroscience and Cognition, F- 59000 Lille, France

² Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41 - UAR 2014 - PLBS, F-59000 Lille, France

³ CHU Lille, Département de Neuroradiologie, F-59000 Lille, France

⁴ CHU Lille, Département de Médecine Nucléaire, F-59000 Lille, France

Editors: Under Review for MIDL 2026

Abstract

Pooling brain T1-weighted MRI from multiple sites increases statistical power but introduces non-biological scanner effects that bias analyses. Generative diffusion models generalize well and train stably, making them attractive for MRI harmonization. However, most prior 3D works use latent-space diffusion for efficiency, and the compression–decoding step can introduce blurring that undermines preservation of fine anatomical detail. We present a patch-level 3D diffusion framework that harmonizes volumes to a chosen target site while preserving individual anatomy. The model is jointly conditioned on an explicit anatomical prior derived from each volume and on a learnable site embedding injected via cross-attention; inference uses manifold-constrained classifier-free guidance to steer site style without displacing anatomy. Working on overlapping high-resolution patches avoids decoder-related blurring and preserves fine structural detail. We train on 1,000 scans pooled from seven public cohorts and evaluate on the SRPBS traveling-subject dataset. Compared to preprocessed images and three published baselines, our method yields modest improvements in image similarity (SSIM 0.864 \rightarrow 0.874) and substantially reduces variance attributable to site (gray matter: 72.9% \rightarrow 50.7%; white matter: 59.9% \rightarrow 21.4%), while subject-level variability remains high, indicating preservation of biological differences. The code and model weights will be made available.

Keywords: MRI harmonization, diffusion models, patch-based synthesis, classifier-free guidance, anatomical representation

1. Introduction

Combining structural magnetic resonance imaging (MRI) from multiple sites increases cohort size and statistical power but introduces non-biological scanner/site effects that bias downstream analyses and reduce reproducibility (Johnson et al., 2007; Abbasi et al., 2024). Classical statistical harmonization methods such as ComBat (Fortin et al., 2017, 2018) and its neuroimaging adaptations effectively remove many site-dependent biases for summary

features (e.g., cortical thickness) but operate on extracted measurements, limiting their use when image-level consistency is needed.

Image-level harmonization methods have been proposed to mitigate scanner-related variability (Shah et al., 2011). Beyond classical statistical approaches, most learning-based solutions rely on U-Net architectures or GAN-style domain translation as DeepHarmony (Dewey et al., 2019), HACA3 (Zuo et al., 2023), STGAN (Liu et al., 2023) or IGUANE (Roca et al., 2025), but these models may generalize poorly or alter anatomy when constraints are insufficient (Cohen et al., 2018).

Recent diffusion-based harmonization methods (Wu et al., 2026; Lan et al., 2025; Pinaya et al., 2022) leverage the strong generative fidelity of diffusion models (Dhariwal and Nichol, 2021), usually in latent space for efficiency. They condition denoising on scanner information to align style while preserving anatomy, but latent diffusion can blur fine structural details: an aspect we specifically address in the present work.

Two practical challenges limit direct application of diffusion models to full-volume MRI harmonization. First, volumetric diffusion models are computationally intensive: 3D architectures and high-resolution volumes demand large memory and compute budgets, complicating training and inference at clinical resolutions (Wu et al., 2026). Second, latent diffusion approaches that compress volumes to a lower-dimensional latent space reduce computational cost but rely on a decoder that may not faithfully reconstruct fine anatomical details critical in neuroimaging applications (Wu et al., 2026; Pinaya et al., 2022). These limitations motivate hybrid strategies that preserve anatomical fidelity while maintaining computational feasibility.

In this work we therefore design a patch-based, anatomy-guided 3D diffusion framework for multi-site T1 brain MRI harmonization. Working on overlapping high-resolution 3D patches enables full-resolution synthesis with manageable GPU memory, while an explicit site-style embedding and a site-invariant anatomical conditioning signal disentangle contrast from structure during denoising. Compared to latent diffusion, our patch-wise approach preserves fine-grained anatomy; compared to image-to-image GANs, diffusion sampling yields more stable, higher-fidelity harmonized outputs (Dhariwal and Nichol, 2021; Wu et al., 2026).

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-extraction and conditioning strategy (including randomized gamma perturbation during training) that prevents leakage of site-specific intensity cues into the structural prior; and (iii) an extensive evaluation on a multi-site traveling-subject dataset (SRPBS) comparing our method to leading baselines (STGAN, HACA3, IGUANE, and non-harmonized preprocessing) using distributional, segmentation-based and mixed-effects metrics.

2. Methodology

2.1. Overview

We aim to harmonize T1-weighted brain MRI volumes from arbitrary source sites into the contrast distribution of a chosen target site while preserving subject-specific anatomy. Our framework operates at the 3D patch level and separates anatomical conditioning from style (site) conditioning. The main stages are: (i) computation of a site-invariant anatomical

representation, (ii) paired 3D patch extraction (brain MRI patch, anatomical patch, coarse anatomical context), (iii) conditional diffusion-based patch generation guided by anatomy and a learnable site embedding, and (iv) patch-wise inference with overlap and Hann-weighted aggregation to reconstruct a full harmonized volume (see Fig. 1).

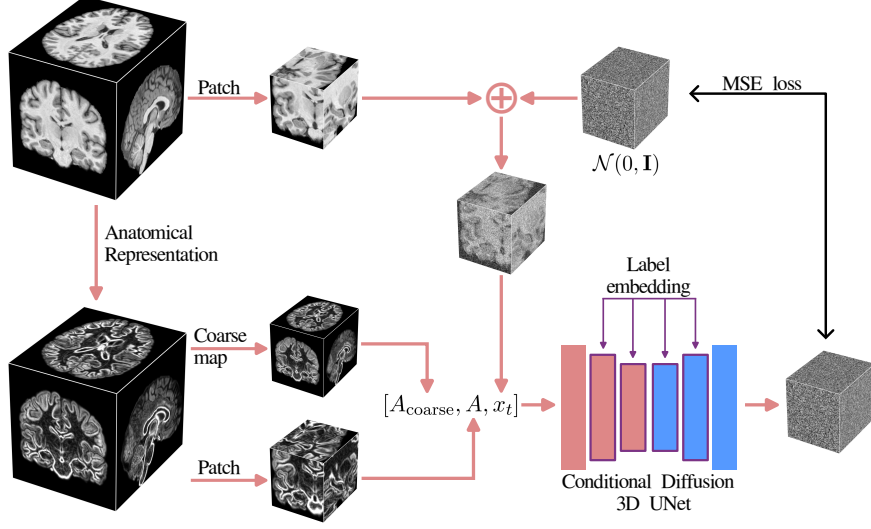


Figure 1: *Overview of the training pipeline.* Anatomical representations are computed from the input volume, and paired patches ($80 \times 96 \times 80$) with coarse anatomical context are extracted for diffusion-based synthesis. Site labels are embedded and injected into the model through cross-attention layers.

2.2. Anatomical Representation

We construct a site-invariant anatomical map by suppressing contrast cues. Before computing anatomy, we apply a random global gamma perturbation inspired by [Cackowski et al. \(2021\)](#) to the full volume to avoid leaking site information into this prior:

$$I^\gamma = \text{GammaTransform}(I; \gamma), \quad \gamma \sim \mathcal{U}(\gamma_{\min}, \gamma_{\max}),$$

with I the whole brain MRI image, $(\gamma_{\min}, \gamma_{\max}) = (-0.5, 0.5)$ during training and $\gamma = 0$ at inference. From I^γ we extract gradient magnitude and high-frequency components, which are combined to capture the global structural layout of the brain while preserving fine-grained anatomical details:

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

where $\mathcal{G}_{\sigma_{\text{hf}}}(\cdot)$ denotes 3D Gaussian smoothing with a kernel size of $(2r+1)^3$ ($r = \max\{1, \lceil 3\sigma_{\text{hf}} \rceil\}$). We combine these terms into a single anatomical map

$$A = \alpha G + (1 - \alpha)H,$$

where α is a fixed hyperparameter balancing global structure and fine details. The anatomical map A is quantile-normalized to $[-1, 1]$, and a coarse version A_{coarse} is produced by trilinear downsampling to provide global context, similarly to the coordinate encoder of Bieder et al. (2023).

2.3. Patch Extraction

We extract patches of size $(80 \times 96 \times 80)$, i.e. 50% of each dimension (1/8 of the volume). During training, patches are randomly sampled inside a brain bounding box; the corresponding A and A_{coarse} patches are extracted at the same coordinates. During inference, patches are placed on a dense grid with $\geq 50\%$ overlap.

2.4. Diffusion Model

Formulation. We adopt the denoising diffusion probabilistic model framework of Ho et al. (2020) 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:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbf{I}),$$

with closed-form marginal

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

We parameterize the denoiser to predict the noise conditioned on anatomy and style:

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

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

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{x_0, t, \epsilon} [\|\epsilon - \epsilon_\theta(x_t, t, c)\|_2^2],$$

where $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$, with x_0 being the brain MRI patch, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

Model input The denoiser is a 3D U-Net with multiscale residual blocks and cross-attention. Inputs are concatenated as $[x_t, A, A_{\text{coarse}}]$. A sinusoidal embedding encodes t . The site-style embedding S is injected to the model at multiple resolutions via the cross-attention layer (see 3.2). The output is a single-channel noise prediction.

2.5. Anatomical and Style Conditioning

Anatomy (A, A_{coarse}). The anatomical signals are always provided during training and inference, both via input concatenation and internal cross-attention, which ensures structural fidelity across local and contextual scales.

Style embedding (S). Each site has a learnable vector S_k . During training, the true site embedding is used with probability 0.85, while in the remaining 0.15 it is replaced by a null embedding $S = \emptyset$, effectively removing style-conditioning. The embedder is optimized jointly to the diffusion model. At inference, harmonization is produced by replacing S with the target embedding S_{target} .

2.6. Classifier-Free Guidance++

We use a manifold-constrained classifier-free guidance update (CFG++) (Chung et al., 2024), adapted to patch-based medical diffusion. Both branches use identical anatomical inputs (A, A_{coarse}), only style changes:

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

CFG++ treats the unconditional prediction as a manifold anchor and the conditional one as a tangent direction restricted to the style subspace. Sampling uses:

$$\hat{\epsilon} = \epsilon_{\text{uncond}} + \lambda(\epsilon_{\text{cond}} - \epsilon_{\text{uncond}}),$$

where anatomical consistency is guaranteed because anatomy does not differ between branches. Compared to classical classifier-free guidance (CFG) (Ho and Salimans, 2022), CFG++ constrains the update to the style-controlled manifold rather than applying a free global interpolation.

2.7. Inference and reconstruction

During inference, we sample patches using the DDIM scheduler (Song et al., 2020), which provides a deterministic approximation of the stochastic diffusion process. For each patch location i , the model produces \hat{x}_i conditioned on $(A_i, A_{\text{coarse}, i}, S_{\text{target}})$. A separable 3D Hann window w_i is applied before fusion, and the final volume is reconstructed via 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 approach removes edge artifacts and ensures spatially consistent harmonized volumes.

3. Experiments

3.1. Datasets

Training dataset: Our training set was constructed from seven publicly available datasets of healthy adults: SALD (Wei et al., 2018), IXI¹, OASIS-3 (LaMontagne et al., 2019), NIMH², AIBL (Ellis et al., 2009), HCP Young Adult³, and ICBM⁴. For each acquisition site, we randomly selected 100 T1-weighted scans, yielding a total of 1,000 volumes (as

1. <https://brain-development.org/ixi-dataset/>
 2. <http://otto.fsm.northwestern.edu/>
 3. <https://www.humanconnectome.org/study/hcp-young-adult>
 4. <https://ida.loni.usc.edu/>

some datasets presents several sites). All images were preprocessed to reduce trivial acquisition variability using: (i) skull stripping, (ii) bias-field correction, (iii) affine registration to 1 mm MNI space, (iv) cropping to $160 \times 192 \times 160$ voxels, and (v) median-based intensity normalization.

To improve generalization across anatomical and scanner variability, we applied on-the-fly data augmentation during training. Each volume underwent one or two random spatial transformations (affine perturbation, elastic deformation, or left-right flip). We further modulated image contrast by sampling a log-gamma intensity factor from a discrete set, producing augmented volumes with diverse global contrast profiles and the corresponding site label was updated accordingly.

Test dataset: For evaluation, we used the SRPBS (Tanaka et al., 2021) traveling-subject dataset, consisting of 10 healthy individuals scanned across 10 MRI sites, for a total of 97 unique T1-weighted scans after removing duplicates. The same preprocessing steps (i)–(v) were applied, but no augmentation was used.

3.2. 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 with 3D convolutions for down-/upsampling, attention in the two middle resolutions, and skip connections. Input patches are $80 \times 96 \times 80$ with three channels (noisy patch, anatomical patch, coarse context), and the network predicts a single denoised output. Cross-attention injects the site label via 512-dim conditioning vectors with 64-dimensional heads. Training uses Adam (learning rate 10^{-4}), batch size 5, for 2058 epochs on a single 48 GB GPU. During inference, we generate harmonized patches using a deterministic DDIM scheduler with 50 timesteps, which provides a fast approximation of the stochastic diffusion process.

3.3. Comparison of CFG and CFG++ Across Guidance Strengths

We assess guidance strength for classical CFG (ω) and manifold-constrained CFG++ variant (λ) using SSIM (Wang et al., 2004) which measures structural similarity between images, and PSNR, which reflects voxel-wise reconstruction fidelity. Metrics are computed between harmonized volumes of the same subject across sites and averaged over all subject-site pairs.

As shown in Table 1, CFG++ achieves higher and more stable SSIM than CFG across all guidance values, avoiding the degradation observed for CFG at larger ω . This indicates that restricting guidance to the style subspace improves robustness and preserves anatomy.

Based on these results, we adopt **CFG++ with $\lambda = 0.6$** for all subsequent experiments, ensuring reliable contrast harmonization without compromising structural fidelity.

4. Metrics and Evaluation of Harmonized Brain MRI Volumes

Baselines. We compare our method against state-of-the-art harmonization models with publicly available code and pretrained weights, namely STGAN, HACA3, and IGUANE, none of which were trained on SRPBS. We additionally include a *non-harmonized* baseline

Table 1: Comparison of CFG and CFG++ across guidance strengths (respectively ω and λ). Values report mean \pm standard deviation of SSIM and PSNR computed over all subject–site pairs in SRPBS. Improvements should be interpreted relative to the preprocessed baseline (SSIM 0.864 ± 0.088 , PSNR 26.78 ± 2.78).

		$\omega = 0.5$	$\omega = 1$	$\omega = 2$	$\omega = 4$	$\omega = 7$
CFG	SSIM	0.871 ± 0.059	0.870 ± 0.060	0.870 ± 0.060	0.865 ± 0.060	0.856 ± 0.061
	PSNR	23.50 ± 2.227	23.51 ± 2.195	23.40 ± 2.213	23.20 ± 2.192	22.80 ± 2.126
		$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1$
CFG++	SSIM	0.874 ± 0.057	0.874 ± 0.058	0.874 ± 0.058	0.871 ± 0.059	0.870 ± 0.059
	PSNR	23.18 ± 2.880	23.22 ± 2.997	23.25 ± 2.949	23.10 ± 3.122	23.04 ± 3.089

consisting of standard preprocessed images. For methods operating on full-head MRI, skull stripping is reapplied to ensure brain-only evaluation.

Tissue volume analysis. Gray matter (GM) and white matter (WM) probability maps are extracted using SPM (Tierney et al., 2025), from which we compute tissue ratios (e.g., tissue/intracranialvolume). To assess residual site effects after harmonization, we select a reference site (KPM from SRPBS) and perform paired t -tests between KPM and each other site. p -values are corrected using the Benjamini–Hochberg FDR procedure. Non-significant differences ($p > 0.05$) indicate effective reduction of scanner-induced bias.

Linear mixed-effects modeling. Using the SPM-derived tissue ratios, we fit a separate linear mixed model for each harmonization method with *site* as a fixed effect and *subject* as a random effect. We report marginal R^2 (variance explained by site only) and conditional R^2 (variance explained by site and subject). Effective harmonization is reflected by low marginal R^2 , indicating minimal remaining site influence, together with higher conditional R^2 dominated by subject-level variability (i.e., preserved biological differences).

5. Results

5.1. Visual assessment of harmonization

The effects of each harmonization method are illustrated in Fig. 2. Visually, the harmonized images exhibit varying degrees of contrast enhancement and suppression depending on the specific deep learning approach employed, with some methods producing subtly refined gray–white matter delineation while others introduce more pronounced intensity shifts. We highlighted inconsistencies such as artifacts or a clear lack of detail in red on the figure.

5.2. Tissue Volume Analysis

We evaluate residual site effects after harmonization using SPM-derived tissue ratios, comparing gray matter and white matter across sites relative to the reference site KPM (SRPBS). Results are summarized in Fig. 3.

Gray matter. Our diffusion-based model slightly reduces the number of sites exhibiting significant differences in GM ratio compared to the preprocessed baseline. In contrast,

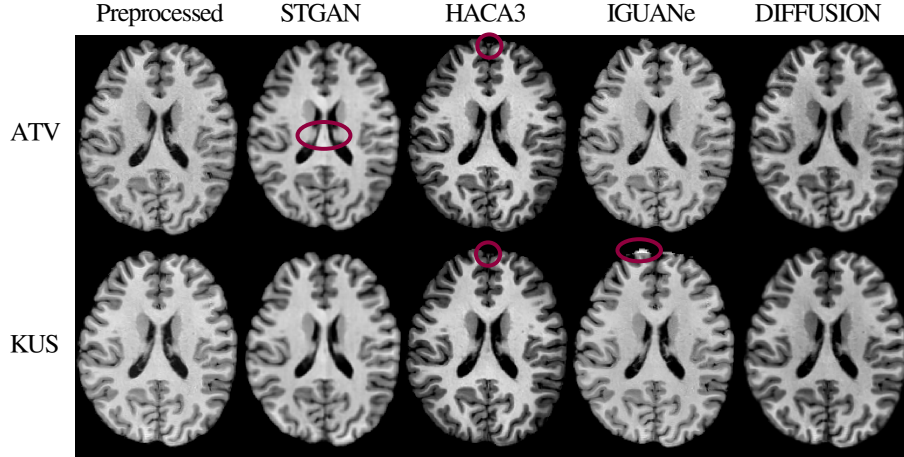


Figure 2: *Visualization of harmonized MR images for each method.* Central axial slice of one SRPBS subject (sub-1) for two different sites (ATV and KUS) are displayed. Inconsistencies (missing details or artifacts) are circled in red.

IGUANE and HACA3 introduce additional significant differences relative to the reference site. Only STGAN reduces the number of significantly different sites to one.

White matter. For WM ratios, our model effectively eliminates all significant site differences, similarly to HACA3. STGAN introduces a new significant difference at site KUT, whereas IGUANE removes one previously significant difference compared to the baseline.

5.3. Linear Mixed-Effects Analysis

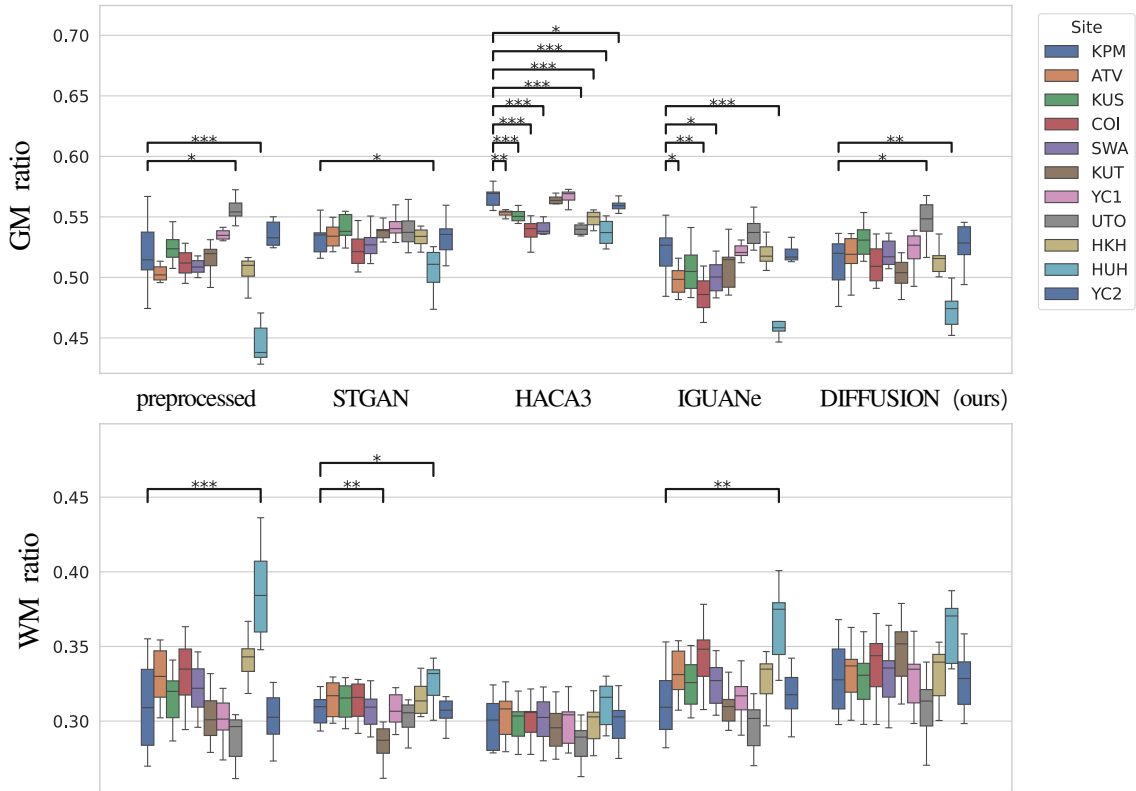
We further assess residual site effects using linear mixed-effects models on SPM-derived tissue ratios, with *site* as a fixed effect and *subject* as a random effect. Table 2 reports marginal R^2 (variance explained by site) and conditional R^2 (variance explained by site and subject) for gray matter and white matter.

Gray matter. For GM, our diffusion-based model reduces marginal R^2 relative to the preprocessed baseline (0.507 vs. 0.729), indicating decreased site influence, while maintaining high conditional R^2 (0.909), reflecting preserved subject-specific variability. STGAN achieves the lowest marginal R^2 (0.359) but at the cost of reduced conditional R^2 (0.771). HACA3 and IGUANE show intermediate reductions in marginal R^2 .

White matter. For WM, the diffusion model also reduces marginal R^2 (0.214) compared to preprocessed images (0.599), effectively diminishing site effects while keeping conditional R^2 high (0.942). HACA3 achieves the lowest marginal R^2 (0.127) and exhibits a very high conditional R^2 (0.965), whereas STGAN and IGUANE show moderate reductions in site variance.

Table 2: Marginal and conditional R^2 values for gray and white matter across harmonization methods.

	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


 Figure 3: *GM and WM ratios for each site and methods.* The asterisks indicate significant t-test comparing the tissue ratio difference between sites among methods (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

6. Discussion and Conclusions

We presented a patch-based, anatomy-guided 3D diffusion framework for multi-site T1 MRI harmonization and evaluated it against several state-of-the-art baselines (STGAN, HACA3, IGUANE) and preprocessed images. Across our experiments the diffusion model consistently reduced site-related variance while preserving subject-specific variability. In the linear mixed-effects analysis (Table 2) the diffusion method yields markedly lower marginal R^2 than the preprocessed baseline for both GM and WM, together with high conditional R^2 , indicating that remaining variability is dominated by subjects rather than scanners.

Our method is not uniformly superior on every metric or tissue. For example, HACA3 attains the lowest marginal R^2 on WM in Table 2, showing that some alternatives can better suppress certain site effects on specific tissues. Conversely, IGUANE and HACA3 sometimes introduce new significant differences in GM ratios (Fig. 3), while STGAN is able to reduce the number of significantly different sites to one for GM but may produce other trade-offs. Overall, the diffusion approach demonstrates a favorable balance: it is stable across guidance strengths (Sec. 2.6) and preserves anatomical fidelity, as evidenced by the high conditional R^2 and SSIM/PSNR trends.

A key factor in our method’s performance is the use of patch-wise processing. Operating on patches allows fine details without artifacts to be preserved more accurately than if latent space compression had been used. This makes the patch-based strategy particularly well-suited for high-resolution MRI harmonization, where preserving subtle structural information is critical.

The method’s robustness stems from explicit anatomical conditioning (multi-scale gradient and high-frequency map), which preserves subject-specific structure, and from CFG++ guidance, which controls site-specific contrast without altering anatomy. Increasing the guidance strength improves inter-site similarity while maintaining structural fidelity.

Limitations remain. Despite randomized gamma perturbations during training, the anatomical representation can still carry residual site cues that the model may exploit to infer source-related information; this leakage may partially explain remaining differences and occasional over-corrections. Computational cost is another practical constraint: 3D patch-wise diffusion enables full-resolution synthesis but increases inference time compared to latent methods. Downstream clinical-task validation is also needed to confirm practical benefits.

In conclusion, our patch-based diffusion harmonizer provides a robust, anatomically faithful solution for multi-site MRI harmonization: it reduces scanner-driven variance, preserves biologically meaningful inter-subject differences, and maintains fine structural details, making it particularly suitable for high-resolution volumetric MRI applications.

References

- Soolmaz Abbasi, Haoyu Lan, Jeiran Choupan, Nasim Sheikh-Bahaei, Gaurav Pandey, and Bino Varghese. Deep learning for the harmonization of structural mri scans: a survey. *BioMedical Engineering Online*, 23:1–42, 12 2024. ISSN 1475925X. doi: 10.1186/S12938-024-01280-6.
- Florentin Bieder, Julia Wolleb, Alicia Durrer, Robin Sandkühler, and Philippe C. Cattin. Memory-efficient 3d denoising diffusion models for medical image processing. *Proceedings of Machine Learning Research*, 3 2023. doi: 10.48550/arXiv.2303.15288.
- Stenzel Cackowski, Emmanuel L. Barbier, Michel Dojat, and Thomas Christen. Imunity: a generalizable vae-gan solution for multicenter mr image harmonization. *Medical Image Analysis*, 88, 9 2021. ISSN 13618423. doi: 10.1016/j.media.2023.102799.
- Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. Cfg++: Manifold-constrained classifier free guidance for diffusion models, 2024.
- Joseph Paul Cohen, Margaux Luck, and Sina Honari. *Distribution Matching Losses Can Hallucinate Features in Medical Image Translation*, page 529–536. Springer International Publishing, 2018. ISBN 9783030009281. doi: 10.1007/978-3-030-00928-1_60.
- Blake E. Dewey, Can Zhao, Jacob C. Reinhold, Aaron Carass, Kathryn C. Fitzgerald, Elias S. Sotirchos, Shiv Saidha, Jiwon Oh, Dzung L. Pham, Peter A. Calabresi, Peter C.M. van Zijl, and Jerry L. Prince. Deepharmony: A deep learning approach to contrast harmonization across scanner changes. *Magnetic Resonance Imaging*, 64:160–170, 12 2019. ISSN 0730-725X. doi: 10.1016/J.MRI.2019.05.041.
- Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021.
- Kathryn A Ellis, Ashley I Bush, David Darby, Daniela De Fazio, Jonathan Foster, Peter Hudson, Nicola T. Lautenschlager, Nat Lenzo, Ralph N. Martins, Paul Maruff, Colin Masters, Andrew Milner, Kerry Pike, Christopher Rowe, Greg Savage, Cassandra Szoek, Kevin Taddei, Victor Villemagne, Michael Woodward, and David Ames. 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, August 2009. ISSN 1041-6102. doi: 10.1017/s1041610209009405.
- Jean-Philippe Fortin, Natalie Cullen, Y. Sheline, W.D. Taylor, I. Aselcioglu, P.A. Cook, P. Adams, C. Cooper, M. Fava, P.J. McGrath, et al. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 152:149–165, 2017. doi: 10.1016/j.neuroimage.2017.03.065.
- Jean-Philippe Fortin, D. Parker, B. Tunc, T. Wager, et al. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167:104–120, 2018. doi: 10.1016/j.neuroimage.2017.11.024.
- Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.

- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.
- W. Evan Johnson, Cheng Li, and Alexander Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007. doi: 10.1093/biostatistics/kxj037.
- Pamela J. LaMontagne, Tammie LS. Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. December 2019. doi: 10.1101/2019.12.13.19014902.
- Haoyu Lan, Bino A. Varghese, Nasim Sheikh-Bahaei, Farshid Sepehrband, Arthur W. Toga, and Jeiran Choupan. Diffusion based multi-domain neuroimaging harmonization method with preservation of anatomical details. *NeuroImage*, 316:121297, 8 2025. ISSN 1053-8119. doi: 10.1016/J.NEUROIMAGE.2025.121297.
- Mengting Liu, Alyssa H. Zhu, Piyush Maiti, Sophia I. Thomopoulos, Shruti Gadewar, Yaqiong Chai, Hosung Kim, and Neda Jahanshad. Style transfer generative adversarial networks to harmonize multisite mri to a single reference image to avoid overcorrection. *Human Brain Mapping*, 44(14):4875–4892, 2023. doi: 10.1002/hbm.26422.
- Walter H.L. Pinaya, Petru Daniel Tudosiu, Jessica Dafflon, Pedro F. Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Brain imaging generation with latent diffusion models. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13609 LNCS:117–126, 2022. ISSN 1611-3349. doi: 10.1007/978-3-031-18576-2_12.
- Vincent Roca, Grégory Kuchcinski, Jean Pierre Pruvo, Dorian Manouvriez, and Renaud Lopes. Iguane: A 3d generalizable cyclegan for multicenter harmonization of brain mr images. *Medical Image Analysis*, 99:103388, 1 2025. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2024.103388.
- Mohak Shah, Yiming Xiao, Nagesh Subbanna, Simon Francis, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Evaluating intensity normalization on mris of human brain with multiple sclerosis. *Medical Image Analysis*, 15:267–282, 4 2011. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2010.12.003.
- Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR 2021 - 9th International Conference on Learning Representations*, 10 2020.
- Saori C. Tanaka, Ayumu Yamashita, Noriaki Yahata, Takashi Itahashi, Giuseppe Lisi, Takashi Yamada, Naho Ichikawa, Masahiro Takamura, Yujiro Yoshihara, Akira Kunitatsu, Naohiro Okada, Ryuichiro Hashimoto, Go Okada, Yuki Sakai, Jun Morimoto, Jin Narumoto, Yasuhiro Shimada, Hiroaki Mano, Wako Yoshida, Ben Seymour, Takeshi Shimizu, Koichi Hosomi, Youichi Saitoh, Kiyoto Kasai, Nobumasa Kato, Hidehiko Takahashi, Yasumasa Okamoto, Okito Yamashita, Mitsuo Kawato, and Hiroshi Imamizu. A multi-site, multi-disorder resting-state magnetic resonance image database. *Scientific Data 2021 8:1*, 8:1–15, 8 2021. ISSN 2052-4463. doi: 10.1038/s41597-021-01004-8.

- Tim M. Tierney, Nicholas A. Alexander, John Ashburner, Nicole Labra Avila, Yaël Balbastre, Gareth Barnes, Yulia Bezsudnova, Mikael Brudfors, Korbinian Eckstein, Guillaume Flandin, Karl Friston, Amirhossein Jafarian, Olivia S. Kowalczyk, Vladimir Litvak, Johan Medrano, Stephanie Mellor, George O’Neill, Thomas Parr, Adeel Razi, Ryan Timms, and Peter Zeidman. Spm 25: open source neuroimaging analysis software. *Journal of Open Source Software*, 10:8103, 6 2025. ISSN 2475-9066. doi: 10.21105/JOSS.08103.
- Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004. ISSN 10577149. doi: 10.1109/TIP.2003.819861.
- Dongtao Wei, Kaixiang Zhuang, Lei Ai, Qunlin Chen, Wenjing Yang, Wei Liu, Kangcheng Wang, Jiangzhou Sun, and Jiang Qiu. Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset. *Scientific Data*, 5(1), July 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.134.
- Mengqi Wu, Minhui Yu, Shuaiming Jing, Pew Thian Yap, Zhengwu Zhang, and Mingxia Liu. Unpaired volumetric harmonization of brain mri with conditional latent diffusion. *Medical Image Analysis*, 107:103849, 1 2026. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2025.103849.
- Lianrui Zuo, Yihao Liu, Yuan Xue, Blake E. Dewey, Samuel W. Remedios, Savannah P. Hays, Murat Bilgel, Ellen M. Mowry, Scott D. Newsome, Peter A. Calabresi, Susan M. Resnick, Jerry L. Prince, and Aaron Carass. Haca3: A unified approach for multi-site mr image harmonization. *Computerized Medical Imaging and Graphics*, 109:102285, 10 2023. ISSN 0895-6111. doi: 10.1016/J.COMPMEDIMAG.2023.102285.