Sample Structure, Refine Contrast: Zero-Shot MR Image Harmonization with a Phase-Preserving Diffusion Prior

Muhammad F. A. Chaudhary[1], Samuel W. Remedios[1], Kathleen M. Bartz[1], Savannah P. Hays[1], Lianrui Zuo[2], Carolyn Koch[3], Sandra D. Cassard[3], Shuwen Wei[1], Scott D. Newsome[3], Ellen M. Mowry[3], Blake E. Dewey[3], Jerry L Prince[1], and Aaron Carass[1]

[1] Image Analysis and Communications Laboratory, Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
mchaud25@jhu.edu

[2] Dept. of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN 37235 USA

[3] Dept. of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA

Abstract. Variability in magnetic resonance (MR) image contrast across scanners leads to inconsistent quantitative measurements which prevents reliable biomarker development for brain disorders. Existing harmonization methods often rely on traveling subjects, paired multicontrast images, or synthetic augmentations; thereby restricting their applicability to unseen domains. Such methods rely on inductive biases, such as contrastive disentanglement objectives, yet lack any principled mechanism for incorporating modality-specific *prior knowledge* for harmonization, leaving the problem significantly underconstrained. We propose a *zero-shot* MR harmonization method based on *structure*-constrained sampling from a *phase-preserving* diffusion prior. While sampling along the *phase-preserving manifold* of the *source* image, our method *refines* its latent representation through controlled low-frequency injection of the target contrast. Instead of standard Gaussian noise, we train our diffusion prior using structured noise obtained from T_1-w MR images. For training our prior, we pooled T_1-w images from three different studies being independently conducted at 47 clinical sites throughout the world. We evaluated our method on two traveling-subjects datasets: an independent, non-overlapping substudy cohort and the fully out-of-distribution (OOD) FTHP dataset. Our method performed significantly better than recent state-of-the-art anatomy-contrast disentanglement approaches with a 10% increase ($p < 0.001$) in PSNR and a 2% increase ($p < 0.001$) in SSIM. The source code and pretrained model weights are publicly available at: https://github.com/MFaizyabAli/zshot-phase-harm.

Keywords: MR Harmonization · Diffusion Models · Domain Shift · Latent Variable Refinement

1 Introduction

Magnetic resonance (MR) imaging is commonly used to characterize neurological disorders owing to its rich soft-tissue contrast and the ability to visualize diverse anatomical structures. Although desired, this flexibility makes MR image acquisition susceptible to *domain shifts* under different scanner types. Such shifts in distribution manifest as *unwanted contrast variations* and inflate biomarker variability by up to 60% [6, 10, 20]. This underscores the need for MR image harmonization [28] to improve contrast consistency without changing the anatomy.

Image-to-image translation has been used to harmonize MR images, but early paired methods rely on traveling-subject datasets acquired under multiple contrasts [5, 18, 23]. Unpaired methods, based on CycleGAN [27], relax this assumption by learning bidirectional mappings $\mathcal{G}_{A\to B}$ and $\mathcal{G}_{B\to A}$ between sites A and B, enforced through cycle-consistency [12, 16]. While effective in two-domain settings, such approaches do not scale well with the number of sites. For example, Roca *et al.* [16] required 11 generators and 20 discriminators to learn mappings across 11 scanners. Multisite harmonization is an $n \to n$ problem, and cycle-consistent methods are impractical for large multicenter studies.

Recent anatomy-contrast disentanglement methods attempt to isolate scanner effects in a latent *contrast* variable while preserving the anatomy. For example, HACA3 relies on the contrastive alignment of paired T_1-w, T_2-w, and FLAIR scans to isolate an anatomy embedding, but demonstrated the best performance when all contrasts were available [30]. Zuo *et al.* used an information bottleneck formulation to disentangle anatomy and contrast for T_1-w harmonization [29, 31]. Similarly, ImUnity used gamma augmentations on T_1-w to encourage contrast-invariant latent spaces [1]. More recently, Scholz *et al.* used a diffusion decoder conditioned on disentangled anatomy-contrast embeddings [19]. These methods assume latent factorization and rely on inductive biases, such as contrastive learning, but lack any principled mechanism for incorporating modality-specific *prior* knowledge into the harmonization process.

Denoising diffusion probabilistic models (DDPMs) have demonstrated remarkable progress in unconditional image synthesis [7]. However, their inherent stochasticity makes it difficult to enforce both *anatomical fidelity* and *target contrast*—requirements that are essential for MR harmonization. Recent constrained-sampling approaches such as iterative latent variable refinement (ILVR) and diffusion posterior sampling (DPS) introduce guidance terms during the reverse diffusion process [3, 4]. Although effective for coarse control, these methods still permit undesired structural drift when applied to medical images. We propose a *phase-preserving diffusion model* trained with structured noise that explicitly maintains the anatomical phase of the source image throughout the diffusion trajectory. During sampling, we use this prior for latent variable refinement of the source image to a desired contrast from a target image. Our contributions can be summarized as follows: **(1)** We develop a **phase-preserving denoising diffusion prior**, trained with structured noise, for anatomically consistent sampling along the diffusion trajectory; **(2)** Based on the structured noise prior, we propose an **iterative latent variable refinement method for zero-shot MR**

image harmonization; **(3)** We demonstrate improved harmonization across **two independent multisite traveling subject** datasets.

2 Methods

We consider a dataset of 2D T_1-w brain MR images $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$ sampled from a diverse multisite population. Our harmonization framework, shown in Fig. 1, comprises two main components: ***(i)*** a *phase-preserving diffusion prior* trained on T_1-w images that enables structure-constrained sampling, and ***(ii)*** a *zero-shot refinement strategy* that alternates between injecting low-frequency target contrast and projecting the sample back onto the structure manifold.

Phase (Structure) Preserving Diffusion Prior. We assume that the anatomical structure is encoded primarily within the *phase* component $e^{j\phi_I}$ of an image I. We train a DDPM using *structured Gaussian noise* from the phase of the input image, thus we retain stochasticity without destroying spatial geometry [26]. Let $\mathbf{x}_0 \in \mathbb{R}^{H\times W}$ be a clean T_1-w image. The standard forward process corrupts $\mathbf{x}_0$ with i.i.d. Gaussian noise over T timesteps, which destroys spatial structure. We instead construct *structured Gaussian noise* by combining the magnitude of standard Gaussian noise A_ϵ ($\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$) with the phase of the input image $e^{j\phi_{\mathbf{x}_0}}$ (see Fig. 1A). If $\mathcal{F}$ denotes the 2D Fourier transform, we can write:

$$\mathcal{F}(\mathbf{x}_0) = A_{\mathbf{x}_0} e^{j\phi_{\mathbf{x}_0}}, \qquad \mathcal{F}(\boldsymbol{\epsilon}) = A_\epsilon e^{j\phi_\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I), \tag{1}$$

and construct structured noise as:

$$\boldsymbol{\epsilon}(\mathbf{x}_0) = \Re\left(\mathcal{F}^{-1}\left(A_\epsilon\, e^{j\phi_{\mathbf{x}_0}}\right)\right), \tag{2}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform and $\Re$ is the real part of a complex number (see Fig. 1B). The forward process using structured noise is defined as:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}(\mathbf{x}_0), \qquad t = 1, \ldots, T, \tag{3}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ is the per-step noise-retention factor and denotes the cumulative product that determines the exact marginal distribution at time t. Standard DDPMs train the network to predict either the diffusion noise $\boldsymbol{\epsilon}_\theta(\cdot, t)$ or the clean sample $\mathbf{x}_{0,\theta}(\cdot, t)$. In contrast, we adopt a *flow estimation* parameterization in which our model predicts the displacement between a clean image $\mathbf{x}_0$ and structured noise $\boldsymbol{\epsilon}(\mathbf{x}_0)$(see Fig. 1C). Given $(\mathbf{x}_0, \boldsymbol{\epsilon}(\mathbf{x}_0))$, the target flow is:

$$\mathbf{u} = \boldsymbol{\epsilon}(\mathbf{x}_0) - \mathbf{x}_0, \tag{4}$$

which is the displacement from the clean image to the structured noise. The time-conditioned network $f_\theta(\mathbf{x}_t, t)$ is then trained with the following objective:

$$\mathcal{L}_{\text{FLOW}} = \mathbb{E}_{\mathbf{x}_0,\, t}\left[\left\| f_\theta(\mathbf{x}_t, t) - \left(\boldsymbol{\epsilon}(\mathbf{x}_0) - \mathbf{x}_0\right)\right\|_2^2\right]. \tag{5}$$

Because $\boldsymbol{\epsilon}(\mathbf{x}_0)$ preserves the phase of $\mathbf{x}_0$, the learned flow field remains anchored to a structure-consistent manifold. During sampling, we follow the DDIM update [22] but replace noise or velocity estimation with our flow prediction.

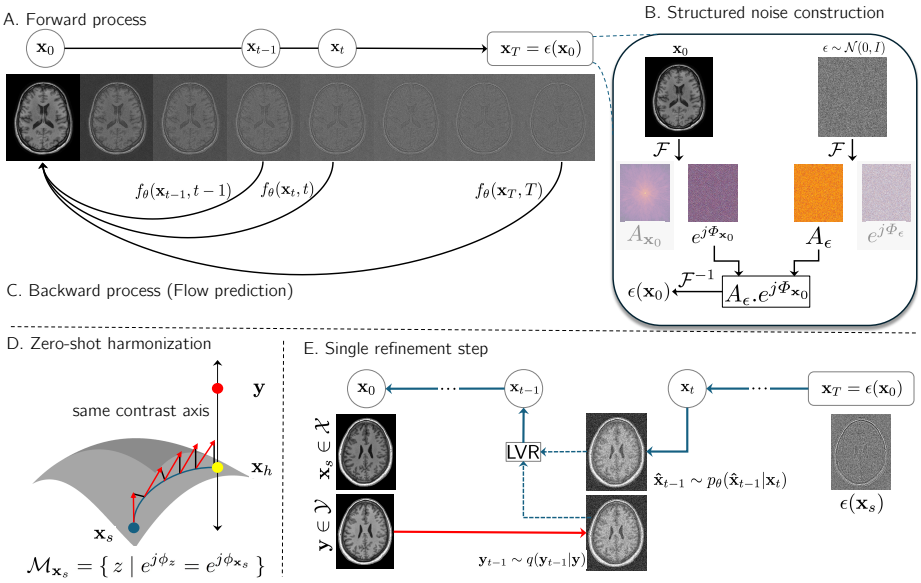

Fig. 1. Overview of our zero-shot harmonization method. (**A**) During forward diffusion, we construct noised samples by adding structured noise $\boldsymbol{\epsilon}(\mathbf{x}_0)$ to the clean image $\mathbf{x}_0$ at different timesteps. (**B**) Structured noise $\boldsymbol{\epsilon}(\mathbf{x}_0)$ is constructed by combining the phase $e^{j\Phi_{\mathbf{x}_0}}$ of the clean image $\mathbf{x}_0$ and the amplitude $A_{\boldsymbol{\epsilon}}$ of standard Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$. (**C**) The backward diffusion f_θ is then learned by a flow prediction objective. (**D**) We sample along a constant-phase manifold to harmonize a source image $\mathbf{x}_s$ to another image $\mathbf{x}_h$ with the same structure but a different contrast. The contrast injection, shown with red arrows, moves the image off the manifold towards the target image $\mathbf{x}_h$, which is then projected back to the source manifold. (**E**) A single refinement step that moves an image from timestep t to $t-1$ along the trajectory based on Eq. 8.

Structure-Consistent Latent Variable Refinement. For zero-shot harmonization, we sample from the structured noise of the *source* image $\boldsymbol{\epsilon}(\mathbf{x}_s)$ to navigate the structural manifold (see Fig. 1D). We define this manifold as:

$$\mathcal{M}_{\mathbf{x}_s} = \{\, z \mid e^{j\phi_z} = e^{j\phi_{\mathbf{x}_s}} \,\}, \tag{6}$$

representing the set of all images sharing the phase of the source while allowing for variations in contrast. Our goal is to reach a harmonized state $\mathbf{x}_h \in \mathcal{M}_{\mathbf{x}_s}$ by progressively injecting the low-frequency components of a target image $\mathbf{y}$, which encodes the desired contrast. We begin the reverse diffusion trajectory at a chosen timestep t_{init}. We draw a structured noise realization $\boldsymbol{\epsilon}(\mathbf{x}_s)$ to ensure the starting point lies on the structure manifold $\mathcal{M}_{\mathbf{x}_s}$ as $\mathbf{x}_{t_{\text{init}}} = \sqrt{\bar{\alpha}_{t_{\text{init}}}}\mathbf{x}_s + \sqrt{1-\bar{\alpha}_{t_{\text{init}}}}\boldsymbol{\epsilon}(\mathbf{x}_s)$. At each timestep t, we define a Gaussian lowpass filter LP_σ with standard deviation σ, and its complement, the high-pass operator $\mathsf{HP}_\sigma = \mathrm{Id} - \mathsf{LP}_\sigma$. During a single reverse step within the trajectory, $t \to t-1$, we first compute a preliminary estimate $\hat{\mathbf{x}}_{t-1}$ as an unconditional DDIM update to maintain consistency with

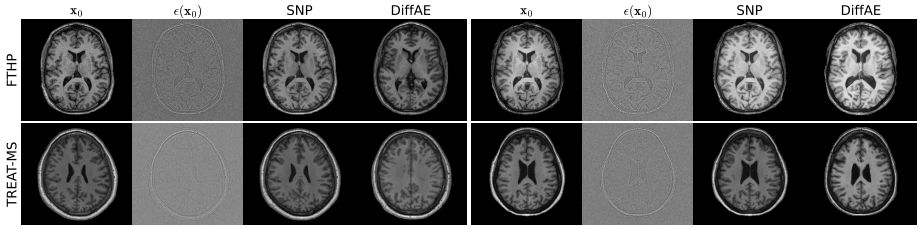

Fig. 2. Structured Noise Prior (SNP) vs. DiffAE. Clean images $\mathbf{x}_0$, their structured noise samples $\boldsymbol{\epsilon}(\mathbf{x}_0)$, SNP (our method) results showing anatomically faithful, contrast-consistent reconstructions, and DiffAE results that are anatomically inconsistent across four different subjects (two subjects in each row).

the learned structural prior. To incorporate the target contrast, we generate noise-matched versions of the target, $\mathbf{y}_t$, and source images, $\mathbf{x}_{s,t}$, at t:

$$\mathbf{y}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{y} + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_{\text{static}}, \qquad \mathbf{x}_{s,t} = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_s + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_{\text{static}}, \tag{7}$$

by using a fixed noise realization $\boldsymbol{\epsilon}_{\text{static}} \sim \mathcal{N}(0, I)$. We then extract the low-frequency components of the estimated $\mathsf{LP}_\sigma(\hat{\mathbf{x}}_{t-1})$ and target image $\mathsf{LP}_\sigma(\mathbf{y}_t)$. To ensure the structural details of the source image are retained, we also compute the high-frequency component of the source image $\mathsf{HP}_\sigma(\mathbf{x}_{s,t})$. The latent variable refinement step blends these components through the mixing weight $\eta(t)$:

$$\mathbf{x}_{t-1} = \eta(t)\,\mathsf{LP}_\sigma(\mathbf{y}_t) + \big(1-\eta(t)\big)\,\mathsf{LP}_\sigma\big(\hat{\mathbf{x}}_{t-1}\big) + \mathsf{HP}_\sigma(\mathbf{x}_{s,t}). \tag{8}$$

In this formulation, the first term injects the target contrast, the second maintains fidelity to the diffusion prior, and the third attempts to preserve the anatomical details from the source image. As t approaches zero, $\eta(t)$ decays to ensure a smooth transition. The resulting final sample, $\mathbf{x}_0 = \mathbf{x}_h$, adopts the contrast of the target image $\mathbf{y}$ and the anatomy of the source image $\mathbf{x}_s$.

Implementation Details. We set total diffusion timesteps to $T = 1000$ [7]. The denoising convolutional neural network $f_\theta(\cdot, t)$ was implemented as a five-level UNet with channels 16, 32, 64, 128, 256 [17]. We optimized all models using the ADAM optimizer with a learning rate of 10^{-4} [9]. We performed zero-shot latent variable updates for 100 timesteps with an initial timestep of $t_{\text{init}} = 99$. At each reverse step, we applied a Gaussian lowpass operator with scale $\sigma = 1$ mm, and mixed target and prior updates using a weight $\eta(t)$ annealed from 1 to 0 over the sampling trajectory. Our models use the MONAI framework [2] and all experiments were conducted on a single NVIDIA A40 48GB GPU.

3 Experiments and Results

Datasets. We trained our structured noise prior using a total of $2,000$ T_1-w MR images drawn from three sources: $1,176$ images from 371 participants from the

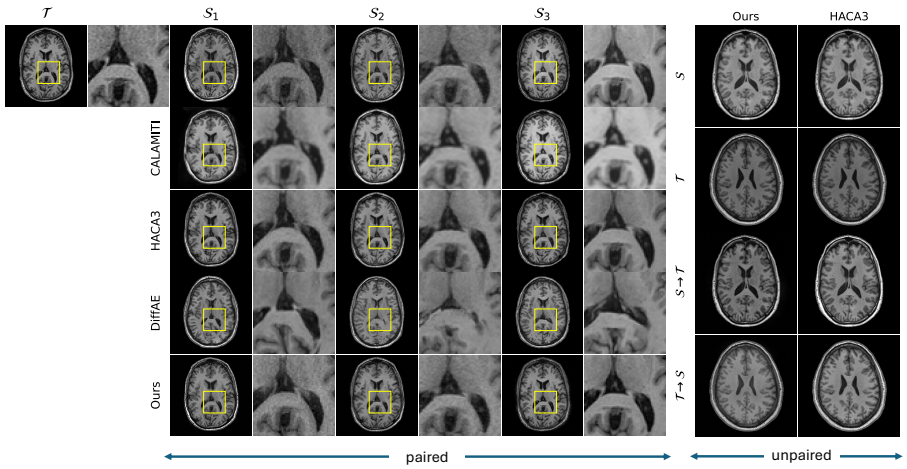

Fig. 3. Qualitative performance evaluation. Target image $\mathcal{T}$ and source images from 3 sites, $\mathcal{S}_i$, show substantial scanner contrast variability. CALAMITI [29] partially alleviates contrast differences while DiffAE [15] suffers from acute anatomical distortions. Our method produces harmonized images that closely match the target contrast while preserving fine structural details of the source image (**left panel**). Harmonization is also visualized in unpaired samples and compared to HACA3 (**right panel**).

TRaditional versus Early Aggressive Therapy for Multiple Sclerosis (TREAT-MS) Trial cohort [13], 780 T_1-w images from the Amsterdam Open MRI collection (AOMIC) [21], and 44 T_1-w images from the cross-modal hierarchical control study (CMHC) [25]. The TREAT-MS study included images acquired at 45 clinical sites with more than 100 scanner configurations, while AOMIC was acquired using a single scanner that was upgraded twice [21]. We evaluated our model on two traveling-subject datasets. The first was a non-overlapping substudy from the TREAT-MS with 14 participants each scanned at 3–5 different sites across the United States [13]. We also evaluated our method on the out-of-distribution (OOD) Frequently Traveling Human Phantom (FTHP) dataset [14]. FTHP contained T_1-w MR images of a single healthy male volunteer from 116 different scanners. For all subjects in both datasets, we designated the first available site as the target domain and harmonized all remaining sites to this reference. Before training, we applied N4 bias correction to all T_1-w images [24] and rigidly registered them to the MNI152 template with 1mm^3 isotropic resolution [11]. Image intensities were then rescaled such that the minimum and 95^{th} percentile were mapped to 0 and 1, respectively, followed by clipping values above 5 to suppress outliers. For all experiments, we extracted 110 central axial slices per subject, covering nearly the full extent of supratentorial brain anatomy.

Structured Noise Prior Enables Superior Structure Preservation. We compared our structured noise prior (SNP) with DiffAE [15] on representative T_1-w images from 2 different sites each from FTHP and TREAT-MS (see Fig. 2).

Table 1. Quantitative evaluation on the FTHP and TREAT-MS datasets. **(A)** Self-reconstruction fidelity of the phase-preserving prior compared with DiffAE. **(B)** MR harmonization performance of our method compared to CALAMITI, DiffAE, and HACA3. Values are mean_{std} computed over all subjects and slices. Best ( ) and second-best ( ) results highlighted per column.

	FTHP		TREAT-MS	
	PSNR (dB) ↑	**SSIM** ↑	**PSNR (dB)** ↑	**SSIM** ↑
(A) Reconstruction Quality				
w/ DiffAE†	$14.65_{1.28}$	$0.620_{0.093}$	$14.28_{1.53}$	$0.626_{0.107}$
w/ ILVR **(Ours)**	$24.79_{1.86}$	$0.938_{0.022}$	$25.92_{1.68}$	$0.949_{0.021}$
(B) MR Image Harmonization				
Unharmonized†	$18.53_{2.33}$	$0.836_{0.055}$	$19.85_{3.13}$	$0.888_{0.062}$
CALAMITI†	$18.53_{2.08}$	$0.835_{0.061}$	$20.79_{2.97}$	$0.899_{0.059}$
DiffAE†	$14.84_{1.03}$	$0.661_{0.081}$	$14.46_{1.47}$	$0.666_{0.099}$
HACA3†	$18.55_{2.32}$	$0.841_{0.055}$	$20.02_{2.93}$	$0.891_{0.059}$
ILVR **(Ours)**	$20.12_{2.08}$	$0.861_{0.046}$	$22.07_{2.54}$	$0.910_{0.047}$

†$p < 0.001$ vs. ILVR **(Ours)** for all reported metrics and both datasets (Wilcoxon signed-rank test, two-sided, $n \approx 8{,}700$ slices per dataset).

Anatomically consistent reconstructions from $\boldsymbol{\epsilon}(\mathbf{x}_0)$ show that our method preserves the source anatomy, while DiffAE failed to retain structural consistency. Quantitative results in Table 1 show that SNP achieved significantly higher SSIM and PSNR values than DiffAE for both the TREAT-MS ($N = 8658$ slices) and FTHP ($N = 8769$ slices) datasets.

Zero-Shot Harmonization Across Multisite and OOD Scanners. Zero-shot harmonization across three source sites $\mathcal{S}_1 - \mathcal{S}_3$ to the target reference $\mathcal{T}$ is shown in Fig. 3. CALAMITI and HACA3 mitigate some contrast variation but often introduce over-smoothing and loss of cortical detail. DiffAE exhibits similar limitations, producing homogenized intensities that obscure deep gray structures while also changing anatomy. Our method yields harmonized images that closely match the target contrast while preserving fine anatomical structure. Across all sites, the outputs maintain gyral morphology, thalamic boundaries, and periventricular sharpness. Quantitatively, our approach achieves the highest SSIM and PSNR on both evaluation datasets (see Table 1).

Downstream Brain Segmentation and Ablation Study. We segmented five anatomical structures of the brain using the spatially localized atlas network tiles (SLANT) framework [8] to assess segmentation consistency before and after harmonization (see Fig. 4). Dice similarity coefficients (DSC) improved after harmonization with our method as well as HACA3, while CALAMITI and DiffAE led to decrease in segmentation agreement. Our model performed

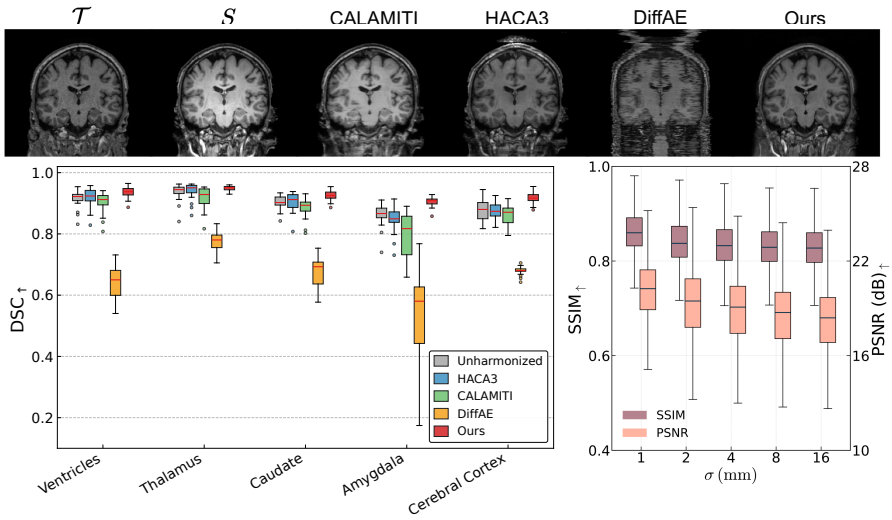

Fig. 4. Multiplanar visualization, downstream evaluation, and ablation study. Coronal views of a representative target $\mathcal{T}$ and source image $\mathcal{S}$ are shown. HACA3 and DiffAE demonstrate slicewise discontinuities which are quite pronounced in the latter. **(Bottom left)** Downstream brain segmentation analysis using SLANT. **(Bottom right)** ablation study for different values of σ for the Gaussian LP_σ filter.

consistently across different σ levels, and produced 3D volumes *without* the slicewise discontinuities observed in HACA3 and DiffAE (see Fig. 4).

4 Discussion and Conclusion

We propose a zero-shot MR harmonization method that relies on sampling from a phase-preserving diffusion model. Training the prior on structured Gaussian noise enables the model to learn a manifold that preserves fine anatomical structure while supporting controlled contrast transformations. On both traveling-subject datasets, our approach outperformed state-of-the-art harmonization methods, including HACA3 [30]. Notably, HACA3 was trained using multicontrast images and benefited from substantially richer information content, yet our method, trained on a single modality, still achieved superior harmonization fidelity. DiffAE demonstrated anatomical distortions, while CALAMITI struggled to achieve complete contrast transfer.

Our method provides a principled approach to incorporating prior knowledge for harmonization, offering a new perspective beyond simply relying on inductive biases such as contrastive disentanglement objectives. However, we would like to note some limitations. In its current form, our method operates on 2D slices but still did not show slice discontinuities. While structured noise provides a strong modality-specific prior, our method may not generalize to other contrasts and

future work could explore adaptive or data-driven noise models tailored to different sequences. Our findings still suggest that structure-constrained diffusion priors offer a principled and scalable path toward robust, zero-shot MR harmonization, with the potential to substantially enhance the comparability and reproducibility of multicenter neuroimaging studies.

Acknowledgments

This research is supported by the following grants: NIH R01EB036013 (Prince), CDMRP W81XWH2010912 (Prince), the National MS Society grant RG-1507-05243 (Pham) and the Patient-Centered Outcomes Research Institute (PCORI) grant MS-1610-37115 (Newsome and Mowry). The statements in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. Data were provided [in part] by the TREAT-MS Traveling Subjects study sites: University of Alabama at Birmingham (PI: William Meador), University of Florida (PI: Torge Rempe), Johns Hopkins University-Nelson (PIs: Mowry and Newsome), Johns Hopkins Outpatient Center (PIs: Mowry and Newsome), Johns Hopkins University (PIs: Mowry and Newsome), Kennedy Krieger Institute (Coordinator: P.A. Calabresi), University of Miami (PI: Leticia Tornes), New York University (PI: Tyler Smith), RadNet (Director of RadNet Research: Kristi A. Maya), and University of South Florida (PI: Derrick Robertson). The authors would like to thank the TREAT-MS Traveling Subjects study participants.

References

1. 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 (2023)
2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14367–14376 (2021)
4. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: The Eleventh International Conference on Learning Representations (2022)
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., et al.: DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. Magnetic Resonance Imaging **64**, 160–170 (2019)
6. Hedges, E.P., Dimitrov, M., Zahid, U., Vega, B.B., Si, S., Dickson, H., McGuire, P., Williams, S., Barker, G.J., Kempton, M.J.: Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, freesurfer version and processing stream. NeuroImage **246**, 118751 (2022)

7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
8. Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvathaneni, P., Bao, S., Bermudez, C., Resnick, S.M., Cutting, L.E., Landman, B.A.: 3D whole brain segmentation using spatially localized atlas network tiles. NeuroImage **194**, 105–119 (2019)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015)
10. Kushol, R., Parnianpour, P., Wilman, A.H., Kalra, S., Yang, Y.H.: Effects of MRI scanner manufacturers in classification tasks with deep learning models. Scientific Reports **13**(1), 16791 (2023)
11. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al.: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**(1412), 1293–1322 (2001)
12. Modanwal, G., Vellal, A., Buda, M., Mazurowski, M.A.: MRI image harmonization using cycle-consistent generative adversarial network. In: Medical Imaging 2020: Computer-Aided Diagnosis. vol. 11314, pp. 259–264. SPIE (2020)
13. Mowry, E.M., Qian, P., Meador, W., Lynch, S., Narayan, R., Borazanci, A., Robertson, D., Qureshi, S., Graves, J., Silversteen, J., et al.: The TRaditional versus Early Aggressive Therapy for MS (TREAT-MS) trial: Design and baseline characteristics of participants. Contemporary Clinical Trials p. 108117 (2025)
14. Opfer, R., Krüger, J., Spies, L., Ostwaldt, A.C., Kitzler, H.H., Schippling, S., Buchert, R.: Automatic segmentation of the thalamus using a massively trained 3D convolutional neural network: higher sensitivity for the detection of reduced thalamus volume by improved inter-scanner stability. European Radiology **33**(3), 1852–1861 (2023)
15. Preechakul, K., Chatthee, N., Wizadwongsa, S., Suwajanakorn, S.: Diffusion autoencoders: Toward a meaningful and decodable representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10619–10629 (2022)
16. Roca, V., Kuchcinski, G., Pruvo, J.P., Manouvriez, D., Lopes, R., et al.: IGUANe: A 3D generalizable CycleGAN for multicenter harmonization of brain MR images. Medical Image Analysis **99**, 103388 (2025)
17. Ronneberger, O., Fischer, P., Brox, T.: UNet: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
18. Saito, Y., Kamagata, K., Andica, C., Maikusa, N., Uchida, W., Takabayashi, K., Yoshida, S., Hagiwara, A., Fujita, S., Akashi, T., et al.: Traveling subject-informed harmonization increases reliability of brain diffusion tensor and neurite mapping. Aging and Disease **15**(6), 2770 (2023)
19. Scholz, D., Erdur, A.C., Holland, R., Ehm, V., Peeken, J.C., Wiestler, B., Rueckert, D.: Contrastive Anatomy-Contrast Disentanglement: A Domain-General MRI Harmonization Method. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 100–110. Springer (2025)
20. Schurink, N.W., van Kranen, S.R., Roberti, S., van Griethuysen, J.J., Bogveradze, N., Castagnoli, F., El Khababi, N., Bakers, F.C., de Bie, S.H., Bosma, G.P., et al.: Sources of variation in multicenter rectal MRI data and their effect on radiomics feature reproducibility. European Radiology **32**(3), 1506–1516 (2022)

21. Snoek, L., van der Miesen, M.M., Beemsterboer, T., Van Der Leij, A., Eigenhuis, A., Steven Scholte, H.: The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. Scientific Data **8**(1), 85 (2021)
22. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2020)
23. Tian, D., Zeng, Z., Sun, X., Tong, Q., Li, H., He, H., Gao, J.H., He, Y., Xia, M.: A deep learning-based multisite neuroimage harmonization framework established with a traveling-subject dataset. NeuroImage **257**, 119297 (2022)
24. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (2010)
25. Yoo, T., Jeon, H.: Cross-modal hierarchical control. OpenNeuro Dataset (2025), https://doi.org/10.18112/openneuro.ds006628.v1.0.1
26. Zeng, Y., Ochoa, C., Zhou, M., Patel, V.M., Guizilini, V., McAllister, R.: NeuralRemaster: Phase-Preserving Diffusion for Structure-Aligned Generation. arXiv preprint arXiv:2512.05106 (2025)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232 (2017)
28. Zuo, L., Liu, Y., Carass, A.: Editorial for the Special Issue on Harmonization Techniques for MRI. NeuroImage **335**, 121979 (2026)
29. Zuo, L., Dewey, B.E., Liu, Y., He, Y., Newsome, S.D., Mowry, E.M., Resnick, S.M., Prince, J.L., Carass, A.: Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. NeuroImage **243**, 118569 (2021)
30. 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., et al.: HACA3: A unified approach for multi-site MR image harmonization. Computerized Medical Imaging and Graphics **109**, 102285 (2023)
31. Zuo, L., Liu, Y., Xue, Y., Han, S., Bilgel, M., Resnick, S.M., Prince, J.L., Carass, A.: Disentangling a single MR modality. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. pp. 54–63. Springer (2022)