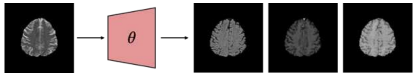
Generation of Multi-modal Brain Tumor MRIs with Disentangled Latent Diffusion Model

Yoonho Na, Kyuri Kim, Hwiyoung Kim, Sung-Joon Ye and Jimin Lee

Introduction

- Prior researches generating multi-modal MRIs rely on Image-to-Image translation and thus require **source image** to obtain fixed structural information.



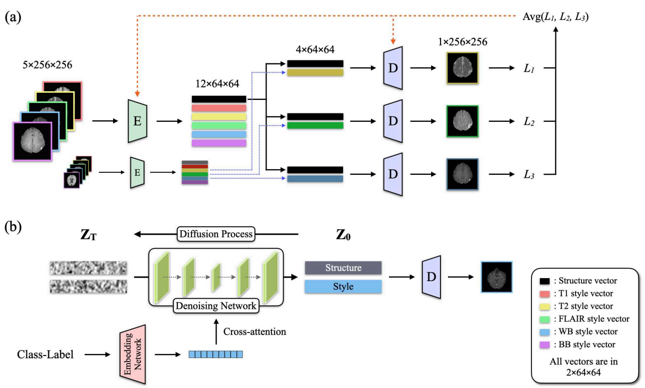
- Image-to-Image translation based approach has limitation that the structural diversity is restricted to the source image.
- We propose a novel approach for generating multi-modal brain tumor MRIs using feature distentanglement and diffusion models.
- Our proposed model, which we call disentangled latent diffusion model (DLDM), is capable of generating **modality-sharing** and **modality-specific** information separately, eliminating the need for source image.

Methods

Key Idea

- Extract modality-sharing information from the multi-modal MRIs.
  → Brain structures
- Decoding the latents with the fixed modality-sharing informations results in images with fixed brain structures.

Overview



DLDM is based on *Latent Diffusion Model* (Rombach et al., 2022).
Therefore our proposed model require two steps for training.
❶ Train autoencoder that maps pixel space to lower dimensional latent space
❷ Train diffusion model that operates in learned latent space of autoencoder

Step 1 : Disentangled Autoencoder

Training Strategy

1. For N number of MRI modalities, set the dimension of latent vector dimension to have $N+1$ channels.
   - Each N channels to have N distinct style vectors z^{style} and the rest channel to have structure vector z^{struct}.
2. Feed z^{struct} and randomly selected single z^{style} to the decoder.
3. Randomly mix z^{style} with other data in mini-batch.
   - Strategy 2 and 3 ensures the encoder to separate multi-modal MRIs into modality-sharing and modality-specific information.
4. Average the loss of selected modalities to ensure every modalities having similar reconstruction quality.

$$L_{Autoencoder} = L_{rec}(x, \mathcal{D}(\mathcal{E}(x))) + L_{reg}(x; \mathcal{E}, \mathcal{D})$$

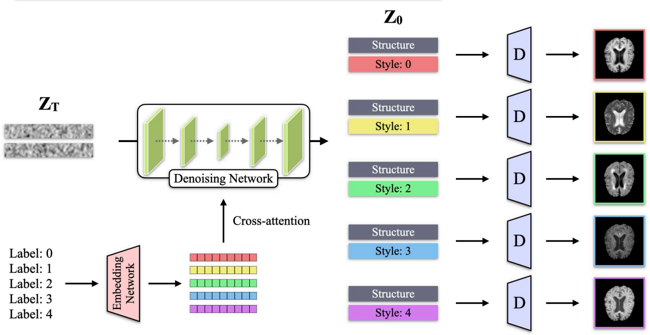
Step 2 : Disentangled Latent Diffusion Model

Training Strategy

1. We use the pair of (z^{struct}, z^{style}) as a input of latent diffusion model.
2. Class-label c, is applied as condition to selectively obtain z^{style} for every MRI modalities.
3. When sampling multi-modal MRIs, we fixed z^{struct} to generate same structural data.

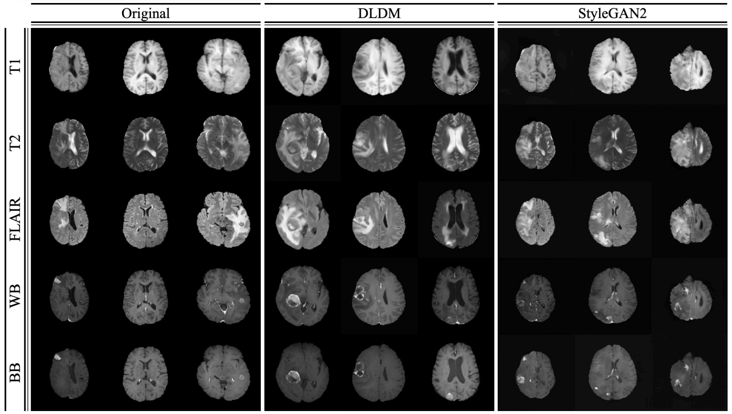
$$L_{DLDM} := \mathbb{E}_{\mathcal{E}(x),c,\epsilon\sim\mathcal{N}(0,1),t}\left[||\epsilon - \epsilon_\theta(z_t^{struct}, z_t^{style}, \tau_\theta(c), t)||_2^2\right]$$

Generation Process of Multi-modal MRIs



Experimental Results

Qualitative Comparison

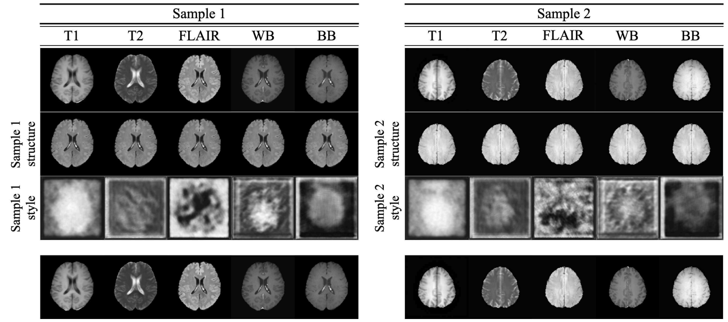


Quantitive Comparison

DLDM						
Modality	Precision↑	Recall↑	Density↑	Coverage↑	FID↓	clean-FID↓
T1	**0.95729**	0.94063	**0.96063**	**0.96563**	18.52	0.00019
T2	0.90104	0.97083	0.81646	0.93438	**13.86**	0.00268
FLAIR	0.87604	0.95417	0.77854	0.93646	25.49	0.00025
WB	0.91146	0.95833	0.79896	0.92708	22.34	**0.00018**
BB	0.91458	**0.97708**	0.82458	0.90833	28.89	0.00022
Average	0.91208	0.96021	0.83583	0.93438	21.82	0.00070

StyleGAN2						
Modality	Precision↑	Recall↑	Density↑	Coverage↑	FID↓	clean-FID↓
T1	**0.91042**	0.72500	0.74125	0.84063	53.55	0.00284
T2	0.27708	0.50000	0.10792	0.25625	169.12	0.00385
FLAIR	0.96250	0.46146	**1.13708**	**0.87188**	102.65	**0.00208**
WB	0.77500	0.75729	0.64937	0.85208	**42.91**	0.00260
BB	0.46354	**0.96771**	0.18625	0.38646	119.40	0.00328
Average	0.67771	0.68229	0.56437	0.64146	97.53	0.00293

Disentanglement Study



Mixed sample: Sample 1 structure + Sample 2 style
Mixed sample: Sample 2 structure + Sample 1 style

- Multi-modal MRIs are successfully disentangled into structure and styles.

Acknowledgments
This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C1161).