WAMS: WAVELET-DOMAIN ADAPTIVE MISSING-MODALITY SYNTHESIS FOR ARBITRARY MISSING-MODALITY MRI

Anonymous authors
Paper under double-blind review

ABSTRACT

Clinical brain MRI is often acquired with incomplete modality sets, making arbitrary missing-modality synthesis essential for robust tumor analysis. Most existing methods handle arbitrary inputs through VAE-style or shared latent representations, where available modalities are compressed into a compact space before reconstruction. Although effective in improving input flexibility, this compression-based paradigm weakens the direct modeling of image structures and limits the fidelity of synthesized modalities. We introduce Wavelet-domain Adaptive Missing-modality Synthesis (WAMS), which is the first framework to leverage wavelet transformation for arbitrary missing-modality synthesis. Instead of reconstructing MRI from a compressed latent space, WAMS decomposes multimodal MRI into low-frequency structural maps and high-frequency detail maps, where the low-frequency representation provides anatomical guidance for high-frequency detail generation. To adapt to different missing patterns, WAMS dynamically generates missing-state-aware priors according to the available modality. During synthesis, it completes only the unavailable modalities in the wavelet domain, while directly retaining the wavelet representations of already acquired modalities for inverse transformation, avoiding unnecessary re-generation of observed scans. Extensive experiments demonstrate that WAMS achieves superior synthesis quality, structural fidelity, and observed-modality preservation over existing methods.

1 INTRODUCTION

Clinical brain MRI is often acquired with incomplete modality sets. T1-weighted (T1), T2-weighted (T2), contrast-enhanced T1-weighted (T1CE), and fluid-attenuated inversion recovery (FLAIR) images provide complementary information about anatomy, edema, tumor boundary, and enhancement patterns Baid et al. (2021). However, contrast-agent contraindications, protocol variation, motion artifacts, acquisition time, and patient tolerance often result in incomplete modality sets. Since repeated acquisition is often impractical, missing-modality synthesis is important for reducing the impact of incomplete protocols and supporting robust tumor analysis from imperfect clinical scans.

A straightforward solution is to synthesize unavailable contrasts from available ones under predefined acquisition protocols. Auto-GAN aggregates a prescribed group of source modalities to synthesize a selected target Cao et al. (2020), while MM-GAN predicts absent sequences from a known input group Sharma & Hamarneh (2020). More recent diffusion-style synthesis also generates specified targets from predefined source configurations Meng et al. (2024). These methods are effective when the source-target route is known in advance. However, real clinical incompleteness is not fixed: different patients may present different observed modality subsets and different missing contrasts. Such patient-specific variability in modality availability requires synthesis models to go beyond predefined source-target protocols and infer unavailable contrasts from arbitrary observed subsets, making arbitrary missing-modality synthesis essential for robust clinical analysis under variable acquisition conditions.

To address this variability, recent studies have explored arbitrary missing-modality synthesis. ReMIC aggregates random observed subsets through multi-domain completion Shen et al. (2021),

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

whereas MMT formulates variable-subset imputation as sequence prediction Liu et al. (2023b). Structure-aware latent diffusion performs arbitrary completion in a learned generative space Zhang et al. (2025a), and efficient modality translation uses arbitrary conditioning to unify multiple translation routes Tokar & Sanner (2026). These methods make arbitrary conditioning practical, but most still rely on a compression-and-reconstruction paradigm that maps heterogeneous evidence from different observed subsets into a shared compact state. For brain-tumor MRI, this compression can weaken anatomical layout, tumor boundaries, and modality-specific contrast cues. The key challenge is therefore not only to accept arbitrary inputs, but also to preserve structural details and acquired observations during synthesis.

This challenge cannot be fully addressed by further compressing multimodal evidence into a shared latent state. Instead, it calls for a representation in which anatomical structure and local detail remain explicitly accessible during synthesis. Multiresolution wavelet analysis separates low-frequency organization from directional high-frequency variation Friedrich et al. (2024), while fixed wavelet analysis and synthesis filters provide an invertible image-coordinate system Tur et al. (2026). This property is well suited to arbitrary missing-modality synthesis, since target details must be inferred from varying source subsets and can benefit from stable low-frequency anatomical guidance. cWDM verifies wavelet diffusion in fixed cross-modality synthesis Friedrich et al. (2024), and WFM initializes wavelet flow from an informed prior Tur et al. (2026). However, their fixed mapping assumption does not hold in arbitrary missing-modality synthesis, where each case may involve a different subset of observed modalities and a different set of missing targets. Therefore, simply applying existing wavelet synthesis models to this task cannot fully exploit their structure-detail advantage. What is needed is a missing-state-adaptive wavelet formulation that adjusts structural and detail priors to the available evidence while avoiding unnecessary re-generation of acquired modalities.

Building on this insight, we introduce **Wavelet-domain Adaptive Missing-modality Synthesis (WAMS)**, the first wavelet-domain framework for arbitrary missing-modality MRI synthesis. Unlike latent reconstruction methods that first compress multimodal evidence into a compact representation, WAMS performs synthesis in the wavelet domain, where anatomical structure and local detail can be explicitly represented. Specifically, WAMS decomposes each modality into low-frequency anatomical structures and high-frequency detail components. The low-frequency representation provides a stable structural scaffold for the missing target modality, while the high-frequency branch focuses on recovering modality-specific boundaries, textures, and local contrast variations under the guidance of this structural information. To cope with different missing patterns, WAMS constructs missing-state-aware wavelet priors from the currently available modalities, allowing both structural and detail estimation to be dynamically conditioned on each observed subset. During synthesis, WAMS performs missing-only synthesis in the wavelet domain by generating unavailable modality components and directly retaining the wavelet representations of acquired modalities for inverse transformation. In this way, WAMS avoids full multimodal re-generation and reformulates arbitrary missing-modality synthesis as adaptive wavelet-domain synthesis guided by available evidence.

Our contributions are summarized as follows:

- We introduce Wavelet-domain Adaptive Missing-modality Synthesis (WAMS), the first framework that leverages wavelet transformation for arbitrary missing-modality MRI synthesis, providing a structure-preserving alternative to compression-based latent reconstruction.
- We design a low-frequency-guided high-frequency generation strategy, where anatomical structure serves as a stable scaffold for recovering modality-specific details under variable observed modality subsets.
- We develop missing-state-aware wavelet priors and a missing-only synthesis mechanism, enabling WAMS to adapt to patient-specific missing patterns while directly retaining acquired modalities for inverse transformation.

2 Related Work

2.1 Arbitrary Missing-Modality MRI Synthesis

Missing-modality MRI synthesis was first studied under predefined source-target protocols. Auto-GAN learns adversarial cross-contrast translation for a selected target modality Cao et al. (2020), and MM-GAN predicts absent sequences from known input modalities Sharma & Hamarneh (2020). Diffusion-based models provide another synthesis route, including denoising diffusion Ho et al. (2020), score-based stochastic differential equations Song et al. (2021), and latent diffusion Rombach et al. (2022). In medical imaging, these generative models have been extended to brain image generation Pinaya et al. (2022), 3D medical synthesis Khader et al. (2023), controlled MRI-to-PET translation Sargood et al. (2026), triplane-aware generation Zhang et al. (2025b), and rectified-flow synthesis Zhao et al. (2026a). Although effective, these methods usually assume predefined input-output routes and are difficult to directly apply when both observed subsets and missing targets vary across patients.

Arbitrary missing-modality synthesis addresses this flexibility by allowing one model to accept different nonempty observed subsets. ReMIC handles random missing inputs through multi-domain completion Shen et al. (2021), while MMT formulates variable-subset imputation as sequence prediction Liu et al. (2023b). Modality-masked diffusion and structure-aware latent diffusion further support arbitrary missing patterns in generative spaces Meng et al. (2024); Zhang et al. (2025a). Efficient modality translation unifies multiple routes through arbitrary conditioning Tokar & Sanner (2026), and SoftVQ-VAE provides a compact continuous-tokenizer alternative Chen et al. (2025). These methods improve input flexibility, but many shared-encoder, VAE/VQ, and latent-space designs still compress heterogeneous observations before reconstruction. This bottleneck can weaken local tumor boundaries and modality-specific details, and it does not explicitly separate measured evidence from generated targets.

Related incomplete-modality segmentation studies also show that the identity of the available subset should be modeled explicitly. M3AE uses heterogeneous-modal aggregation Liu et al. (2023a), while dual-stage missing-modality learning adapts to incomplete inputs Zhao et al. (2026b). Other methods exploit modality decoupling Wang et al. (2024), absence-aware fusion Yao et al. (2024), modality-specific encoders Dai et al. (2024), or semantic-guided masked learning Liang et al. (2025). However, these discriminative objectives do not synthesize the missing MR contrasts required here. WAMS addresses arbitrary missing-modality synthesis by completing only unavailable coefficients in invertible wavelet coordinates, while observed coefficients bypass generation under a missing-state-aware formulation.

2.2 Wavelet Transform for Image Synthesis

The discrete wavelet transform provides a fixed multiresolution analysis that separates low-frequency content from directional high-frequency subbands. Its inverse reconstructs the image without a learned decoder, making the transform useful when measured information should remain directly accessible. In MRI, low-frequency (LL) coefficients form a stable anatomical scaffold across contrasts, whereas LH, HL, and HH coefficients retain target-dependent boundaries and local texture. Frequency-guided diffusion exploits related decompositions for missing-MRI synthesis Xiao et al. (2025), and FCDiffusion controls image translation through frequency-specific components Gao et al. (2024). More directly, cWDM performs cross-modality synthesis with wavelet diffusion Friedrich et al. (2024), while WFM initializes wavelet-domain flow from an informed prior Tur et al. (2026). General flow matching provides an efficient transport view for generation Lipman et al. (2023), and recent rectified or symmetrical flow variants show that flow-based objectives can support fast image generation and transfer Liu et al. (2023c); Caetano et al. (2026).

Existing wavelet synthesis methods nevertheless assume fixed or predefined source configurations. Arbitrary completion requires a different formulation: the prior must change with the observed subset, LL anatomy must guide target-specific HF detail, and measured coefficients must remain outside the generative path. WAMS realizes these three requirements through a missing-state-aware LL/HF prior, prior-guided missing-coefficient flow, and evidence-preserving projection.

3 METHOD

3.1 PROBLEM FORMULATION

Let $X = \{x^k\}_{k=1}^K$ denote a complete multimodal MRI set, where $K = 4$ corresponds to T1, T2, T1CE, and FLAIR. In arbitrary missing-modality synthesis, only a subset of modalities is available for each case. We use a binary availability mask $\mathbf{m} \in \{0,1\}^K$ to indicate the observed subset, where $m_k = 1$ means that modality k is observed and $m_k = 0$ means that it is missing. Given the observed subset $X_{\text{obs}} = \{x^k \mid m_k = 1\}$ and the mask $\mathbf{m}$, the goal is to synthesize all unavailable modalities while preserving the acquired scans. During training, WAMS samples all 14 nontrivial observed subsets, excluding the empty-observed and all-observed cases. During inference, the model receives only X_{obs} and $\mathbf{m}$. The complete mask list and modality order are provided in the supplementary material.

WAMS formulates this task in the wavelet domain. For each modality, we apply a one-level discrete wavelet transform (DWT) $\mathcal{W}$ and reconstruct it with inverse DWT (IDWT) $\mathcal{W}^{-1}$:

$$C^k = \mathcal{W}(x^k) = \left[C^k_{LL}, C^k_{LH}, C^k_{HL}, C^k_{HH}\right], \tag{1}$$

$$x^k = \mathcal{W}^{-1}(C^k), \tag{2}$$

where C^k_{LL} denotes the low-frequency anatomical component, while C^k_{LH}, C^k_{HL}, and C^k_{HH} are grouped as the high-frequency representation C^k_{HF}, which contains directional boundary and texture information. In this work, $\mathcal{W}$ is instantiated as a one-level db4 wavelet transform. Let $C = [C_{LL}, C_{HF}]$ denote the stacked coefficients of all modalities. The modality mask $\mathbf{m}$ is expanded to a wavelet-level mask B by copying each m_k to the corresponding one low-frequency channel and three high-frequency channels, resulting in a 16-channel coefficient layout. The detailed wavelet layout and mask-broadcasting convention are provided in the supplementary material.

With $\odot$ denoting element-wise multiplication, $B \odot C$ represents the available wavelet evidence, while the components indicated by $\bar{B}$ are the synthesis targets. Thus, WAMS learns to complete only the missing wavelet components from the observed ones, rather than the entire multimodal set.

3.2 OVERVIEW OF WAMS

Based on the above formulation, arbitrary missing-modality synthesis is converted into a wavelet-domain completion problem. The model uses the observed wavelet evidence $B \odot C$ to estimate the unavailable components indicated by $\bar{B}$, while the observed components should remain unchanged. WAMS is designed to solve this problem without compressing heterogeneous multimodal evidence into a single shared latent state.

As shown in Figure 1, WAMS consists of three modules. First, the wavelet-domain structure-detail representation separates each modality into low-frequency anatomical components and high-frequency detail components. This provides an explicit representation for structure-aware synthesis and allows acquired modalities to be reconstructed through the fixed DWT–IDWT path. Second, the missing-state-aware prior generator estimates source-set-specific low-frequency and high-frequency priors. The low-frequency prior first recovers anatomical structure, which then guides high-frequency target-detail generation. Third, the prior-initialized missing-only flow refines unavailable modality components, while the evidence-preserving projection restores acquired components throughout inference.

3.3 WAVELET-DOMAIN FORMULATION

The fixed discrete wavelet transform and inverse discrete wavelet transform (DWT–IDWT) pair serves as the coordinate system for WAMS. In this coordinate system, multimodal MRI is decomposed into structure-dominant low-frequency components and detail-sensitive high-frequency components before synthesis. This allows WAMS to decouple anatomical support estimation from local detail generation, instead of mixing heterogeneous multimodal evidence in a compact latent state. The low-frequency representation provides the spatial support for missing-modality completion, while the high-frequency representation retains modality-dependent boundary, texture, and contrast variations that are refined by the subsequent generation module.

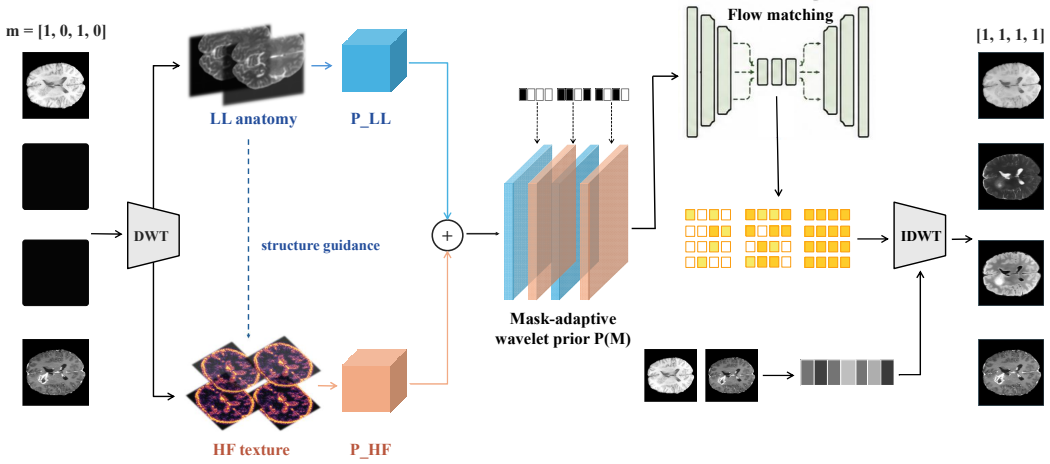


Figure 1: Overview of WAMS. Given an observed modality subset and its availability mask $\mathbf{m}$, WAMS applies discrete wavelet transform (DWT) to obtain wavelet-domain representations, constructs missing-state-aware low-frequency and high-frequency priors from the available modalities, refines unavailable modality components with a conditional flow model, and preserves acquired modalities through mask-based replacement before inverse discrete wavelet transform (IDWT).

The invertibility of the DWT–IDWT pair further enables an explicit separation between acquired evidence and synthesis targets. Given the wavelet-level mask B, components selected by B are treated as fixed measurements, whereas components selected by $\bar{B} = 1 - B$ are treated as variables to be estimated. Therefore, different missing patterns change only the mask-defined completion region, not the underlying wavelet coordinate system. This design provides a unified representation for all observed subsets and makes evidence preservation an architectural constraint rather than a reconstruction penalty imposed after generation.

3.4 Mask-Adaptive LL/HF Prior

The available evidence for a target modality changes with the observed subset. For example, synthesizing T1CE from T1 and T2 provides different structural and contrast cues from synthesizing T1CE from FLAIR alone. Therefore, WAMS constructs a missing-state-aware wavelet prior $P(\mathbf{m}) = [P_{LL}, P_{HF}]$ instead of using a fixed prior or a separate translator for each source-target route.

The prior is estimated in a low-to-high manner. The low-frequency branch first aggregates observed anatomical evidence and predicts the missing low-frequency prior:

$$P_{LL} = f_{LL}(B_{LL} \odot C_{LL}, \mathbf{m}). \tag{3}$$

This branch estimates the anatomical scaffold that is plausible under the current observed subset.

The high-frequency branch then predicts missing directional details under the guidance of the low-frequency prior:

$$P_{HF} = f_{HF}(B_{HF} \odot C_{HF}, P_{LL}, \mathbf{m}). \tag{4}$$

Conditioning on P_{LL} guides the high-frequency branch to recover target-specific boundaries, textures, and contrast transitions consistent with the predicted anatomy. This low-to-high ordering does not imply that low-frequency information alone determines the missing image. Rather, it provides a stable spatial scaffold while leaving the high-frequency branch to model modality-specific details.

The resulting prior is both subset-dependent and target-specific. Mask conditioning controls which source modalities are aggregated and which target modalities need to be estimated. Measured components and the predicted missing prior are then combined to form the flow starting point:

$$P_0 = B \odot C + \bar{B} \odot P(\mathbf{m}). \tag{5}$$

Thus, observed positions start from exact measurements, whereas unavailable positions start from source-conditioned low-frequency and high-frequency estimates.

M	Methods	T1		T2		T1CE		FLAIR	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
1	cWDM	15.261/19.335	0.681/0.818	22.319/22.473	0.831/0.836	25.931/21.070	0.877/0.818	18.938/20.539	0.750/0.806
	WFM	16.604/15.414	0.730/0.722	17.642/14.859	0.762/0.656	18.707/17.726	0.775/0.744	18.289/16.959	0.758/0.727
	SoftVQ-VAE	22.504/22.737	0.759/0.846	24.512/23.237	0.862/0.844	28.059/23.787	0.903/0.841	23.795/21.372	0.787/0.812
	DiffM4RI-LDM	15.375/17.336	0.609/0.607	17.043/17.527	0.762/0.576	19.910/15.259	0.805/0.324	15.307/14.722	0.684/0.414
	XLSTM-HVED	18.445/19.864	0.701/0.581	17.479/21.720	0.740/0.798	22.908/22.069	0.850/0.806	19.670/19.630	0.683/0.662
	Ours	**25.421/32.898**	**0.833/0.903**	**26.651/32.832**	**0.905/0.904**	**29.171/32.564**	**0.927/0.889**	**25.708/31.379**	**0.839/0.875**
2	cWDM	16.148/21.627	0.688/0.868	23.895/23.664	0.866/0.875	27.207/21.704	0.903/0.851	21.205/21.671	0.773/0.834
	WFM	18.060/16.160	0.737/0.751	17.925/15.012	0.769/0.665	19.334/19.680	0.788/0.799	19.876/18.748	0.766/0.751
	SoftVQ-VAE	22.784/23.439	0.766/0.856	24.811/23.465	0.866/0.847	28.209/24.025	0.905/0.846	24.335/21.569	0.795/0.815
	DiffM4RI-LDM	15.375/17.792	0.609/0.620	17.043/17.687	0.762/0.577	19.910/15.441	0.805/0.326	15.306/14.946	0.684/0.412
	XLSTM-HVED	18.489/20.339	0.702/0.556	17.457/22.249	0.740/0.814	23.570/22.315	0.854/0.814	20.475/19.750	0.701/0.661
	Ours	**26.818/33.146**	**0.857/0.913**	**27.663/33.212**	**0.917/0.910**	**30.571/33.054**	**0.937/0.896**	**27.532/31.825**	**0.869/0.882**
3	cWDM	15.790/22.800	0.689/0.889	25.044/23.752	0.885/0.886	28.718/21.868	0.916/0.864	20.782/21.567	0.780/0.847
	WFM	18.742/16.315	0.741/0.769	18.127/15.051	0.779/0.674	19.551/20.875	0.807/0.838	20.342/19.641	0.776/0.780
	SoftVQ-VAE	22.906/23.931	0.772/0.861	24.935/23.635	0.867/0.850	28.553/24.205	0.906/0.849	24.468/21.591	0.797/0.817
	DiffM4RI-LDM	15.375/18.087	0.609/0.631	17.043/17.754	0.762/0.576	19.910/15.551	0.805/0.328	15.307/15.148	0.684/0.417
	XLSTM-HVED	18.576/20.673	0.703/0.647	17.465/22.558	0.743/0.821	23.322/22.456	0.854/0.818	20.649/19.971	0.709/0.657
	Ours	**27.718/33.069**	**0.868/0.921**	**28.467/33.577**	**0.926/0.915**	**31.254/33.590**	**0.940/0.904**	**28.632/32.299**	**0.884/0.889**

Table 1: Paired target-only PSNR/SSIM under the LDM12 arbitrary-missing protocol. Each entry reports BraTS2021 2D/BraTS-GLI 3D. Best results are bold; second-best results are underlined.

3.5 Prior-Initialized Missing-Only Flow

The missing-state-aware prior provides an informed estimate, but it may still deviate from the target wavelet-domain distribution. WAMS refines this prior using a conditional flow model. Unlike noise-start generative models or full-channel reconstruction methods, the flow starts from P_0 and transports only unavailable components toward their target wavelet representations.

Following flow matching Lipman et al. (2023), we define a linear interpolation path between P_0 and the complete wavelet representation C:

$$C_t = (1-t)P_0 + tC, \qquad t \sim \mathcal{U}(0,1). \tag{6}$$

The supervised displacement is restricted to missing positions:

$$u = \bar{B} \odot (C - P_0). \tag{7}$$

During training, perturbation is applied only to the missing positions while measured coefficients remain fixed:

$$\widetilde{C}_t = B \odot C + \bar{B} \odot \big(C_t + \sigma(t)\epsilon\big), \qquad \epsilon \sim \mathcal{N}(0, I), \tag{8}$$

where $\sigma(t) = \sigma_{\max}\sqrt{t(1-t)+10^{-8}}$. A residual encoder–decoder predicts the conditional velocity field:

$$v_\theta = v_\theta(\widetilde{C}_t, t, \mathbf{m}, P_0), \tag{9}$$

where t, $\mathbf{m}$, and P_0 provide temporal, missing-state, and prior-start conditions. Implementation details of the velocity network, coefficient-state conditioning, and perturbation settings are provided in the supplementary material.

This prior-initialized path separates coarse estimation from distributional refinement. Since P_0 already provides source-conditioned structure and detail estimates at missing positions, the flow focuses on residual corrections toward the target coefficient distribution. Conditioning each velocity evaluation on $\mathbf{m}$ and P_0 further preserves source-set identity, allowing the same target modality to follow different refinement paths under different observed subsets.

3.6 Training Objective

The training objective combines masked velocity supervision, prior supervision, reconstruction constraints, and high-frequency regularization:

$$\begin{aligned} \mathcal{L}_{\text{flow}} &= \mathbb{E}\|\bar{B} \odot (v_\theta - u)\|_2^2 + \lambda_a \mathbb{E}\|B \odot v_\theta\|_2^2, \\ \mathcal{L} &= \mathcal{L}_{\text{flow}} + \lambda_p \mathcal{L}_{\text{prior}} + \lambda_r \mathcal{L}_{\text{rec}} + \lambda_h \mathcal{R}_{HF}. \end{aligned} \tag{10}$$

The first term in $\mathcal{L}_{\text{flow}}$ learns target-directed transport for missing components, while the second term encourages near-zero velocity on observed components. $\mathcal{L}_{\text{prior}}$ supervises the low-frequency and high-frequency priors on missing positions. $\mathcal{L}_{\text{rec}}$ constrains the refined wavelet representation,

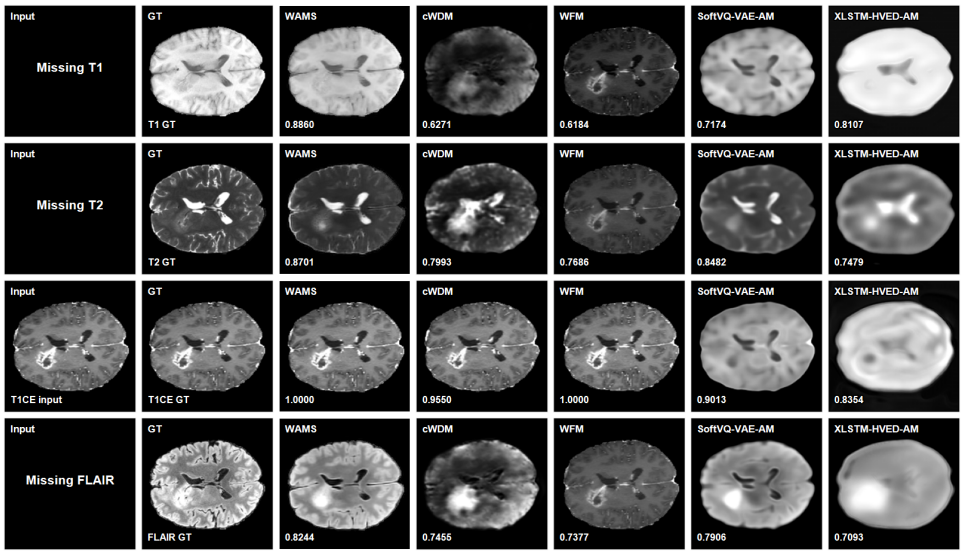


Figure 2: Qualitative comparison with only T1CE observed. T1, T2, and FLAIR are missing; SSIM is shown for each output.

and $\mathcal{R}_{HF}$ regularizes high-frequency directional details. All coefficient-wise prior and reconstruction terms are masked to unavailable modalities, keeping optimization aligned with missing-only synthesis. The active loss weights, optimization schedule, and checkpoint selection rule are listed in the supplementary material.

3.7 Evidence-Preserving Projection

During inference, WAMS starts from the prior-initialized state $Z_0 = P_0$ and integrates the learned velocity field with a 50-step Heun solver. The evidence-preserving projection is applied after both predictor and corrector updates:

$$\Pi_B(Z) = C_{\text{obs}} + \bar{B} \odot Z, \tag{11}$$

where $C_{\text{obs}} = B \odot C$ denotes the wavelet representations of acquired modalities. This projection restores all observed components to their measured values and allows only unavailable components to evolve along the generated trajectory. Therefore, even if the velocity field produces nonzero updates on observed positions, they are removed before the next sampling step.

After the final sampling step, the completed wavelet representation is obtained by

$$\widetilde{C} = \Pi_B(Z_N), \qquad \hat{X} = \mathcal{W}^{-1}(\widetilde{C}). \tag{12}$$

As a result, acquired modalities follow an exact DWT-IDWT path, while unavailable modalities are synthesized in the wavelet domain. Evidence preservation is therefore enforced as an inference constraint rather than only encouraged by a reconstruction loss. Additional implementation details of the sampler are provided in the supplementary material.

4 Experiments

4.1 Experimental Settings

BraTS2021-PT is used as the primary benchmark Baid et al. (2021). Each case contains T1, T2, T1CE, and FLAIR. The main benchmark follows the LDM12 arbitrary-missing protocol, where $M = 1, 2, 3$ denotes the number of available modalities. For every nonempty and non-complete observed subset, the model receives available modalities and synthesizes the remaining targets. PSNR and SSIM are computed only on missing modalities, while acquired modalities are evaluated separately with drift diagnostics. BraTS-GLI is used only as a paired 3D transfer check. We

Setting	Model	Train mask	PSNR	Miss. SSIM
Fixed 3-to-1	cWDM	Fixed sources	22.648	0.814
Fixed 3-to-1	WFM	Fixed sources	18.541	0.747
Same $M = 3$	WAMS	Arbitrary masks	29.018	0.905
Fixed 3-to-1	WAMS	Fixed sources	29.570	0.910

Table 2: Fixed 3-to-1 comparison on BraTS2021-PT. cWDM and WFM use fixed-source protocols. WAMS is tested with arbitrary-mask and fixed-mask training.

compare WAMS with cWDM Friedrich et al. (2024), WFM Tur et al. (2026), an MRI adaptation of SoftVQ-VAE Chen et al. (2025), DiffM4RI-LDM Zhang et al. (2025a), and XLSTM-HVED Liu et al. (2023b). All methods use fixed patient-level splits, observed-modality masks, target definitions, preprocessing, and metric code. WAMS uses one shared mask-adaptive model for all arbitrary observed subsets, with one-level db4 DWT, a 16-channel coefficient tensor, and 50-step Heun sampling. Full protocol and implementation details are provided in the supplementary material.

4.2 Comparison with State-of-the-Art Methods

Table 1 reports target-only PSNR/SSIM. Across the 12 BraTS2021 target cells, WAMS averages 27.967 dB and 0.892 SSIM, improving over the strongest listed baseline by 2.978 dB and 0.060 SSIM. The advantage is retained for $M = 1, 2, 3$ and all four target contrasts, including single-source cases. Gains on both T1/T2 and T1CE/FLAIR are consistent with the intended division between LL anatomy and HF detail. The paired BraTS-GLI results show the same overall ordering in 3D.

Table 2 separates arbitrary conditioning from fixed-source accuracy. At $M = 3$, arbitrary-mask WAMS exceeds the native fixed-source cWDM and WFM runs, while fixed-mask WAMS provides only a small additional gain. This comparison shows that one shared WAMS model covers all 14 nontrivial observed subsets without sacrificing performance in the fixed 3-to-1 setting. WAMS has 77.534M trainable parameters, compared with 81.517M for the reproduced WFM model; the fixed DWT and IDWT introduce no trainable parameters. Additional fixed-source control details are provided in the supplementary material.

4.3 Qualitative Analysis

The qualitative and diagnostic results test two properties that aggregate target scores cannot separate: structural fidelity of synthesized contrasts and stability of acquired evidence.

Figure 2 considers the challenging T1CE-only setting. WAMS yields the highest displayed SSIM for all three missing targets while preserving ventricular anatomy and the lesion region. The observed T1CE remains identical to the input at SSIM 1.0000, illustrating that target synthesis and evidence retention are handled separately.

Figure 3(a) reports all 14 observed subsets. Missing-target SSIM varies with the source-set identity, indicating that synthesis difficulty depends on both the number and identity of available modalities. WAMS remains strongest across the masks. Methods that reconstruct observed channels show nonzero drift; WAMS stays on the exact-preservation line, with maximum available-channel error 4.77×10^{-7} at numerical transform precision. This result shows that target synthesis can improve without redrawing acquired modalities. The supplementary diagnostics further compare the no-replacement variant and confirm that coefficient replacement is responsible for preserving observed evidence.

Figure 3(b) isolates the representation before generation. DWT–IDWT retains the selected anatomical boundary and local texture, whereas VAE-f8 smooths the same region. This diagnostic supports the use of an invertible wavelet-coordinate representation for detail-sensitive synthesis. Additional qualitative cases for the remaining observed-subset masks are included in the supplementary material.

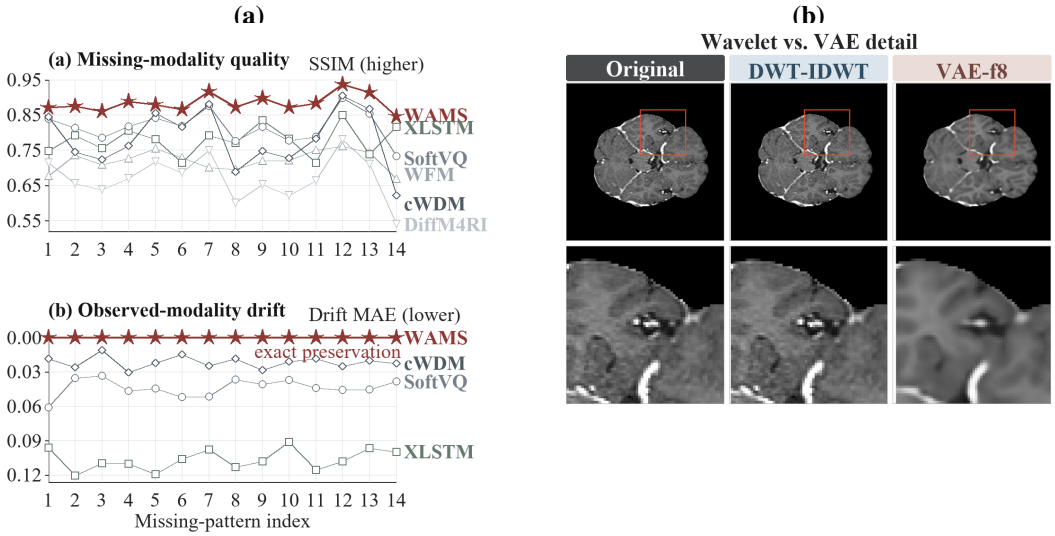


Figure 3: Diagnostic analysis. (a) Evidence preservation across 14 observed subsets: missing-target SSIM and observed-modality drift. (b) Representation-level comparison between wavelet reconstruction and learned VAE reconstruction.

Variant	Native wavelet	LL/HF split	Adaptive prior	Prior start	Missing SSIM
Image-space flow	–	–	–	✓	0.868
No LL/HF split	✓	–	✓	✓	0.871
Unified prior	✓	✓	–	✓	0.841
Zero-prior start	✓	✓	✓	–	0.849
Full WAMS	✓	✓	✓	✓	**0.905**

Table 3: Ablations on BraTS2021-PT. Each row removes one design component.

4.4 Ablation Studies

Table 3 shows that every component contributes to missing-modality synthesis. Image-space flow and removal of the LL/HF split reduce Missing SSIM to 0.868 and 0.871, while a unified prior drops it to 0.841. The unified-prior result indicates that one shared initialization cannot adequately adapt to different observed modality subsets. Zero-prior initialization reaches 0.849, confirming the value of the mask-adaptive starting point and showing that informed initialization is more effective than refining from an uninformed state. Together, the ablations support native wavelet coordinates, explicit structure-detail modeling, adaptive priors, and prior-initialized flow.

5 Conclusion

We presented WAMS, a wavelet-domain framework for arbitrary missing-modality MRI synthesis. Instead of compressing heterogeneous multimodal evidence into a shared latent state or reconstructing the entire multichannel input, WAMS formulates synthesis as evidence-preserving coefficient completion. It uses fixed wavelet coordinates to separate low-frequency anatomical structure from high-frequency directional details, constructs mask-adaptive LL/HF priors according to the observed modality subset, and refines only unavailable coefficients through prior-initialized flow. During inference, evidence-preserving projection keeps acquired coefficients outside the generated trajectory, allowing observed modalities to follow an exact DWT–IDWT path. Experiments across arbitrary observed subsets show that WAMS consistently improves target-only synthesis quality while maintaining near-zero observed-channel drift. The qualitative analysis, representation diagnosis, fixed-source comparison, and ablation studies further support the role of explicit structure–detail modeling, source-set-dependent initialization, and missing-only transport. These results demonstrate that wavelet coefficient completion provides a direct and effective alternative to full multimodal reconstruction for incomplete brain MRI.

REFERENCES

Ujjwal Baid et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. doi: 10.48550/arXiv.2107.02314.

Francisco Caetano, Christiaan Viviers, Peter H. N. de With, and Fons van der Sommen. Symmetrical flow matching: Unified image generation, segmentation, and classification with score-based generative models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 2498–2506, 2026. doi: 10.1609/aaai.v40i4.37236.

Bing Cao, Han Zhang, Nannan Wang, Xinbo Gao, and Dinggang Shen. Auto-GAN: Self-supervised collaborative learning for medical image synthesis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 10486–10493, 2020. doi: 10.1609/aaai.v34i07.6619.

Hao Chen, Ze Wang, Xiang Li, Ximeng Sun, Fangyi Chen, Jiang Liu, Jindong Wang, Bhiksha Raj, Zicheng Liu, and Emad Barsoum. SoftVQ-VAE: Efficient 1-dimensional continuous tokenizer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28358–28370, 2025.

Qi Dai, Dong Wei, Hong Liu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. Federated modality-specific encoders and multimodal anchors for personalized brain tumor segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1445–1453, 2024. doi: 10.1609/aaai.v38i2.27909.

Paul Friedrich, Alicia Durrer, Julia Wolleb, and Philippe C. Cattin. cWDM: Conditional wavelet diffusion models for cross-modality 3D medical image synthesis. *arXiv preprint arXiv:2411.17203*, 2024. doi: 10.48550/arXiv.2411.17203.

Xiang Gao, Zhengbo Xu, Junhan Zhao, and Jiaying Liu. Frequency-controlled diffusion model for versatile text-guided image-to-image translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1824–1832, 2024. doi: 10.1609/aaai.v38i3.27951.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Firas Khader et al. Denoising diffusion probabilistic models for 3D medical image generation. *Scientific Reports*, 13:7303, 2023. doi: 10.1038/s41598-023-34341-2.

Guoyan Liang, Qin Zhou, Zhe Wang, Jingyuan Chen, Lin Gu, Chang Yao, Sai Wu, Bingcang Huang, and Kai Chen. Semantic-guided masked mutual learning for multi-modal brain tumor segmentation with arbitrary missing modalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5137–5145, 2025. doi: 10.1609/aaai.v39i5.32545.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Hong Liu, Dong Wei, Donghuan Lu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. M3AE: Multimodal representation learning for brain tumor segmentation with missing modalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 1657–1665, 2023a. doi: 10.1609/aaai.v37i2.25253.

Jiang Liu, Srivathsa Pasumarthi, Ben Duffy, Enhao Gong, Keshav Datta, and Greg Zaharchuk. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. *IEEE Transactions on Medical Imaging*, 42(9):2577–2591, 2023b. doi: 10.1109/TMI.2023.3261707.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023c.

Xiangxi Meng, Kaicong Sun, Jun Xu, Xuming He, and Dinggang Shen. Multi-modal modality-masked diffusion network for brain MRI synthesis with random modality missing. *IEEE Transactions on Medical Imaging*, 43(7):2587–2598, 2024. doi: 10.1109/TMI.2024.3368664.

Walter H. L. Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F. da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Brain imaging generation with latent diffusion models. In *Deep Generative Models*, volume 13609 of *Lecture Notes in Computer Science*, pp. 117–126. Springer, 2022. doi: 10.1007/978-3-031-18576-2_12.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjoern Ommer. 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. doi: 10.1109/CVPR52688.2022.01042.

Alec Sargood, Lemuel Puglisi, James H. Cole, Neil P. Oxtoby, Daniele Ravi, and Daniel C. Alexander. CoCoLIT: Controlnet-conditioned latent image translation for MRI-to-amyloid PET synthesis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 8778–8786, 2026. doi: 10.1609/aaai.v40i11.37831.

Anmol Sharma and Ghassan Hamarneh. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Transactions on Medical Imaging*, 39(4):1170–1183, 2020. doi: 10.1109/TMI.2019.2945521.

Liyue Shen, Wentao Zhu, Xiaosong Wang, Lei Xing, John M. Pauly, Baris Turkbey, Stephanie Anne Harmon, Thomas Hogue Sanford, Sherif Mehralivand, Peter Choyke, Bradford Wood, and Daguang Xu. Multi-domain image completion for random missing input data. *IEEE Transactions on Medical Imaging*, 40(4): 1113–1122, 2021. doi: 10.1109/TMI.2020.3046444.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Tomas Tokar and Scott Sanner. Efficient modality translation via arbitrary conditioning and wasserstein regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25949–25957, 2026. doi: 10.1609/aaai.v40i31.39795.

Yalcin Tur, Mihajlo Stojkovic, and Ulas Bagci. WFM: 3D wavelet flow matching for ultrafast multi-modal MRI synthesis. In *Proceedings of the 9th International Conference on Medical Imaging with Deep Learning*, volume 315 of *Proceedings of Machine Learning Research*, pp. 3779–3796. PMLR, 2026. URL `https://proceedings.mlr.press/v315/tur26a.html`.

Hao Wang, Shengda Luo, Guosheng Hu, and Jianguo Zhang. Gradient-guided modality decoupling for missing-modality robustness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15483–15491, 2024. doi: 10.1609/aaai.v38i14.29474.

Xiaojiao Xiao, Qinmin Vivian Hu, and Guanghui Wang. FgC2F-UDiff: Frequency-guided and coarse-to-fine unified diffusion model for multi-modality missing MRI synthesis. *arXiv preprint arXiv:2501.03526*, 2025. doi: 10.48550/arXiv.2501.03526.

Wenfang Yao, Kejing Yin, William K. Cheung, Jia Liu, and Jing Qin. DrFuse: Learning disentangled representation for clinical multi-modal fusion with missing modality and modal inconsistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16416–16424, 2024. doi: 10.1609/aaai.v38i15.29578.

Xinzhe Zhang, Junjie Liang, Peng Cao, Jinzhu Yang, and Osmar R. Zaiane. Structure-aware MRI translation: Multi-modal latent diffusion model with arbitrary missing modalities. In *Medical Image Computing and Computer Assisted Intervention – MICCAI*, pp. 508–518, 2025a.

Zhenkai Zhang, Krista A. Ehinger, and Tom Drummond. TCAM-Diff: Triplane-aware cross-attention medical diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22732–22740, 2025b. doi: 10.1609/aaai.v39i21.34433.

Can Zhao, Pengfei Guo, Dong Yang, Yufan He, Yucheng Tang, Benjamin Simon, Mason Belue, Stephanie Harmon, Baris Turkbey, and Daguang Xu. MAISI-v2: Accelerated 3D high-resolution medical image synthesis with rectified flow and region-specific contrastive loss. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13088–13098, 2026a. doi: 10.1609/aaai.v40i15.38309.

Yunpeng Zhao, Cheng Chen, Qing You Pang, Yibing Fu, Quanzheng Li, Carol Tang, Beng Ti Ang, and Yueming Jin. Tackling dual-stage missing modalities in brain tumor segmentation via robust modality reconstruction and prompt-guided modality adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13314–13322, 2026b. doi: 10.1609/aaai.v40i16.38334.