Conditional Generation of 3D Brain Tumor Regions via VQGAN and Temporal-Agnostic Masked Transformer

Meng Zhou[1,2] and Farzad Khalvati[1,2,3]

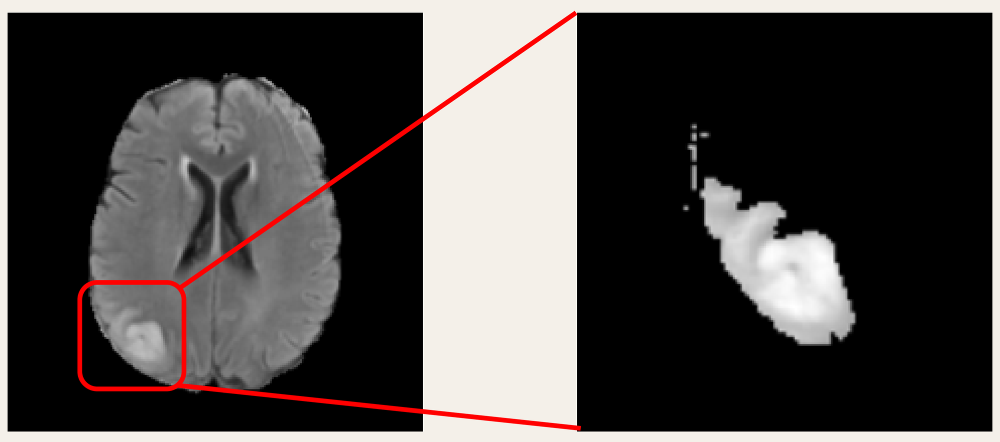
[1]Neurosciences & Mental Health Research Program, SickKids Research Institute
[2]Department of Computer Science, University of Toronto
[3]Department of Medical Imaging, University of Toronto
IMICS Lab, https://imics.ca/

MIDL Paris 2024

INTODUCTION

Context:

- Neuroradiology studies often suffer from **lack of sufficient data** to properly train deep learning models, especially for **identifying brain tumor pathology types (HGG and LGG)**.
- GANs can be used for synthetic image generation, however they sometimes are unstable and struggle to produce high-fidelity images [1].
- Diffusion Probabilistic Models [2,3] sometimes require extensive computational resources.
- ROI volumes are more effective than whole brain images when classifying tumor types.



Challenges:

- Is there a way to generate high-resolution, high-fidelity images that is computationally friendly?
- Previous work [4] focuses on generating only one tumor type, can we extend to other types?

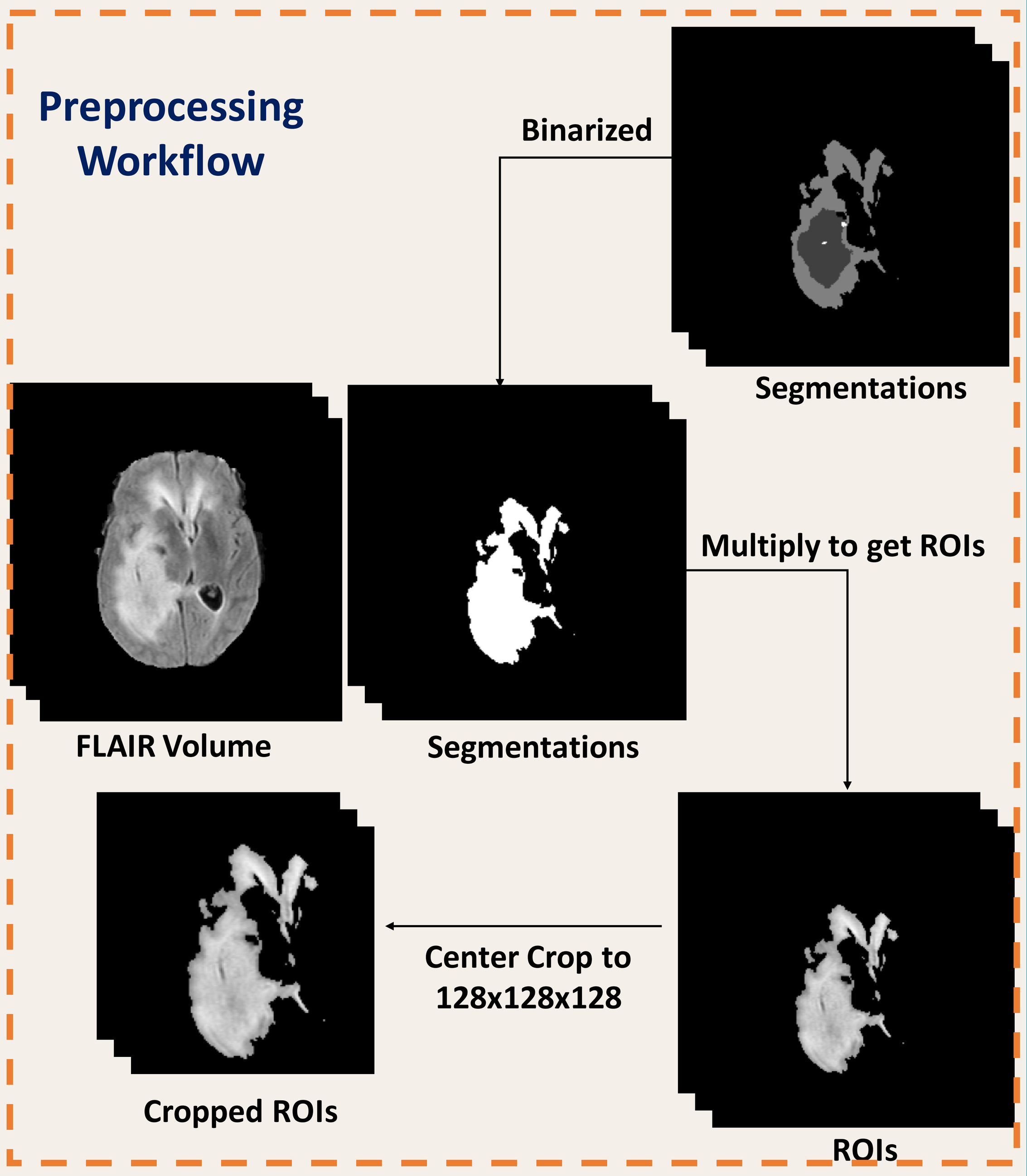
Contributions:

- We present the first **class-conditional** generation framework for different tumor types based on the given class label.
- We learn an **importance score map** for encoded region features.
- We propose a novel temporal-agnostic hybrid masking strategy which **uses the importance score to mask tokens** when training the transformer model.
- Experiments show our proposed method outperforms several baselines in both image generation quality and the downstream classification task.

MATERIALS

Dataset and Preprocessing:

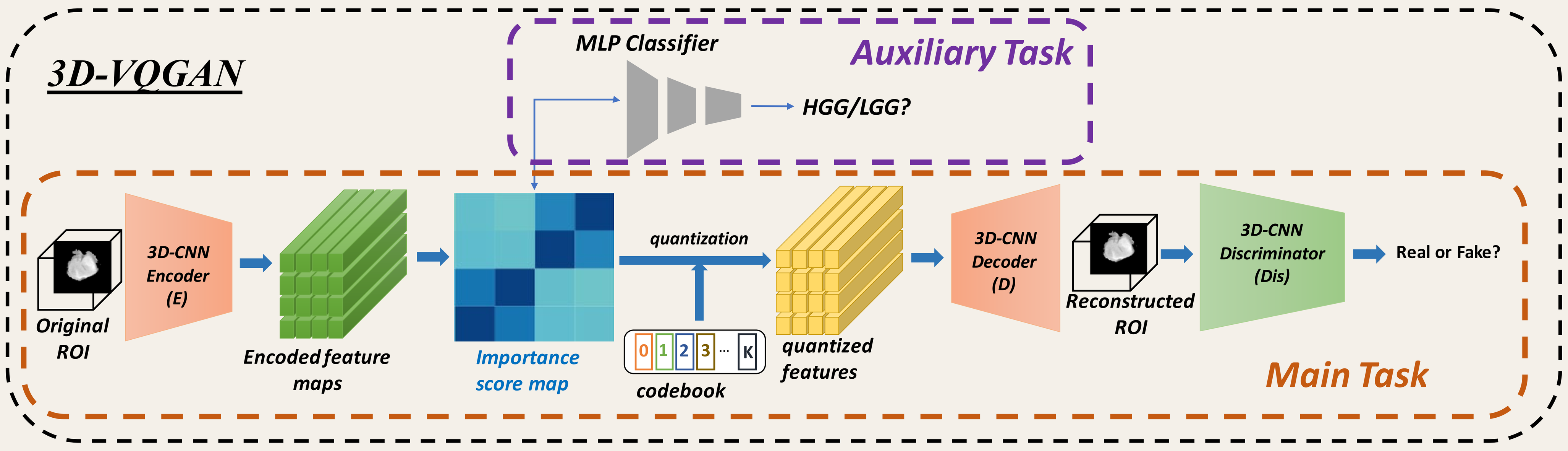
- BraTS 2019 Dataset with MRI FLAIR sequence is used in this work
- Reshape from whole brain volume 240x240x155 to ROI-based volume 128x128x128
- Normalized to [-1,1]



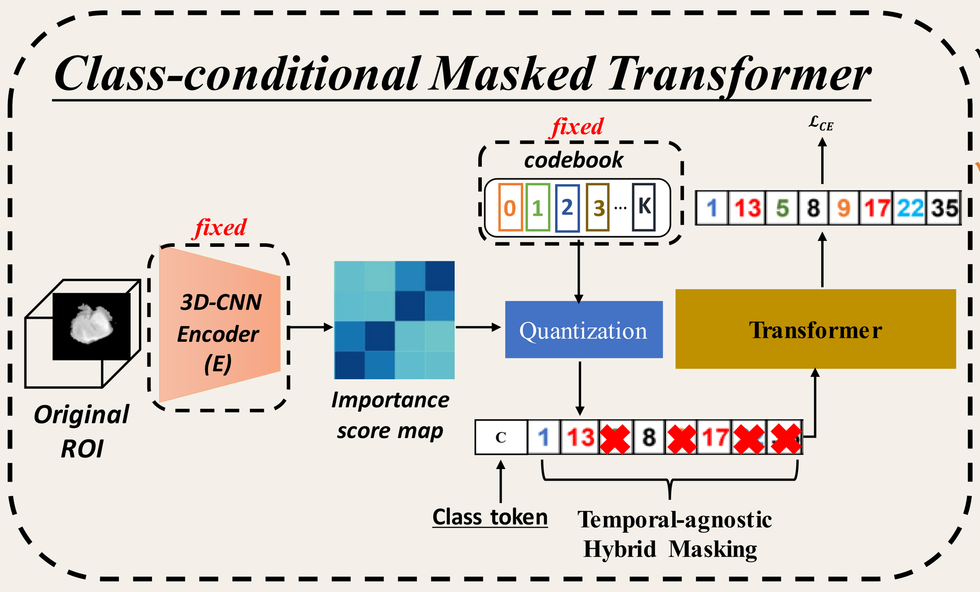
METHODS

Two-Stage Model:

- **Stage 1:** 3D-VQGAN, learning an efficient data representation in the codebook [4,5]
  - learn the region importance in the latent space

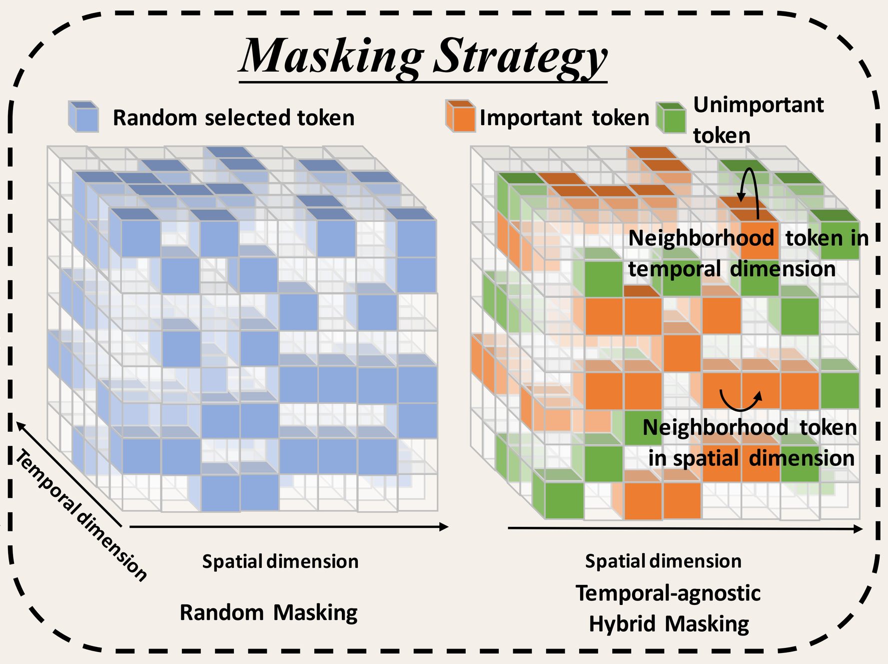


- **Stage 2:** Masked Transformer, learning the long-term interrelations of the semantic token compositions
  - Proposed Masking strategy: important tokens and their neighboring tokens masked with **[MASK]**, unimportant tokens been replaced by **other random tokens** in the codebook.



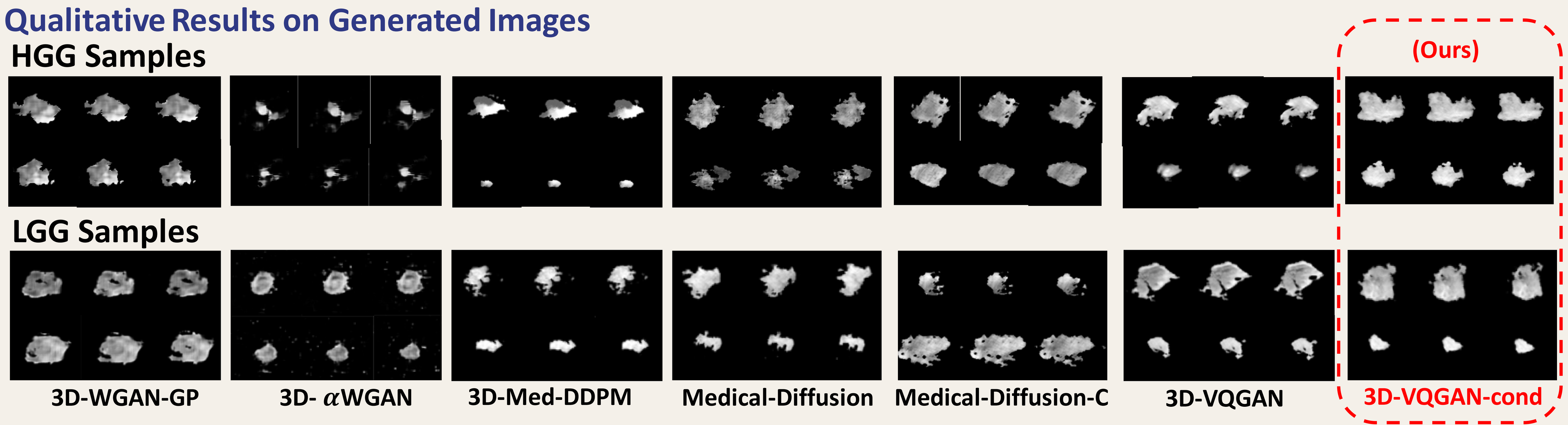
Key Takeaways

- ✓ Masking introduces noise to the model, hence **improves the robustness**.
- ✓ **Important vs. Unimportant tokens:** Important tokens contain rich semantic information; mask their neighbors with **[MASK]** could **prevent information leakage**. Unimportant tokens are selected to be replaced by other random tokens to **retain enough randomness** in the model.



RESULTS

Qualitative Results on Generated Images



Quantitative Results on Downstream Classification task

- LGG/HGG Pathology type ROI-based classification

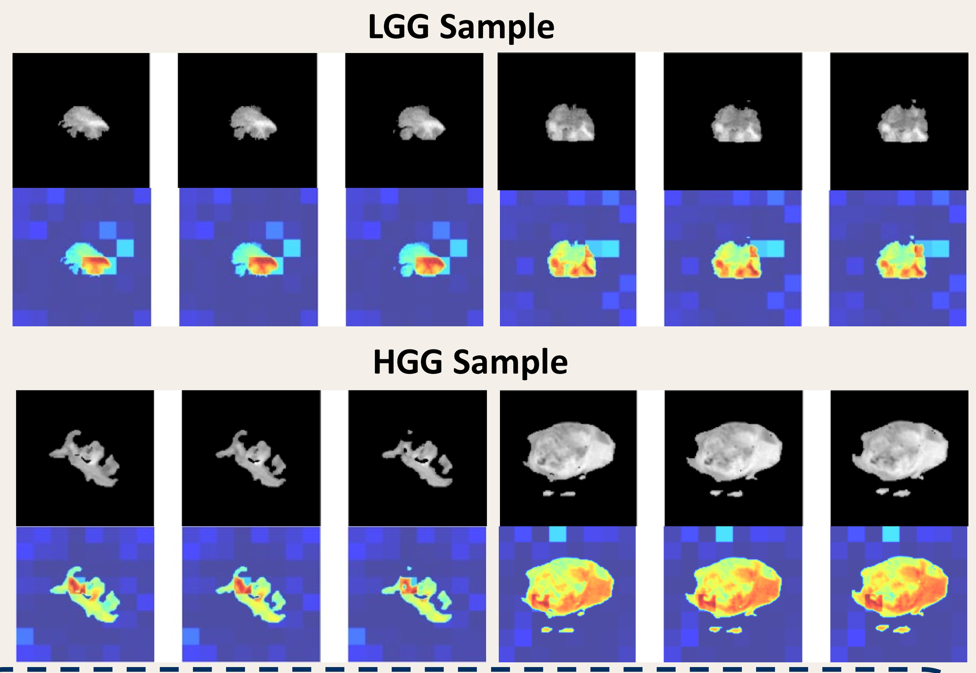
	AUC	F1-Score	Acc.	Prec.	Rec.
3D-αWGAN	0.71$\pm$0.04	0.65$\pm$0.02	0.64$\pm$0.06	0.65$\pm$0.08	0.68$\pm$0.10
3D-Med-DDPM	0.71$\pm$0.09	0.69$\pm$0.02	0.65$\pm$0.03	0.62$\pm$0.04	0.77$\pm$0.04
Med-Diffusion	0.73$\pm$0.04	0.67$\pm$0.02	0.61$\pm$0.02	0.55$\pm$0.04	0.84$\pm$0.11
Med-Diffusion-C	0.75$\pm$0.03	0.68$\pm$0.02	0.66$\pm$0.03	0.64$\pm$0.04	0.73$\pm$0.04
3D-VQGAN	0.78$\pm$0.04	0.71$\pm$0.06	0.70$\pm$0.06	**0.69$\pm$0.06**	0.72$\pm$0.07
Ours	**0.80$\pm$0.02**	**0.74$\pm$0.02**	**0.70$\pm$0.01**	0.66$\pm$0.05	**0.87$\pm$0.13**

Ablations

- The ratio of important tokens to all tokens, which controls the trade-off between diversity and image

Top-k Ratio	LGG MMD(10^4 ↓)	LGG MS-SSIM	HGG MMD(10^4 ↓)	HGG MS-SSIM
0%	1.67	89.4 (4.1)	1.46	89.7 (1.1)
25%	1.72	87.9 (2.6)	1.57	88.5 (0.1)
50%	2.14	87.3 (2.0)	1.80	89.2 (0.6)

Visualization of Importance Score



Key Takeaways

- Visually, ROIs generated from our method contains the **detailed attributes of the tumor** and **exhibit high image fidelity**
- Using generated ROIs as additional data, our method yields the best performance compared to other methods, which also implies synthetic ROI data can be **effectively used to pre-train** the deep models and only a small amount is needed for finetuning.

Conclusion

Summary:

- We propose the first class-conditional generation framework for LGG and HGG brain tumor types. Experiments show our method outperforms several baselines.
- This unified framework to generate different types of tumor regions saves a large amount of time and resources.

Future Work:

- Incorporate text condition, we can generate tumor regions based on radiology reports.
- Instead of generating ROIs, we may generate both whole brain volumes and segmentation masks at the same time.

References

[1] Kwon G, Han C, Kim DS. Generation of 3D brain MRI using auto-encoding generative adversarial networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2019 Oct 10 (pp. 118-126). Cham: Springer International Publishing.
[2] Dorjsembe Z, Pao HK, Odonchimed S, Xiao F. Conditional diffusion models for semantic 3D medical image synthesis. Authorea Preprints. 2023 Oct 31.
[3] Peng W, Adeli E, Bosschieter T, Park SH, Zhao Q, Pohl KM. Generating realistic brain mris via a conditional diffusion probabilistic model. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2023 Oct 1 (pp. 14-24). Cham: Springer Nature Switzerland.
[4] Zhou M, Wagner MW, Tabori U, Hawkins C, Ertl-Wagner BB, Khalvati F. Generating 3D Brain Tumor Regions in MRI using Vector-Quantization Generative Adversarial Networks. arXiv preprint arXiv:2310.01251. 2023 Oct 2.
[5] Esser P, Rombach R, Ommer B. Taming transformers for high-resolution image synthesis. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021 (pp. 12873-12883).

Acknowledgements

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada