POWDR: Pathology-preserving Outpainting with Wavelet Diffusion for 3D MRI

Fei Tan[1] (iD)   FEI.TAN@GEHEALTHCARE.COM
[1] *GE HealthCare, San Ramon, CA, USA*

Ashok Vardhan Addala[1]   ASHOKVARDHAN.ADDALA@GEHEALTHCARE.COM
Bruno Astuto Arouche Nunes[1] (iD)   BRUNO.NUNES@GEHEALTHCARE.COM
Xucheng Zhu[2] (iD)   XUCHENG.ZHU@GEHEALTHCARE.COM
[2] *GE HealthCare, Menlo Park, CA, USA*

Ravi Soni[1]   RAVI.SONI@GEHEALTHCARE.COM

Abstract

Medical imaging datasets often suffer from class imbalance and limited availability of pathology-rich cases, which constrains the performance of machine learning models for segmentation, classification, and vision–language tasks. To address this challenge, we propose POWDR, a pathology-preserving outpainting framework for 3D MRI based on a conditioned wavelet diffusion model. Unlike conventional augmentation or unconditional synthesis, POWDR retains real pathological regions while generating anatomically plausible surrounding tissue, enabling diversity without fabricating lesions.

Our approach leverages wavelet-domain conditioning to enhance high-frequency detail and mitigate blurring common in latent diffusion models. We introduce a random connected mask training strategy to overcome conditioning-induced collapse and improve diversity outside the lesion. POWDR is evaluated on brain MRI using BraTS datasets and extended to knee MRI to demonstrate tissue-agnostic applicability. Quantitative metrics (FID, SSIM, LPIPS) confirm image realism, while diversity analysis shows significant improvement with random-mask training (cosine similarity reduced from 0.9947 to 0.9580; KL divergence increased from 0.00026 to 0.01494). Clinically relevant assessments reveal gains in tumor segmentation performance using nnU-Net, with Dice scores improving from 0.6992 to 0.7137 when adding 50 synthetic cases. Tissue volume analysis indicates no significant differences for CSF and GM compared to real images.

These findings highlight POWDR as a practical solution for addressing data scarcity and class imbalance in medical imaging. The method is extensible to multiple anatomies and offers a controllable framework for generating diverse, pathology-preserving synthetic data to support robust model development.

Keywords: Diffusion Models, Outpainting, 3D MRI, Pathology Preservation, Medical Image Synthesis

1. Introduction

The development of medical imaging analysis methodologies using deep learning is hindered by two major challenges: the scarcity of pathological data and pronounced class imbalance. Acquiring large-scale, diverse datasets is resource-intensive, and many pathological conditions occur at low prevalence, limiting the availability of representative samples. These

constraints impede the training of robust and generalizable models. Consequently, developing strategies to augment the availability of pathology-rich images without incurring additional acquisition costs is critical.

Generative models have emerged as a promising solution to bridge this gap. Among them, diffusion models have demonstrated strong capabilities in synthesizing realistic medical images (Khader et al., 2023). While unconditional diffusion models can generate plausible images, they often fail to capture rare pathological patterns (Dar et al., 2025). Existing approaches such as inpainting have primarily focused on reconstructing healthy regions (Kofler et al., 2024) or performing super-resolution across slices (Kang et al., 2021). In contrast, outpainting—the process of expanding an image beyond its original boundaries while preserving a given region—remains underexplored in medical imaging. Outpainting offers a unique advantage: it allows retention of real pathological regions while generating surrounding healthy tissue, thereby creating diverse and clinically relevant training samples. This work leverages conditional diffusion models to achieve this goal.

Latent Diffusion Models (LDMs) (Rombach et al., 2022) have become the standard for image synthesis; however, they struggle to recover high-frequency details (Lozupone et al., 2025; Falck et al., 2025), often producing blurry outputs. Wavelet-based diffusion models (Gerdes et al., 2024; Friedrich et al., 2025) address this limitation by decomposing images into high- and low-frequency components and conditioning generation on these channels, thereby enhancing fine structural details. Although wavelet diffusion has been applied to contrast synthesis, its potential for outpainting tasks in medical imaging remains untapped.

In this study, we introduce POWDR, Pathology-preserving Outpainting with conditioned Wavelet Diffusion for 3D MRI. The main contributions of our work are threefold:

1. We propose a novel outpainting framework that conditions on pathological regions and synthesizes surrounding anatomy, enabling generation of full MRI volumes with real pathology.

2. We demonstrate the effectiveness of POWDR through a comprehensive evaluation using quantitative image quality metrics (FID, SSIM, LPIPS) and clinically relevant assessments, including segmentation accuracy and tissue volume comparisons against real patient data.

3. We demonstrate that our approach is tissue-agnostic, validating its applicability on brain and knee MRI and discussing its extensibility to other anatomical regions.

2. Methods

2.1. Network Design

2.1.1. Training

Our architecture (Figure 1) builds upon the conditional wavelet diffusion framework (Friedrich et al., 2024) with several enhancements for 3D medical image synthesis. In the forward diffusion process, the original 3D MRI volume $y_0 \in \mathbb{R}^{H \times W \times D}$ is decomposed using a digital wavelet transform (DWT) with Haar kernel (Stanković and Falkowski, 2003) into

$x_0 \in \mathbb{R}^{8 \times \frac{H}{2} \times \frac{W}{2} \times \frac{D}{2}}$ which has eight subbands:

$$x_0 = DWT(y_0) = \{LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH\}$$

representing low- and high-frequency components across different orientations. Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ is progressively added according to a linear β-schedule ranging from 0.0001 to 0.02. Resulting in the noised image $x_t = \sqrt{(\bar{\alpha}_t)}x_0 + (1 - \bar{\alpha}_t)\epsilon$, where $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, $\alpha_t = 1 - \beta_t$.

For conditional denoising, we employ a 3D ResUNet (Jha et al., 2019) architecture. The conditioning image is also transformed into wavelet subbands and concatenated with the noisy target subbands $[\widetilde{x}_t, c]$ at each denoising step, forming 16 input channels. The conditioning image during training can be either a masked pathology region or a randomly masked region, as detailed in the Dataset section. The network outputs $\widetilde{x}_{t-1}$ and estimates $\widetilde{x}_0$, containing 8 channels, and the mean squared error (MSE) in the wavelet domain is adopted as loss function

$$L = MSE_{wavelet} = \|x_0 - \widetilde{x}_0\|^2$$

Other key training parameters include, dropout rate = 0.1, training 200,000 iterations, diffusion time step = 1000, ResUNet base channel size = 64, channel multipliers = [1, 2, 2, 4, 4], AdamW optimizer (Loshchilov and Hutter, 2019), learning rate = 1×10^{-5}. All experiments were conducted on an NVIDIA H100 GPU (80 GB VRAM).
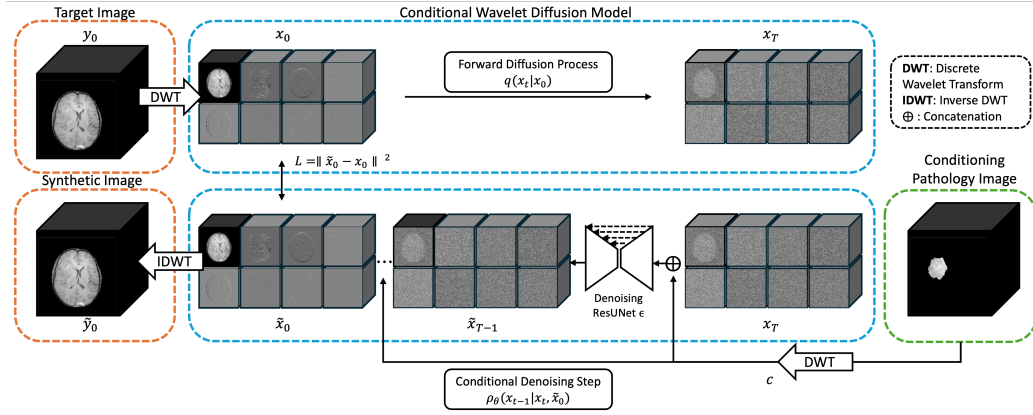

Figure 1: Architecture of POWDR: a pathology-preserving outpainting framework using conditioned wavelet diffusion model for 3D MRI. The pipeline begins with wavelet decomposition and Gaussian noise addition during forward diffusion. In the conditional denoising step, the wavelet-transformed masked pathology is concatenated with the noisy input and processed by an attention-based ResUNet. The inverse wavelet transform recombines frequency components to reconstruct the synthesized image

2.1.2. Sampling

During inference, masked pathology region conditioning is achieved by element-wise multiplication of the pathology mask with the corresponding intensity volume (e.g., tumor mask with brain image or cartilage mask with knee image). This approach ensures that pathological regions are preserved while enabling synthesis of surrounding anatomy. During inference, we apply 1,000 denoising steps following the Denoising Diffusion Probabilistic Model (DDPM) sampling strategy (Ho et al., 2020). The generated sample, initially in latent space, is then reconstructed into the image domain using the Inverse Discrete Wavelet Transform (IDWT).

$$\widetilde{y}_0 = IDWT(\widetilde{x}_0)$$

2.2. Datasets

2.2.1. Brain MRI

We use the BraTS 2023 dataset for training (1251 cases, T1n contrast) and BraTS 2024 additional cases for evaluation (100 cases for sampling, 171 cases for segmentation testing) (Karargyris et al., 2023; Baid et al., 2021; Bakas et al., 2017; Menze et al., 2015). The same approach can be expanded to other contrasts (T2w, T1 contrast-enhanced, T2-FLAIR).

The default training strategy utilizes masked tumor region as training condition. To increase conditioning diversity, we introduce additional training strategy, which generates random connected masks multiplied by image volume as training conditions. These masks are 6-connected and sampled based on tumor volume distribution from the training set to match realistic tumor sizes. All images are intensity-normalized to the 1st-99th percentile, and cropped into $224 \times 224 \times 155$ matrix size and 1mm isotropic resolution. During sampling for both training scenarios, only the tumor region is used for conditioning to ensure pathology preservation.

2.2.2. Knee MRI

We further validate our approach on a knee MRI dataset comprising 768 3D volumes collected across multiple scanners and two distinct MRI sequences. The first subset is a public dataset acquired using Double Echo Steady State (DESS) sequences on Siemens scanners, originally described by (Balamoody et al., 2010) and used in a segmentation challenge (Desai et al., 2021). This subset includes 176 MRI volumes from 88 patients, with manual 3D segmentation masks for four tissue compartments—patellar, tibial, and femoral cartilage, as well as the menisci—generated by a single expert reader (Desai et al., 2021).

The second subset is an internal dataset of 592 studies from 275 unique patients. These volumes were acquired using CUBE, a high-resolution 3D Fast Spin Echo (FSE) sequence developed by GE Healthcare, and previously utilized by (Astuto et al., 2021) for anomaly and object detection. Scans were obtained on five 3-T MRI scanners (GE Healthcare, Waukesha, WI) between 2006 and 2018, spanning multiple scanner models and sites.

All knee MRI volumes were preprocessed to ensure consistency: cropped to a common field-of-view ($144 \times 144 \times 80$ mm), resized to $288 \times 288 \times 160$ voxels, and resampled to 0.5 mm isotropic resolution. Cartilage and meniscus masks were combined and dilated by 3 pixels using a spherical kernel to account for thin structures, and element-wise multiplied with

image volume as condition. Both healthy and pathological cases were included in training, while sampling was conditioned on meniscus tear pathology. This setup demonstrates that the model can leverage both healthy and pathological images during training while enabling pathology-conditioned synthesis—a capability particularly valuable when pathological cases are scarce.

Approval for data use was obtained from the Osteoarthritis Initiative (OAI) (https://nda.nih.gov/oai) for the DESS dataset and from the University of California, San Francisco (UCSF) Institutional Review Board for all other datasets. All datasets were de-identified, and informed consent was obtained from all patients through these institutions. All methods were performed in accordance with relevant guidelines and regulations.

2.3. Evaluation

2.3.1. QUANTITATIVE METRICS

We report standard generative metrics: Fréchet Inception Distance (FID) (Heusel et al., 2018) measuring how close the distribution of generated images is to real images, Multi-Scale Structural Similarity (MS-SSIM) (Wang et al., 2003) evaluating preservation of anatomical details, and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018) assessing how visually similar two images appear to a human observer. FID was computed using Med3D network (Chen et al., 2019) features for medical relevance. MS-SSIM and LPIPS, calculated separately inside and outside the tumor region to assess pathology preservation and anatomical diversity.

To quantify sample diversity from repeated sampling of the same conditioning input, we computed pixel-wise mean and standard deviation across samples to assess spatial variation in generated outputs, as well as pairwise similarity metrics across all unique sample pairs. Cosine similarity was calculated by flattening each 3D volume into a 1D vector and computing the normalized dot product between sample pairs, where lower values indicate greater diversity. KL divergence was computed by binning intensity values into 50-bin histograms to form probability distributions for each sample, followed by calculating the Kullback–Leibler divergence between all pairs using the formula: $\mathrm{KL}(P \parallel Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$. Higher KL values indicate more diverse intensity distributions. For N samples, this resulted in $N(N-1)/2$ pairwise comparisons, and the mean across all pairs was reported as the final diversity metric.

2.3.2. CLINICAL RELEVANCE

We evaluate downstream utility via whole tumor segmentation using nnU-Net (Isensee et al., 2021). The baseline uses 1251 BraTS 2023 cases + 100 BraTS 2024 Additional cases. Synthetic augmentation experiments add 10, 50, or 100 generated images conditioned on real pathology extracted from the same 100 BraTS 2024 cases. This training set selection ensures the performance variation across models are introduced by the synthetic POWDR images, instead of additional pathologies in the training set. nnU-Net is trained for 1000 epochs with default parameters, and Dice scores are computed on 171 unseen test cases from BraTS 2024.

Additionally, we assess anatomical plausibility by comparing brain tissue volumes (cerebral spinal fluid: CSF, gray matter: GM, and white matter: WM) between real and syn-

thetic images using FSL FAST segmentation (Zhang et al., 2001). We analyze 100 synthetic and 100 real cases, reporting volume statistics.

3. Results

3.1. Brain MRI

3.1.1. Qualitative Results

Figure 2 illustrates three-plane views of the conditioning pathology image, the sampled synthetic image, and the original image where the conditioning pathology is extracted. The synthetic image differs substantially from the target image while preserving the pathology region. Brain structures are clearly visible, and tumors are seamlessly integrated without sharp edges or artifacts, indicating realistic anatomical synthesis. We also trained a conditional latent diffusion outpainting model (cLDM) (Rombach et al., 2022) as a baseline for comparison. However, as shown in Figure 2D, it failed to generate anatomically plausible images and was therefore excluded from the following quantitative analysis.
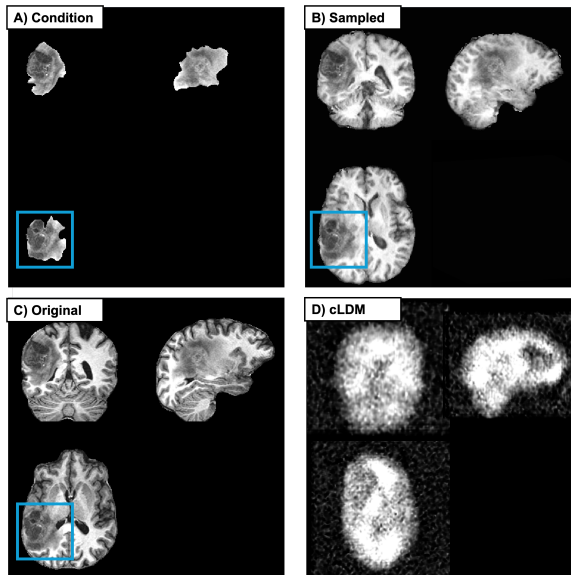

Figure 2: Qualitative example of brain MRI outpainting. Shown are A) the input condition (masked tumor), B) the synthetic image generated by POWDR, and C) the original image from which the tumor was extracted. Highlighted boxes indicate tumor regions preserved during synthesis. D) Synthetic image generated using a conditional Latent Diffusion outpainting model failed to converge.

3.1.2. Quantitative Results

We evaluate similarity to real images using FID, SSIM, and MSE (Table 1). Low FID scores of 0.0042 between generated and original cases indicates that synthesized images closely

match the distribution of real data. For reference, the baseline FID score computed between two sets of real images is 0.0005. LPIPS analysis reveals high similarity within tumor regions (LPIPS-in 0.0030), confirming pathology preservation, and greater dissimilarity outside tumors (LPIPS-out 0.1484), suggesting successful augmentation of non-pathological areas. High MS-SSIM of 0.9998 inside tumor, 0.8195 outside of tumor further supports structural consistency within tumor region and diversity yet realism outside of tumor region.

Table 1: Quantitative evaluation of image realism and diversity. Metrics include FID, MS-SSIM, and LPIPS, computed inside and outside the tumor region. Downward arrows indicate lower values are better; upward arrows indicate higher values are better; a dash ($-$) denotes mid-to-high values are optimal.

Metrics	Mean (Std) - Real vs. Synthetic	Baseline - Real vs. Real
FID $\downarrow$	0.0042	0.0005
MS SSIM - inside tumor $\uparrow$	0.9998 (0.0003)	-
MS SSIM - outside tumor -	0.8195 (0.0517)	-
LPIPS - inside tumor $\downarrow$	0.0030 (0.0017)	-
LPIPS - outside tumor $\uparrow$	0.1484 (0.0311)	-

To assess diversity, we perform repeated sampling from the same conditioning input. Figure 3A shows that when trained with masked tumor conditioning, generated samples exhibit small variation—an issue commonly observed in conditioned diffusion models. The mean sample across 20 repetitions appears nearly identical, and the standard deviation is low, indicating potential overfitting. To address this, we introduce random connected masks during training. Figure 3B shows a substantial increase in standard deviation across repeated samples in regions outside the pathology, while remaining minimal within the conditioned pathology area, when trained with random connected masked region conditioning. The two example outputs further illustrate markedly variability in anatomical structures, especially in anatomical structures such as CSF cavities and brain boundaries. Figure 3C further proves that strategy B significantly improves diversity: Cosine similarity between pairwise samples decreases from 0.9947 ± 0.00003 to 0.9580 ± 0.0066. KL divergence increases from 0.00026 to 0.01494.

This diversity enhancement is critical for downstream tasks. With tumor-mask conditioning (Strategy A), the maximum augmentation equals the number of unique conditions N, doubling the dataset at best. In contrast, random-mask training (Strategy B) enables generating nN non-redundant samples, where n is user-defined, which can be optimized to balance flexible augmentation without overwhelming the real data in downstream tasks.

3.1.3. Clinical Relevance

We assess utility for tumor segmentation using nnU-Net. To isolate the effect of POWDR as data augmentation, synthetic images are generated from the same 100 BraTS 2024 cases used in segmentation training. Dice scores improve from 0.6992 ± 0.3009 (baseline) to 0.7015 ± 0.2955 with 10 synthetic cases, plateauing at 0.7137 ± 0.2868 for 50 cases and
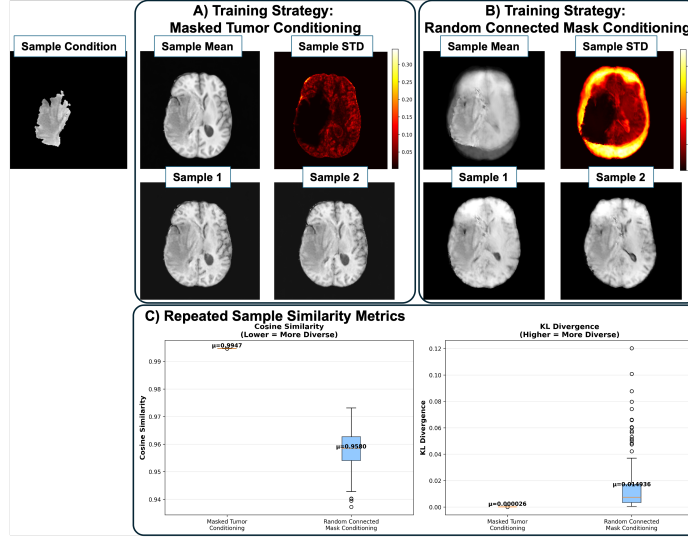
Figure 3: Sampling diversity under different training strategies. A–B) Mean, standard deviation, and example outputs from 20-case repeated sampling using A) masked tumor conditioning and B) random connected mask conditioning. C) Similarity metrics cosine similarity and KL divergence quantify diversity improvements.

0.7135 ± 0.2866 for 100 cases (Figure 4). This suggests that moderate augmentation yields optimal gains.
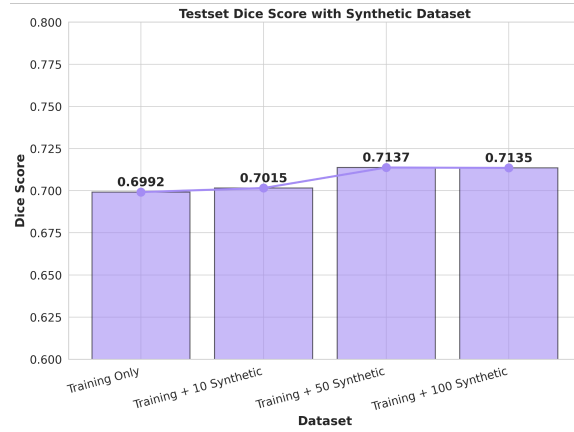

Figure 4: Impact of synthetic augmentation on tumor segmentation. Dice scores for nnU-Net segmentation under varying numbers of synthetic cases (10, 50, 100) compared to baseline.

For anatomical credibility, we perform brain tissue segmentation using FSL FAST algorithm. As shown in Figure 5, volumes of CSF (Synthetic: 323.428 ± 32.001 mL, Real: 325.397 ± 49.228 mL) and GM (Synthetic: 534.475 ± 29.312 mL, Real: 524.283 ± 66.766 mL) show no significant difference from real images. WM volume is slightly lower in synthetic images, likely due to tumor presence affecting segmentation.
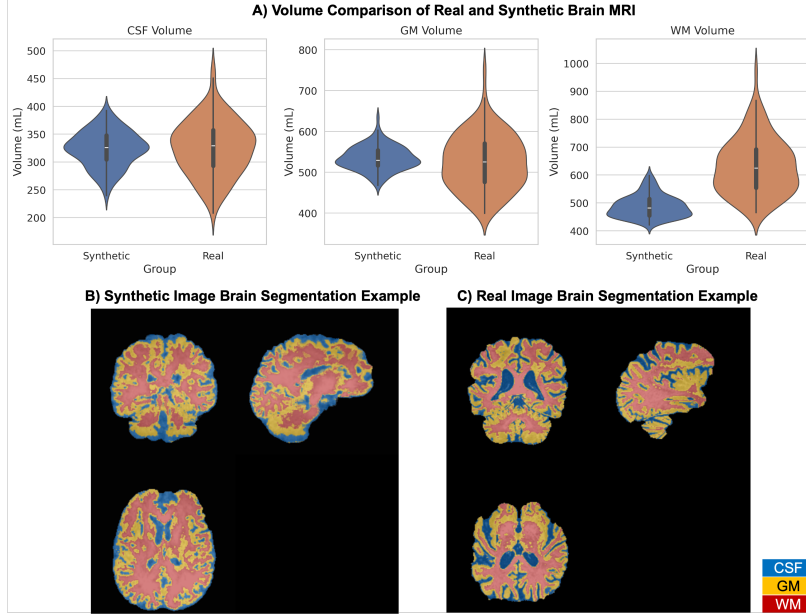

Figure 5: Comparison of brain tissue volumes between real and synthetic images. Volumes of CSF, gray matter, and white matter segmented using FSL FAST for 100 synthetic and 100 real cases.

3.1.4. KNEE MRI

Figure 6 shows qualitative results for 3D knee MRI outpainting. Despite larger matrix size and longer training time, the model successfully preserves meniscus tears while generating realistic surrounding anatomy. This demonstrates the model is anatomy agnostic and highlights the method's potential for pathology augmentation in low-prevalence conditions. We note that the contrast of the sampled images differs noticeably from the original images, likely due to variations in MR image contrast and scanner vendors present in the diffusion model's training set.
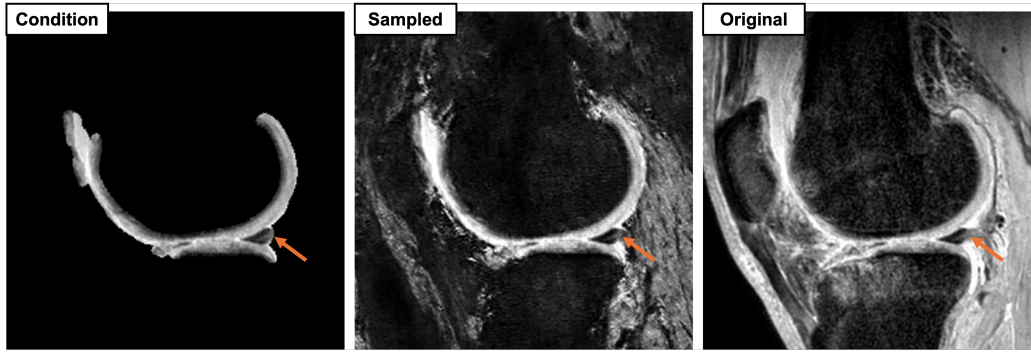
Figure 6: Qualitative results for knee MRI outpainting. The model preserves meniscus tears while generating realistic surrounding anatomy, demonstrating the framework is tissue agnostic.

4. Conclusion

In this work, we introduced POWDR, a pathology-preserving outpainting framework for 3D MRI based on a conditioned wavelet diffusion model. To our knowledge, this is among the first approaches to tackle outpainting in MRI while explicitly preserving real pathological regions. We conducted a comprehensive evaluation, including conventional generative metrics (FID, SSIM, LPIPS), diversity analysis through repeated sampling, and clinically relevant assessments such as tumor segmentation performance and brain tissue volume comparisons. Furthermore, we demonstrated the tissue-agnostic nature of our method by successfully applying it to knee MRI, highlighting its potential for broader anatomical domains. These results suggest that POWDR can serve as a practical solution for addressing data scarcity and class imbalance in 3D medical imaging, with wide-ranging applications in segmentation, classification, localization, and vision–language tasks.

References

Bruno Astuto, Io Flament, Nikan K. Namiri, Rutwik Shah, Upasana Bharadwaj, Thomas M. Link, Matthew D. Bucknor, Valentina Pedoia, and Sharmila Majumdar. Automatic Deep Learning–assisted Detection and Grading of Abnormalities in Knee MRI Studies. *Radiology: Artificial Intelligence*, 3(3):e200165, April 2021. ISSN 2638-6100. doi: 10.1148/ryai.2021200165. URL http://pubs.rsna.org/doi/10.1148/ryai.2021200165.

Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C. Kitamura, Sarthak Pati, Luciano M. Prevedello, Jeffrey D. Rudie, Chiharu Sako, Russell T. Shinohara, Timothy Bergquist, Rong Chai, James Eddy, Julia Elliott, Walter Reade, Thomas Schaffter, Thomas Yu, Jiaxin Zheng, Ahmed W. Moawad, Luiz Otavio Coelho, Olivia McDonnell, Elka Miller, Fanny E. Moron, Mark C. Oswood, Robert Y. Shih, Loizos Siakallis, Yulia Bronstein, James R. Mason, Anthony F. Miller, Gagandeep Choudhary, Aanchal Agarwal, Cristina H. Besada, Jamal J. Derakhshan, Mariana C. Diogo, Daniel D. Do-Dai, Luciano Farage, John L. Go, Mohiuddin Hadi, Virginia B. Hill, Michael Iv, David Joyner, Christie Lincoln, Eyal Lotan, Asako Miyakoshi, Mariana Sanchez-Montano, Jaya Nath, Xuan V. Nguyen, Manal Nicolas-Jilwan, Johanna Ortiz Jimenez, Kerem Ozturk, Bojan D. Petrovic, Chintan Shah, Lubdha M. Shah, Manas Sharma, Onur Simsek, Achint K. Singh, Salil Soman, Volodymyr Statsevych, Brent D. Weinberg, Robert J. Young, Ichiro Ikuta, Amit K. Agarwal, Sword C. Cambron, Richard Silbergleit, Alexandru Dusoi, Alida A. Postma, Laurent Letourneau-Guillon, Gloria J. Guzman Perez-Carrillo, Atin Saha, Neetu Soni, Greg Zaharchuk, Vahe M. Zohrabian, Yingming Chen, Milos M. Cekic, Akm Rahman, Juan E. Small, Varun Sethi, Christos Davatzikos, John Mongan, Christopher Hess, Soonmee Cha, Javier Villanueva-Meyer, John B. Freymann, Justin S. Kirby, Benedikt Wiestler, Priscila Crivellaro, Rivka R. Colen, Aikaterini Kotrotsou, Daniel Marcus, Mikhail Milchenko, Arash Nazeri, Hassan Fathallah-Shaykh, Roland Wiest, Andras Jakab, Marc-Andre Weber, Abhishek Mahajan, Bjoern Menze, Adam E. Flanders, and Spyridon Bakas. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, September 2021. URL http://arxiv.org/abs/2107.02314. arXiv:2107.02314 [cs].

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1):170117, September 2017. ISSN 2052-4463. doi: 10.1038/sdata.2017.117. URL https://www.nature.com/articles/sdata2017117.

Sharon Balamoody, Tomos G Williams, John C Waterton, Michael Bowes, Richard Hodgson, Chris J Taylor, and Charles E Hutchinson. Comparison of 3T MR scanners in regional cartilage-thickness analysis in osteoarthritis: a cross-sectional multicenter, multivendor study. *Arthritis Research & Therapy*, 12(5):R202, October 2010. ISSN 1478-6354. doi: 10.1186/ar3174. URL https://arthritis-research.biomedcentral.com/articles/10.1186/ar3174.

Sihong Chen, Kai Ma, and Yefeng Zheng. Med3D: Transfer Learning for 3D Medical Image Analysis, July 2019. URL http://arxiv.org/abs/1904.00625. arXiv:1904.00625 [cs].

Salman Ul Hassan Dar, Marvin Seyfarth, Isabelle Ayx, Theano Papavassiliu, Stefan O. Schoenberg, Robert Malte Siepmann, Fabian Christopher Laqua, Jannik Kahmann, Norbert Frey, Bettina Baeßler, Sebastian Foersch, Daniel Truhn, Jakob Nikolas Kather, and Sandy Engelhardt. Unconditional latent diffusion models memorize patient imaging data. *Nature Biomedical Engineering*, August 2025. ISSN 2157-846X. doi: 10.1038/s41551-025-01468-8. URL https://www.nature.com/articles/s41551-025-01468-8.

Arjun D. Desai, Francesco Caliva, Claudia Iriondo, Aliasghar Mortazi, Sachin Jambawalikar, Ulas Bagci, Mathias Perslev, Christian Igel, Erik B. Dam, Sibaji Gaj, Mingrui Yang, Xiaojuan Li, Cem M. Deniz, Vladimir Juras, Ravinder Regatte, Garry E. Gold, Brian A. Hargreaves, Valentina Pedoia, Akshay S. Chaudhari, on behalf of the IWOAI Segmentation Challenge Writing Group, Naji Khosravan, Drew Torigian, Jutta Ellermann, Mehmet Akcakaya, Radhika Tibrewala, Io Flament, Matthew O'Brien, Sharmila Majumdar, Kunio Nakamura, and Akshay Pai. The International Workshop on Osteoarthritis Imaging Knee MRI Segmentation Challenge: A Multi-Institute Evaluation and Analysis Framework on a Standardized Dataset. *Radiology: Artificial Intelligence*, 3(3):e200078, May 2021. ISSN 2638-6100. doi: 10.1148/ryai.2021200078. URL http://pubs.rsna.org/doi/10.1148/ryai.2021200078.

Fabian Falck, Teodora Pandeva, Kiarash Zahirnia, Rachel Lawrence, Richard Turner, Edward Meeds, Javier Zazo, and Sushrut Karmalkar. A Fourier Space Perspective on Diffusion Models, May 2025. URL http://arxiv.org/abs/2505.11278. arXiv:2505.11278 [stat].

Paul Friedrich, Alicia Durrer, Julia Wolleb, and Philippe C. Cattin. cWDM: Conditional Wavelet Diffusion Models for Cross-Modality 3D Medical Image Synthesis, November 2024. URL http://arxiv.org/abs/2411.17203. arXiv:2411.17203 [eess].

Paul Friedrich, Julia Wolleb, Florentin Bieder, Alicia Durrer, and Philippe C. Cattin. WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis. pages 11–21. 2025. doi: 10.1007/978-3-031-72744-3_2. URL http://arxiv.org/abs/2402.19043. ISSN: 0302-9743, 1611-3349 arXiv:2402.19043 [eess].

Mathis Gerdes, Max Welling, and Miranda C. N. Cheng. GUD: Generation with Unified Diffusion, October 2024. URL http://arxiv.org/abs/2410.02667. arXiv:2410.02667 [cs].

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, January 2018. URL http://arxiv.org/abs/1706.08500. arXiv:1706.08500 [cs].

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020. URL http://arxiv.org/abs/2006.11239. arXiv:2006.11239 [cs].

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, February 2021. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-020-01008-z. URL https://www.nature.com/articles/s41592-020-01008-z.

Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D. Johansen. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255, San Diego, CA, USA, December 2019. IEEE. ISBN 978-1-7281-5606-4. doi: 10.1109/ISM46123.2019.00049. URL https://ieeexplore.ieee.org/document/8959021/.

Seung Kwan Kang, Seong A. Shin, Seongho Seo, Min Soo Byun, Dong Young Lee, Yu Kyeong Kim, Dong Soo Lee, and Jae Sung Lee. Deep learning-Based 3D inpainting of brain MR images. *Scientific Reports*, 11(1):1673, January 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-80930-w. URL https://www.nature.com/articles/s41598-020-80930-w.

Alexandros Karargyris, Renato Umeton, Micah J. Sheller, Alejandro Aristizabal, Johnu George, Anna Wuest, Sarthak Pati, Hasan Kassem, Maximilian Zenk, Ujjwal Baid, Prakash Narayana Moorthy, Alexander Chowdhury, Junyi Guo, Sahil Nalawade, Jacob Rosenthal, David Kanter, Maria Xenochristou, Daniel J. Beutel, Verena Chung, Timothy Bergquist, James Eddy, Abubakar Abid, Lewis Tunstall, Omar Sanseviero, Dimitrios Dimitriadis, Yiming Qian, Xinxing Xu, Yong Liu, Rick Siow Mong Goh, Srini Bala, Victor Bittorf, Sreekar Reddy Puchala, Biagio Ricciuti, Soujanya Samineni, Eshna Sengupta, Akshay Chaudhari, Cody Coleman, Bala Desinghu, Gregory Diamos, Debo Dutta, Diane Feddema, Grigori Fursin, Xinyuan Huang, Satyananda Kashyap, Nicholas Lane, Indranil Mallick, FeTS Consortium, BraTS-2020 Consortium, AI4SafeChole Consortium, Pietro Mascagni, Virendra Mehta, Cassiano Ferro Moraes, Vivek Natarajan, Nikola Nikolov, Nicolas Padoy, Gennady Pekhimenko, Vijay Janapa Reddi, G. Anthony Reina, Pablo Ribalta, Abhishek Singh, Jayaraman J. Thiagarajan, Jacob Albrecht, Thomas Wolf, Geralyn Miller, Huazhu Fu, Prashant Shah, Daguang Xu, Poonam Yadav, David Talby, Mark M. Awad, Jeremy P. Howard, Michael Rosenthal, Luigi Marchionni, Massimo Loda, Jason M. Johnson, Spyridon Bakas, and Peter Mattson. Federated benchmarking of medical artificial intelligence with MedPerf. *Nature Machine Intelligence*, 5(7):799–810, July 2023. ISSN 2522-5839. doi: 10.1038/s42256-023-00652-2. URL https://www.nature.com/articles/s42256-023-00652-2. Publisher: Springer Science and Business Media LLC.

Firas Khader, Gustav Mueller-Franzes, Soroosh Tayebi Arasteh, Tianyu Han, Christoph Haarburger, Maximilian Schulze-Hagen, Philipp Schad, Sandy Engelhardt, Bettina Baessler, Sebastian Foersch, Johannes Stegmaier, Christiane Kuhl, Sven Nebelung, Jakob Nikolas Kather, and Daniel Truhn. Medical Diffusion: Denoising Diffusion Probabilistic Models for 3D Medical Image Generation, January 2023. URL http://arxiv.org/abs/2211.03364. arXiv:2211.03364 [eess].

Florian Kofler, Felix Meissen, Felix Steinbauer, Robert Graf, Stefan K. Ehrlich, Annika Reinke, Eva Oswald, Diana Waldmannstetter, Florian Hoelzl, Izabela Horvath, Oezguen Turgut, Suprosanna Shit, Christina Bukas, Kaiyuan Yang, Johannes C. Paetzold, Ezequiel de da Rosa, Isra Mekki, Shankeeth Vinayahalingam, Hasan Kassem, Juexin Zhang, Ke Chen, Ying Weng, Alicia Durrer, Philippe C. Cattin, Julia Wolleb, M. S. Sadique, M. M. Rahman, W. Farzana, A. Temtam, K. M. Iftekharuddin, Maruf Adewole, Syed Muhammad Anwar, Ujjwal Baid, Anastasia Janas, Anahita Fathi Kazerooni, Dominic LaBella, Hongwei Bran Li, Ahmed W. Moawad, Gian-Marco Conte, Keyvan Farahani, James Eddy, Micah Sheller, Sarthak Pati, Alexandros Karagyris, Alejandro Aristizabal, Timothy Bergquist, Verena Chung, Russell Takeshi Shinohara, Farouk Dako, Walter Wiggins, Zachary Reitman, Chunhao Wang, Xinyang Liu, Zhifan Jiang, Elaine Johanson, Zeke Meier, Ariana Familiar, Christos Davatzikos, John Freymann, Justin Kirby, Michel Bilello, Hassan M. Fathallah-Shaykh, Roland Wiest, Jan Kirschke, Rivka R. Colen, Aikaterini Kotrotsou, Pamela Lamontagne, Daniel Marcus, Mikhail Milchenko, Arash Nazeri, Marc-André Weber, Abhishek Mahajan, Suyash Mohan, John Mongan, Christopher Hess, Soonmee Cha, Javier Villanueva-Meyer, Errol Colak, Priscila Crivellaro, Andras Jakab, Abiodun Fatade, Olubukola Omidiji, Rachel Akinola Lagos, O. O. Olatunji, Goldey Khanna, John Kirkpatrick, Michelle Alonso-Basanta, Arif Rashid, Miriam Bornhorst, Ali Nabavizadeh, Natasha Lepore, Joshua Palmer, Antonio Porras, Jake Albrecht, Udunna Anazodo, Mariam Aboian, Evan Calabrese, Jeffrey David Rudie, Marius George Linguraru, Juan Eugenio Iglesias, Koen Van Leemput, Spyridon Bakas, Benedikt Wiestler, Ivan Ezhov, Marie Piraud, and Bjoern H. Menze. The Brain Tumor Segmentation (BraTS) Challenge: Local Synthesis of Healthy Brain Tissue via Inpainting, September 2024. URL http://arxiv.org/abs/2305.08992. arXiv:2305.08992 [eess].

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization, January 2019. URL http://arxiv.org/abs/1711.05101. arXiv:1711.05101 [cs].

Gabriele Lozupone, Alessandro Bria, Francesco Fontanella, Frederick J. A. Meijer, Claudio De Stefano, and Henkjan Huisman. Latent Diffusion Autoencoders: Toward Efficient and Meaningful Unsupervised Representation Learning in Medical Imaging, April 2025. URL http://arxiv.org/abs/2504.08635. arXiv:2504.08635 [cs].

Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-Andre Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Cagatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa,

Mauricio Reyes, and Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, October 2015. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2014.2377694. URL http://ieeexplore.ieee.org/document/6975210/.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models, April 2022. URL http://arxiv.org/abs/2112.10752. arXiv:2112.10752 [cs].

Radomir S. Stanković and Bogdan J. Falkowski. The Haar wavelet transform: its status and achievements. *Computers & Electrical Engineering*, 29(1):25–44, January 2003. ISSN 00457906. doi: 10.1016/S0045-7906(01)00011-8. URL https://linkinghub.elsevier.com/retrieve/pii/S0045790601000118.

Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, pages 1398–1402, Pacific Grove, CA, USA, 2003. IEEE. ISBN 978-0-7803-8104-9. doi: 10.1109/ACSSC.2003.1292216. URL http://ieeexplore.ieee.org/document/1292216/.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, April 2018. URL http://arxiv.org/abs/1801.03924. arXiv:1801.03924 [cs].

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, January 2001. ISSN 02780062. doi: 10.1109/42.906424. URL http://ieeexplore.ieee.org/document/906424/.