

FMM-Diff: A Feature Mapping and Merging Diffusion Model for MRI Generation with Missing Modality

Wenjin Zhong¹, Cong Cong¹, Zihan Wang², Zeya Yan², Antonio Di Ieva³, and Sidong Liu^{1*}

¹ Centre for Health Informatics, Macquarie University, Sydney, Australia

² Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

³ Computational NeuroSurgery Lab, Macquarie University, Sydney, Australia
Sidong.Liu@mq.edu.au

Abstract. Brain disease diagnosis and treatment planning rely on complementary information from multiple MRI modalities. Compared to routine modalities (RM) such as T1, T2, and FLAIR, modalities like DWI and T1ce provide unique diagnostic information but are less commonly used due to longer scan times, higher costs, or the need for contrast agents. To mitigate this, multi-modal MRI synthesis methods are proposed to generate advanced MRIs from routine MRIs. However, in clinical practice, missing modality is a known issue in MRI generation which degrades the synthesis quality. Existing methods typically use shared encoders and masking strategies to compensate for missing modality. However, as the number of missing modalities increases, it becomes harder to capture the inter-modal correlations, causing a sharp performance drop. To address this, we propose the Feature Mapping and Merging Diffusion Model (FMM-Diff). Instead of using a shared encoder, we introduce dedicated mapping encoders for each modality. When a modality is missing, its latent representation is inferred from the available ones via its dedicated encoder. This ensures complete latent representations, allowing the Merge Module to selectively extract and fuse inter-modal correlations, significantly improving synthesis performance. Evaluated on two public MRI datasets, including CGGA and BraTS2021, FMM-Diff not only outperforms the state-of-the-art models by 4.35% in terms of Structural Similarity Index Measure (SSIM) while demonstrating exceptional stability, with less than a 1.0% SSIM drop, which is significantly lower than the 2.0–3.45% drop observed with other methods, across various missing modality scenarios. The source code will be available at: <https://github.com/ZJohnWenjin/FMMDIFF.git>

Keywords: Diffusion Models · Multiple Modalities · MRI Synthesis · Latent Space · Missing Modalities.

* Sidong.Liu@mq.edu.au

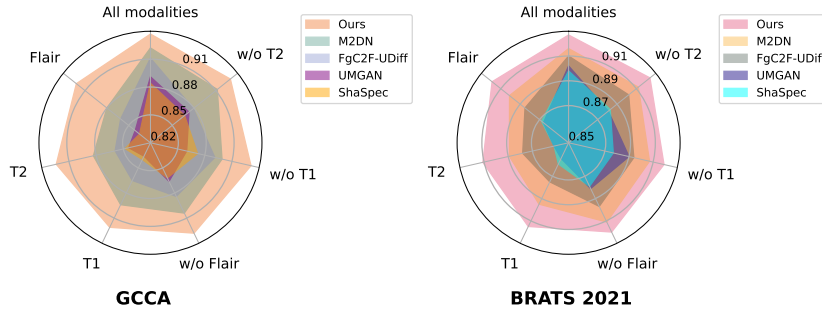


Fig. 1. SSIM comparison across different input modalities on two datasets.

1 Introduction

Magnetic Resonance Imaging (MRI) is pivotal in diagnosing brain diseases, offering various imaging sequences that accentuate distinct tissue characteristics [9,7]. Routine modalities, such as T1, T2, and FLAIR, are extensively utilised. However, advanced modalities (AM), such as Diffusion-Weighted Imaging (DWI) and T1-weighted contrast-enhanced Imaging (T1ce), provide essential diagnostic information beyond the capabilities of routine MRI (RM) [12,13]. Despite their diagnostic advantages, these advanced techniques are often costlier and more technically demanding, limiting their routine clinical application.

Recently, generative models have shown remarkable success in multi-modal MRI generation [8,21,1,16], inspiring us to explore their potential for generating AM from RM. Among them, Diffusion Models have demonstrated superior training stability and the ability to generate high-quality images [4,6,10,11,23], outperforming conventional GAN-based methods [3,21,19]. However, a key challenge in data generation is the missing modality issue. Therefore, previous models that rely on complete RM inputs experience significant performance degradation due to their inability to capture information from missing modalities. To address the challenge of missing MRI modalities, several approaches have been developed. ShaSpec [15] utilises a shared encoder to learn representations from all available modalities, effectively approximating any absent ones. Additionally, FgC2F-UDiff [17] and the Multi-Modal Modality-Masked Diffusion Network (M2DN) [10] employ modality masking strategies to enhance adaptability to incomplete inputs. Zhang *et al.* [22] propose a GAN-based model to generate missing modalities from any combination of existing ones. While effective, their performance significantly deteriorates as the number of missing modalities increases, as shown in Fig. 1. This decline occurs because the shared encoder extracts less information with each missing modality, and the masking strategy struggles to capture inter-modality correlations, leading to a sharp drop in performance.

Accordingly, we propose the Feature Mapping and Merge Diffusion Model (FMM-Diff), which comprises the Feature Mapping Module (FMM) and the Multi-Modal Feature Share and Merge Module (MFSM). The FMM includes a

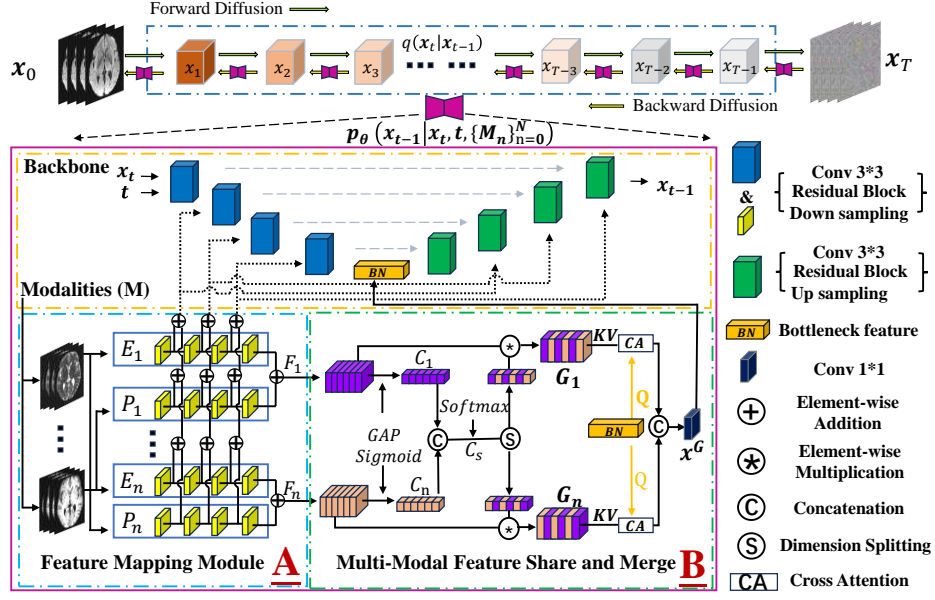


Fig. 2. Architecture of the proposed FMM-Diff model.

dedicated encoder and a mapping encoder for each input modality. When one or more modalities are missing, their mapping encoders utilise the available modalities to reconstruct their latent representations, effectively compensating for the missing information. Once the FMM obtains the latent features for all modalities, the MFSM employs an attention mechanism to efficiently extract inter-modal correlations and fuse features from different modalities accordingly. The synergy between these two modules allows FMM-Diff to remain effective and stable, even when only a single input modality is available. Our main contributions include:

- Technical innovation: we propose a novel diffusion-based model (FMM-Diff) for MRI generation with missing modalities. It comprises a Feature Mapping Module (FMM) to reconstruct latent representations of missing modalities using the existing modalities, and a Multi-Modal Feature Share and Merge (MFSM) module to effectively fuse the correlative information between modalities.
- Extensive evaluation: extensive experiments were preformed on two public datasets, i.e., CGGA and BraTS2021, and FMM-Diff’s performance was compared to four state-of-the-art (SOTA) methods across various scenarios with missing input modalities. FMM-Diff demonstrates superior performance in all tested input modality scenarios, and remains effective even when only one single input modality is available.

2 Methods

As shown in Fig. 2, FMM-Diff is a U-Net-based diffusion framework. During training, we employ a two-stage training strategy. In the first stage, we train our proposed Feature Mapping Module (FMM) (Section 2.2 (a)) using the complete set of RM. Once FMM is trained, it encodes RM as conditional input for the second stage (Section 2.2 (b)), where the standard reverse diffusion process is performed. Meanwhile, Multi-Modal Feature Share and Merge (MFSM) (Section 2.3) is employed in the bottleneck layer to model inter-modal correlations. During testing, when certain modalities are missing, FMM infers their latent representations from the available modalities, and MFSM selectively merges them based on correlation. This ensures FMM-Diff to remain highly effective for missing modality multi-modal generation of AM.

2.1 Diffusion model

Forward Diffusion Process The diffusion model [5] follows a Markov chain-based approach over T timesteps. The forward process is defined as the gradual addition of Gaussian noise to an initial AM sample, $x_{t=0}$, over T timesteps until it becomes a fully corrupted noise image, $x_{t=T}$. Formally, this process is modelled as a conditional probability distribution, $q(x_{t+1}|x_t)$, where each forward step can be explicitly expressed as:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (1)$$

Here, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, where $\alpha_t = 1 - \beta_t$ and β_t is the variance of the additive prescheduled noise at the current timestep t .

Reverse Diffusion Process The reverse process gradually reconstructs $x_{t=0}$ by iteratively denoising $x_{t=T}$ through a learned conditional distribution. In this work, FMM-Diff is employed for this reverse process. It takes the state of current timestep x_t , time embedding t , and N available modalities $\{M_n\}_{n=0}^N$ as inputs to obtain the previous state x_{t-1} with $p_\theta(x_{t-1} | x_t, t, \{M_n\}_{n=0}^N)$, where θ is the learnable parameters in the backbone model (μ_θ) and any reverse process step can be written explicitly as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \mu_\theta(x_t, t, \{M_n\}_{n=0}^N) \right) + \sqrt{1 - \alpha_t} \mu_\theta(x_t, t, \{M_n\}_{n=0}^N), \quad (2)$$

During training, $\mu_\theta(x_t, t, \{M_n\}_{n=0}^N)$ optimised by minimizing the mean squared error between the mean of predicted noise and the added noise using:

$$L_{mse} = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \mu_\theta(x_t, t, \{M_n\}_{n=0}^N)\|^2] \quad \epsilon \sim \mathcal{N}(0, I) \quad (3)$$

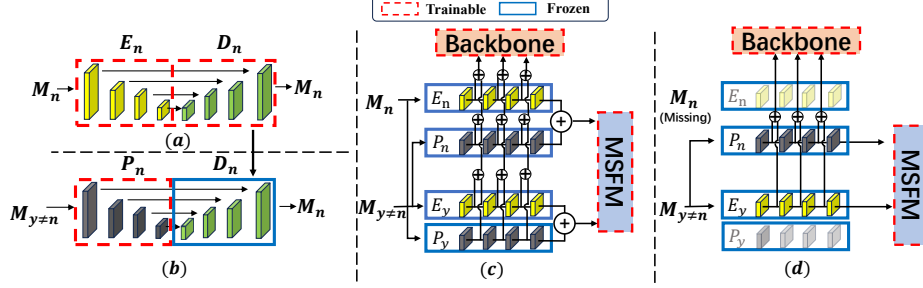


Fig. 3. FMM: Module pretraining shown as (a,b) and employing (c,d).

2.2 Feature Mapping Module

a) FMM Pretraining. As illustrated in Fig. 3, for each input modality M_n , we introduce a dedicated encoder E_n , and a mapping encoder P_n . Both share the same architectural structure as the backbone and utilize the same decoder D_n . Specifically, E_n and D_n are trained by minimizing the reconstruction error of M_n using Mean Squared Error loss (MSE):

$$L_{\text{mse}} = \mathbb{E} [\|M_n - D_n(E_n(M_n))\|^2] \quad (4)$$

To train P_n , we first freeze the parameters of D_n and use P_n to encode the remaining modalities M_y (where $y \in \{1, 2, \dots, N\}$ and $y \neq n$). Then, D_n reconstructs M_n from the features generated by P_n , as illustrated in Fig. 3 (b). This training process is formulated in Equation 5:

$$L_{\text{mse}} = \mathbb{E} [\|M_n - D_n^*(P_n(M_y))\|^2] \quad (5)$$

where D_n^* represents the dedicated encoder with frozen parameters. The above training is conducted before μ_θ 's training and it ensures that P_n encodes the M_y into the latent space of M_n even when M_n is missing.

b) FMM Deployment When all the $\{E_n\}_{n=1}^N$ and $\{P_n\}_{n=1}^N$ are fully trained, we freeze their parameters and leverage them to facilitate the training of μ_θ which is equipped with MSFM. To effectively integrate information from different modalities into the backbone, we perform element-wise summation on the outputs from each intermediate layer of $\{P_n\}_{n=1}^N$ and $\{E_n\}_{n=1}^N$ and integrate them into the corresponding layers of the backbone, as shown in Part A of Fig. 2. Moreover, the final outputs of FMM, $\{F_n\}_{n=1}^N$, are used as inputs to MSFM. During training, when all modalities are available, the features from $P_n(M_{y \neq n})$ are added to those from $E_n(M_n)$, serving as a feature enhancement (Fig. 3 (c)). In contrast, during testing, when a modality M_n is missing, P_n leverages the available modalities M_y to compensate the feature for missing modalities M_n , shown as Fig. 3 (d). This ensures that MSFM consistently receives features from complete modalities, thereby improving overall performance.

2.3 Multi-Modal Feature Share and Merge Module

Multi-Modal Feature Share and Merge Module (MSFM) is proposed to extract and integrate $\{F_n\}_{n=1}^N$ generated by FMM. MSFM not only recalibrates feature importance both within each modality and across modalities, but also leverages time-step-specific bottleneck features as queries to selectively retrieve relevant cross-modal features for effective fusion, as shown in Part B of Fig. 2.

Local Feature Aggregation. MSFM first aggregates features within each modality. Each output feature $\{F_n\}_{n=1}^N$ undergoes Global Average Pooling followed by the Sigmoid activation to compute the average channel-wise weights and map them into the range $[0, 1]$, producing local aggregated feature $\{C_n\}_{n=1}^N$.

Inter-modal Feature Aggregation. Next, to capture inter-modal relationships, $\{C_n\}_{n=1}^N$ are concatenated along the last dimension and apply a Softmax operation [18] to refine the weight distribution across modalities, forming C_s . Then, C_s is split and element-wise multiplied with their corresponding features F_n , yielding $\{G_n\}_{n=1}^N$. Finally, as the bottleneck layer b_n contains highly condensed information, we resort to cross-attention [14] which treats the bottleneck feature as a query (Q_b) to selectively extract relevant information from $\{G_n\}_{n=1}^N$:

$$x^G = \text{Cat} \left(\sum_{n=0}^N \frac{(Q_b K_{G_n}^\top)}{\sqrt{d_{K_{G_n}}}} V_{G_n} \right) \quad (6)$$

Here, **Cat** stands for concatenation. After obtaining x^G , we use a convolutional layer with a kernel size of 1 to adjust the dimensionality before integration into the bottleneck layer. Then, backbone continues the decoding process with this modality-fused information.

3 Experiments and Results

3.1 Datasets and Experimental Setup

The CGGA database [20] contains 2,426 patients, each with four modalities: T1, T2, FLAIR, and DWI (b=1000). The task involves generating DWI (AM) from T1, T2, and FLAIR (RM). The dataset is randomly divided into 1,941 patients (80%) for training and 524 patients (20%) for testing. During preprocessing, all modalities are registered to the T2-weighted images to ensure proper alignment.

The BRATS 2021 dataset [2] includes MRI scans from 1,251 glioma patients, each with four modalities: T1, T2, T1ce, and FLAIR. The task involves generating T1ce (AM) from T1, T2, and FLAIR (RM). The dataset is randomly split into 1,000 patients (80%) for training and 251 patients (20%) for testing.

Experimental Setup The models were developed using PyTorch 1.12.1 on an NVIDIA A100 GPU. We employed a diffusion process with $T = 1000$ steps, a learning rate of 3×10^{-5} , and a batch size of 4. The noise variances β ranged from $1e-4$ to 0.02, with a linear noise schedule applied. Following [17,10], model performance was evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

Table 1. Performance comparison on CGGA database and BRATS 2021 datasets.

Modalities			CGGA Database									
			ShaSpec		UMGAN		M2DN		FgC2F-UDiff		FMM-Diff	
T1	T2	FL	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
•	◦	◦	24.4	83.6	25.6	83.6	28.6	86.5	30.7	89.4	32.8	92.1
◦	•	◦	25.2	84.8	25.0	84.4	27.3	85.9	29.4	88.3	32.6	92.4
◦	◦	•	24.1	83.5	24.6	84.1	28.2	86.1	29.1	88.0	32.6	92.2
•	•	◦	27.8	86.2	27.2	86.6	29.8	88.4	31.7	90.4	33.7	92.8
◦	•	•	28.9	87.2	27.7	86.0	30.2	89.2	30.9	89.9	33.8	93.1
•	◦	•	27.7	86.7	28.6	87.3	28.8	87.9	32.2	91.2	34.0	93.0
•	•	•	29.1	88.4	29.7	89.1	32.0	91.1	32.9	92.2	34.4	93.7

Modalities			BRATS 2021									
			ShaSpec		UMGAN		M2DN		FgC2F-UDiff		FMM-Diff	
T1	T2	FL	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
•	◦	◦	28.0	86.8	27.8	86.6	30.6	88.5	32.3	90.1	33.6	91.9
◦	•	◦	26.7	85.2	26.3	85.2	30.2	87.8	30.8	88.9	32.8	91.1
◦	◦	•	26.7	85.9	26.8	85.4	29.4	87.5	31.6	89.5	33.1	91.5
•	•	◦	30.4	87.9	30.1	88.1	31.6	89.7	32.9	91.1	33.5	92.0
◦	•	•	29.8	87.7	30.8	89.0	31.2	89.4	31.7	90.7	33.1	91.9
•	◦	•	30.7	88.3	30.4	88.1	31.9	90.2	33.2	91.2	34.1	92.4
•	•	•	31.1	89.8	31.8	90.2	32.4	91.0	33.3	91.6	34.2	92.7

3.2 Experimental Results

Quantitative Results We conducted experiments on two datasets, using RM (T1, T2, and FLAIR) to generate AM (T1ce or DWI). We compared FMM-Diff with four state-of-the-art multi-modal learning approaches for missing modality generation: ShaSpec [15], UMGAN [22], M2DN [10] and FgC2F-UDiff [17]. Table 1 demonstrates that FMM-Diff achieves the best performance in all missing modality scenarios. On the CGGA dataset, FMM-Diff outperforms the best-competing models, FgC2F-UDiff, with a 1.3 dB improvement in PSNR and a 1.5% increase in SSIM when all RM modalities are available. Even in extreme cases with only one input modality is available (e.g., only FL), FMM-Diff maintains the best performance, surpassing FgC2F-UDiff by 3.5 dB in PSNR and 4.2% in SSIM. Furthermore, when transitioning from full-modality scenarios to multi-modal missing scenarios, FMM-Diff experiences an average drop of only 1.12 dB in PSNR and 1.10% in SSIM, respectively, whereas other models exhibit declines more than twice as large. Similarly, on the BraTS2021 dataset, FMM-Diff exhibits an average drop of just 0.83 dB in PSNR and 0.90% in SSIM. This stability is further illustrated in Fig. 1, where FMM-Diff remains consistently stable across various missing modality scenarios. We attribute this resilience to the integration of FMM and MFSM, which effectively compensate for missing modalities and preserve inter-modal correlations.

Qualitative Results Fig. 4 presents visualisations of the generated AM along with SSIM values using FMM-Diff. We compare results across different missing

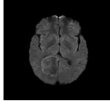
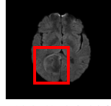
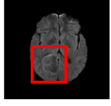
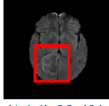
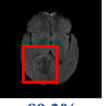
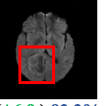
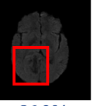
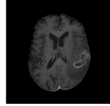
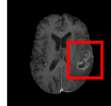
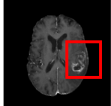
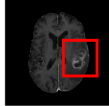
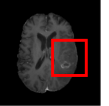
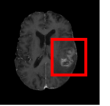
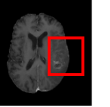
	Ground Truth	ALL Modalities		w/o T1		Only T1	
		FMM-Diff	FgC2F-UDiff	FMM-Diff	FgC2F-UDiff	FMM-Diff	FgC2F-UDiff
DWI							
		($\Delta 2.0$) 94.1%	92.1%	($\Delta 4.1$) 93.4%	89.3%	($\Delta 6.9$) 93.2%	86.3%
T1c							
		($\Delta 1.1$) 93.7%	92.6%	($\Delta 2.5$) 92.9%	90.4%	($\Delta 4.3$) 92.6%	88.3%

Fig. 4. Visualisation and SSIM comparison between FMM-Diff and FgC2F-UDiff under different missing modality scenarios.

Table 2. Ablation study on CGGA database and BRATS 2021 datasets.

Available modalities	CGGA Database (DWI)						BRATS 2021 (T1ce)					
	w/ Both		w/o FMM		w/o MSFM		w/ Both		w/o FMM		w/o MSFM	
	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%	PSNR	SSIM%
All available	34.2	93.7	32.4	92.1	31.6	90.6	34.4	92.7	32.2	90.9	32.1	90.9
w/o one modality	33.8	93.0	30.7	90.1	32.0	91.2	33.6	92.1	30.9	89.2	32.7	91.1
w/o two modality	32.7	92.2	28.8	87.0	32.1	90.9	33.2	91.5	30.4	88.3	31.4	90.5

modality scenarios, including all modalities (**All**), one missing modality (**w/o T1**), and a single available modality (**only T1**). FMM-Diff consistently produces high-quality results. As shown in Fig. 4, even as the number of input modalities decreases, FMM-Diff is still able to generate complete and detailed results in lesion regions. In contrast, FgC2F-UDiff exhibits a noticeable decline in lesion areas as the number of available RM modalities decreases.

Ablation Study We evaluated the effectiveness of each component in FMM-Diff on two datasets with different missing modalities. The average performance over different modality combinations is reported in Table 2. The results indicate that each component contributes to overall performance improvement. Specifically, FMM provides a maximum PSNR gain of 3.9 dB and an SSIM increase of 5.2%, while MFSM improves PSNR by up to 2.6 dB and an SSIM by 3.1%. Furthermore, the results indicate that as more input modalities are missing, the FMM provides greater performance gains. This suggests that the module effectively maps available modalities into the latent space of missing modalities, enabling the model to retain comprehensive input modality information.

4 Conclusion

This paper introduces FMM-Diff, a diffusion-based framework for missing modality MRI synthesis. It leverages the Feature Mapping Module to infer features of

missing modalities from available modalities and employs the Multi-Modal Feature Share and Merge Module to integrate inter-modal correlations, ensuring robust inference even in the presence of missing inputs. Experimental results demonstrate that FMM-Diff outperforms state-of-the-art models in various scenarios with missing input modalities. In conclusion, FMM-Diff serves as an effective MRI synthesis tool, generating high-quality modality-specific images without requiring full set of training MRI sequences.

Acknowledgments. This work was partially supported by an NHMRC Ideas Grant (GNT202035).

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

References

1. Azad, R., Khosravi, N., Dehghanmanshadi, M., Cohen-Adad, J., Merhof, D.: Medical image segmentation on mri images with missing modalities: A review. arXiv preprint arXiv:2203.06217 (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. Conte, G.M., Weston, A.D., Vogelsang, D.C., Philbrick, K.A., Cai, J.C., Barbera, M., Sanvito, F., Lachance, D.H., Jenkins, R.B., Tobin, W.O., et al.: Generative adversarial networks to synthesize missing t1 and flair mri sequences for use in a multisequence brain tumor segmentation model. *Radiology* **299**(2), 313–323 (2021)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **34**, 8780–8794 (2021)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
6. 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)
7. Jiang, Y., Ma, D., Keenan, K.E., Stupic, K.F., Gulani, V., Griswold, M.A.: Repeatability of magnetic resonance fingerprinting t1 and t2 estimates assessed using the ismrm/nist mri system phantom. *Magnetic resonance in medicine* **78**(4), 1452–1457 (2017)
8. Lei, X., Yu, X., Bai, M., Wu, C.: Modalities guided latent diffusion model for brain mri synthesis. In: *2024 36th Chinese Control and Decision Conference (CCDC)*. pp. 4440–4444. IEEE (2024)
9. Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J.L., Duerk, J.L., Griswold, M.A.: Magnetic resonance fingerprinting. *Nature* **495**(7440), 187–192 (2013)
10. Meng, X., Sun, K., Xu, J., He, X., Shen, D.: Multi-modal modality-masked diffusion network for brain mri synthesis with random modality missing. *IEEE Transactions on Medical Imaging* (2024)

11. Müller-Franzes, G., Niehues, J.M., Khader, F., Arasteh, S.T., Haarbuerger, C., Kuhl, C., Wang, T., Han, T., Nolte, T., Nebelung, S., et al.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports* **13**(1), 12098 (2023)
12. Parsons, M.W., Barber, P.A., Chalk, J., Darby, D.G., Rose, S., Desmond, P.M., Gerraty, R.P., Tress, B.M., Wright, P.M., Donnan, G.A., et al.: Diffusion-and perfusion-weighted mri response to thrombolysis in stroke. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **51**(1), 28–37 (2002)
13. Parsons, M.W., Yang, Q., Barber, P.A., Darby, D.G., Desmond, P.M., Gerraty, R.P., Tress, B.M., Davis, S.M.: Perfusion magnetic resonance imaging maps in hyperacute stroke: relative cerebral blood flow most accurately identifies tissue destined to infarct. *Stroke* **32**(7), 1581–1587 (2001)
14. Vaswani, A.: Attention is all you need. *Advances in Neural Information Processing Systems* (2017)
15. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: Multi-modal learning with missing modality via shared-specific feature modelling. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15878–15887 (2023)
16. Wang, J., Fan, L., Jing, W., Di, D., Song, Y., Liu, S., Cong, C.: Hypergraph tversky-aware domain incremental learning for brain tumor segmentation with missing modalities. *arXiv preprint arXiv:2505.16809* (2025)
17. Xiao, X., Hu, Q.V., Wang, G.: Fgc2f-udiff: Frequency-guided and coarse-to-fine unified diffusion model for multi-modality missing mri synthesis. *IEEE Transactions on Computational Imaging* (2024)
18. Xie, L., Li, C., Wang, Z., Zhang, X., Chen, B., Shen, Q., Wu, Z.: Shisrc-net: Super-resolution and classification network for low-resolution breast cancer histopathology image. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 23–32. Springer (2023)
19. Yurt, M., Özbey, M., Dar, S.U., Tinaz, B., Oguz, K.K., Çukur, T.: Progressively volumetrized deep generative models for data-efficient contextual learning of mr image recovery. *Medical Image Analysis* **78**, 102429 (2022)
20. Z, Z., KN, Z., Q, W., et al.: Chinese glioma genome atlas (cgga): A comprehensive resource with functional genomic data from chinese glioma patients. *Genomics, Proteomics & Bioinformatics* **19**(1), 1–12 (Feb 2021). <https://doi.org/10.1016/j.gpb.2020.10.005>
21. Zhan, B., Li, D., Wu, X., Zhou, J., Wang, Y.: Multi-modal mri image synthesis via gan with multi-scale gate mergece. *IEEE Journal of Biomedical and Health Informatics* **26**(1), 17–26 (2021)
22. Zhang, Y., Peng, C., Wang, Q., Song, D., Li, K., Zhou, S.K.: Unified multi-modal image synthesis for missing modality imputation. *IEEE Transactions on Medical Imaging* (2024)
23. Zhong, W., Cong, C., Azemi, G., Tabassum, M., Di Ieva, A., Liu, S.: Multi-sequence mri to multi-tracer pet generation via diffusion model. In: *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. pp. 1–4. IEEE (2025)