Towards a Universal 3D Medical Multi-modality Generalization via Learning Personalized Invariant Representation

Supplementary Material

A. Brief descriptions of different modalities

Magnetic Resonance Imaging (MRI) scans [56]: use strong magnetic fields and radiofrequency currents yielding distinct sequences. Typically, MRI has different modalities, include T1, T2, T1ce and Flair.

Computed Tomography (CT) scans [35, 54] employ X-rays to measure its attenuation.

Positron Emission Tomography (PET) scans are expensive functional imaging scans that employ radiotracers emitting gamma rays to visualize and measure metabolic processes. Thus, PET scans have a large percentage of background areas.

B. Related work

Medical generalization tasks. Most current work focuses on homogeneous generalization, introducing tasks such as modality transfer and missing modality segmentation. The most commonly employed structural modalities — Flair, T1, T2, and T1ce of MRI — are used for brain tumor segmentation [56], or between MRI and CT [54] for modality transfer. [36] propose an approach for heterogeneous generalization in terms of modality transfer, but only tailored for transferring PET to CT.

Self-supervised medical pre-train models for medical generalization. Our approach aims to learn the $\mathbb{X}_h$ through pre-training. We list related medical pre-training work [9, 25, 42, 49] here. A notable work among them is [25], which extracts class-specific anatomical invariance. However, they only focus on a single modality. Such single-modality approaches may not be able to construct $\mathbb{X}_h$ for improving the generalization across modalities.

Alignment in multi-domain generalization. The issue of cross-modality generalization is similar to the problem of multi-domain generalization, which aims to extract domain invariant representations [16, 22, 29, 30, 41]. Most of these approaches focus on learning invariance across different domains, which may not fit the scope of personalization.

Generalization for medical translation. Typical modality transfer approaches are based on GAN models [15, 24, 28, 37, 58]. In contrast to these GAN-based approaches, some work adopts transformer models [32, 40], while others, such as [13, 27, 35, 50], explore diffusion-based approaches. The methods such as MedM2G [54] further incorporate textual information for modality transfer. Additionally, UNET-like architectures, which can also be applied to these tasks, are highlighted in [19, 20]. Most current modality transfer research focuses on improving synthesis quality. Our approach, however, demonstrates that full-modality transfer, when accompanied by specific constraints, not only enhances generation but also improves downstream generalization.

Generalization for medical segmentation. There are three main types of approaches to missing modality segmentation. Knowledge distillation-based approaches transfer knowledge from models with complete modality information (teachers) to models with missing modality information (students) [6, 45]. [14, 55] recover missing information by leveraging the multimodal latent feature space. Domain adaptation-based methods aim to reduce the gap between models with complete and incomplete modalities by aligning their domains [46]. One prominent shared latent space method, MmFormer [55], exploits intra- and inter-modality dependencies for feature fusion, which is closely related to our work. Our work reveals that our pre-train model with basic segmentation tuning exceeds these approaches.

C. Limitations, challenges, and future work

To enhance the validation of our approach, we adhere to commonly used settings during the tuning stage. Exploring alternative strategies, such as knowledge distillation, could further improve downstream performance. Our approach requires datasets where all modalities are instance-level matched, which can be a stringent condition and may be unattainable for certain modalities. Future research should explore methods to achieve personalized invariance without relying on instance-level matched datasets. Additionally, we advocate for the availability of more open-source multimodal medical datasets, particularly for functional modalities, as these are not widely accessible to researchers.

D. Social impact

This work presents an approach to tackle multi-modality generalization through personalization. We hope our work can encourage the community to work towards practical, personalized medical models with border generalization ability.

E. Downstream segmentation ablation study

The effectiveness of our proposed components is demonstrated alongside an exploration of the methodology employed to develop an individual-invariant representation. Experimental results for downstream segmentation tasks

Table 7. **Ablation study - Segmentation results of using different pre-train models** on AutoPET-II: Comparison between the pre-train models with different settings and ours. The best results are highlighted in blue and cyan.

ID	Pretrian	**DICE↑**	DICE-↑	TPR↑	TNR↑	FNR↓	FPR↓
1	+ Contrastive + Decomposition + Equivariance	40.85	55.79	81.72	69.09	18.28	30.91
2	+ Contrastive + Decomposition + Invariance	44.34	48.63	77.42	91.82	22.58	**8.18**
3	+ Contrastive + Decomposition + Equivariance + Invariance	42.42	60.67	**89.25**	63.64	**10.75**	36.36
4	+ Contrastive + Decomposition + Equivariance + $\mathbb{O}$	46.31	55.77	83.87	82.73	16.13	17.27
5	+ Contrastive + Decomposition + Invariance + $\mathbb{O}$	44.42	57.80	88.17	74.55	11.83	25.45
6	**+ Contrastive + Decomposition + Equivariance + Invariance + $\mathbb{O}$**	**48.20**	**61.16**	**88.17**	**77.27**	**11.83**	22.72

and visualizations of the pre-trained models are presented in Tab. 7. All experiments are conducted under consistent settings to ensure a fair comparison.

Using all constraints together with $\mathbb{O}$ yields the best results. Consistent with Sec. 3.1.2, the results indicate that using different constraints alone may not guarantee improvements; however, incorporating all constraints along with $\mathbb{O}$ results in the best outcomes. This validates the plausibility of the $\mathbb{X}_h$ Hypothesis and demonstrates that achieving a good approximation of it significantly enhances generalization.

Using prior $\mathbb{O}$ with decomposition constraint improves the model performance for different settings. Despite different settings, additionally using $\mathbb{O}$ with decomposition improves the downstream model performance. Combined with the improvements from modality transfer results in Tab. 4, it suggests that $\mathbb{O}$ helps in better obtaining anatomical structure.

The invariance and equivariance constraints can not be applied to the same feature. It needs to be highlighted that invariance and equivariance constraints cannot be applied to the same features as they conflict with each other. As shown in task 3, without $\mathbb{O}$, invariance and equivariance constraints are applied to the latent feature simultaneously, leading to a significant performance drop. In comparison, applying equivariance constraint before using $\mathbb{O}$ and applying the invariance constraint after using $\mathbb{O}$ yields the best results. This is because the geometrical transformation contained in z_h^i needs to be accomplished by fetching other possible geometrical transformation information from $\mathbb{O}$ and then fusing it to be invariant.

F. Experimental details

The model and data loaders are built by using MONAI `https://docs.monai.io/en/stable/index.html`. Please refer to all the details of the implementation in the code. We present some key implementations below.

F.1. Overall training procedure

A pseudo-code is provided for our approach. The loss calculation for **Pre-training** procedure is simplified as Algorithm 1 and **Downstream tuning** as Algorithm 2. It is notable that the empirical procedure is flexible as long as the $\mathbb{O}$ is properly used to construct $\mathbb{X}_h'$ and those constraints are applied to $\mathbb{X}_h'$.

F.2. Homogeneous generalization: structural modalities in MRI

F.2.1. Pre-training and Modality transfer.

Experimental settings. Four A100 GPUs are employed for training. The learning rate we used for the modality transfer is set to 0.0002, and the training epoch is set to 1000. Both the number of input and out channels is set as 4.

Training details. For the model, both the input and output channels are set to 4, corresponding to the four MRI modalities. All modalities are loaded and cropped to a size of 96 × 96 × 96 simultaneously. Following [27], we also normalize each MRI modality to have zero mean and unit variance. During training, the background is excluded for modal generation. A single modality is repeated four times to create four channels during training to obtain ${\mathbb{X}_h^i}'$. The training loss follows the $\mathcal{L}_{pre}$, whose calculation details during the training phase can be seen in Algorithm 1.

F.2.2. Missing modality segmentation.

Four A100 GPUs are employed for tuning. The learning rate we used for the modality transfer is set to 0.0002, and the training epoch is set to 1000. Both the number of input and out channels is set as 4.

Training details. Following [40], we also normalize each MRI modality to zero mean and unit variance. For the fine-tuning, we employ Dice loss, the weighted cross-entropy loss that is adopted by [40], and the additional $\mathcal{L}_{inv}$.

F.3. Heterogeneous generalization: PET and CT modalities

F.3.1. Modality transfer

All models are trained using A100 GPUs. **Training details.** All models are trained under the same situations, using the same data pre-processing transforms.

F.3.2. Downstream segmentation

Training details. All training and fine-tuning experiments use the same losses, while the approaches with our pre-train

Algorithm 1: Calculate losses during one step for pre-training

```
Data: $X \in \mathcal{X}, epoch$
Initialize learnable $\mathbb{O}$ $\mathcal{E}(\cdot), \mathcal{D}(\cdot)$;
while $i \neq epoch$ do
    $X_h' \leftarrow None$;
    for $h \in \mathcal{H}$ do
        for $i \in \mathcal{M}$ do
            $\mathcal{L}_{pre} \leftarrow 0$;
            $X_h^i \sim X, \phi^i \sim \Phi$;
            $X_h^{i+}, X_h^{i-} = Augment(\phi^i(X_h^i))$;
            $(z_h^i, x_h^i), (z_h^{i-}, x_h^{i-}), (z_h^{i+}, x_h^{i+}) \leftarrow$
             $\mathcal{E}(\phi^i(X_h^i)), \mathcal{E}(X_h^{i-}), \mathcal{E}(X_h^{i+})$;
            Calculate $\mathcal{L}_{contr}(z_h^i, z_h^{i+}, z_h^{i-})$,
             $\mathcal{L}_{pre} += \mathcal{L}_{contr}$;
            $\mathcal{F}(z_h^i) \rightarrow \phi^{i\prime}$;
            Calculate $\mathcal{L}_{equ}(\phi^{i\prime}, \phi^i)$, $\mathcal{L}_{pre} += \mathcal{L}_{equ}$;
            $z_h^{i\prime} := Attn(query : z_h^i, key :$
             $\mathbb{O}, value : \mathbb{O})$;
            $X_h^{i\prime} := Conv(z_h^{i\prime}, z_h^i)$ ;
            if $X_h'$ is not None ;   /* For saving
             memory */
             then
                Calculate $\mathcal{L}_{inv}(X_h^{i\prime}, X_h')$,
                 $\mathcal{L}_{pre} += \mathcal{L}_{inv}$;
                $X_h' := (X_h + X_h^{i\prime})/2$;
            else
                $X_h' := X_h^{i\prime}$ ;
            end
            $X_h^{i\prime} := \mathcal{D}(X_h^{i\prime}, x_h^i)$;
            Calculate $\mathcal{L}_{decom}(\phi^{i-1}(X_h^{i\prime}), X_h)$,
             $\mathcal{L}_{pre} += \mathcal{L}_{decom}$;
        end
    end
end
```

additionally use $\mathcal{L}_{inv}$ for downstream fine-tuning. Moreover, we also compare the original architecture of SwinUNETR using our pre-trained weights with fully using our architecture and our weights for fine-tuning.

F.4. Fine-tuning special case: Tuning from heterogeneous to homogeneous generalization with domain gap

Training details. For the fine-tuning stage, we use the decoder architecture of SwinUNETR, which is randomly initialized. The training procedure is similar to the above modality transfer experiments, with the primary difference being that the input and output channels are set to two. Additionally, we reproduced the results of UNETR and SwinUNETR for comparison, ensuring that the same loss functions were applied across models.

Algorithm 2: Calculate losses during one step for fine-tuning

```
Data: $(X, Y) \in (\mathcal{X}, \mathcal{Y}), epoch$
Load pre-trained $\mathbb{O}$ $\mathcal{E}(\cdot), \mathcal{D}(\cdot)$;
while $i \neq epoch$ do
    $X_h' \leftarrow None$;
    for $h \in \mathcal{H}$ do
        for $i \in \mathcal{M}$ do
            $\mathcal{L}_{down} \leftarrow 0$;
            $(X_h^i, Y_h) \sim X, Y$ ;
            $(z_h^i, x_h^i) \leftarrow \mathcal{E}(X_h^i)$;
            $z_h^{i\prime} := Attn(query : z_h^i, key :$
             $\mathbb{O}, value : \mathbb{O})$;
            $X_h^{i\prime} := Conv(z_h^{i\prime}, z_h^i)$ ;
            if $X_h'$ is not None ;   /* For saving
             memory */
             then
                Calculate $\mathcal{L}_{inv}(X_h^{i\prime}, X_h')$,
                 $\mathcal{L}_{down} += \mathcal{L}_{inv}$;
                $X_h' := (X_h + X_h^{i\prime})/2$;
            else
                $X_h' := X_h^{i\prime}$ ;
            end
            $Y_h' := \mathcal{D}(X_h^{i\prime}, x_h^i)$;
            Calculate $\mathcal{L}_{ori}(Y_h', Y_h)$,
             $\mathcal{L}_{down} += \mathcal{L}_{ori}$;
        end
    end
end
```

G. More results and analysis

G.1. More analysis on learnable biological prior

Analysis of $\mathbb{O}$. We show that using $\mathbb{O}$ for $\mathbb{X}_h$ mainly accomplishes the personalized knowledge of each sequence from MRI modalities. Those modalities are mainly focused on the physical anatomy. For the Flair modality in MRI, which mainly highlights the lesion but suppresses structures like bones, Fig. 8 shows that without $\mathbb{O}$, the main difference between the generated images and ground truth (GT) images is the personalized structure. Prior $\mathbb{O}$ for $\mathbb{X}_h$ accomplishes and refines the personal level anatomical information, mitigating the gap between them with the GT, so it can be better transferred to other structural focusing modalities.

G.2. Segmentation results on AutoPET-II.

Detailed metrics results of AutoPET-II are presented in Tab. 8. The results indicate that with proper model archi-

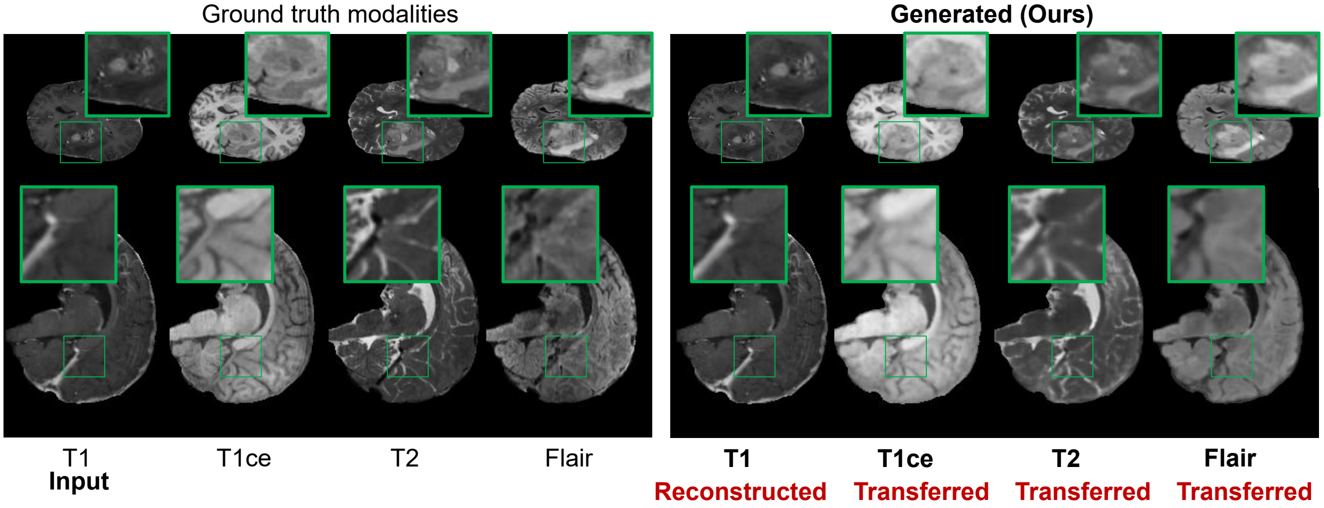


Figure 7. Visualizations of generated modalities with T1 as input of our method, which allows the capturing of subtle structures.

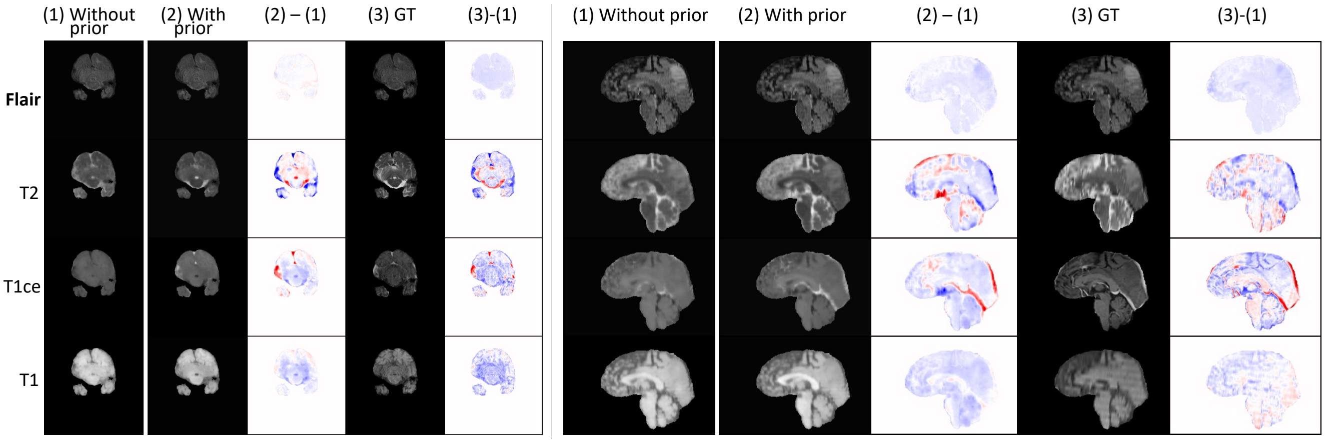


Figure 8. Visualization of the efficacy of prior $\mathbb{O}$. Displayed are the generated modalities on the input Flair modality of a testing sample on the BTATS21 dataset. Columns show: the generated images of the model (1) without prior $\mathbb{O}$ and (2) with prior $\mathbb{O}$ are aligned with (3) the GT images. Typically, the differences between without and with prior $\mathbb{O}$ (the (2)-(1) column) are visualized to compare with the differences between without $\mathbb{O}$ and GT (the (3)-(1) column). Red and blue refer to the positive (accomplishment) and negative (refinement) values of the differences, respectively.

tecture, such as SwinUNETR, using both two modalities usually outperforms solely using PET. It can be observed that models using our pre-train improve the results across all metrics. Typically, SwinUNETR using our pre-train significantly exceeds it without our pre-trained model, indicating the personalized invariant learned by our pre-train generalizes to the downstream well and can boost the downstream tasks. Moreover, using our proposed components with the pre-train leads to the best DICE and DICE-. This validates that using the prior further emphasizes the personalized invariant, which yields the most segmentation improvements.

G.3. Modality transfer results on BRATS22.

Tab. 9 and Tab. 10 presents the generation result with standard derivations. The results of our method and SwinUNETR are produced by ourselves, while the rest of the results are gathered from [27]. Generated examples are presented in Figs. 10 to 12.

G.4. Missing modality segmentation results on BRATS18.

We provide detailed segmentation results on BRATS18 as Tab. 11.

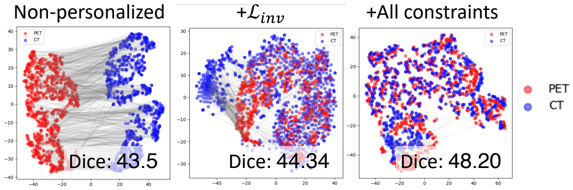
Method	**Dice↑**	Dice-↑	TPR↑	TNR↑	FNR↓	FPR↓
		From scratch				
nnUnet [23]	33.10	-	-	-	-	-
SwinUNETR [19]	43.45	**62.60**	**90.32**	62.73	**9.68**	37.27
		SwinUNETR with different pre-train				
With pre-train in [42]	44.06	57.79	**89.25**	73.64	**10.75**	26.36
With **ours**	**48.20**	**61.16**	88.17	**77.27**	11.83	**22.72**

Table 8. **Segmentation results of PET and CT** on AutoPET-II: Comparison between the previous method and ours. The best results are highlighted in blue.

Task		**T1→T2**			**T2 → Flair**		
Dimension	Method	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑
2D	Pix2Pix	24.624 ± 0.962	0.109 ± 0.028	0.874 ± 0.015	24.361 ± 1.061	0.117 ± 0.021	0.846 ± 0.019
	CycleGAN	23.535 ± 1.334	0.155 ± 0.035	0.837 ± 0.028	23.418 ± 0.944	0.164 ± 0.033	0.825 ± 0.035
	NICEGAN	23.721 ± 1.136	0.148 ± 0.029	0.840 ± 0.029	23.643 ± 1.045	0.148 ± 0.022	0.829 ± 0.033
	RegGAN	24.884 ± 0.991	0.094 ± 0.024	0.881 ± 0.017	24.576 ± 1.073	0.112 ± 0.022	0.852 ± 0.028
	ResViT	25.578 ± 0.812	0.088 ± 0.021	0.895 ± 0.018	24.825 ± 1.030	0.108 ± 0.018	0.861 ± 0.021
3D	CycleGAN	25.181 ± 0.861	0.097 ± 0.031	0.887 ± 0.012	24.602 ± 1.181	0.113 ± 0.021	0.854 ± 0.018
	Pix2Pix	23.740 ± 1.198	0.138 ± 0.032	0.835 ± 0.019	23.508 ± 1.301	0.152 ± 0.039	0.822 ± 0.024
	EaGAN	24.884 ± 0.991	0.094 ± 0.024	0.881 ± 0.017	24.576 ± 1.073	0.112 ± 0.022	0.852 ± 0.028
	MS-SPADE	25.818 ± 0.857	0.079 ± 0.016	0.904 ± 0.012	25.074 ± 1.085	0.098 ± 0.021	0.867 ± 0.018
	Ours	**30.756** ± 1.950	**0.065** ± 0.034	**0.944** ± 0.031	**32.224** ± 2.518	**0.046** ± 0.029	**0.941** ± 0.025

Table 9. **Modality transfer results of MRI** on BRATS23: Comparison between previous methods and our method. The best results are highlighted in blue.

Figure 9. TSNE of latent features of PET and CT images obtained under different applied constraints. Downstream performances are noted.



G.5. Comparison with non-personalized methods.

We provide further visual evidence in Fig. 9 for an in-depth analysis. Fig. 9 shows that applying our constraints aligns CT and PET representations more closely, indicating smaller $d_{\mathcal{G}\Delta\mathcal{G}}(\mathcal{M}|\mathcal{H}^{\mathcal{S}}, \mathcal{M}|\mathcal{H}^{\mathcal{U}})$ in Paper Eq. (10), yielding tightened bounds and better downstream performance (Dice:48.20), in comparison to the non-personalized baseline (Dice:43.5). This further supports our theoretical analysis.

	Target	T1			T1ce			T2			Flair		
Source		PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑	PSNR↑	NMSE↓	SSIM↑
T1	**SwinUNETR**	**32.815**	**0.092**	**0.941**	**31.655**	**0.202**	**0.912**	**24.650**	**0.361**	**0.857**	**27.593**	**0.202**	**0.883**
	Std.	*0.968*	*0.043*	*0.049*	*1.062*	*0.067*	*0.052*	*1.008*	*0.069*	*0.077*	*1.144*	*0.072*	*0.050*
	MS-SPADE	**29.001**	**0.055**	**0.942**	**26.119**	**0.078**	**0.912**	**25.818**	**0.103**	**0.904**	**24.842**	**0.113**	**0.859**
	Std.	*0.643*	*0.025*	*0.022*	*0.816*	*0.022*	*0.015*	*0.857*	*0.030*	*0.014*	*0.728*	*0.034*	*0.019*
	Ours	**43.472**	**0.003**	**0.996**	**34.547**	**0.045**	**0.955**	**30.756**	**0.065**	**0.944**	**31.693**	**0.049**	**0.937**
	Std.	*2.495*	*0.004*	*0.011*	*1.956*	*0.030*	*0.018*	*1.950*	*0.034*	*0.031*	*2.287*	*0.024*	*0.019*
T1ce	**SwinUNETR**	**32.456**	**0.100**	**0.929**	**33.001**	**0.156**	**0.926**	**25.125**	**0.366**	**0.859**	**27.699**	**0.211**	**0.882**
	Std.	*1.018*	*0.044*	*0.048*	*0.889*	*0.055*	*0.051*	*0.964*	*0.071*	*0.074*	*1.129*	*0.071*	*0.049*
	MS-SPADE	**26.228**	**0.076**	**0.922**	**28.759**	**0.060**	**0.937**	**25.990**	**0.092**	**0.907**	**25.204**	**0.092**	**0.881**
	Std.	*0.794*	*0.027*	*0.033*	*0.885*	*0.019*	*0.015*	*0.859*	*0.032*	*0.908*	*0.811*	*0.050*	*0.037*
	Ours	**34.077**	**0.020**	**0.962**	**46.663**	**0.003**	**0.996**	**30.775**	**0.063**	**0.942**	**32.224**	**0.046**	**0.941**
	Std.	*2.484*	*0.012*	*0.017*	*3.240*	*0.004*	*0.008*	*1.812*	*0.030*	*0.028*	*2.518*	*0.029*	*0.025*
T2	**SwinUNETR**	**30.102**	**0.171**	**0.896**	**30.354**	**0.283**	**0.883**	**26.831**	**0.268**	**0.887**	**27.234**	**0.242**	**0.872**
	Std.	*1.405*	*0.056*	*0.050*	*1.249*	*0.086*	*0.054*	*1.144*	*0.054*	*0.075*	*1.154*	*0.073*	*0.051*
	MS-SPADE	**25.422**	**0.085**	**0.908**	**25.234**	**0.087**	**0.895**	**29.230**	**0.048**	**0.942**	**25.074**	**0.098**	**0.867**
	Std.	*0.852*	*0.026*	*0.020*	*1.152*	*0.034*	*0.025*	*0.720*	*0.018*	*0.915*	*1.085*	*0.021*	*0.018*
	Ours	**32.646**	**0.028**	**0.955**	**33.857**	**0.051**	**0.949**	**43.653**	**0.006**	**0.991**	**32.224**	**0.046**	**0.941**
	Std.	*2.391*	*0.028*	*0.028*	*1.925*	*0.040*	*0.027*	*3.467*	*0.024*	*0.038*	*2.518*	*0.029*	*0.025*
Flair	**SwinUNETR**	**31.371**	**0.135**	**0.916**	**31.285**	**0.240**	**0.905**	**25.579**	**0.338**	**0.867**	**29.092**	**0.148**	**0.923**
	Std.	*1.198*	*0.051*	*0.054*	*1.161*	*0.077*	*0.053*	*0.956*	*0.064*	*0.073*	*0.974*	*0.055*	*0.049*
	MS-SPADE	**25.186**	**0.090**	**0.905**	**25.899**	**0.094**	**0.906**	**26.146**	**0.086**	**0.913**	**28.608**	**0.058**	**0.938**
	Std.	*0.759*	*0.028*	*0.048*	*1.039*	*0.025*	*0.027*	*0.636*	*0.028*	*0.944*	*0.769*	*0.025*	*0.028*
	Ours	**32.752**	**0.026**	**0.951**	**33.471**	**0.055**	**0.944**	**30.571**	**0.068**	**0.940**	**43.624**	**0.004**	**0.995**
	Std.	*2.399*	*0.020*	*0.022*	*1.634*	*0.035*	*0.021*	*1.951*	*0.035*	*0.034*	*2.441*	*0.008*	*0.013*

Table 10. **Modality transfer results of MRI** on BRATS23: The averaged results with standard derivations of metrics between all modalities.

Missing Num		=3				=2						=1				=0
Modality	flair	■				■	■			■		■	■	■		■
	T1		■			■		■	■			■	■		■	■
	T1ce			■			■		■		■	■		■	■	■
	T2				■			■		■	■		■	■	■	■
Tumour Core	SPA	65.86	65.27	78.26	66.4	72.99	83.23	70.66	81.25	70.66	80.63	83.22	73.89	83.36	82.05	83.4
	M3AE	69.4	65.45	79.12	71.84	79.9	70.45	82.79	81.17	71.62	73.35	81.78	82.42	73.31	81.61	82.22
	mmFormer	67.8	77.32	64.56	64.08	81.51	79.43	69.14	70.63	68.6	80.75	81.75	70.92	81.74	81.55	82.23
	RFNET	64.03	74.53	58.63	61.95	79.2	77.45	69.25	67.48	67.98	78.85	80.15	70.75	79.4	80.15	80.29
	M2F	65.79	63.29	77.31	63.64	70.38	79.93	68.01	79.62	67.68	79.37	80.65	69.73	80.01	79.53	80.34
	Ours	75.83	71.2	75.29	75.71	80.66	83.6	79.23	74.83	79.51	79.52	83.92	82.78	86.65	81.22	86.72
Enhancing tumour	SPA	39.85	41.39	70.43	41.72	45.99	73.07	45.25	72.87	45.25	72.59	73.52	47.56	73.01	73.55	73.65
	M3AE	37	38.41	75.8	44.22	78.09	45.2	79.36	78.16	41.71	48.12	79.14	80.06	47.63	79.31	79.91
	mmFormer	40.08	72.19	38.89	37.23	73.11	73.06	40.64	42.27	43.65	75.56	43.34	81.74	73.36	75.31	73.4
	RFNET	38.69	69.22	30.89	33.56	71.4	70.9	38.53	41.91	40.9	69.51	71.61	43.37	71.17	74.2	73.79
	M2F	37.99	37.79	71.74	39.28	43.37	74.66	45.42	73.48	43.5	73.48	73.56	45.93	73.15	74.03	75.26
	Ours	67.45	54.83	70.86	47.63	69.38	52.91	70.1	59.45	67.44	63.91	70.79	57.78	69.42	59.76	70.64
Whole tumour	SPA	85.77	72.69	71.95	80.4	87.82	87.97	88.27	75.57	88.27	81.8	88.3	88.78	89.06	82.87	89.27
	M3AE	87.78	74.69	74.91	84.43	76.09	84.48	89.63	84.4	88.64	88.91	84.04	89.29	88.58	88.45	88.26
	mmFormer	84.09	72.85	73.37	85.6	85.97	76.93	87.09	86.09	87.55	87.94	88.36	88.16	88.74	85.96	89.03
	RFNET	80.52	67.06	68.42	82.96	82.57	71.97	85.82	83.25	86	84.94	86.06	86.53	86.34	83.61	86.82
	M2F	85.72	72.48	71.78	82.53	87.73	87.66	84.35	76.03	87.69	84.27	88.17	88.22	88.47	84.32	88.72
	Ours	89.23	81.73	82.26	87.45	89.74	89.03	88.00	81.92	89.72	86.27	89.12	90.5	91.25	87.1	91.19

Table 11. **Missing modality segmentation results of MRI** on BRATS18: Num denotes the number of missing modalities for different settings. The used modalities are highlighted with gray boxes and the missing ones remain as blank. The results of each setting are presented accordingly.

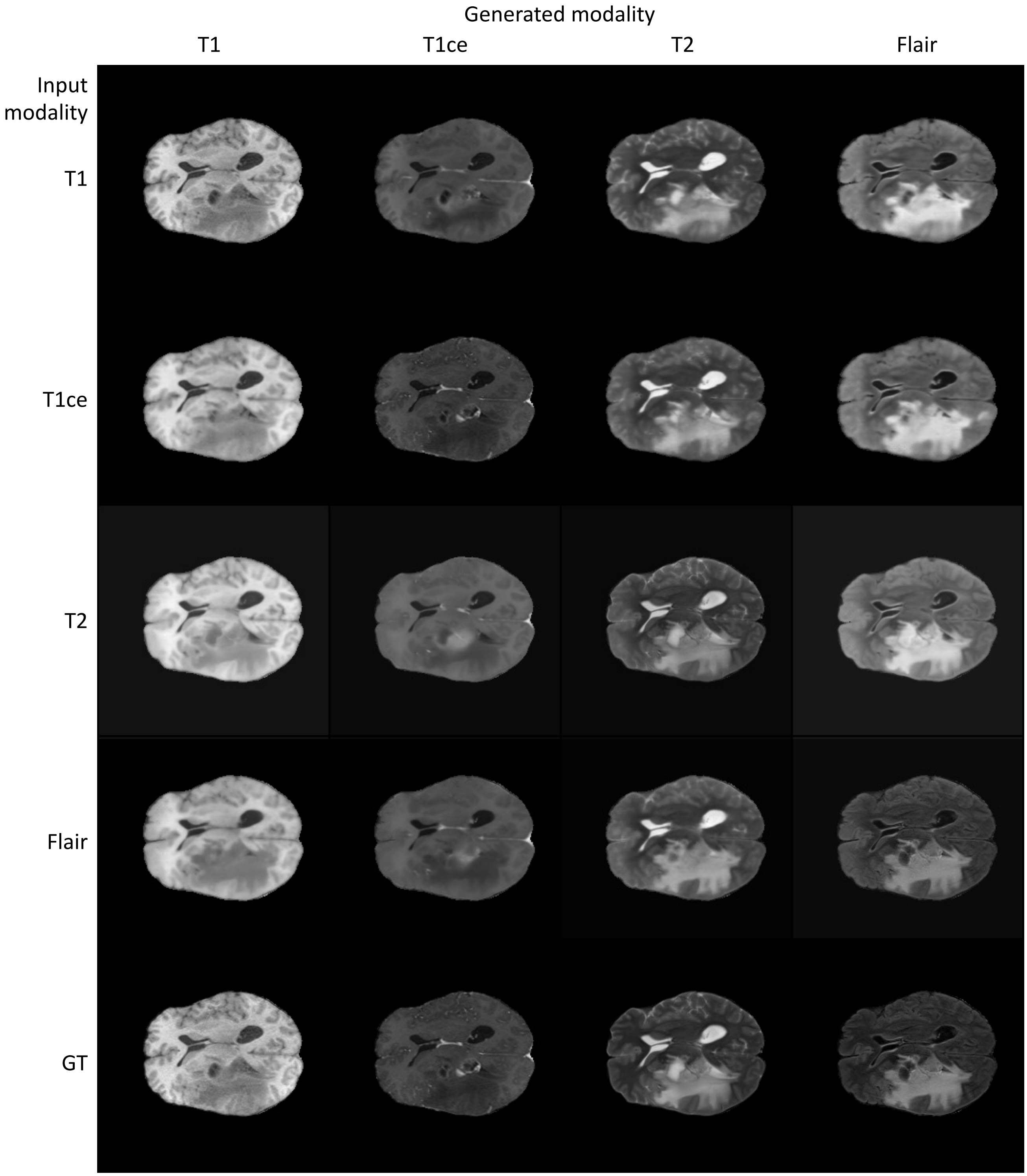


Figure 10. Generated images of our proposed method: slices across ventricles.

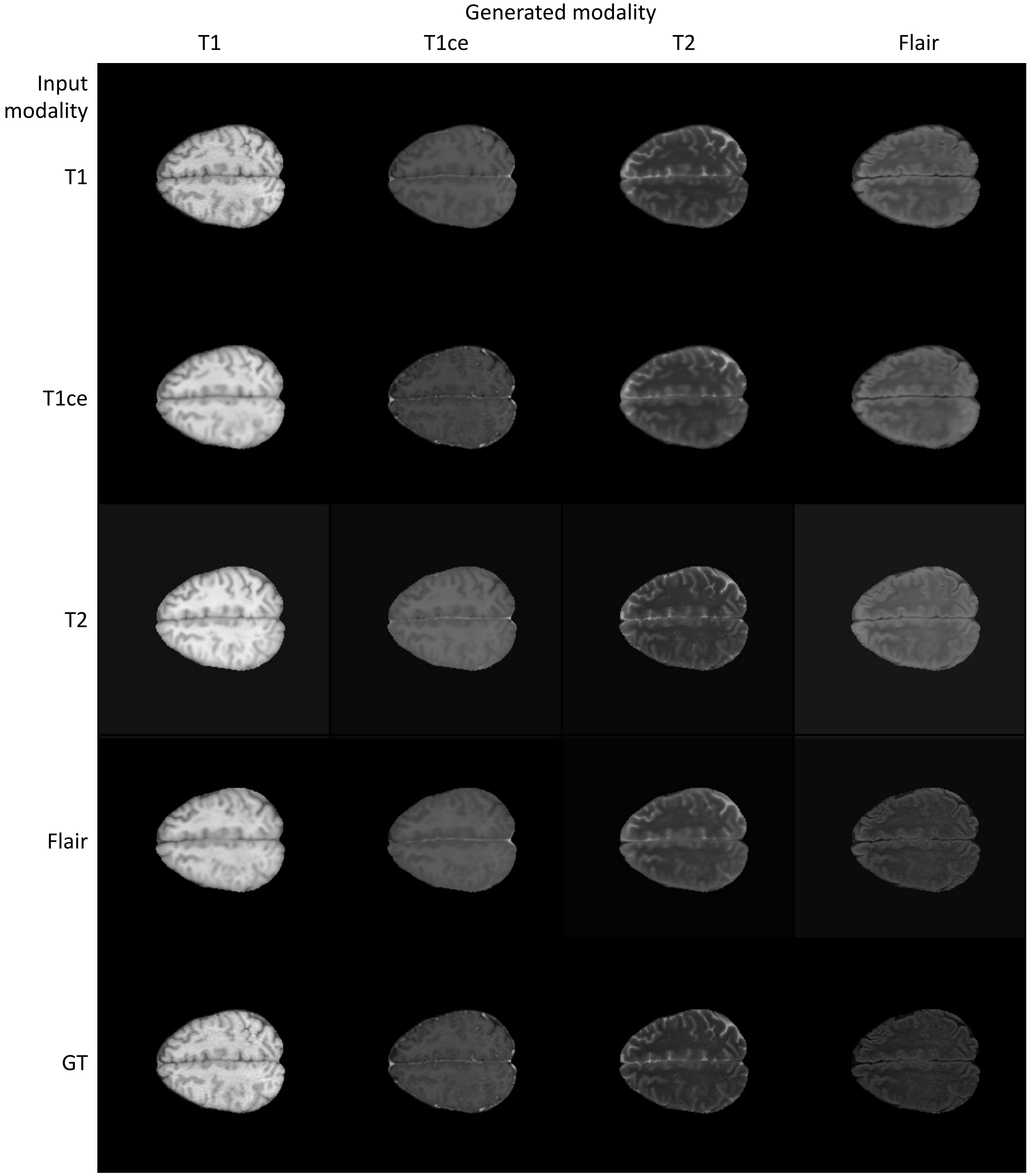


Figure 11. Generated images of our proposed method: slices across cerebral sulcus.

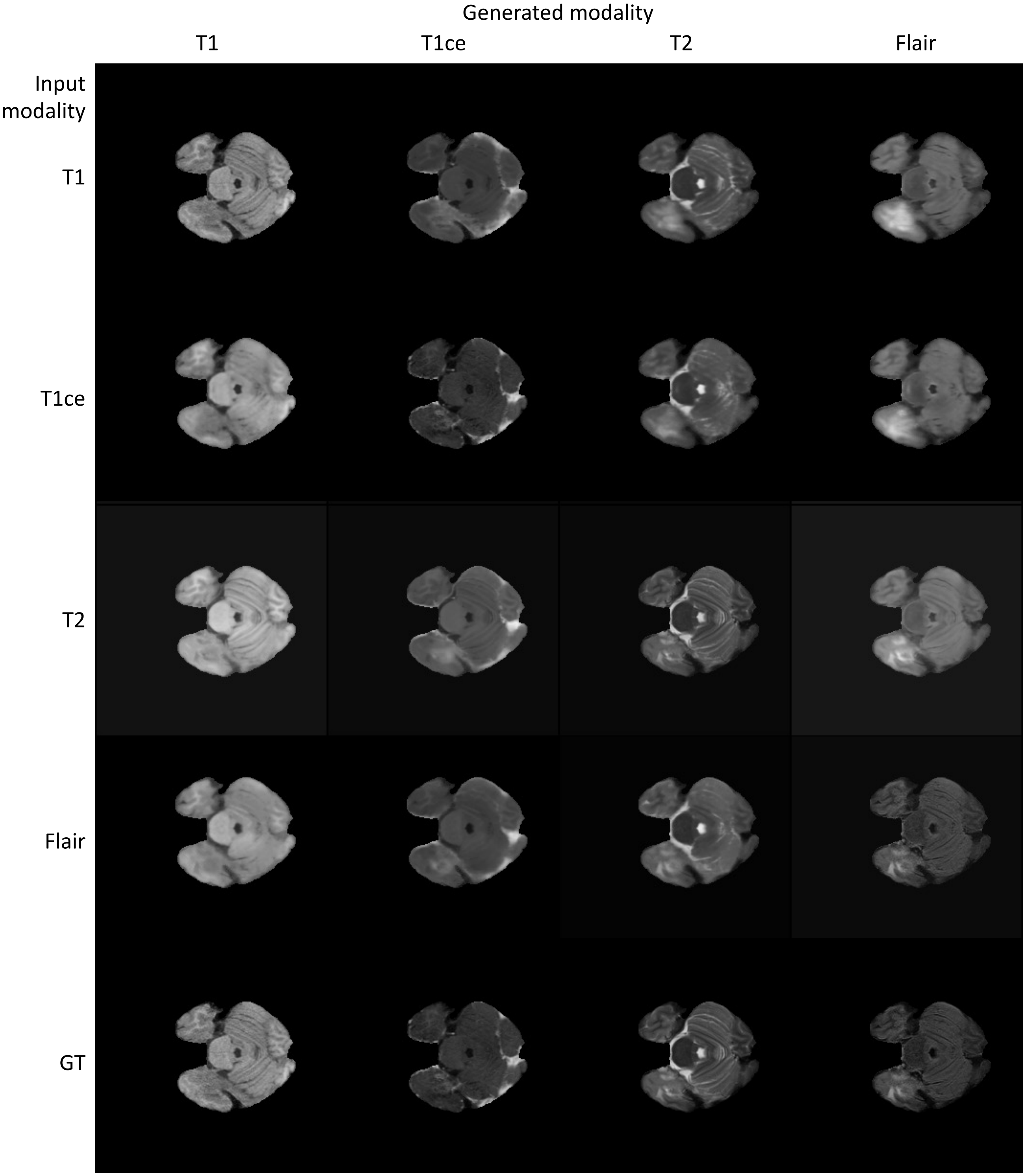


Figure 12. Generated images of our proposed method: slices across the cerebellar hemisphere. Our method can generate defined cerebellar folia.