Structure-aware MRI Translation: Multi-Modal Latent Diffusion Model with Arbitrary Missing Modalities

Xinzhe Zhang[1,2,†], Junjie Liang[1,2,†], Peng Cao[1,2,3(✉)], Jinzhu Yang[1,2,3], and Osmar R. Zaiane[4]

[1] Computer Science and Engineering, Northeastern University, Shenyang, China
[2] Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, Shenyang, China
[3] National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Shenyang, China
caopeng@cse.neu.edu.cn
[4] Amii, University of Alberta, Edmonton, Alberta, Canada

Abstract. Multi-modal Magnetic Resonance Imaging (MRI) plays a crucial role in clinical diagnosis by providing complementary anatomy and pathology information. However, incomplete acquisitions remain common due to practical constraints such as cost, scan time and image corruption. Recently, the diffusion model has shown significant potential in the medical image-to-image translation task. However, most existing diffusion-based synthesis models are constrained to fixed input-output modality pairs, lacking the flexibility to handle arbitrary missing scenarios. Furthermore, these approaches inevitably sacrifice anatomical structures consistency and degrade critical texture details during generation, potentially leading to the misdiagnosis of subtle pathological patterns. To address these issues, we propose **MISA-LDM**, the first many-to-many MRI synthesis framework with modality-invariant structure awareness based on the latent diffusion model. Our approach enables the synthesis of missing modalities within a single model by utilizing any available combinations of modalities. Meanwhile, we introduce a Structure-Preserving Module (SPM) that employs a disentanglement strategy to obtain modality-invariance structural representation and use high-frequency information as a supplement. We use the anatomical priors obtained by SPM to guide the diffusion process, preserving anatomical structures integrity. Extensive experiments conducted on the BraTS2020 and BraTS2021 datasets demonstrate the superiority of our method. The result confirms the necessity of introducing more comprehensive anatomical priors for preserving generation consistency in multi-modal MRI translation. The source code is available at https://github.com/yichen-byte/misa-ldm.

Keywords: Multi-modal MRI Synthesis · Latent Diffusion Models · Structure Preservation · Medical Image Translation

[†] The two authors contribute equally to this work.

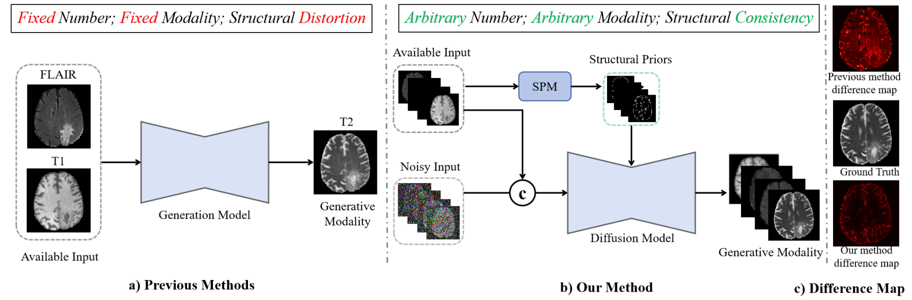

Fig. 1. Comparison between the previous synthesis methods and our proposed method.

1 Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique widely used in diagnosing and treating brain diseases. Multi-modal MRI scans, including T1-weighted (T1w), T2-weighted (T2w), contrast-enhanced T1-weighted (T1CE) and fluid Attenuated Inversion Recovery (FLAIR) images, provide complementary views of anatomical structures, each highlighting distinct soft tissue characteristics. These diverse perspectives are crucial for effective multimodal medical image segmentation and detection [16,20,21,13,26,12]. However, due to limitations in scanning time and cost, obtaining a complete set of multi-modal images can be challenging in clinical settings. To address this issue, there has been growing interest in multi-modal generative models, which leverage available data from accessible modalities to synthesize the missing modalities. Early studies for multi-modal MRI translation predominantly relied on Generative Adversarial Networks (GANs). GAN-based models [18,19,5,7] are designed to learn a generator-discriminator framework that directly maps available modalities to the missing ones. Despite their success, these methods suffer from several limitations: premature convergence, mode collapse [3], and the inability to generate high-quality, structurally consistent images.

Recently, diffusion models, with the advantages of a stable training process and superior generation quality [6], have gradually replaced GANs as the mainstream approach for generative tasks. The flexibility and stable training of the diffusion model make it more suitable for medical image synthesis [9,4,23]. For instance, Ozbey et al. [15] first proposed an adversarial diffusion model for unsupervised medical image translation. Xing et al. [24] proposed a cross-conditioned diffusion model that conditions the generation process on the distribution of target modalities. However, as shown in Fig. 1, existing diffusion generative approaches inevitably lead to the degradation of fine-grained detail and anatomical structures [8], negatively impacting the accuracy of disease diagnoses. This issue arises from the iterative process of noise addition and denoising in diffusion models, particularly during the early or late stages of denoising. The issue is further exacerbated when dealing with multi-modal data, as the correlations among different modalities are more intricate to model. Until now, research on many-to-many MRI translation based on the diffusion model has still not been explored.

To tackle the challenges associated with multi-modal MRI synthesis, we introduce MISA-LDM, a novel **M**odality-**I**nvariant **S**tructure-**A**ware MRI synthesis model based on the **L**atent **D**iffusion **M**odel (LDM) [17], as shown in Fig. 2. Unlike traditional models [10,25,18], MISA-LDM is a many-to-many synthesis model specifically designed to handle arbitrary input-output modality combinations, providing flexibility in scenarios where certain modalities may be missing. To alleviate the structure loss during the MRI synthesis, MISA-LDM incorporates a dedicated Structure-Preserving Module (SPM) that enhances the preservation of anatomical structures for improving the synthesis ability of high-quality MRI. Specifically, SPM utilizes a disentanglement strategy to extract modality-invariant structural representations, which serve as anatomical priors, guiding the generation process and ensuring that core structural features are maintained. While SPM ensures the integrity of high-level anatomical structures, it may not fully capture the fine-grained texture details necessary for accurate image synthesis. To solve it, we propose a High-Frequency Compensation Module (HFCM) for serving as a powerful complement to SPM by incorporating high-frequency information into the diffusion process, which enhances the sharpness of edges and textural patterns. This combination ensures that the model not only preserves overall anatomical consistency but also generates images with fine-grained, realistic textures. Our main contributions can be summarized as follows:

1. **Structural and anatomical consistency preserving.** A Structure-Preserving scheme is proposed to ensure the preservation of critical anatomical structures by extracting modality-invariant structural representations to serve as anatomical priors. To complement this, we propose a High-Frequency Compensation Module, which enhances the fine-grained details of generated images by incorporating high-frequency information into the diffusion process.
2. **Flexible diffusion-based multi-modal MRI translation.** MISA-LDM enables many-to-many synthesis using a diffusion model to flexibly handle arbitrary combinations of input and output modalities, which is especially advantageous in clinical scenarios.
3. We conducted experiments on two public multi-modal MRI datasets, including the BraTS2020 dataset and BraTS2021 dataset. The results show that our method can not only synthesis high-quality medical images but also better preserve texture details and anatomical structures.

2 Method

2.1 Overview of MISA-LDM

As shown in Fig. 2, MISA-LDM consists of two core components: (a) a multi-modal latent diffusion model designed for arbitrary modality combination translation, and (b) a Structure-Preserving Module (SPM) for preserving structural

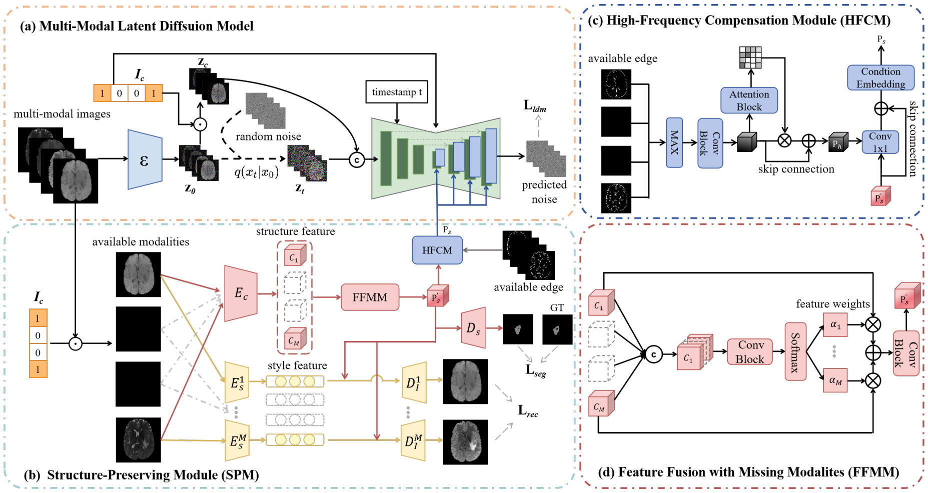

Fig. 2. The overview of our proposed MISA-LDM for multi-model MRI synthesis.

consistency. Building on LDM [17], MISA-LDM extends this paradigm to multi-modal MRI synthesis. We first train a general variational autoencoder (VAE) which consists of an encoder ε and a decoder D to establish a unified latent space across modalities. The diffusion process is then conducted in this latent space to reduce computational complexity. Fig. 2 (a) illustrates the training process of multi-modal latent diffusion model. To be specific, each modality is individually encoded into its corresponding latent representation through ε, and they are then concatenated into a unified representation z_0, which can produce its noisy version z_t at time step t. To enable our model to handle arbitrary missing scenarios, we randomly mask an arbitrary number of modalities (at least one modality remains available) as a conditional variable z_c. Our model receives z_t and z_c as inputs to predict the originally injected Gaussian noise ϵ at time step t. Furthermore, to explicitly guide the generation of missing target modalities, inspired by [17,9], we incorporate a binary indicator vector (I_c) into the diffusion model via a cross-attention mechanism. The loss function is defined as follows:

$$\mathcal{L}_{\text{ldm}} := \mathbb{E}_{\mathcal{E}(x),\epsilon\sim\mathcal{N}(0,1),t}\left[\left\|\epsilon-\epsilon_\theta\left(z_{\text{t}}\ ,z_c,I_c,\boldsymbol{P}_s,t\right)\right\|_2^2\right] \tag{1}$$

where $\epsilon_\theta(\circ,t)$ represents our neural backbone which is implemented as a time-conditional UNet. $\mathbf{P}_s$ denotes the anatomical priors described in Section 2.2.

2.2 Structure-Preserving Module

Disentanglement for coarse modality-invariant structural representation As shown in Fig. 2 (b), to preserve anatomical consistency in the synthesized images, we propose a disentanglement strategy to extract modality-invariant structural representation from the available modalities for guiding the generation process. Specifically, we propose a dual-encoder framework consisting of modality-specific style encoders and a shared structure encoder. For each modality i, a dedicated style encoder E_s^i is developed to extract its unique stylistic vector s_i, while the shared structure encoder E_c captures modality-invariant, high-level structural representation $\mathbf{C}_i$. To effectively leverage useful information, we propose a Feature Fusion with Missing Modalities (FFMM) module that dynamically adjusts feature weighting through a sigmoid gating mechanism, as illustrated in Fig. 2 (d). All structure features (with missing values zero-imputed) are integrated through FFMM into a coarse structural representation $\mathbf{P}_s^{'} = \mathrm{Conv}\left(\sum_{i=1}^{M} \alpha_i \cdot \mathbf{C}_i\right)$, preserving essential semantic information across modalities, where α_i denotes the feature weighting and M denotes the total number of modalities. To encourage successful disentanglement, we employ two distinct decoder architectures: (1) a set of modality-specific decoders D_I^i to reconstruct the available modalities from $\mathbf{P}_s^{'}$ and s_i, and (2) a segmentation decoder D_s that predicts brain tumor masks directly from $\mathbf{P}_s^{'}$. The reconstruction loss $\mathcal{L}_{rec}$ is a pixel-wise L_1 loss measuring the difference between available modalities and their reconstructed images, i.e. $\mathcal{L}_{\mathrm{rec}} = \frac{1}{\overline{M}} \sum_{i=1}^{\overline{M}} \left\| D_I^i\left(\boldsymbol{P}_s^{'}, s_i\right) - x_i \right\|_1$, where $\overline{M}$ denotes the available number of modalities. Meanwhile, a segmentation loss with Dice is applied to guide the anatomical structures extraction. The overall loss of our proposed method is formulated as: $\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{ldm} + \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{seg}$.

Supplementing with high-frequency information While the coarse modality-invariant structural representation preserves anatomical coherence in synthesized modalities, its emphasis on high-level semantics may lead to compromised fine-grained, low-level details. To address this limitation, we design a High-Frequency Compensation Module (HFCM) that explicitly preserves and enhances edge patterns, as depicted in Fig. 2 (c). Specifically, we employ the Laplacian operator [22] to extract edge information from the available modalities, and then retain the most informative high-frequency features through a max-pooling operation. Subsequently, a self-attention mechanism is constructed to focus on critical regions (e.g., lesion boundaries), establishing dynamic weight allocation for detail features. The coarse structural representation $\mathbf{P}_s^{'}$ is incorporated with the optimized high-frequency representation $\mathbf{P}_h$ via residual compensation (Eq.(2)), producing enhanced anatomical priors $\mathbf{P}_s$, i.e. $\mathbf{P}_s = \mathrm{Conv}(\mathbf{P}_s^{'} + \mathbf{P}_h) + \mathbf{P}_s^{'}$. Finally, multi-scale priors information is injected into the diffusion model's denoising process via a learnable conditional embedding layer, achieving better anatomical plausibility and detail fidelity simultaneously.

$$\mathbf{P}_h = \mathrm{Attn}(\mathrm{Conv}(\mathrm{Max}(e_1, ..., e_M))) \tag{2}$$

where e_i represents the edge map extracted from the i-th available modality, with missing values filled by zero.

Table 1. Quantitative comparison on the BraTS2020 and BraTS2021 datasets. $\overline{M}$. denotes the number of available modalities. Blue indicates the best result.

$\overline{M}$.	Methods	T1		T2		T1CE		FLAIR	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
1	MMGAN	25.379/25.032	0.920/0.919	25.468/25.571	0.918/0.913	28.128/27.351	0.930/0.921	25.096/24.172	0.908/0.895
	ReMIC	25.761/25.337	0.921/0.917	26.017/25.135	0.920/0.915	28.387/27.187	0.931/0.926	25.314/24.438	0.910/0.907
	MMT	27.587/26.820	0.939/0.929	27.214/26.987	0.928/0.922	29.048/29.128	0.940/0.935	27.008/25.731	0.925/0.913
	Ours	27.978/27.532	0.943/0.932	28.087/27.794	0.932/0.927	29.917/29.430	0.946/0.937	27.187/25.963	0.929/0.915
2	MMGAN	26.034/26.035	0.932/0.923	26.203/26.438	0.925/0.916	29.318/28.157	0.938/0.928	26.327/25.051	0.913/0.908
	ReMIC	26.181/26.417	0.930/0.921	26.882/26.567	0.926/0.919	29.575/28.651	0.940/0.931	26.158/25.607	0.914/0.913
	MMT	28.186/27.781	0.944/0.938	28.354/28.104	0.937/0.925	30.107/29.889	0.948/0.941	28.119/26.439	0.933/0.920
	Ours	28.824/28.156	0.949/0.939	29.713/28.278	0.941/0.931	31.682/30.478	0.955/0.943	28.678/27.036	0.936/0.922
3	MMGAN	27.827/27.034	0.935/0.926	28.324/27.147	0.931/0.920	30.162/30.149	0.941/0.935	27.287/26.152	0.917/0.912
	ReMIC	27.586/27.354	0.933/0.925	28.076/27.735	0.930/0.922	30.245/30.510	0.943/0.938	27.373/26.515	0.921/0.915
	MMT	29.063/28.481	0.947/0.941	29.568/29.078	0.942/0.931	31.621/31.132	0.956/0.948	28.806/27.358	0.936/0.926
	Ours	29.966/29.608	0.954/0.947	30.882/29.890	0.951/0.939	33.173/31.889	0.963/0.956	29.691/27.987	0.941/0.929

3 Experiments and Results

3.1 Datasets and Implementation

The BraTS2020[14,2] dataset includes a total of 369 multi-modal 3D brain MRI volumes. The BraTS2021 [1] dataset contains 1251 multi-modal 3D brain MRI volumes. For both datasets, each case consists of four structural MRI scans T1, T2, T1CE, and FLAIR images. During preprocessing, we slice the 3D volumes into a collection of 2D images along the z-axis (including segmentation labels), and resize each image to 256×256. We implement our model using PyTorch 2.1.0 framework with 3× NVIDIA A30 GPUs. In training phase, we train 100 epochs across all datasets with batch size of 12. The model is optimized using the Adam optimizer with a learning rate = 1e-4, decay rate = 1e-5, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. All datasets were evaluated using 5-fold cross-validation.

3.2 Comparison with State-of-the-Art Methods

Quantitative Analysis. In this section, we compare the proposed MISA-LDM with several multi-modal MRI synthesis methods including MMGAN [18], ReMIC [19] and MMT [11]. Table 1 shows the results of comparative synthesis methods on the BraTS2020 and BraTS2021 datasets. Our method achieves the highest average Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) scores on both datasets. These results quantitatively validate the effectiveness of our approach. Notably, the observed gradual improvement in image synthesis quality with increasing numbers of available modalities demonstrates our model's capacity to effectively harness complementary information across different modalities. To further demonstrate that our method ensures structural consistency, we evaluate the preservation of anatomical structures in the synthesized images with a trained segmentation model. Specifically, we trained a multi-modal segmentation model that takes T1, T2, T1CE and FLAIR images as inputs on the training set and produces tumor masks on the test set. The result is shown in the first row of Table 2. Then, we consider the scenario where a single modality is missing and synthesize the missing modality via different

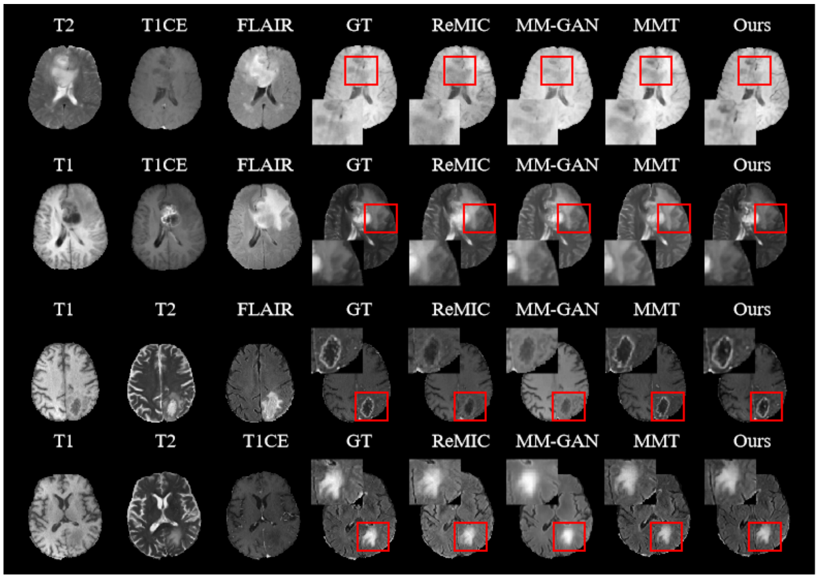

Fig. 3. Visual comparisons of proposed MISA-LDM and advanced methods.

Table 2. Tumor segment evaluation on the BraTS2020 dataset.

Methods	T1			T2			T1CE			FLAIR		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
Complete	0.939	0.915	0.889	0.939	0.915	0.889	0.939	0.915	0.889	0.939	0.915	0.889
MMGAN	0.854	0.832	0.812	0.827	0.802	0.781	0.734	0.721	0.704	0.809	0.788	0.771
ReMIC	0.839	0.816	0.797	0.851	0.842	0.827	0.742	0.723	0.710	0.821	0.806	0.779
MMT	0.901	0.871	0.846	0.908	0.894	0.863	0.824	0.808	0.783	0.886	0.862	0.834
Ours	0.915	0.895	0.873	0.913	0.892	0.873	0.846	0.823	0.809	0.883	0.856	0.838

methods. We feed the imputed multi-modality image sequences of test set into the same trained segmentation model and produce the tumor masks. We evaluate the segmentation results in terms of Dice score of the whole tumor (WT), tumor core (TC), and enhancing tumor (ET). As shown in Table 2, our MISA-LDM outperforms the other methods in terms of Dice score, indicating that it exhibits excellent ability in preserving anatomical structures during the synthesis process. Our results also suggest that preserving the structural consistency during the image synthesis is crucial for improving image quality.

Qualitative Analysis. Qualitative results are presented in Figure. 3, which compares synthetic MRI images generated by our MISA-LDM and various comparison methods on the BraTS2020 datasets. We generate each missing modality using the other three available modalities through different synthesis models and obtain the corresponding visual results. As shown in the highlighted region-of-interest (ROI) in the zoomed-in red box, our method produces more accurate brain tumor regions and more realistic tissue textures compared to other ap-

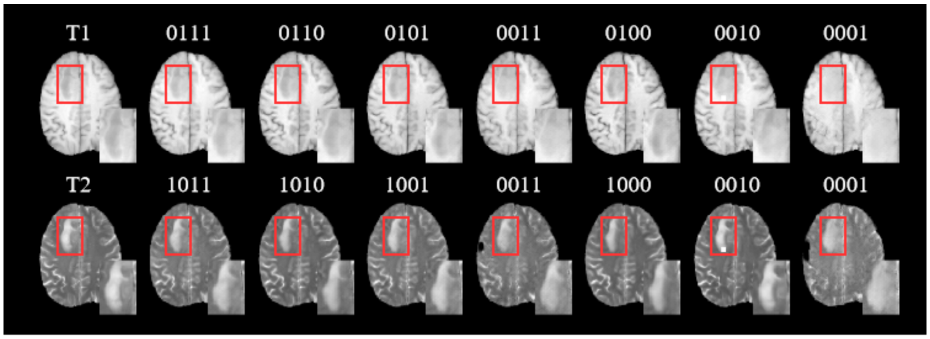

Fig. 4. Visual examples of generative images produced by MISA-LDM. The four-bit digits represent the availability of T1, T2, T1CE and FLAIR modalities, where "0" and "1" denote the "missing" and "available", respectively.

Table 3. Ablation study on BraTS2020 dataset.

Methods	T1			T1CE		
	PSNR↑	SSIM↑	DICE↑	PSNR↑	SSIM↑	DICE↑
Baseline	28.661	0.946	0.894	31.871	0.953	0.812
Baseline+SR	29.128	0.953	0.913	32.564	0.959	0.839
Baseline+HFCM	28.255	0.945	0.896	32.024	0.954	0.816
Baseline+SPM	29.966	0.954	0.915	33.173	0.963	0.846

proaches. These visual observations validate that MISA-LDM not only excels in robust cross-modality synthesis across diverse clinical scenarios but also outperforms others in preserving structural consistency. Fig. 4 illustrates MRI images synthesized by MISA-LDM under different conditions of available modalities. It can be observed that more available modalities during synthesis leads to higher-quality target modality images. The results also demonstrate that MISA-LDM efficiently exploit the complementary information in the multi-modal images.

Ablation Study. In this section, we conducted an ablation study on the BraTS2020 dataset to rigorously evaluate the individual contributions of our key technical components. Following a progressive ablation protocol, we systematically investigated three critical elements: 1) the Structure-Preserving Module (SPM), 2) the High-Frequency Compensation Module (HFCM), and 3) the corse modality-invariance Structural Representation (SR) obtained by disentanglement strategy within SPM. We mask one target modality at a time while utilizing the remaining three available modalities for imputation. Quantitative results are summarized in Table 3. We use PSNR and SSIM to evaluate image fidelity and perceptual quality, while the Dice score of the whole tumor (WT) reflects structural consistency with the ground truth. The results show that each module contributes to performance improvements, with the complete SPM achieving the best overall results, indicating its effectiveness in enhancing both synthesis quality and preserving anatomical structures.

4 Conclusion

In this work, we present MISA-LDM, a novel latent diffusion framework for flexible multi-modal MRI synthesis that can handle arbitrary combinations of missing modalities. By integrating a Structure-Preserving Module with modality-invariant structural feature disentanglement and adaptive high-frequency compensation, our method successfully preserves both global anatomical structures and fine-grained pathological details during cross-modal generation. Comprehensive experiments on BraTS2020 and BraTS2021 datasets demonstrate the superiority of our proposed MISA-LDM. Moreover, the ablation studies are conducted to verify the effectiveness of each module.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (No.62076059), the Science and Technology Joint Project of Liaoning province (2023JH2/101700367, ZX20240193). Osmar Zaiane acknowledges the funding from NSERC and the Canada CIFAR AI Chairs Program.

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

References

1. 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)
2. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)
3. Berard, H., Gidel, G., Almahairi, A., Vincent, P., Lacoste-Julien, S.: A closer look at the optimization landscapes of generative adversarial networks. arXiv preprint arXiv:1906.04848 (2019)
4. Choo, K., Jun, Y., Yun, M., Hwang, S.J.: Slice-consistent 3d volumetric brain ct-to-mri translation with 2d brownian bridge diffusion model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 657–667. Springer (2024)
5. 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)
6. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
7. 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)

8. 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)
9. Kim, J., Park, H.: Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7604–7613 (2024)
10. 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)
11. Liu, J., Pasumarthi, S., Duffy, B., Gong, E., Datta, K., Zaharchuk, G.: One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. IEEE Transactions on Medical Imaging **42**(9), 2577–2591 (2023)
12. Liu, L., Aviles-Rivero, A.I., Schönlieb, C.B.: Contrastive registration for unsupervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems (2023)
13. Maqsood, S., Damaševičius, R., Maskeliūnas, R.: Multi-modal brain tumor detection using deep neural network and multiclass svm. Medicina **58**(8), 1090 (2022)
14. 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)
15. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)
16. Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D.: Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11. pp. 455–463. Springer (2018)
17. 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)
18. 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)
19. Shen, L., Zhu, W., Wang, X., Xing, L., Pauly, J.M., Turkbey, B., Harmon, S.A., Sanford, T.H., Mehralivand, S., Choyke, P.L., et al.: Multi-domain image completion for random missing input data. IEEE transactions on medical imaging **40**(4), 1113–1122 (2020)
20. Staartjes, V.E., Seevinck, P.R., Vandertop, W.P., van Stralen, M., Schröder, M.L.: Magnetic resonance imaging–based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept. Neurosurgical focus **50**(1), E13 (2021)
21. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: Transbts: multimodal brain tumor segmentation using transformer, medical image computing and computer

assisted intervention-miccai 2021. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–119 (2021)

22. Wang, X.: Laplacian operator-based edge detectors. IEEE transactions on pattern analysis and machine intelligence **29**(5), 886–890 (2007)
23. Wang, Z., Zhang, L., Wang, L., Zhang, Z.: Soft masked mamba diffusion model for ct to mri conversion. arXiv preprint arXiv:2406.15910 (2024)
24. Xing, Z., Yang, S., Chen, S., Ye, T., Yang, Y., Qin, J., Zhu, L.: Cross-conditioned diffusion model for medical image to image translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 201–211. Springer (2024)
25. Yurt, M., Dar, S.U., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T.: mustgan: multi-stream generative adversarial networks for mr image synthesis. Medical image analysis **70**, 101944 (2021)
26. Zhao, J., Xing, Z., Chen, Z., Wan, L., Han, T., Fu, H., Zhu, L.: Uncertainty-aware multi-dimensional mutual learning for brain and brain tumor segmentation. IEEE Journal of Biomedical and Health Informatics **27**(9), 4362–4372 (2023)