Medical Image Data Analysis
: 3D MRI Generation via Flow Matching Model

Wonyoung Jang

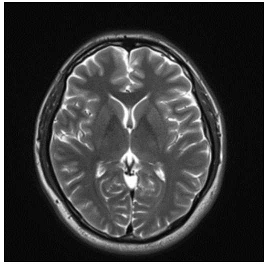
Seoul National University
Graduate School of Data Science

We Need More Medical Imaging Data

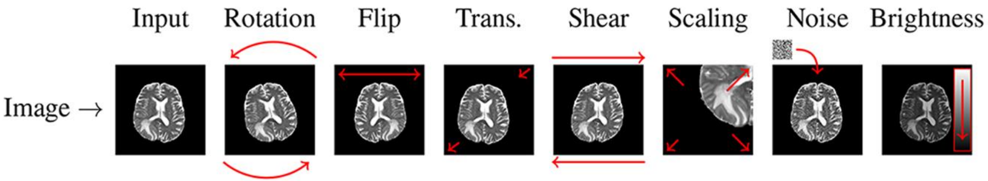
- Medical video/image is…
- Hard to collect at scale.
- Labeling costs a lot.

- Aggressive data augmentation/transformation could harm semantics.
- Hard to collect at scale.

Label: ???

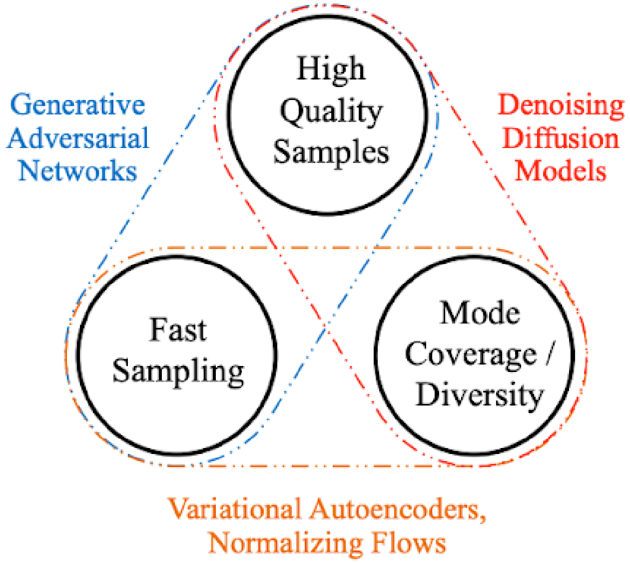


Dangerous



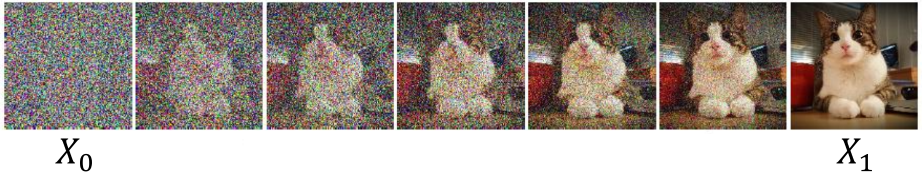
Medical Image Generation

- Diffusion & Flow matching model could help
- Can do semantic preserved generation with conditional training/sampling.

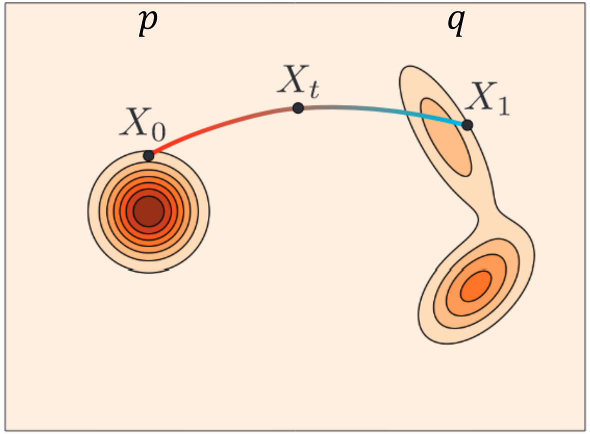


What is Flow Matching?

- **GOAL**: Build a flow ψ_t that transforms a source sample $X_0 \sim p$ into a target sample $X_1 \coloneqq \psi_1(X_0)$ such that $X_1 \sim q$ where $0 \le t \le 1$.



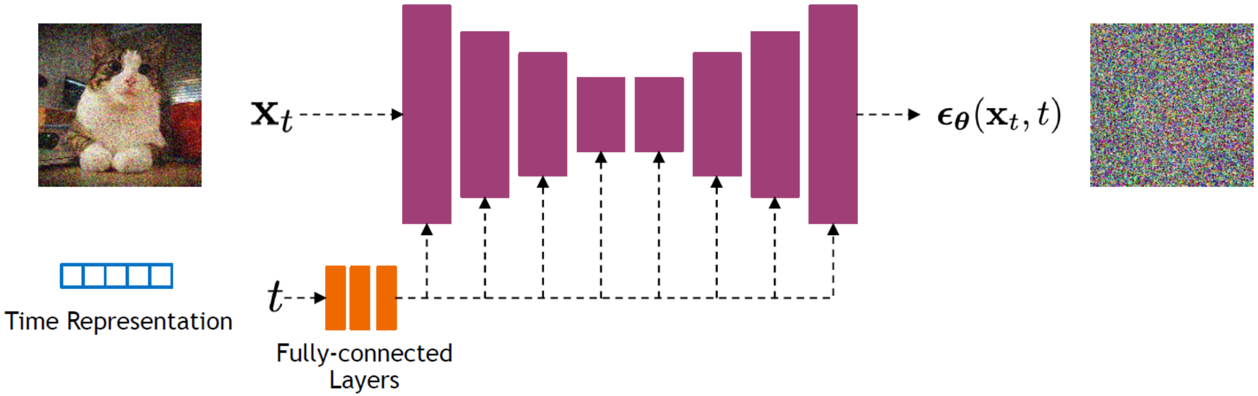
$$X_t = tX_1 + (1 - t)X_0$$



Image=Noise=Velocity Prediction Network

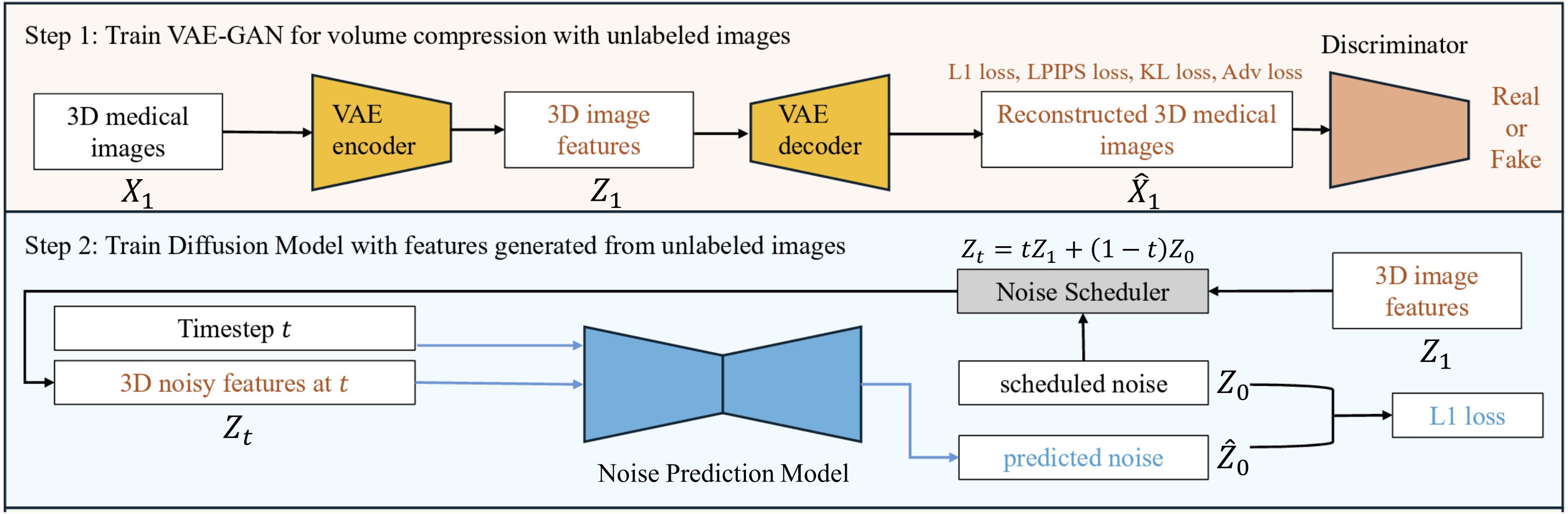
- Use U-Net architectures with ResNet blocks and self-attention layers to construct $\epsilon_\theta(\mathbf{x}_t, t)$.
- Input: noisy image x_t and corresponding time point t.
- Output: noise x_0 applied to input noisy image x_t. (=noise prediction)

$$X_t = tX_1 + (1-t)X_0$$



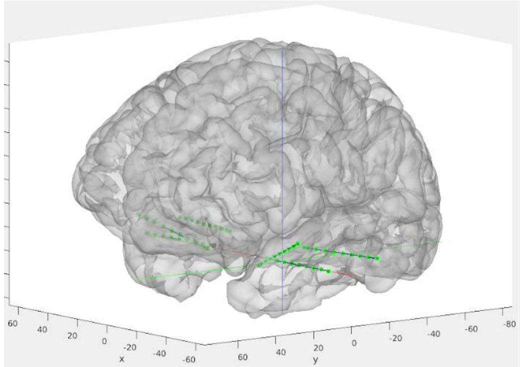
Latent Model Architecture

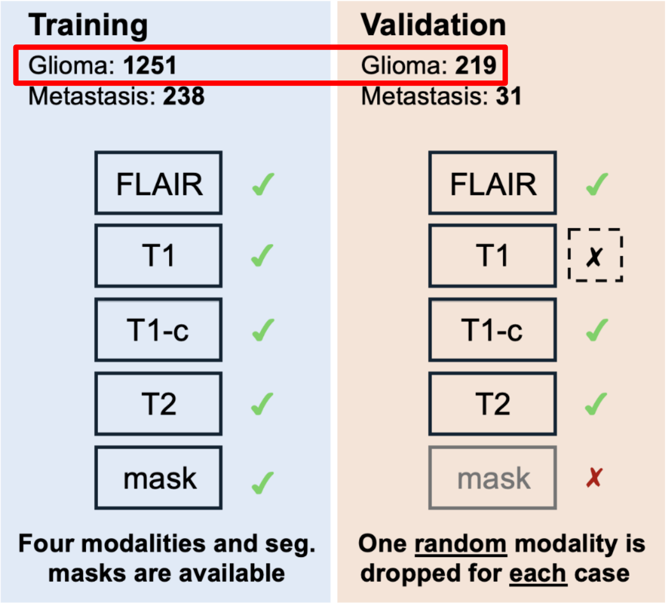
- **GOAL**: lower computational demands of training high-resolution image.
- **Solution**: explicitly separate image compression stage & generative learning stage.



BraTS Dataset

- Brain Tumor Segmentation Dataset
  - 3D spatial MRI
  - (1x240x240x155) raw matrix each
  - Grayscale channel
  - Modality: T1, T1-c, T2, Flair and mask.

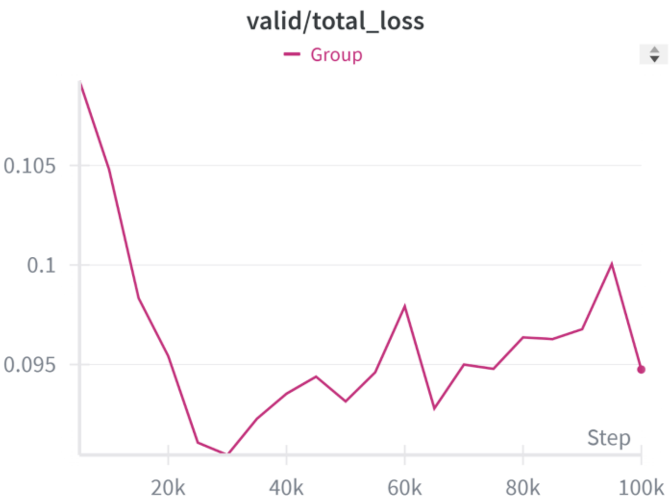




BraTS Training

VAE

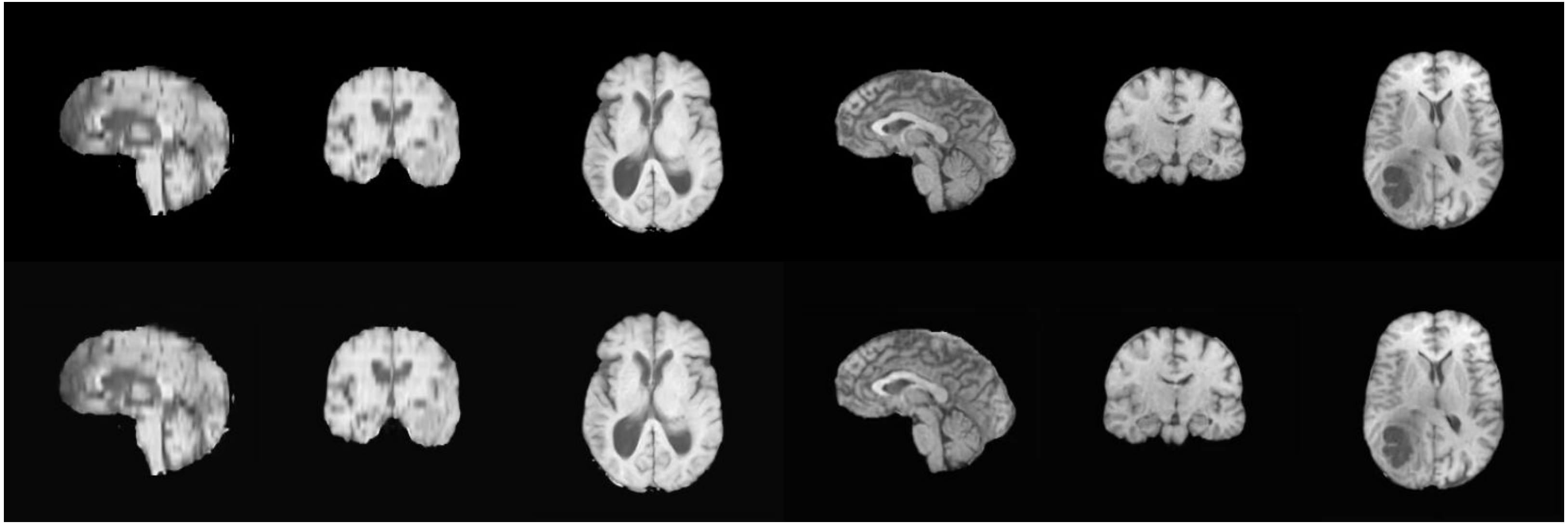
UNET



BraTS (VAE Reconstruction)

Real
(input)

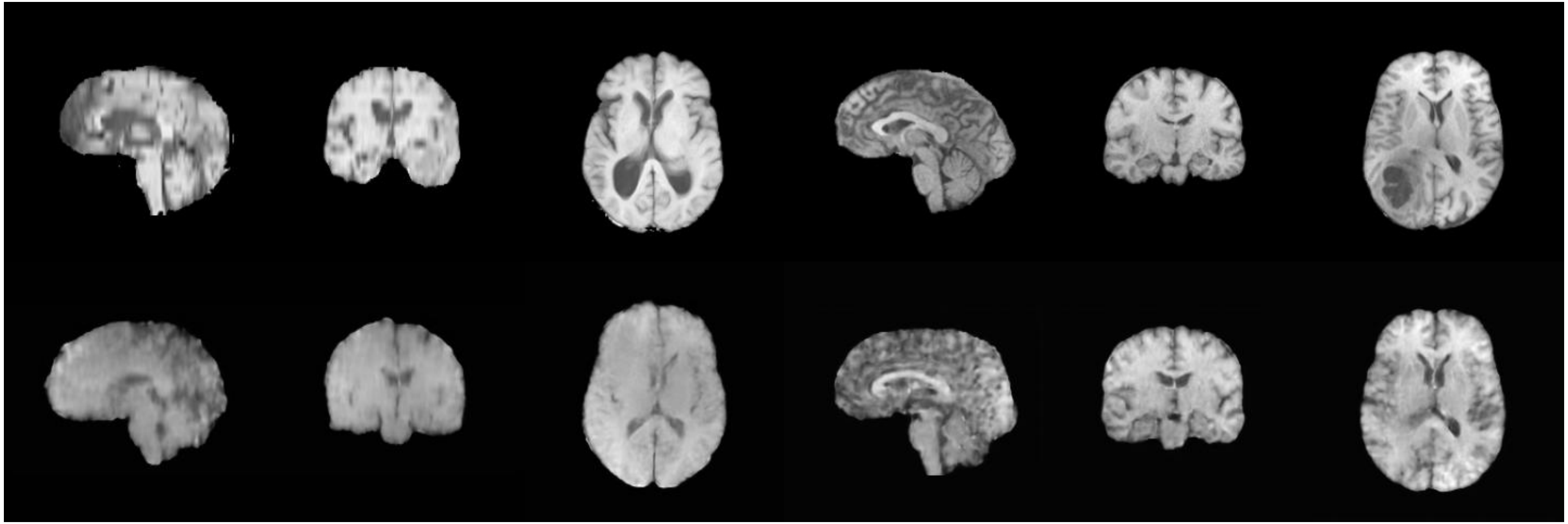
Recon
(output)



BraTS (Unet Generation)

Real

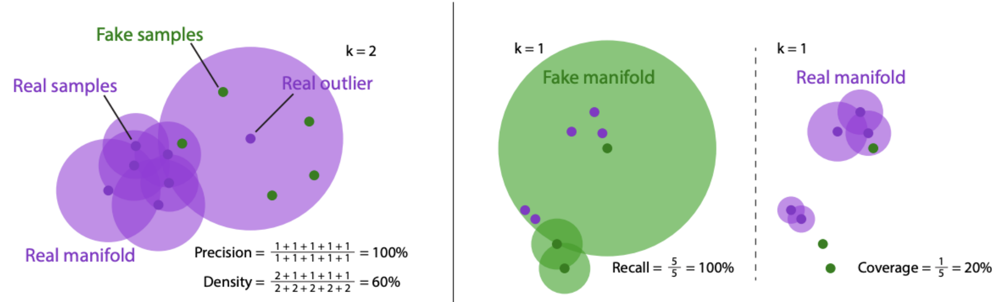
Generated



BraTS (Evaluation)

- Applied bootstrapping for sampling. (200 vs. 200 / 20 times)

BraTS (T1n)	Real vs. Recon	Real vs. Real (other)	Real vs. Gen (ckpt-100K)
FID↓	0.0000 ± 0.0000	0.0021 ± 0.0003	0.0006 ± 0.0001
Precision↑	0.9734 ± 0.0104	0.7421 ± 0.0260	0.7832 ± 0.0215
Recall↑	0.9749 ± 0.0105	0.8193 ± 0.0168	0.5251 ± 0.0257
Density↑	1.1717 ± 0.0326	0.8497 ± 0.0329	1.1223 ± 0.0382
Coverage↑	0.9949 ± 0.0056	0.8015 ± 0.0213	0.7810 ± 0.0225
LPIPS↓	0.019		
PSNR↑	36.79		
SSIM↑	0.9866		



(a) Precision versus density.

(b) Recall versus coverage.

Useful Resources

- Flow Matching Guide and Code: https://arxiv.org/abs/2412.06264
- MAISI paper: https://arxiv.org/abs/2409.11169
- High-Resolution Image Synthesis with Latent Diffusion Models: https://arxiv.org/pdf/2112.10752
- BraTS website: https://www.synapse.org/Synapse:syn64153130/wiki/630130
- MONAI: https://github.com/Project-MONAI/MONAI
- MAISI tutorial: https://github.com/Project-MONAI/tutorials/tree/main/generation/maisi

Thank You!