This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

AsynDiff: Asynchronous-Timestep Diffusion with Noise-Aware Attention for Multi-Sequence MRI Synthesis

Luyi Han[1,2,3], Pengxiang Min[4], Xinghe Xie[1], Tianyu Zhang[2,3], Xin Wang[2,3], Yuan Gao[2,3], Chunyao Lu[2,3], Xinglong Liang[2,3], Yaofei Duan[1,2,3], Antonio Portaluri[2,3], Nika Rasoolzadeh[2,3], Jarek van Dijk[2,3], Muzhen He[3], Yue Sun[1], Chan-Tong Lam[1], Ritse Mann[2,3], and Tao Tan[1(✉)]

[1] Faculty of Applied Sciences, Macao Polytechnic University, 999078, Macao, China
[2] Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Geert Grooteplein 10, 6525 GA, Nijmegen, The Netherlands
[3] Department of Radiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands
[4] Department of Medical Biochemistry and Microbiology, Uppsala University Biomedical Center, Uppsala University, Uppsala, Sweden
taotan@mpu.edu.mo

Abstract. Recent diffusion-based approaches have achieved high-fidelity synthesis of missing sequences from partially observed multi-sequence MRI. However, most existing methods rely on a one-way conditional paradigm, in which observed sequences are used as fixed conditions while denoising pre-specified targets, which limits adaptability to arbitrary missing patterns. A more flexible formulation should encode the observed-missing asymmetry within the diffusion process itself. Hence, we propose AsynDiff, an asynchronous-timestep diffusion framework for multi-sequence MRI fusion and synthesis. At each training iteration, we assign sequence-specific timesteps, deliberately inducing inter-sequence noise disparities that promote dependable guidance from low-noise observed sequences to high-noise missing targets. To exploit this asymmetry, we propose cross-sequence noise-aware attention, which adaptively modulates attention weights based on relative noise levels, amplifying reliable features while suppressing noisy and potentially misleading cues. Experiments on BraTS2021 and BraTS-GLI (2024) demonstrate that AsynDiff outperforms comparisons in missing-sequence synthesis. Further analyses across sampling strategies and attention visualizations under varying noise schedules corroborate both the effectiveness and interpretability of the proposed design, highlighting asynchronous-timestep diffusion as a principled mechanism for robust multi-sequence MRI completion. Our code is publicly available [5].

Keywords: Multi-Sequence MRI Synthesis · Diffusion Models · Asynchronous Timestep · Cross-Sequence Noise-Aware Attention

[5] https://github.com/fiy2W/AsynDiff

1 Introduction

Multi-sequence magnetic resonance imaging (MRI) provides complementary tissue contrasts that are critical for diagnosis, treatment planning, and quantitative analysis [5, 21, 20]. In clinical practice, however, acquiring a complete set of sequences is often infeasible, as sequences may be missing or unusable due to scan-time and cost constraints, motion and susceptibility artifacts, contraindications to contrast agents, and protocol heterogeneity across scanners and institutions [4]. Such incompleteness can substantially degrade downstream pipelines, such as segmentation, registration, and radiomic analysis, thereby motivating methods that synthesize missing sequences from available data to improve robustness and harmonization [24].

Deep generative models have been extensively explored for MRI synthesis, evolving from GAN-based image-to-image translation [24, 15, 28, 17, 6, 11, 10] to diffusion-based generation [13, 26, 8, 27]. Although GANs can produce visually plausible images, they often suffer from unstable training, mode collapse, and limited mechanisms for uncertainty characterization [7]. Diffusion models, such as DDPM [12] and subsequent refinements including EDM [14] and latent diffusion [23], instead learn an iterative denoising process and have demonstrated improved stability and high-fidelity synthesis in medical imaging. Nevertheless, multi-sequence MRI synthesis poses challenges beyond single-input translation, as the model must fuse a variable subset of observed sequences and propagate anatomy-consistent structure to unobserved targets. Most existing methods [13, 26, 8, 27] follow a one-way conditional paradigm, using observed sequences as static conditions while denoising pre-specified targets, which restricts adaptation to diverse missing patterns. This design also complicates controlled cross-sequence transfer when multiple targets are synthesized jointly, potentially leading to anatomical inconsistencies across sequences under incomplete inputs. These issues suggest that the key challenge is not only conditioning but also regulating information flow based on what is observed versus what is missing during denoising. A natural solution is to embed this observed-missing asymmetry into the diffusion dynamics, so that reliable sources are explicitly favored when guiding uncertain targets.

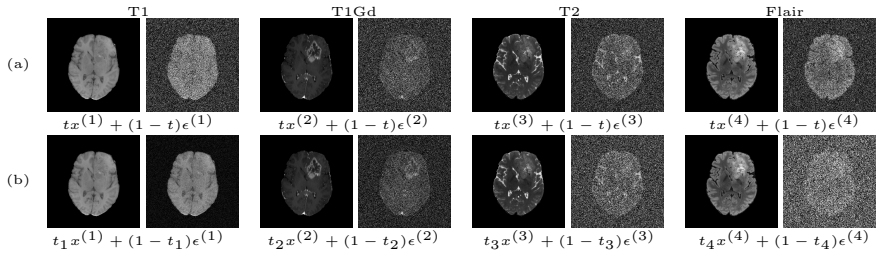

Fig. 1. Comparison between synchronous and asynchronous noise injection. (a) Synchronous: all sequences are perturbed at the same noise timestep t. (b) Asynchronous: the sequences are perturbed at different noise timesteps $t_i, i = 1, 2, 3, 4$.

To address these limitations, we propose AsynDiff, an asynchronous-timestep diffusion framework for multi-sequence MRI fusion and synthesis. As shown in Fig. 1 (b), different sequences are assigned different timesteps at each training iteration, deliberately inducing controlled inter-sequence noise disparities that prioritize reliable guidance from less-corrupted observed sequences to more-corrupted missing targets. Our primary contributions are as follows: (1) We introduce a novel paradigm that applies asynchronous-timestep diffusion to multi-sequence MRI synthesis. (2) We propose cross-sequence noise-aware attention that adapts fusion patterns according to relative noise levels, promoting reliable information transfer from observed sequences to missing targets. (3) We explore various sampling strategies for robust inference under arbitrary missing-sequence patterns and analyze sampling-step trade-offs. (4) We demonstrate consistent gains on the BraTS2021 and BraTS-GLI (2024) datasets with comprehensive comparisons, ablations, and attention visualizations that highlight both effectiveness and interpretability.

2 Methods

2.1 Asynchronous-Timestep Diffusion

Synchronous Timestep. A common formulation for multi-sequence diffusion represents the multi-sequence state as a concatenated tensor, $x = [x^{(1)}, \ldots, x^{(M)}]$, where each sequence $x^{(m)} \in \mathbb{R}^{H\times W}$ (2D) or $x^{(m)} \in \mathbb{R}^{H\times W\times D}$ (3D), and applies a shared diffusion timestep t under a linear noise schedule [1, 18, 19]:

$$x_t = tx_0 + (1-t)\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{1}$$

where x_0 denotes the clean multi-sequence sample, x_t is its noisy version at timestep $t \in [0, 1]$, and ϵ is the Gaussian noise with identity covariance I. Under such a synchronized schedule, the denoiser observes all sequences at the same corruption level, which can hinder the effective extraction and propagation of informative cues from the observed sequences.

Asynchronous Timestep. We decouple noise levels across sequences by assigning each sequence its own timestep t_m at each training iteration. Specifically, we independently sample $\{t_m\}_{m=1}^{M}$, which induces controlled inter-sequence noise disparities. This design exposes the denoiser to asymmetric corruption patterns and enables it to learn how to leverage relatively cleaner sequences to guide the reconstruction of more heavily corrupted targets. For each sequence m, we apply a sequence-specific forward noising process:

$$x_{t_m}^{(m)} = t_m x_0^{(m)} + (1-t_m)\epsilon^{(m)}, \quad \epsilon^{(m)} \sim \mathcal{N}(0, I). \tag{2}$$

Training Objective. Following the observation in JiT [16] that directly predicting clean data (x-prediction) is advantageous in high-dimensional pixel/patch

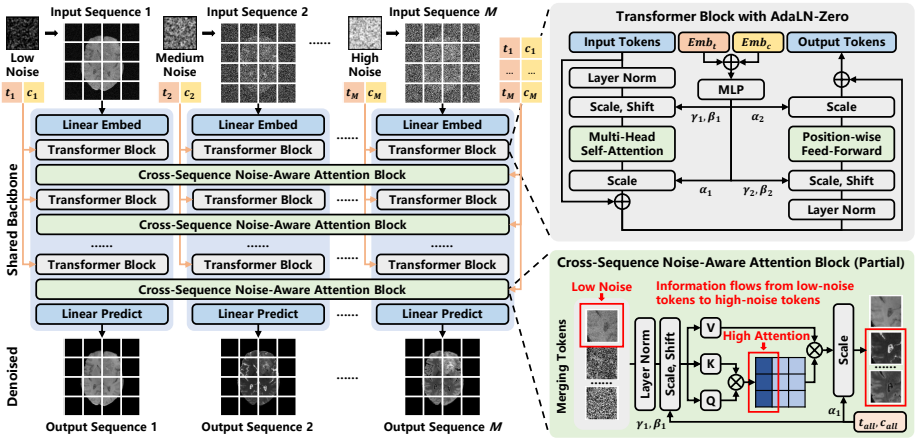

Fig. 2. An overview of AsynDiff architecture. Asynchronous noisy images at timesteps t_m, conditioned on the sequence class c_m, are fed into AsynDiff for denoising.

space, we parameterize the denoiser as,

$$\{\hat{x}_0^{(m)}\}_{m=1}^{M} = f_\theta\left(\{x_{t_m}^{(m)}\}_{m=1}^{M},\ \{t_m\}_{m=1}^{M}\right), \tag{3}$$

where f_θ denotes a neural network with parameters θ, $\hat{x}_0^{(m)}$ is the predicted clean sample for sequence m, and $\{\cdot\}_{m=1}^{M}$ indicates the collection over all M sequences. An asynchronous flow velocity $v^{(m)}$ is defined as the time-derivative of $x_{t_m}^{(m)}$, that is, $v^{(m)} = (x_0^{(m)} - x_{t_m}^{(m)})/(1-t_m)$. We train the model by minimizing the following flow-based objective:

$$\mathcal{L}(\theta) = \sum_{m=1}^{M} \lambda_m \left\| \hat{v}^{(m)} - v^{(m)} \right\|^2, \tag{4}$$

where $\hat{v}^{(m)}$ denotes the predicted flow velocity based on $\hat{x}_0^{(m)}$ for sequence m, and λ_m is a sequence-specific weighting factor. To accommodate samples with incomplete sequences, we set $\lambda_m = 0$ and $t_m = 0$ (i.e., $x_{t_m}^{(m)} \equiv \epsilon^{(m)}$) for any missing data; otherwise, we use $\lambda_m = 1$.

2.2 AsynDiff Architecture

Our asynchronous-timestep diffusion mechanism is model-agnostic and can be seamlessly integrated into existing diffusion backbones. In Fig. 2, we instantiate AsynDiff by augmenting the JiT [16] with additional cross-sequence noise-aware attention blocks, retaining the simplicity and scalability of JiT while enabling explicit multi-sequence interaction under sequence-asynchronous noise schedules.

Intra-Sequence Transformer Block. Let $H_{\text{in}}^{(m)} \in \mathbb{R}^{B\times N\times L}$ be the token features of sequence m from a linear image embedding, where B, N, and L denote batch size, number of spatial tokens, and hidden size. The intra-sequence block performs denoising within each sequence using a Transformer block equipped with adaptive layer normalization (AdaLN-Zero) [22]:

$$\begin{aligned} H_{\text{mid}}^{(m)} &= H_{\text{in}}^{(m)} + \alpha_1^{(m)}\,\text{Attn}\Big((1+\gamma_1^{(m)})\,\text{LN}(H_{\text{in}}^{(m)}) + \beta_1^{(m)}\Big), \\ H_{\text{out}}^{(m)} &= H_{\text{mid}}^{(m)} + \alpha_2^{(m)}\,\text{FFN}\Big((1+\gamma_2^{(m)})\,\text{LN}(H_{\text{mid}}^{(m)}) + \beta_2^{(m)}\Big), \end{aligned} \tag{5}$$

where $\text{Attn}(\cdot)$ denotes multi-head self-attention, $\text{FFN}(\cdot)$ is a position-wise feed-forward network, and $\text{LN}(\cdot)$ presents a layer normalization. The modulation parameters are predicted from:

$$\big(\alpha_1^{(m)}, \gamma_1^{(m)}, \beta_1^{(m)}, \alpha_2^{(m)}, \gamma_2^{(m)}, \beta_2^{(m)}\big) = \text{MLP}\Big(\text{Emb}_t\big(t_m\big) + \text{Emb}_c\big(c_m\big)\Big), \tag{6}$$

where $\text{Emb}_t(\cdot)$ and $\text{Emb}_c(\cdot)$ are embedding layers for the noise timestep t_m and the sequence class c_m, respectively, and $\text{MLP}(\cdot)$ is a modulation network, implemented as a two-layer perceptron with a SiLU nonlinearity.

Cross-Sequence Noise-Aware Attention Block. To enable cross-sequence information flow, we insert a fusion module between intra-sequence Transformer blocks. Given per-sequence token features $\{H^{(m)}\}_{m=1}^{M}$, we align them by spatial index and reshape them into $H^{\text{all}} \in \mathbb{R}^{(BN)\times M\times L}$, such that each spatial token performs self-attention over the M sequences. Formally, this module is obtained by directly modifying Eq. 5 and Eq. 6 such that the intra-sequence feature $H^{(m)}$ is replaced with H^{all}, and the per-sequence conditions (t_m, c_m) are replaced with a joint condition $(t_{\text{all}}, c_{\text{all}})$ shared across all sequences:

$$t_{\text{all}} = \text{Concat}\big(t_1, \ldots, t_M\big), \qquad c_{\text{all}} = \text{Concat}\big(c_1, \ldots, c_M\big). \tag{7}$$

Conditioning on $(t_{\text{all}}, c_{\text{all}})$ enables the block to adapt cross-sequence information passing according to relative noise levels. Finally, we reshape the output back into $\{H^{(m)}\}_{m=1}^{M}$ and proceed with subsequent intra-sequence Transformer blocks.

3 Experiments

3.1 Experimental Settings

Datasets and Evaluation Metrics. We evaluate missing-sequence synthesis on the BraTS2021 [2, 3, 21], which contains four brain MRI sequences per subject (T1, T1Gd, T2, and Flair) for pre-treatment glioma. We use 1,001 subjects for training, 250 for validation, and 219 for testing. For external validation, we evaluate on 188 subjects from the BraTS-GLI (2024) [25], which focuses on post-treatment glioma. All volumes are intensity-clipped at the 0.5th and 99.5th percentiles, normalized to $[-1, 1]$, and center-cropped to $128 \times 192 \times 160$. During training, we randomly sample an input subset of sequences and treat the remaining ones as targets. For validation and testing, input-target combinations are fixed. We report synthesis quality using PSNR, SSIM, and LPIPS [11].

Table 1. Quantitative results for missing-sequence synthesis given a subset of inputs.

N	Methods	BraTS2021 (Internal)			BraTS-GLI (External)		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
1	MM-GAN	23.1±1.9	0.858±0.040	9.84±3.27	17.6±3.1	0.731±0.096	17.1±4.39
	ResViT	22.6±1.7	0.850±0.040	10.1±3.16	17.9±3.2	0.732±0.095	17.0±4.33
	TSF-Seq2Seq	23.3±2.0	0.862±0.044	9.17±3.47	18.0±3.6	0.742±0.114	15.2±4.13
	CoLa-Diff	22.7±2.0	0.888±0.022	10.3±2.47	19.6±2.8	0.837±0.046	14.1±3.20
	MU-Diff	23.0±1.9	0.891±0.023	8.74±2.43	19.9±2.7	0.844±0.049	12.7±3.25
	AsynDiff	**23.3±2.0**	**0.899±0.024**	**8.50±2.24**	**20.1±3.1**	**0.846±0.052**	**12.0±2.94**
	(w/o Asyn)	19.0±1.8	0.737±0.038	21.6±3.79	18.1±1.6	0.733±0.031	22.7±3.22
2	MM-GAN	23.9±2.2	0.878±0.037	8.21±2.75	18.1±3.7	0.747±0.109	15.3±4.43
	ResViT	23.7±2.0	0.877±0.036	8.51±2.50	18.6±3.8	0.751±0.108	15.0±4.48
	TSF-Seq2Seq	**24.2±2.1**	0.881±0.037	8.23±2.59	18.2±4.1	0.747±0.124	14.9±4.63
	CoLa-Diff	23.3±2.3	0.904±0.022	8.68±2.15	19.5±3.4	0.843±0.055	12.9±3.43
	MU-Diff	23.8±2.2	0.910±0.025	7.25±2.13	19.8±3.4	0.850±0.058	11.4±3.52
	AsynDiff	24.1±2.4	**0.918±0.024**	**7.17±2.15**	**20.0±3.8**	**0.853±0.062**	**11.1±3.20**
	(w/o Asyn)	19.1±1.4	0.708±0.045	22.0±4.44	18.0±1.2	0.697±0.036	22.8±3.30
3	MM-GAN	24.4±3.1	0.899±0.043	6.89±3.18	17.0±3.2	0.720±0.096	16.5±3.13
	ResViT	24.1±2.6	0.899±0.040	7.26±2.76	17.7±3.3	0.731±0.092	16.1±3.33
	TSF-Seq2Seq	24.5±2.7	0.900±0.043	6.84±3.00	17.2±3.7	0.724±0.120	15.7±3.91
	CoLa-Diff	23.8±2.5	0.919±0.022	7.15±2.01	18.8±2.8	0.830±0.049	13.0±2.25
	MU-Diff	24.6±2.7	0.928±0.026	5.84±2.11	19.0±2.8	0.833±0.048	11.8±2.25
	AsynDiff	**24.8±2.8**	**0.936±0.024**	**5.60±2.03**	**19.2±3.0**	**0.837±0.055**	**11.8±2.11**
	(w/o Asyn)	12.2±1.0	0.491±0.004	67.9±3.47	12.3±0.7	0.485±0.006	67.0±2.87

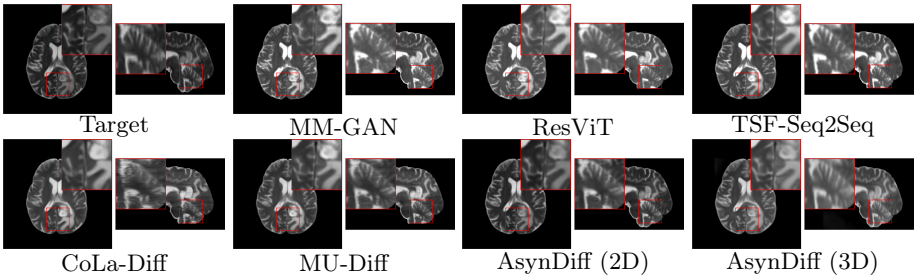

Fig. 3. Examples of synthetic T2 of methods input with T1, T1Gd, and Flair.

Implementation Details. All models are implemented in PyTorch and trained on a single NVIDIA H200 GPU. The backbone configurations are summarized in Table 2. For 2D models, we use a patch resolution of 192×160 with a batch size of 32; for 3D models, we use a patch resolution of $80 \times 80 \times 80$ with a batch size of 4. We train for 250,000 steps using AdamW with a learning rate of 10^{-4}, a linear warm-up of 1,250 steps, and a cosine learning-rate schedule. For inference, we use a rectified-flow sampler and an Euler ODE solver with 5 steps (Fig. 5), and further compare timestep scheduling strategies as summarized in Fig. 4.

Table 2. Details and an ablation study of the backbone architecture.

Backbone	Dim	Patch	Depth	Hidden	Head	PSNR↑	SSIM↑	LPIPS↑
JiT-B	2D	16×16	12	768	12	23.8±2.4	0.912±0.027	7.65±2.53
	3D	8×8×8	12	792	12	22.3±3.1	**0.923±0.024**	7.53±2.53
JiT-L	2D	16×16	24	1024	16	24.0±2.5	0.920±0.028	**7.01±2.51**
	3D	8×8×8	24	1056	16	22.8±2.9	0.921±0.024	8.04±2.60
JiT-H	2D	16×16	32	1280	16	**24.3±2.4**	0.918±0.027	7.18±2.52
	3D	8×8×8	32	1344	16	22.6±3.0	0.923±0.025	7.85±2.71

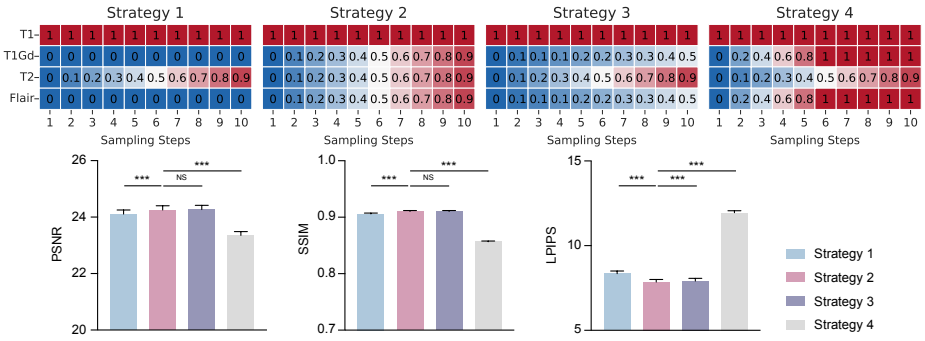

Fig. 4. Four t_m sampling strategies for synthesizing T2 images from T1 inputs (JiT-B 2D). Bar plots compare PSNR, SSIM, and LPIPS across strategies (mean±SEM). Statistical significance was assessed using repeated-measures one-way ANOVA.

3.2 Results

Alternative Comparisons. We compare AsynDiff with GAN-based (MM-GAN [24], ResViT [6], TSF-Seq2Seq [11]) and diffusion-based (CoLa-Diff [13], MU-Diff [8]) methods under varying numbers of available input sequences N. Fig. 3 visualizes qualitative examples of T2 synthesis from the input combination of T1, T1Gd, and Flair. Table 1 illustrates that GAN-based methods tend to yield more stable intensity fidelity (higher PSNR), whereas diffusion-based approaches better preserve structural and semantic details (better SSIM and LPIPS). Our AsynDiff maintains the diffusion advantage while remaining competitive in PSNR. Note that PSNR values are lower overall because they are highly sensitive to intensity scaling and normalization [9].

Ablation Study. Removing asynchronous timesteps (w/o Asyn) causes a substantial performance drop across all settings (Table 1), demonstrating that decoupling noise levels is critical for effective cross-sequence utilization. Specifically, "w/o Asyn" keeps the same backbone and cross-sequence noise-aware attention blocks, but replaces the sequence-specific timesteps with a shared synchronized timestep for all sequences. From Table 2, the 2D backbones show an overall scaling trend (e.g., JiT-L outperforming JiT-B), while the gain from JiT-H is

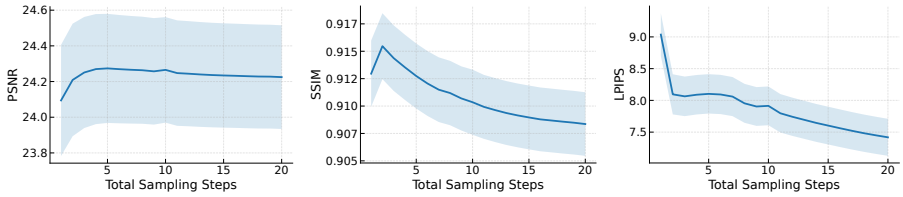

Fig. 5. Performance curves with 95% confidence interval (CI) of T1-to-T2 image synthesis under Strategy 2 (JiT-B 2D), evaluated across different total sampling steps.

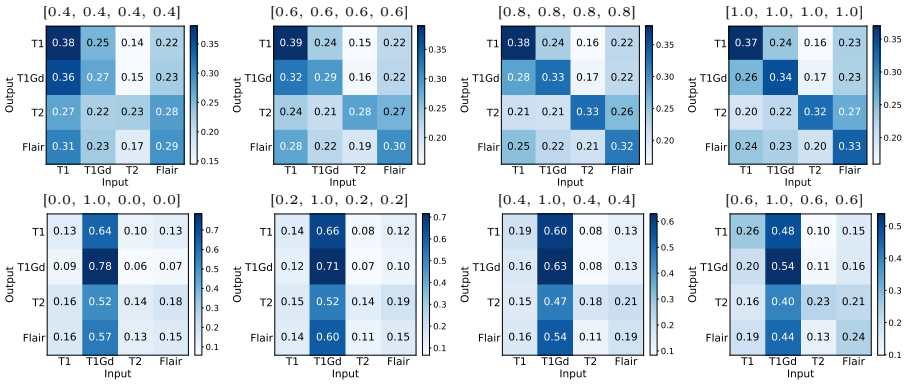

Fig. 6. Attention map in the first cross-sequence noise-aware attention block of AsynDiff (JiT-B 2D) under different noise timesteps $[t_1, t_2, t_3, t_4]$.

marginal, likely due to its substantially larger capacity requiring more training data to exploit fully. For 3D variants, the PSNR drop can be attributed to patch-wise prediction introducing local-contrast bias, a phenomenon also observed in other 3D designs. Nevertheless, 3D backbones exhibit better structural details (higher SSIM) than their 2D counterparts.

Inference Strategies. Fig. 4 compares four timestep scheduling strategies during sampling: (1) denoising only the target; (2) synchronous denoising of all missing sequences; (3) denoising other missing sequences slower than the target; and (4) denoising them faster. Strategy 2 performs best overall (significantly lower LPIPS than Strategy 3), whereas Strategies 1 and 4 degrade markedly, indicating that stable synthesis benefits from coordinated denoising trajectories and avoids unreliable sequences, which may weaken cross-sequence guidance. Fig. 5 further shows performance versus the total number of sampling steps. When considering all metrics, 5 steps provide the best trade-off for T1-to-T2 synthesis.

Noise-Aware Attention. To examine how AsynDiff routes information under asynchronous noise, Fig. 6 shows attention maps from the first cross-sequence noise-aware attention block under different timestep settings. With uniformly

high noise, attention is broadly distributed across sequences; with uniformly low noise, it becomes more concentrated within each sequence, indicating intra-sequence refinement. Under asynchronous noise, information transfers toward lower-noise (more reliable) sequences to higher ones, consistent with supporting the intended mechanism for missing-sequence synthesis.

4 Conclusion

In this work, we propose AsynDiff, an asynchronous-timestep diffusion framework for multi-sequence MRI synthesis that enables noise-aware cross-sequence information transfer. Extensive experiments on BraTS2021 and BraTS-GLI (2024) demonstrate consistent improvements over state-of-the-art baselines, and attention visualizations provide interpretability of the learned fusion behavior.

Acknowledgments. This study is supported by Shenzhen Medical Research Fund (D2501013), Macao Polytechnic University Grant (RP/FCA-08/2024), the grant from Science and Technology Development Fund of Macao (0014/2025/RIC), and the Science and Technology Development Fund, Macau SAR (File no. 0004/2025/ASJ) under the FDCT-FAPESP Joint Funding Scheme.

Disclosure of Interests. The authors declare no competing interests.

References

1. Albergo, M.S., Vanden-Eijnden, E.: Building normalizing flows with stochastic interpolants. arXiv preprint arXiv:2209.15571 (2022)
2. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
4. Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A.: Multimodal mr synthesis via modality-invariant latent representation. IEEE transactions on medical imaging **37**(3), 803–814 (2017)
5. Chen, J.H., Su, M.Y.: Clinical application of magnetic resonance imaging in management of breast cancer patients receiving neoadjuvant chemotherapy. BioMed research international **2013** (2013)
6. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: Residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
7. Dayarathna, S., Islam, K.T., Uribe, S., Yang, G., Hayat, M., Chen, Z.: Deep learning based synthesis of mri, ct and pet: Review and analysis. Medical image analysis **92**, 103046 (2024)

8. Dayarathna, S., Wu, Y., Cai, J., Wong, T.T., Law, M., Islam, K.T., Peiris, H., Chen, Z.: Mu-diff: a mutual learning diffusion model for synthetic mri with application for brain lesions. npj Artificial Intelligence **1**(1), 11 (2025)
9. Dohmen, M., Klemens, M., Baltruschat, I., Truong, T., Lenga, M.: Similarity metrics for mr image-to-image translation. arXiv e-prints pp. arXiv–2405 (2024)
10. Han, L., Tan, T., Zhang, T., Huang, Y., Wang, X., Gao, Y., Teuwen, J., Mann, R.: Synthesis-based imaging-differentiation representation learning for multi-sequence 3d/4d mri. Medical Image Analysis **92**, 103044 (2024)
11. Han, L., Zhang, T., Huang, Y., Dou, H., Wang, X., Gao, Y., Lu, C., Tan, T., Mann, R.: An explainable deep framework: towards task-specific fusion for multi-to-one mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 45–55. Springer (2023)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
13. Jiang, L., Mao, Y., Wang, X., Chen, X., Li, C.: Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 398–408. Springer (2023)
14. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems **35**, 26565–26577 (2022)
15. Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B.: Diamondgan: unified multi-modal generative adversarial networks for mri sequences synthesis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. pp. 795–803. Springer (2019)
16. Li, T., He, K.: Back to basics: Let denoising generative models denoise. arXiv preprint arXiv:2511.13720 (2025)
17. Li, W., Xiao, H., Li, T., Ren, G., Lam, S., Teng, X., Liu, C., Zhang, J., Lee, F.K.h., Au, K.h., et al.: Virtual contrast-enhanced magnetic resonance images synthesis for patients with nasopharyngeal carcinoma using multimodality-guided synergistic neural network. International Journal of Radiation Oncology* Biology* Physics **112**(4), 1033–1044 (2022)
18. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
19. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
20. Mann, R.M., Cho, N., Moy, L.: Breast mri: state of the art. Radiology **292**(3), 520–536 (2019)
21. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
22. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
23. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)

24. Sharma, A., Hamarneh, G.: Missing mri pulse sequence synthesis using multi-modal generative adversarial network. IEEE transactions on medical imaging **39**(4), 1170–1183 (2019)
25. de Verdier, M.C., Saluja, R., Gagnon, L., LaBella, D., Baid, U., Tahon, N.H., Foltyn-Dumitru, M., Zhang, J., Alafif, M., Baig, S., et al.: The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri. arXiv preprint arXiv:2405.18368 (2024)
26. Zhan, C., Lin, Y., Wang, G., Wang, H., Wu, J.: Medm2g: Unifying medical multi-modal generation via cross-guided diffusion with visual invariant. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11502–11512 (2024)
27. Zhong, W., Cong, C., Wang, Z., Yan, Z., Di Ieva, A., Liu, S.: Fmm-diff: A feature mapping and merging diffusion model for mri generation with missing modality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 227–236. Springer (2025)
28. Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.: Hi-net: hybrid-fusion network for multi-modal mr image synthesis. IEEE transactions on medical imaging **39**(9), 2772–2781 (2020)