

MRI Potato Head: Feature-Guided Latent Diffusion for Anatomically Consistent 3D Brain MRI Editing

Nghi C. D. Truong¹, Chandan Ganesh Bangalore Yogananda¹, Benjamin C. Wagner¹, Niloufar Saadat¹, James M. Holcomb¹, Divya Reddy¹, Jason Bowerman¹, Kimmo J. Hatanpaa², Toral R. Patel³, Baowei Fei^{1,4}, Matthew D. Lee⁵, Rajan Jain^{5,6}, Richard J. Bruce⁷, Marco C. Pinho¹, Ananth J. Madhuranthakam⁸, and Joseph A. Maldjian^{1,*}

¹Department of Radiology, UT Southwestern Medical Center, Texas, USA

²Department of Pathology, UT Southwestern Medical Center, Texas, USA

³Department of Neurological Surgery, UT Southwestern Medical Center, Texas, USA

⁴Department of Bioengineering, University of Texas at Dallas, Texas, USA

⁵Department of Radiology, NYU Grossman School of Medicine, New York, USA

⁶Department of Neurosurgery, NYU Grossman School of Medicine, New York, USA

⁷Department of Radiology, University of Wisconsin-Madison, Wisconsin, USA

⁸Department of Radiology, Mayo Clinic, Rochester, MN, USA

ABSTRACT

Medical image editing enables simulation of rare disease presentations and counterfactual examples but often struggles to preserve subject-specific anatomy. We propose a zero-shot, prompt-driven framework for anatomically consistent editing of three-dimensional (3D) multi-sequence brain MRI using pretrained latent diffusion models. Our method introduces a feature injection strategy, where intermediate features from residual and attention blocks of a guidance image are incorporated into the denoising process alongside text or mask conditions. This design preserves local structural fidelity and global contextual coherence during editing. We demonstrate three applications: (1) insertion of brain tumors into healthy brains, (2) manipulation of IDH molecular status with mask- and text-guided prompts, and (3) tumor removal with anatomically coherent reconstruction of healthy tissue. Results show anatomically faithful outputs that maintain patient-specific brain structure while performing semantically precise edits. The framework is broadly applicable for simulating disease subtypes, augmenting training datasets, and generating counterfactuals in medical imaging research.

Keywords: Latent diffusion model, Generative models, Brain tumor imaging, Synthetic data, Image Editing

1. INTRODUCTION

Medical image editing, the ability to precisely modify pathological features within existing MRI scans, holds notable potential for advancing biomedical research and artificial intelligence (AI) development. By enabling the controlled manipulation of imaging attributes such as lesions, molecular disease markers, or brain aging patterns, image editing can enhance the diversity and utility of available imaging datasets. Such capabilities facilitate the synthesis of rare or underrepresented disease scenarios, creation of counterfactual examples (e.g., the same brain with and without a tumor), and simulation of disease progression, regression, or therapeutic outcomes.

Despite these potential benefits, achieving anatomically coherent and semantically precise edits in multi-sequence volumetric MRI remains technically challenging. Traditional generative approaches frequently struggle to preserve anatomical consistency between original and edited images and often lack robust semantic control for clinically complex scenarios. Recent advances in prompt-based generative models have shown promising results for semantic editing, enabling modifications guided by textual descriptions while preserving original image

Further author information: (Send correspondence to Joseph A. Maldjian)

Joseph A. Maldjian: E-mail: Joseph.Maldjian@UTSouthwestern.edu

structures. However, most of these studies focus primarily on 2D natural images, with limited applications to multi-contrast, volumetric medical imaging.

A few recent studies tackled the problem of medical image editing. For instance, MedEdit¹ enables the insertion of a stroke lesion in brain MRI scans at the location provided by the input conditional mask. Another high-impact study introduced Latent Drifting,² a technique that conditions pre-trained diffusion models at inference time, enabling zero-shot generation of counterfactual medical images, such as simulating aging or inducing diseases like Alzheimer's. Similarly, SkEditTumor³ offers sketch-based interaction for precise tumor progression edits in 3D imaging. Multi-Channel Fusion Diffusion (MCGFDiffusion)⁴ enhances brain tumor data augmentation by fusing healthy and pathological channels, resulting in improved classification and segmentation results.

In this study, we introduce a zero-shot, prompt-driven latent diffusion framework⁵⁻⁸ designed specifically for anatomically consistent editing of 3D multi-sequence brain MRI volumes. Our approach leverages pre-trained conditional latent diffusion models used in our prior works,⁹⁻¹¹ which leveraged text conditioning to enable flexible manipulation of images without retraining. Specifically, our method allows: (1) insertion of clinically meaningful lesions into healthy images, (2) modification of the molecular status of existing lesions without altering patient-specific anatomical features, and (3) seamless removal of pathological lesions to reconstruct healthy anatomy. Although we demonstrate our approach using IDH mutation status as a representative application, the framework can be broadly adaptable to other disease markers, neurological conditions, tumor subtypes, and modeling structural transformations such as aging or neurodegeneration, highlighting its wide-ranging utility across medical imaging research.

2. METHODOLOGY

2.1 Latent Diffusion Models

We employed pretrained three-dimensional (3D) latent diffusion models (LDMs) previously developed in our prior work.⁹⁻¹¹ Two variants of the model were utilized: (i) a text-conditioned 3D LDM, and (ii) a hybrid 3D LDM conditioned jointly on tumor masks and text prompts. The former allows semantic control via text condition, while the latter provides spatially guided edits over pathological regions, ensuring precise manipulation of the brain tumor regions while maintaining global anatomical structure.

2.2 UNet Backbone for Denoising

The denoising process within the LDM is parameterized by a UNet architecture comprising residual blocks, self-attention blocks, and cross-attention layers. Residual blocks enable efficient hierarchical representation learning and stable gradient propagation. Self-attention modules capture long-range spatial dependencies within the 3D MRI volumes. Cross-attention layers integrate external conditioning signals, aligning semantic information from the text conditions with image features to ensure semantically consistent edits.

2.3 Editing Mechanism

Given an input 3D MRI volume and an associated textual condition, the objective of our framework is to synthesize a modified volume that preserves the input-specific anatomical structure while altering only the targeted region in accordance with the conditions. As illustrated in Figure 1, the first stage involves extracting intermediate features from the input image during its diffusion process. Specifically, the input MRI is projected into latent space and passed through the UNet denoiser, from which multi-scale feature maps are collected. Features are extracted not only from residual pathways but also from attention mechanisms, thereby capturing both localized structural patterns and global contextual information. These features constitute structural priors that reflect the original patient-specific anatomy.

In the second stage, feature maps obtained from the guidance image are injected into the denoising trajectory of the generation process. This injection is performed at multiple levels of the UNet, ensuring that both fine-grained structural cues and high-level contextual dependencies are preserved. Simultaneously, text conditions

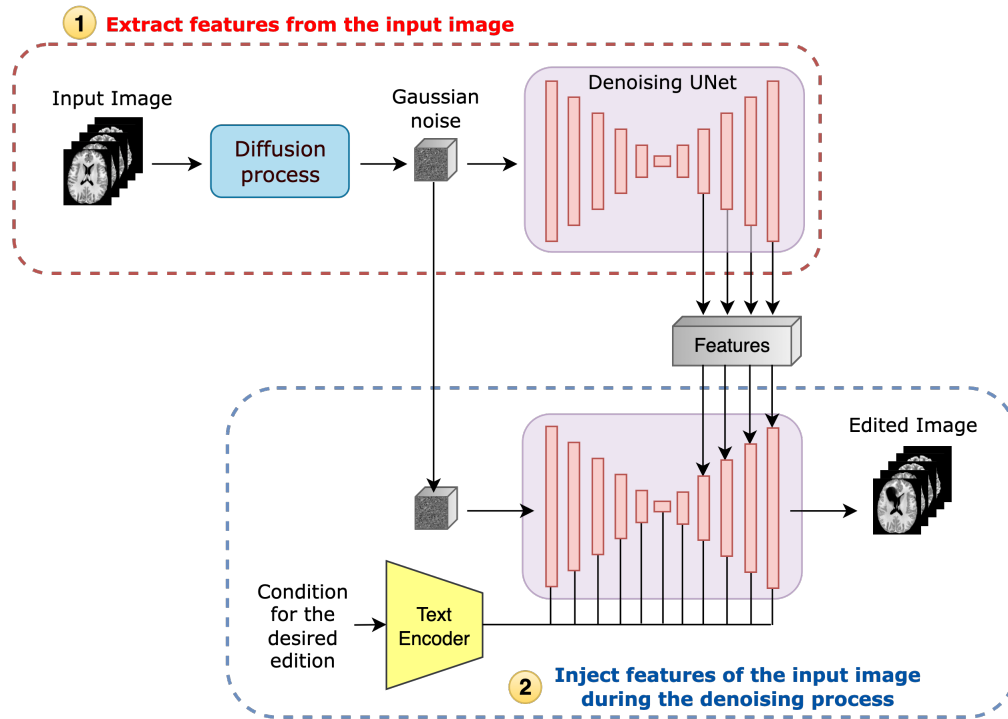


Figure 1: Overview of the proposed anatomically consistent editing framework using pretrained 3D latent diffusion models.

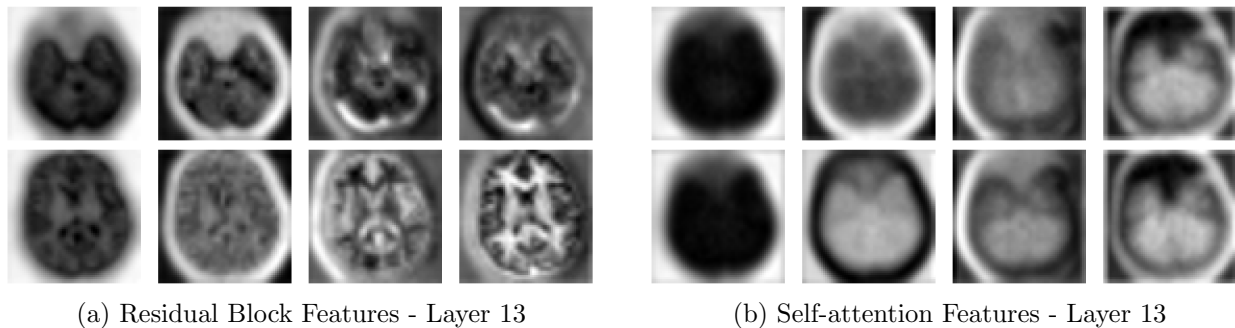


Figure 2: Visualization of intermediate feature representations extracted from the denoising U-Net at layer 13, comparing (a) residual block features and (b) self-attention features. Each panel shows representative feature maps (channels) from the same input slices. Residual block features emphasize localized anatomical textures and edges, whereas self-attention features capture more global, low-frequency contextual patterns across the image.

are encoded via a transformer-based text encoder and incorporated into the denoising process through cross-attention. The joint conditioning on input-derived features and text prompts constrains the generative process to produce anatomically faithful yet semantically modified outputs.

Figure 2 presents intermediate feature representations extracted from the denoising U-Net of the 3D LDM at layer 13 (high resolution). Panels (a) and (b) show features from the residual block and self-attention block, respectively. To visualize the high-dimensional feature maps, principal component analysis (PCA) was applied, and the first four principal components are retained for visualization. Residual block features exhibit pronounced structural patterns that resemble anatomical details of the brain, indicating their role in preserving fine-grained

spatial and anatomical fidelity. In contrast, self-attention features appear more spatially diffuse and globally organized across examples, consistent with their function in modeling long-range contextual dependencies within the 3D MRI volume.

3. RESULTS

3.1 Tumor Insertion into Healthy Brain MRI

We first evaluated our framework in the task of introducing synthetic tumors into healthy brain MRI volumes using the text condition. Healthy brain MRI data was conditioned with prompts specifying either IDH mutated or IDH wildtype. As shown in Figure 3, the framework successfully added the tumor across multiple MRI sequences while preserving the surrounding anatomical structure. The inserted tumors are consistent across contrasts, demonstrating the model’s ability to perform anatomically coherent, zero-shot image editing without retraining.

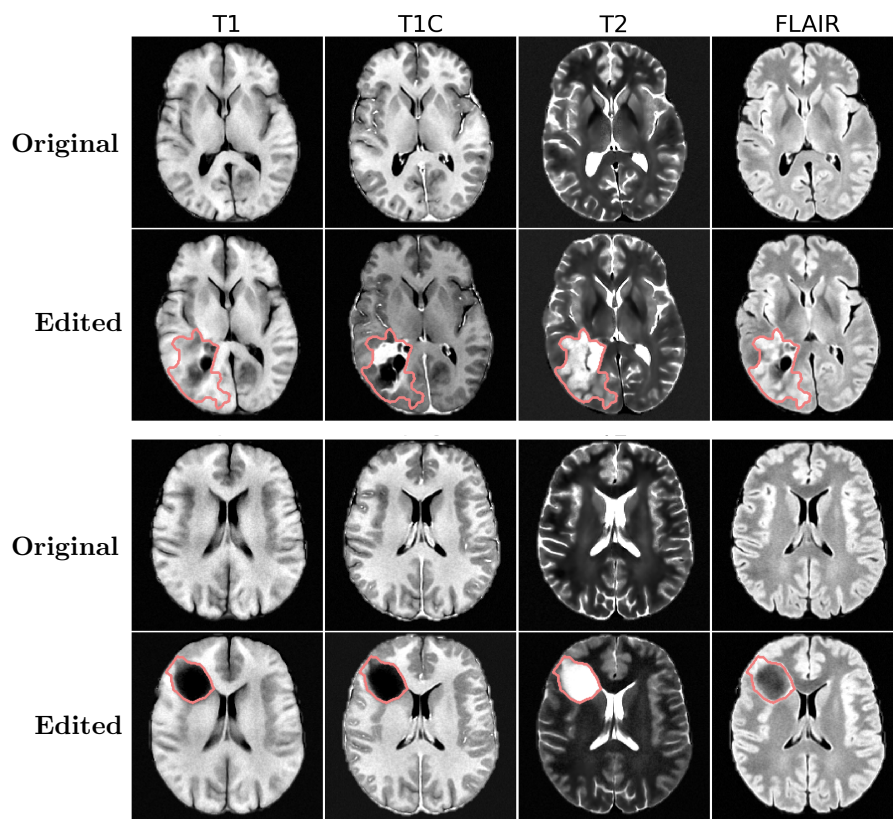


Figure 3: Tumor insertion into healthy brain MRI using text-conditioned LDMs.

3.2 IDH Molecular Status Manipulation of Existing Tumors

We next evaluated our framework in the task of modifying the IDH mutation status of pre-existing tumors while preserving patient-specific anatomy. This experiment leveraged the pretrained LDM with both tumor mask and text conditions. As shown in Figure 4, the framework successfully modified tumor appearance in accordance with the specified molecular status, demonstrating the ability to disentangle semantic representations of molecular subtypes.

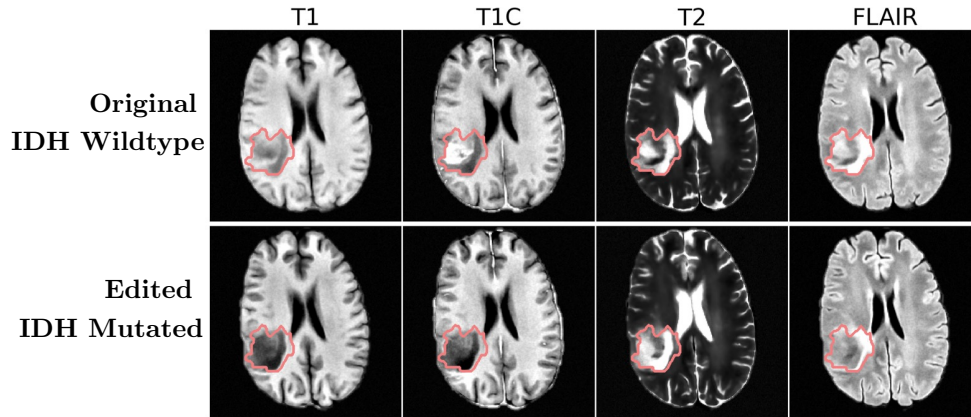


Figure 4: Manipulation of IDH molecular status using tumor mask and text conditions.

3.3 Tumor Removal for Healthy Brain Reconstruction

Finally, we assessed the ability of the framework to remove tumors and reconstruct underlying healthy brain structures. Conditioned on the prompt “no tumor,” the model eliminated pathological tissue and restored anatomically coherent brain parenchyma (Figure 5). Moreover, the model also corrected for the mass effect of the tumor.

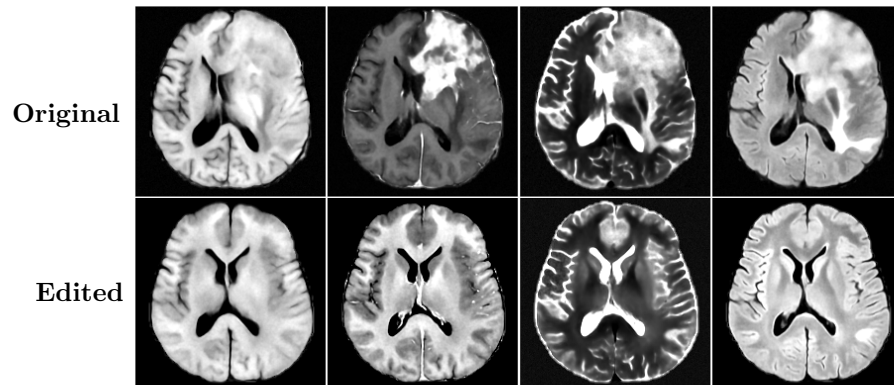


Figure 5: Tumor removal and healthy brain reconstruction.

4. CONCLUSIONS

We introduced a feature-guided, prompt-driven latent diffusion framework for 3D brain MRI editing that preserves anatomical integrity while enabling semantically meaningful modifications. Through tasks of tumor insertion, molecular status manipulation, and tumor removal, the method demonstrated flexibility and anatomical fidelity without retraining. Beyond IDH mutation status, the approach is extensible to other tumor subtypes, neurological disorders, and structural transformations, offering a powerful tool for data augmentation and counterfactual reasoning in neuroimaging.

ACKNOWLEDGMENTS

This research was supported by the NIH/NCI R01CA260705 (JAM).

DISCLOSURES

The authors declare no conflict of interest.

REFERENCES

- [1] Alaya, M. B., Lang, D. M., Wiestler, B., Schnabel, J. A., and Bercea, C. I., “MedEdit: Counterfactual Diffusion-based Image Editing on Brain MRI,” (July 2024).
- [2] Yeganeh, Y., Farshad, A., Charisiadis, I., Hasny, M., Hartenberger, M., Ommer, B., Navab, N., and Adeli, E., “Latent Drifting in Diffusion Models for Counterfactual Medical Image Synthesis,” (Apr. 2025).
- [3] Huang, G., Jin, R., Tang, Y., Zhao, C., Harada, T., Li, X., and Lin, G., “Interactive Tumor Progression Modeling via Sketch-Based Image Editing,” (Mar. 2025).
- [4] Zuo, C., Xue, J., and Yuan, C., “Multi channel fusion diffusion models for brain tumor MRI data augmentation,” *Scientific Reports* **15**, 22459 (July 2025).
- [5] Tumanyan, N., Geyer, M., Bagon, S., and Dekel, T., “Plug-and-Play Diffusion Features for Text-Driven Image-to-Image Translation,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 1921–1930 (2023).
- [6] Wu, J. Z., Ge, Y., Wang, X., Lei, S. W., Gu, Y., Shi, Y., Hsu, W., Shan, Y., Qie, X., and Shou, M. Z., “Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation,” in [*Proceedings of the IEEE/CVF International Conference on Computer Vision*], 7623–7633 (2023).
- [7] Qi, C., Cun, X., Zhang, Y., Lei, C., Wang, X., Shan, Y., and Chen, Q., “FateZero: Fusing Attentions for Zero-shot Text-based Video Editing,” in [*Proceedings of the IEEE/CVF International Conference on Computer Vision*], 15932–15942 (2023).
- [8] Mo, S., Mu, F., Lin, K. H., Liu, Y., Guan, B., Li, Y., and Zhou, B., “FreeControl: Training-Free Spatial Control of Any Text-to-Image Diffusion Model with Any Condition,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 7465–7475 (2024).
- [9] Truong, N. C. D., Yogananda, C. G. B., Wagner, B. C., Holcomb, J. M., Reddy, D., Saadat, N., Hatanpaa, K. J., Patel, T. R., Fei, B., Lee, M. D., Jain, R., Bruce, R. J., Pinho, M. C., Madhuranthakam, A. J., and Maldjian, J. A., “Synthesizing 3D multicontrast brain tumor MRIs using tumor mask conditioning,” in [*Medical Imaging 2024: Imaging Informatics for Healthcare, Research, and Applications*], **12931**, 116–120, SPIE (Apr. 2024).
- [10] Truong, N. C. D., Yogananda, C. G. B., Wagner, B. C., Saadat, N., Holcomb, J. M., Reddy, D., Lodhi, S., Bowerman, J., Hatanpaa, K. J., Patel, T. R., Fei, B., Lee, M. D., Jain, R., Bruce, R. J., Pinho, M. C., Madhuranthakam, A. J., and Maldjian, J. A., “Mitigating data scarcity in the classification of glioma molecular subtypes: The power of generative imaging,” in [*Medical Imaging 2025: Imaging Informatics*], **13411**, 115–121, SPIE (Apr. 2025).
- [11] Truong, N. C. D., Bangalore Yogananda, C. G., Wagner, B. C., Holcomb, J. M., Reddy, D. D., Saadat, N., Bowerman, J., Hatanpaa, K. J., Patel, T. R., Fei, B., Lee, M. D., Jain, R., Bruce, R. J., Madhuranthakam, A. J., Pinho, M. C., and Maldjian, J. A., “Categorical and phenotypic image synthetic learning as an alternative to federated learning,” *Nature Communications* **16**, 9384 (Oct. 2025).