No More Slice Wars: Towards Harmonized Brain MRI Synthesis for the BraSyn Challenge

Omar Carpentiero*, Kevin Marchesini*,
Costantino Grana, and Federico Bolelli ✉

University of Modena and Reggio Emilia, Italy
{*name.surname*}@unimore.it

Abstract. The synthesis of missing MRI modalities has emerged as a critical solution to address incomplete multi-parametric imaging in brain tumor diagnosis and treatment planning. While recent advances in generative models, especially GANs and diffusion-based approaches, have demonstrated promising results in cross-modality MRI generation, challenges remain in preserving anatomical fidelity and minimizing synthesis artifacts. In this work, we build upon the Hybrid Fusion GAN (HF-GAN) framework, introducing several enhancements aimed at improving synthesis quality and generalization across tumor types. Specifically, we incorporate z-score normalization, optimize network components for faster and more stable training, and extend the pipeline to support multi-view generation across various brain tumor categories, including gliomas, metastases, and meningiomas. Our approach focuses on refining 2D slice-based generation to ensure intra-slice coherence and reduce intensity inconsistencies, ultimately supporting more accurate and robust tumor segmentation in scenarios with missing imaging modalities. Our source code is available at https://github.com/AImageLab-zip/BraSyn25.

Keywords: Image Synthesis · MRI · Multimodal · BraTS · Brain Tumor Imaging · GANs · Medical Imaging

1 Introduction

In recent years, deep learning has significantly advanced medical image analysis, particularly for tasks such as segmentation and classification across various imaging modalities [8, 10, 20, 23, 39, 41]. Moreover, generative models have also emerged as a powerful technique to produce fully synthetic datasets or expand existing ones, thereby increasing data variability, mitigating class imbalance, and supporting the development of more robust and generalizable deep learning models [11, 17, 24, 33, 38]. In this context, while certain applications involve distinct anatomical structures that can be accurately analyzed using a single image modality [7, 29], many clinical scenarios require multi-modal imaging to

*Equal contribution. Authors are allowed to list their name first on their CVs.
✉ Corresponding authors: federico.bolelli@unimore.it.

effectively capture complex anatomical and pathological variations, lesion heterogeneity, and enhance tissue contrast [37]. Among the latter, the diagnosis and monitoring of brain tumors rely on multi-parametric Magnetic Resonance Imaging (MRI), considered the standard due to its superior capability in delineating tumor boundaries, quantifying tumor volumes, and guiding therapeutic decisions [4, 7]. Specifically, clinical practice typically employs four complementary MRI sequences: T1-weighted images (T1), T1-weighted images with contrast enhancement (T1c), T2-weighted images (T2), and Fluid-Attenuated Inversion Recovery (FLAIR). Each modality highlights distinct tumor sub-regions, facilitating comprehensive analysis. However, acquiring all four MRI modalities is not always feasible in clinical practice due to constraints such as differing acquisition protocols, scanner limitations, or patient-specific issues like allergies to contrast agents (in the case of T1c modality). This absence of modalities, which can compromise the accuracy of diagnostic tasks, in particular tumor segmentation, has motivated extensive research into synthesizing missing MRI modalities from available ones. Recently, generative approaches spanning from Generative Adversarial Networks (GANs) to diffusion models, have been proposed to preserve the informative characteristics of each modality [22].

GAN-based modality synthesis. GANs have been widely used for cross-modality MRI translation, yielding promising results in producing realistic missing scans. Early works focused on paired image-to-image translation, adapting state-of-the-art general frameworks, such as Pix2Pix [6,14,40,45], to the MRI domain. Several other works demonstrated that GANs can produce anatomically plausible MRI sequences, if integrated with specific losses, such as a cycle-consistency loss [12, 28], an edge-aware loss [45], a frequency loss [6], or masked versions of common losses to penalize more the errors in tumor regions [6]. Authors proved that synthetic modalities produced by GAN methods retain critical tumor information, leading to improved segmentation performance [34, 42].

Diffusion models. Diffusion models have recently emerged as a strong alternative to GANs for cross-modality MRI synthesis, offering higher fidelity via explicit likelihood modeling and gradual denoising [35]. Approaches include latent-space diffusion [21, 46], which conditions on compressed representations to save memory, and modality-masked diffusion, like M2DN [30], which treats missing channels as noise for inpainting. The second and third place teams in the BraSyn 2024 challenge [16, 18] used wavelet-domain diffusion, showing that denoising in wavelet space improves full-volume reconstruction and reduces 3D artifacts.

Hybrid and Multi-Stage Methods. Recent work explored hybrid architectures and cascades to improve synthesis quality [19, 22, 35, 36]. Hybrid Fusion GAN (HF-GAN) [22], the basis of our model, uses a hybrid generator with attention-based fusion to integrate modality-specific features, which are then mapped to the target sequence via a modality infuser. The BraSyn 2024 winner [22] extended HF-GAN with an intensity encoder for global context and a 3D Refiner to reduce artifacts and improve tumor segmentation.

In this work, we refine the HF-GAN framework by adding z-score normalization, optimizing network components, and adapting the training pipeline for

multiview generation of tumors such as gliomas, metastases, and meningiomas. We focus on improving 2D generation to produce coherent slices and reduce intra-slice artifacts like intensity discrepancies.

2 Method

2.1 Preliminaries

We adopted HF-GAN [22] as our baseline model, using a lighter 2D pipeline to improve training and inference performance. The framework consists of a generator that synthesizes 2D brain slices from preprocessed 3D volumes and a discriminator for GAN-style adversarial learning. To enable the synthesis using a unified network independently from the missing modality scenario, the generator is composed of 4 modality-specific late-fusion encoders (one for each modality), an early fusion encoder that takes as input all the available modalities, a channel attention feature fusion module, a modality infuser, and a decoder. The encoder-decoder architecture is based on a U-Net structure.

The late fusion encoders, composed of residual convolutional blocks, with SiLU activation [15] and group normalization [44], are used for modality-specific feature extraction. The early-fusion encoder, architecturally identical to the specific encoders, accepts a stacked 4-channel image to extract complementary information from all the modalities, masking the missing ones.
Then, a feature fusion module integrates global and modality-specific information, with channel attention using global average pooling and softmax. A modality-infuser, made of Transformer blocks [43], infuses information about the missing modality into the hidden space. The decoder expands the feature maps using upsampling blocks characterized by a nearest neighbor interpolation layer followed by a 3x3 2D-convolution layer to smooth the image.

As for the standard U-Net architecture, the model incorporates skip connections between corresponding layers of the encoder and decoder to preserve spatial information and facilitate gradient flow.

2.2 Dataset & Preprocessing

Data. The Brain Tumor Segmentation (BraTS) challenge series has been organized annually since 2012, providing standardized multimodal MRI datasets and benchmarks that have driven progress in AI-based brain tumor analysis [3, 27, 31]. The BraSyn-2025 dataset is based on datasets containing different tumor cases, i.e., the BraTS-GLI 2023 (Glioma, GLI), BraTS-METS 2023 (Metastasis, MET) [32], and BraTS-MEN (Meningioma, MEN) [26]. The resulting dataset contains a retrospective collection of brain tumor mpMRI (multiparametric MRI) scans acquired from multiple institutions under standard clinical conditions but with different equipment and imaging protocols, resulting in a vastly heterogeneous image quality reflecting diverse clinical practice across different institutions. The training set is composed of 1,251 complete sequences from

Table 1: MRI intensity values of the training set before and after applying the 99.5th percentile clipping and normalization strategies.

Value	Clipp.	Norm.	T1c	T1n	T2f	T2w
Max.	✗	✗	2,120,538	155,724	612,368	4,563,634
Max.	✓	✗	8,664	7,315	8,842	8,233
Avg.	✓	✗	1,066.34	781.22	510.99	673.44
Std.	✓	✗	1,301.70	944.34	769.42	804.39
Min.	✓	✓	-0.8192	-0.8273	-0.6641	-0.8372
Max.	✓	✓	5.8367	6.9189	10.8277	9.3979

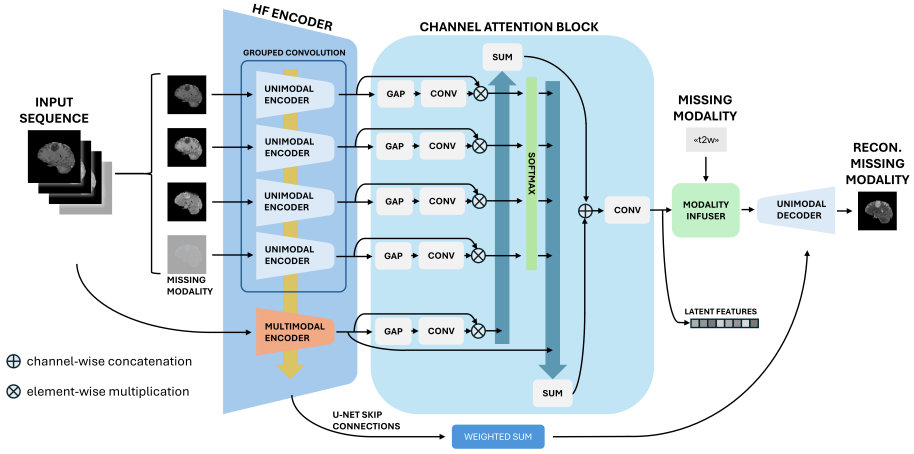
the BraTS-GLI dataset and 238 complete sequences from the BraTS-METS. All samples are annotated with a segmentation mask with 3 tumor structures: Enhancing Tumor (ET), Non-Enhancing Tumor Core (NETC), and peritumoral EDema (ED) [5]. The evaluation is always performed on an aggregation of these classes, namely Whole Tumor (WT) = ET + NETC + ED and Tumor Core (TC) = NETC + ET, as well as on the Enhancing Tumor (ET) region individually. The validation set is composed of 219 complete sequences from the BraTS-GLI dataset and 31 complete sequences from the BraTS-METS. Finally, the test set is composed of 219 complete sequences from the BraTS-GLI dataset, 59 complete sequences from the BraTS-METS, and 283 sequences from BraTS-MEN.

In contrast to previous editions, this year's challenge subtask introduces MET cases into both the training and validation sets and includes GLI, MET and MEN cases in the test set. This design aims to evaluate the models' ability to generalize across different tumor types. During the validation and test phases, ground-truth segmentation masks are not provided, and one of the four imaging modalities is randomly withheld ("modality dropout") for each subject.

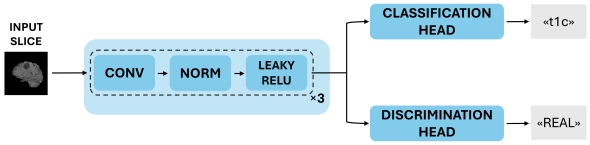
Data Preprocessing. Our contribution begins by modifying the original HF-GAN data preprocessing pipeline. Since MRI intensities are unbounded and often contain extreme outliers, we first applied intensity clipping at the 99.5th percentile after setting all negative values to zero and excluding zero-valued background pixels from the calculation. This step helps mitigate the influence of outlier voxels while preserving the meaningful dynamic range of the brain tissue. The original maximum values across modalities in the training before and after the clipping are reported in Tab. 1.

Moreover, instead of linearly projecting the data into the $[-1, 1]$ range as done in the original work [22], we adopt *dataset-wise* z*-score normalization*, computing the global mean (Avg.) and standard deviation (Std.) across all training volumes from both the GLI and MET datasets. These statistics are computed after applying a 99.5th percentile intensity clipping and are reported in Tab. 1, along with the resulting voxel intensity ranges post-normalization.

To maintain compatibility with the original HF-GAN framework, we standardized background voxel values across all modalities. Specifically, all background voxels (i.e., those originally equal to zero) were reassigned a constant value of -1, which also serves as the placeholder for masked input slices.



(a) Generator



(b) Discriminator

Fig. 1: Architecture of the proposed model.

To adapt our 2D framework to the dataset's 3D nature, we extracted axial, sagittal, and coronal slices, discarding those with fewer than 2,000 foreground pixels per modality ($< 3.47\%$ brain tissue). Sagittal and coronal slices were symmetrically padded to 240×240; padding was precomputed for training and applied on-the-fly at inference.

2.3 The Proposed Solution

As a GAN-based architecture, our framework consists of a Generator and a Discriminator, with the addition of a 2D Segmenter to guide the generation process. The generator receives three available modalities and a masked placeholder to reconstruct the missing one; this synthetic image is then combined with the originals to form a four-channel input for the tumor segmenter, which performs tumor segmentation. The resulting segmentation is used to compute a task-specific loss to enhance segmentation metrics on the reconstructed volumes. Following the typical GAN setup, the reconstructed modality is also passed to the Discriminator, which distinguishes between real and generated images, but also classifies the modality type to enforce modality-specific feature learning. The details regarding each module are reported below.

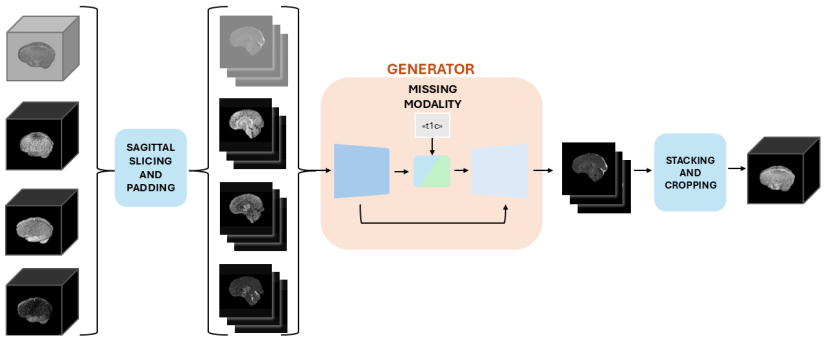


Fig. 2: Full inference pipeline. The input is split into slices, processed by the generator, then stacked to form the 3D volume. Padding and cropping are applied only to non-square sagittal slices.

Generator. Our model incorporates four modality-specific encoders and a shared fusion encoder. Each encoder consists of five downsampling stages with channel dimensions of $[64, 128, 256, 512, 640]$. Each stage includes a convolutional residual block, composed of normalization, SiLU activation, and a 3×3 convolution (stride 1, padding 1), followed by a downsampling block implemented as a 3×3 convolution with stride 2 and no padding. The fusion encoder is implemented using grouped convolutions to parallelize modality processing, replacing the previous sequential approach. Feature fusion, performed by the channel attention module is followed by a FlashAttention-compatible multi-head attention module, composed of four stages, replacing the previous custom implementation. This module is responsible for infusing modality information into the hidden space and can optionally incorporate view-specific information for multi-view generation tasks. Due to the use of z-score normalization, which preserves intensity structure while improving robustness to inter-slice variation, the intensity encoding modules were removed. The decoder mirrors the encoder with five upsampling stages, each comprising a convolutional residual block (as above) and an upsampling block using nearest-neighbor interpolation followed by a 3×3 convolution. Skip connections link each encoder stage to its corresponding decoder stage, merged by a weighted sum of the feature vectors: encoders corresponding to available modalities share the contribution equally, so that the total sum of coefficients is 1; encoders of missing modalities contribute 0, while the early-fusion encoder always contributes with full weight (1). The complete architecture of the generator is represented in Fig. 1a.

Segmenter. To guide the generation process and improve segmentation accuracy, we trained a lightweight 2D segmentation model to segment brain tumor subregions. Specifically, we adopted a nnU-Net-based architecture with four downsampling-upsampling stages and feature dimensions of [32, 64, 128, 256], employing SiLU as the activation function.

The model was trained on a subset of the training data consisting exclusively of slices containing at least 0.1% tumor tissue, thereby focusing the learning

process on informative regions. To enhance generalizability, the segmenter was trained across all anatomical views (axial, coronal, and sagittal) and tumor types (GLI, MET), with the view and tumor type information incorporated into the input via one-hot encoding. This design enabled a compact yet effective model comprising only 1.79 million parameters. The resulting model achieved robust performance, with validation Dice scores of 0.74, 0.80, and 0.82 on the NETC, ED, and ET classes, respectively. The model was trained using the Focal Tversky loss [1], due to its ability to provide better performance in contexts where the different classes are highly imbalanced. The segmenter was independently trained, and then frozen during the training of the generation pipeline.

Discriminator. The discriminator, shown in Fig. 1b, is based on the PatchGAN architecture [13], which classifies local image patches rather than the whole image. It uses three downsampling layers and starts with 32 filters. Each block includes a 4×4 convolution with stride 2 (except the last, which uses stride 1), group normalization, and LeakyReLU activations. The final convolution outputs a single-channel map for real/fake classification. An auxiliary classifier branch predicts 4 class logits per image, one for each modality.

Final Generation. To correctly reconstruct a volume from 2D slices, the generator processes batches of slices, and then the outputs stacked (Fig. 2). After stacking, the image is de-normalized, and background pixels are set to zero. Padding is removed for 3D volumes generated from the sagittal view.

To enable a faster pipeline for training and evaluation, we rely on efficient GPU implementations of image processing algorithms [2, 9]. The entire process, including data loading, preprocessing, and saving, requires, on average, 13.4 seconds per sample on an RTX5000 16GB.

Table 2: List of loss components.

Term	Role
$\mathcal{L}_{\text{recon}}$	Enforces pixel-level accuracy.
$\mathcal{L}_{\text{adv}}$	Encourages realistic outputs.
$\mathcal{L}_{\text{class}}$	Encourages modality specific features.
$\mathcal{L}_{\text{feat}}$	Promotes consistency in the latent space.
$\mathcal{L}_{\text{cycle}}$	Promotes cycle consistency.
$\mathcal{L}_{\text{SSIM}}$	Improves structural fidelity.
$\mathcal{L}_{\text{Tversky}}$	Promotes better downstream segmentation.

2.4 The loss

Our model leverages a combination of different loss functions used to guide the architecture toward learning more robust and balanced representations. The overall loss function is:

$$\mathcal{L}_{\text{total}} = 10 \cdot \mathcal{L}_{\text{recon}} + 0.25 \cdot \mathcal{L}_{\text{adv}} + 0.25 \cdot \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{feat}} + \mathcal{L}_{\text{cycle}} + 5 \cdot \mathcal{L}_{\text{SSIM}} + 5 \cdot \mathcal{L}_{\text{Tversky}}$$

Each loss targets a specific aspect of the task: a summary is reported in Tab. 2 and detailed in the following. The reconstruction loss $\mathcal{L}_{\text{recon}}$, based on the mean absolute error (MAE), serves as the dominant loss term, prioritizing pixel fidelity between the synthesized and ground-truth images. To enhance perceptual quality, we incorporate Structural Similarity Index (SSIM) losses $\mathcal{L}_{\text{SSIM}}$, computed as 1-SSIM on a per-pixel basis, and aggregated only over the regions of interest. We implement two variants of this loss: one that considers the entire

Table 3: Characterization of settings employed in different experiments. Unmentioned losses are always present.

RunID	Losses					Train		View	
	$\mathcal{L}_{\text{cycle}}$	$\mathcal{L}_{\text{feat}}$	$\mathcal{L}_{\text{SSIM\_whole}}$	$\mathcal{L}_{\text{SSIM\_dual}}$	$\mathcal{L}_{\text{Tversky}}$	GLI	MET	Axial	Sagit.
R1	✓	✓	✗	✗	✗	✓	✗	✓	✗
R2	✓	✓	✗	✓	✗	✓	✗	✓	✗
R3	✓	✓	✗	✓	✗	✓	✗	✓	✗
R4	✓	✓	✓	✗	✗	✓	✗	✓	✗
R5	✗	✗	✗	✗	✗	✓	✗	✓	✗
R6	✓	✓	✗	✓	✓	✓	✗	✓	✗
R7	✓	✓	✗	✓	✓	✓	✗	✗	✓
R8A	✓	✓	✗	✓	✗	✓	✓	✓	✓
R8S	✓	✓	✗	✓	✗	✓	✓	✓	✓
R9	✓	✓	✗	✓	✓	✓	✗	✓	✓
R10	✓	✓	✗	✓	✗	✓	✓	✓	✓

volume, and another that produces two separate terms, one for healthy brain tissue and one for tumor tissue, which are equally weighted during training. The adversarial loss $\mathcal{L}_{\text{adv}}$ is used to encourage realism in the generated outputs and to suppress common synthesis artifacts. Cycle consistency is enforced through a two-step generation process. In the first step, the generator synthesizes a missing modality using ground truth modalities. In the second step, this previously generated output replaces the corresponding placeholder, one of the available modalities is randomly masked (selected from a uniform distribution, so each maskable modality has equal probability of being chosen), and the generator is applied again using the two remaining ground truth modalities along with the generated one. The cycle loss $\mathcal{L}_{\text{cycle}}$ is defined as the MAE between the output of the second generation and the ground truth of the corresponding modality. To further promote consistency at the representational level, a feature loss $\mathcal{L}_{\text{feat}}$ is included, defined as the cosine similarity between the hidden features of the bottleneck layer extracted during the first and second generation passes. Finally, the Tversky-focal loss [1] $\mathcal{L}_{\text{Tversky}}$ is introduced to guide the generator to produce outputs that are easier to segment, particularly in the presence of class imbalance. As in the default HF-GAN framework, forward propagation is performed twice, for cycle consistency, to generate each modality, resulting in a total of 8 forward passes on the same network (not along different paths). Gradients from these multiple forwards are aggregated by combining the corresponding losses into a single scalar and performing backpropagation once.

3 Experiments & Results

3.1 Assessment Metrics

Different algorithms are assessed using a combination of image quality and segmentation metrics. The evaluation relies on multiple quantitative metrics.

Table 4: Experimental results obtained combining different losses, leveraging different training sets and 2D views. The Run IDs are explained in Tab. 3.

Metric	Val.Set	Class	Run ID										
			R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8A**	**R8S**	**R9**	**R10**
SSIM	ALL	WT	99.73	**99.77**	99.76	99.74	99.71	99.76	99.75	99.76	**99.77**	99.75	99.76
		HT	93.31	94.07	94.05	93.82	93.15	**94.09**	93.91	93.79	93.89	94.07	93.79
	GLI	WT	99.72	**99.76**	**99.76**	99.74	99.70	99.75	99.74	99.75	99.75	99.74	99.75
		HT	93.69	94.47	94.44	94.22	93.52	**94.49**	94.28	94.10	94.20	94.47	94.11
	MET	WT	99.79	99.81	99.82	99.80	99.78	99.82	99.82	**99.84**	**99.84**	99.82	**99.84**
		HT	90.62	91.23	91.23	90.94	90.52	91.29	91.27	91.59	**91.73**	91.27	91.59
DICE	ALL	ET	53.62	57.13	54.16	54.16	52.43	57.02	57.70	55.98	**57.86**	57.00	56.42
		TC	73.86	76.49	75.48	75.48	73.89	76.88	75.11	75.28	**77.55**	76.88	75.28
		WT	71.74	73.63	71.89	71.89	70.41	74.07	74.97	73.57	**74.45**	74.07	73.58
	GLI	ET	52.94	56.67	53.82	53.82	51.87	56.58	57.40	55.90	**57.14**	56.58	55.90
		TC	72.28	75.27	74.24	74.24	72.55	75.59	73.55	73.88	**76.16**	75.58	73.85
		WT	69.75	71.91	70.01	70.01	68.47	72.25	**73.26**	71.79	72.50	72.25	71.79
	MET	ET	58.44	60.36	56.51	56.51	56.40	60.11	59.77	56.52	**62.92**	59.98	60.09
		TC	85.03	85.06	84.28	84.28	83.35	86.05	86.14	85.17	**87.38**	86.00	85.33
		WT	85.78	85.83	85.15	85.15	84.16	86.93	87.03	86.13	**88.26**	86.92	86.18
NSD	ALL	ET	54.06	56.48	57.66	54.38	52.55	56.96	**58.03**	56.81	57.92	56.96	56.81
		TC	63.27	66.80	68.55	65.83	63.06	67.46	65.73	65.79	**68.30**	67.48	65.79
		WT	52.80	55.54	56.51	53.02	51.32	56.11	**57.30**	55.41	56.64	56.11	55.41
	GLI	ET	52.68	55.41	56.86	53.43	51.34	55.84	**57.08**	55.51	56.40	55.85	55.51
		TC	61.11	65.04	**66.94**	64.22	61.26	65.32	63.42	63.65	66.25	65.35	63.65
		WT	49.19	52.17	53.20	49.60	47.90	52.35	**53.79**	51.78	52.88	52.36	51.78
	MET	ET	63.82	64.05	63.25	61.12	61.10	64.87	64.74	66.00	**68.64**	64.81	66.00
		TC	78.47	79.24	79.88	77.22	75.73	82.62	82.08	80.89	**82.82**	82.58	80.89
		WT	78.30	79.34	79.91	77.15	75.48	82.65	82.13	81.12	**83.23**	82.64	81.12

Structural Similarity Index Measure (SSIM), as an image quality metric, it is employed to measure the realism of the reconstructed volume. It is independently computed on the Whole Tumor (WT) area and on the Healthy Tissue (HT) part of the brain, resulting in two scores for each test subject.

Dice score and Normalized Surface Distance (NSD) are used to evaluate the effect of reconstructed volumes on segmentation masks and boundaries. Metrics are computed for three tumor structures: Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT). Segmentation pseudo-labels for the validation set are generated using state-of-the-art algorithms from the BraTS python package [25], then compared to segmentations from sequences where one modality was masked and reconstructed.

3.2 Results

All experimental results from our analysis, on the validation set, are summarized in Tab. 4, where each run is identified by a unique identifier (e.g., R1, R2) and corresponds to a specific configuration of the generator's training process. The

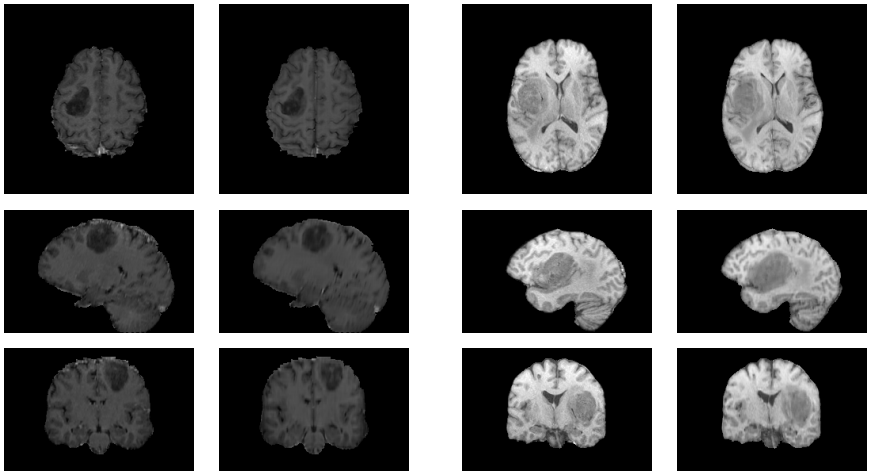
Table 5: Quantitative results on the hidden test set for the submitted best-performing run (R8S), as provided by the challenge organizers.

Dataset		DICE			NSD			SSIM
		ET	TC	WT	ET	TC	WT	Whole
GLI	Mean	74.06	80.86	90.46	53.44	46.97	47.63	93.37
	Std.	31.13	27.79	12.04	29.52	29.36	20.16	5.43
MEN	Mean	72.26	73.95	78.42	56.65	56.87	53.54	93.41
	Std.	38.54	37.14	32.28	34.75	34.54	28.02	2.05
ALL	Mean	72.34	77.50	85.29	53.47	49.54	48.34	93.33
	Std.	33.97	31.87	22.83	31.17	31.28	23.43	4.49

details of each configuration are provided in Tab. 3. All runs are trained for an equivalent number of steps, corresponding to 10 epochs on the GLI dataset, with the exception of R6 and R7. Notably, R1, R2, R4, and R6 differ from the others in how the input modalities are provided to the network, since the number of available modalities varies between 1, 2, or 3, whereas in the other runs it is fixed to 3. Runs R6 and R7 resume training from the checkpoint of R3 and are further optimized for 3 additional epochs, adding the Tversky loss. Runs R8A, R8S, and R10 share the same model, which is trained on multiple views and on the combined GLI+MET dataset, but are evaluated under three different settings: R8A generates axial views, R8S generates sagittal views, and R10 combines both views by averaging the outputs. Similarly, R9 aggregates the axial predictions from R6 and the sagittal predictions from R7. All models were evaluated on the combined GLI+MET validation set to assess their generalization capabilities across different tumor types. Tab. 5 reports the results on the hidden test set, provided by the challenge organizers. Additionally, Fig. 3 reports a comparison of the original and reconstructed slices sampled from all four modalities of random patients. The figure shows that although no 3D refinement was used, the slices show little to no striping artifacts, which are typical of 2D generation approaches.

3.3 Discussion

All of our experiments demonstrate strong MRI synthesis performance and good generalization to unseen data. In particular, models trained exclusively on the GLI dataset still perform reasonably well when tasked with generating MET volumes, highlighting the robustness of the framework. R1, even without the SSIM loss term, achieves strong perceptual quality. Subsequent runs focus on incremental improvements. Introducing the SSIM loss on R2, split into two terms for healthy and tumor tissue, improves structural similarity metrics. In contrast, R3 applies the SSIM loss over the entire volume without differentiating between tissue types, which leads to worse performance, highlighting the importance of spatially targeted perceptual losses. The original framework supports cases with 1, 2, or 3 missing modalities. By always providing three known modalities as input, convergence is faster, but final scores showed minimal variation, as evidenced by the comparison between R2, which used variable input modalities,



T1c - BraTS-GLI-01724-000 **T1 - BraTS-GLI-00333-000**

Fig. 3: Comparisons of real (left) and reconstructed (right) for two different patients sampled from the GLI validation dataset.

and R3, where three known modalities were always provided. We also experiment with removing the cycle consistency and feature losses on R5. This leads to performance degradation, confirming their contribution to training stability and reconstruction quality. However, the inclusion of cycle consistency roughly doubles training time and memory utilization. Incorporating the Tversky loss, intended to improve segmentation quality, did not yield significant gains. Similarly, using a simple mean of multiview predictions did not enhance performance. Our best-performing model, R8S, is trained using both axial and sagittal views and does not include the Tversky loss. It produces slices in the sagittal plane and benefits from the increased diversity of input representations.

4 Conclusion

We proposed an enhanced 2D MRI synthesis framework based on HF-GAN, introducing z-score normalization, architectural optimizations, and multi-view training to improve synthesis quality and generalization across tumor types. Our method achieves high structural fidelity and supports robust tumor segmentation, even with missing modalities. Results show strong cross-dataset performance, confirming the effectiveness of our design.

Acknowledgments. This work was supported by the University of Modena and Reggio Emilia and Fondazione di Modena through the "Fondo di Ateneo per la Ricerca - FAR 2024" (CUP E93C24002080007) and FARD-2024.

Disclosure of Interests. The authors have no conflicts of interest to declare.

References

1. Abraham, N., Khan, N.M.: A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. In: IEEE 16th International Symposium on Biomedical Imaging (2019)
2. Allegretti, S., Bolelli, F., Cancilla, M., Pollastri, F., Canalini, L., Grana, C.: How Does Connected Components Labeling with Decision Trees Perform on GPUs? In: International Conference on Computer Analysis of Images and Patterns (2019)
3. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F., Pati, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv preprint arXiv:2107.02314 (2021)
4. Baid, U., et al. (eds.): Brain Tumor Segmentation, and Cross-Modality Domain Adaptation for Medical Image Segmentation, Lecture Notes in Computer Science, vol. 14669 (2023)
5. Bakas, Spyridon and Akbari, Hamed and Sotiras, Aristeidis and Bilello, Michel and Rozycki, Martin and Kirby, Justin S and Freymann, John B and Farahani, Keyvan and Davatzikos, Christos: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data **4**(1) (2017)
6. Baltruschat, I.M., Janbakhshi, P., Lenga, M.: BraSyn 2023 Challenge: Missing MRI Synthesis and the Effect of Different Learning Objectives. In: International Challenge on Cross-Modality Domain Adaptation for Medical Image Segmentation (2023)
7. Bolelli, F., Marchesini, K., van Nistelrooij, N., Lumetti, L., Pipoli, V., Ficarra, E., Vinayahalingam, S., Grana, C.: Segmenting Maxillofacial Structures in CBCT Volumes. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025)
8. Bontempo, G., Bolelli, F., Porrello, A., Calderara, S., Ficarra, E.: A Graph-Based Multi-Scale Approach with Knowledge Distillation for WSI Classification. IEEE Transactions on Medical Imaging (2023)
9. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
10. Cipriano, M., Allegretti, S., Bolelli, F., Di Bartolomeo, M., Pollastri, F., Pellacani, A., Minafra, P., Anesi, A., Grana, C.: Deep Segmentation of the Mandibular Canal: a New 3D Annotated Dataset of CBCT Volumes. IEEE Access **10** (2022)
11. Cipriano, M., Allegretti, S., Bolelli, F., Pollastri, F., Grana, C.: Improving Segmentation of the Inferior Alveolar Nerve through Deep Label Propagation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)
12. Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cukur, T.: Image Synthesis in Multi-Contrast MRI With Conditional Generative Adversarial Networks. IEEE Transactions on Medical Imaging **38**(10) (2019)
13. Demir, U., Unal, G.: Patch-Based Image Inpainting with Generative Adversarial Networks. arXiv preprint arXiv:1803.07422 (2018)
14. Eker, A.G., Pehlivanoğlu, M.K., Duru, N., Duendar, T.T.: BrainPixGAN: Generating intraoperative MRI images with mask-based generative networks. Engineering Science and Technology, an International Journal **58** (2024)
15. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural networks (2018)

16. Ferreira, A., Luijten, G., Puladi, B., Kleesiek, J., Alves, V., Egger, J.: Brain Tumour Removing and Missing Modality Generation using 3D WDM. arXiv preprint arXiv:2411.04630 (2024)
17. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. Neurocomputing (2018)
18. Friedrich, P., Durrer, A., Wolleb, J., Cattin, P.C.: cWDM: Conditional Wavelet Diffusion Models for Cross-Modality 3D Medical Image Synthesis. arXiv preprint arXiv:2411.17203 (2024)
19. Hamghalam, M., Lei, B., Wang, T.: High Tissue Contrast MRI Synthesis Using Multi-Stage Attention-GAN for Segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34 (2020)
20. Huang, S.C., Pareek, A., Jensen, M., Lungren, M.P., Yeung, S., Chaudhari, A.S.: Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digital Medicine **6**(1) (2023)
21. 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 (2023)
22. Jihoon, C., Jonghye, W., Jinah, P.: A Unified Framework for Synthesizing Multi-sequence Brain MRI via Hybrid Fusion. Medical Image Analysis (2024)
23. Karargyris, A., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Wuest, A., Pati, S., Kassem, H., Zenk, M., Baid, U., et al.: Federated benchmarking of medical artificial intelligence with MedPerf. Nature machine intelligence **5**(7) (2023)
24. Kazeminia, S., et al.: GANs for Medical Image Analysis. AIME **109** (2020)
25. Kofler, F., Rosier, M., Astaraki, M., Baid, U., Möller, H., Buchner, J.A., Steinbauer, F., Oswald, E., de la Rosa, E., Ezhov, I., et al.: BraTS orchestrator : Democratizing and Disseminating state-of-the-art brain tumor image analysis. arXiv preprint arXiv:2506.13807 (2025)
26. LaBella, D., Adewole, M., Alonso-Basanta, M., Altes, T., Anwar, S.M., Baid, U., Bergquist, T., Bhalerao, R., Chen, S., Chung, V., et al.: The ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2023: Intracranial Meningioma. arXiv preprint arXiv:2305.07642 (2023)
27. Li, H.B., Conte, G.M., Hu, Q., Anwar, S.M., Kofler, F., Ezhov, I., van Leemput, K., Piraud, M., Diaz, M., Cole, B., et al.: The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn). ArXiv (2024)
28. Li, Hongwei and Paetzold, Johannes C and Sekuboyina, Anjany and Kofler, Florian and Zhang, Jianguo and Kirschke, Jan S and Wiestler, Benedikt and Menze, Bjoern: DiamondGAN: Unified Multi-modal Generative Adversarial Networks for MRI Sequences Synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2019)
29. Lumetti, L., Capitani, G., Ficarra, E., Grana, C., Calderara, S., Porrello, A., Bolelli, F.: U-Net Transplant: The Role of Pre-training for Model Merging in 3D Medical Segmentation. In: 28th International Conference on Medical Image Computing and Computer Assisted Intervention (2025)
30. 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 **43**(7) (2024)
31. Menze, B.H., Jakab, A., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10) (2014)

32. Moawad, A.W., Janas, A., Baid, U., Ramakrishnan, D., Saluja, R., Ashraf, N., Maleki, N., Jekel, L., Yordanov, N., Fehringer, P., et al.: The Brain Tumor Segmentation - Metastases (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI. ArXiv (2024)
33. Morelli, N., Marchesini, K., Lumetti, L., Santi, D., Grana, C., Bolelli, F.: Enhancing Testicular Ultrasound Image Classification Through Synthetic Data and Pretraining Strategies. In: Image Analysis and Processing – ICIAP 2025 (2025)
34. Osuala, R., Joshi, S., Tsirikoglou, A., Garrucho, L., Pinaya, W.H., Diaz, O., Lekadir, K.: Pre- to Post-Contrast Breast MRI Synthesis for Enhanced Tumour Segmentation. In: Medical Imaging 2024: Image Processing. vol. 12926 (2024)
35. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Cukur, T.: Unsupervised Medical Image Translation With Adversarial Diffusion Models. IEEE Transactions on Medical Imaging **42**(12) (2023)
36. Pan, S., Eidex, Z., Safari, M., Qiu, R., Yang, X.: Cycle-guided Denoising Diffusion Probability Model for 3D Cross-modality MRI Synthesis. In: Medical Imaging 2025: Clinical and Biomedical Imaging. vol. 13410 (2025)
37. Pipoli, V., Saporita, A., Marchesini, K., Grana, C., Ficarra, E., Bolelli, F.: IM-Fuse: A Mamba-based Fusion Block for Brain Tumor Segmentation with Incomplete Modalities. In: 28th International Conference on Medical Image Computing and Computer Assisted Intervention (2025)
38. Pollastri, F., et al.: Improving Skin Lesion Segmentation with Generative Adversarial Networks. In: IEEE 31st International Symposium on Computer-Based Medical Systems (2018)
39. Pollastri, F., et al.: Confidence Calibration for Deep Renal Biopsy Immunofluorescence Image Classification. In: 2020 25th International Conference on Pattern Recognition (2021)
40. Preetha, C.J., Meredig, H., Brugnara, G., Mahmutoglu, M.A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., et al.: Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. The Lancet Digital Health **3**(12) (2021)
41. Sistaninejhad, B., Rasi, H., Nayeri, P.: A Review Paper about Deep Learning for Medical Image Analysis. Computational and Mathematical Methods in Medicine **2023**(1) (2023)
42. Thomas, M.F., Kofler, F., Grundl, L., Finck, T., Li, H., Zimmer, C., Menze, B., Wiestler, B.: Improving Automated Glioma Segmentation in Routine Clinical Use Through Artificial Intelligence-Based Replacement of Missing Sequences With Synthetic Magnetic Resonance Imaging Scans. Investigative Radiology **57**(3) (2022)
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need. Advances in Neural Information Processing Systems **30** (2017)
44. Wu, Y., He, K.: Group Normalization. International Journal of Computer Vision (2020)
45. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis. IEEE Transactions on Medical Imaging **38**(7) (2019)
46. Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L.: Make-A-Volume: Leveraging Latent Diffusion Models for Cross-Modality 3D Brain MRI Synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2023)