Retrieval-Augmented Wavelet Diffusion for Local Synthesis of Healthy Brain Tissue

Sanna Persson[1 [0009−0000−3035−0165]], Simone Bendazzoli[2,3 [0000−0001−6673−1314]], and Rodrigo Moreno[1 [0000−0001−5765−2964]]

[1] KTH, Royal Institute of Technology, Department of Biomedical Engineering and Health Systems, Sweden
[2] Linköping University, Department of Science and Technology, Sweden
[3] Karolinska University Hospital, Department of Diagnostic and Research IT, Sweden

Abstract. Reconstructing healthy tissue in a voided brain MRI is hardest in the interior of large voids, where the missing structure is underdetermined by the surrounding context. To address this, we propose a retrieval-augmented wavelet diffusion model that conditions the reconstruction on an estimate of the missing tissue constructed from other subjects. For each case, we retrieve the most similar real brains and average the nearest donors into a prior, and the prior is concatenated to the wavelet-domain input and fine-tuned end to end. To improve the generation of healthy tissue, we further use lesion-free brains from the Human Connectome Project both as additional training data and for the retrieval donor pool. On a held-out split of the BraTS 2026 inpainting data, the merged prior improves over a strong unconditioned diffusion baseline, reaching an SSIM of 0.879; a size-resolved analysis shows that the residual error is concentrated in the interior of large voids. Our results indicate that a merged retrieved-donor prior is a promising strategy to supply the non-redundant structure that context alone cannot recover.

Keywords: Brain MRI inpainting · Wavelet diffusion model · Retrieval-augmented generation · Structural prior · BraTS.

1 Introduction

Many automated tools for analyzing brain MRI, from tissue segmentation to atlas registration, cortical parcellation, and brain extraction, are developed and validated on structurally intact brains, and their output degrades on images that contain a tumor and its mass effect [23]. This hinders the application of such tools to these patients. One suggested solution is to synthesize a plausible healthy version of the pathological region, so that downstream methods can be applied to a lesion-free reconstruction.

Our data comes from the BraTS challenge cluster, the largest publicly available collection of expert-curated, annotated brain-tumor MRI. It is built on the adult-glioma benchmark [28,4,1,5] with pre-operative annotations of the TCGA glioma collections [3,2], and has since expanded to further tumor types, missing-modality synthesis, and related tasks [22,21,29,11,25,24,26,37,6,20].

Healthy-tissue synthesis is posed as a three-dimensional image inpainting problem [23]: given a T1-weighted scan in which the tumor region, together with an additional tumor-shaped region of healthy tissue, has been voided, a method must synthesize plausible healthy tissue inside the void. Two inpainting strategies are common. Regression models, typically 3D U-Nets [33], predict the conditional mean of the missing tissue and score well on the pixel-level metrics [38,39]. Generative models based on denoising diffusion [18] instead sample from the distribution of healthy tissue and produce sharper, more plausible fills [12]. Operating in the wavelet domain keeps high-resolution 3D diffusion tractable [31,14]. All of these strategies, however, share a failure mode in that the interior of large voids is reconstructed poorly because the missing tissue there is under-determined by the surrounding context.

In this work, we propose a retrieval-augmented wavelet diffusion model that targets this failure mode directly. Augmenting a generative model with content retrieved from a database of real examples is an established idea for image synthesis [9], which we adapt to 3D healthy-tissue inpainting where the retrieved examples are whole donor brains. Previous work [27,19] has also shown that curating data can be a more efficient approach than architectural innovations. We therefore augment the training signal with lesion-free brains from the Human Connectome Project [36,16] and use these subjects as part of the retrieval pool. We evaluate the approach with a size-stratified analysis and an oracle-prior bound on the task, and submit it to the BraTS 2026 Local Synthesis of Healthy Brain Tissue (Inpainting) challenge under the Team ID "The Brat-tein Shakes".

2 Methods

Our method reconstructs voided healthy tissue in T1-weighted brain MRI with a conditional 3D wavelet diffusion model, conditioned on a prior retrieved and merged from a pool of real donor brains and trained on BraTS glioma data enlarged with lesion-free brains (see Figure 1 for an overview).

2.1 Wavelet diffusion backbone

Our base generator is a conditional 3D wavelet diffusion model in the style of fastWDM3D [12,14]. A single-level orthonormal Haar transform maps each volume to eight half-resolution coefficient channels, an eightfold spatial reduction that lets a 3D denoising diffusion model [18] run at the full field of view within memory and its inverse reconstructs the volume without loss. The model is implemented as a MONAI `DiffusionModelUNet` [10] (channel widths 64 to 256, self-attention at the two coarsest scales) which receives the noised target wavelet concatenated with the transforms of the voided image and the mask [34], 24 channels in total, and predicts the target wavelet directly (x_0-prediction). At inference we use a four-step DDIM sampler [35], average n independent samples, and apply flip-and-offset test-time augmentation, which moves the prediction toward the conditional mean.

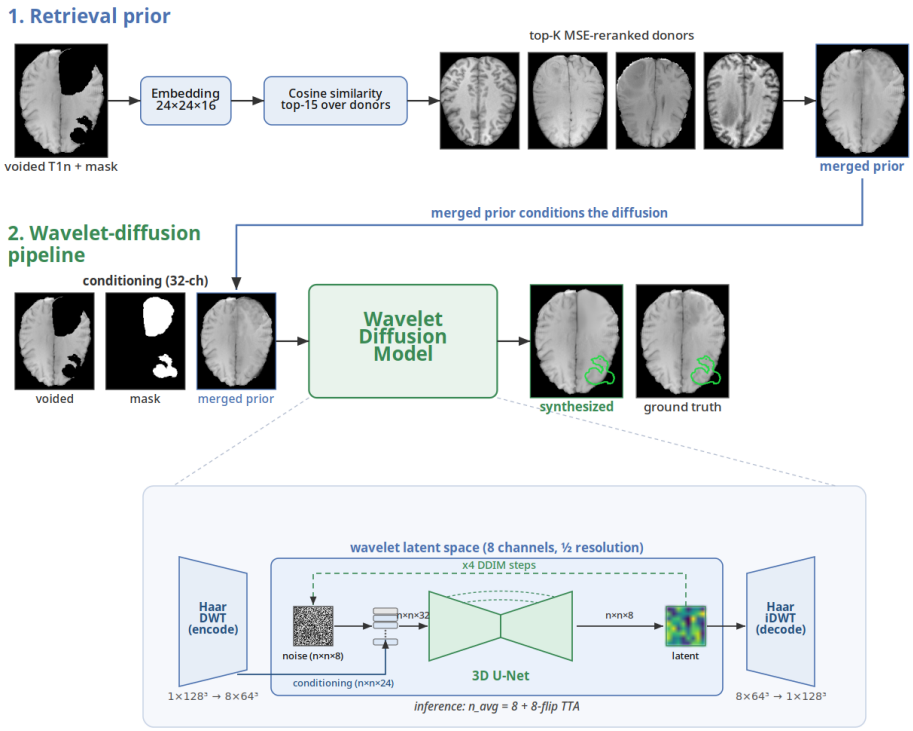

Fig. 1: Overview of the method. *Stage 1:* the voided query is embedded, shortlisted by cosine similarity over the 1909 donors, re-ranked at full resolution, and the top-K=10 nearest averaged into a merged prior. *Stage 2:* the prior is concatenated with the voided image and mask and conditions the Haar-wavelet diffusion model that synthesizes the void.

2.2 Data

We train on the 1251 adult-glioma (GLI) cases of the BraTS collection [28,4,1], each a skull-stripped T1-weighted volume of shape $240 \times 240 \times 155$ at 1 mm isotropic resolution in the BraTS-SRI24 reference space [32]. In each volume the tumor region and an additional tumor-shaped healthy region, the decoy, are voided. Our model takes the voided volume and the binary mask of these regions as input and reconstructs the missing healthy tissue [23]. Ground-truth healthy tissue exists only under the decoy, so evaluation is confined to that region. We hold out a fixed subset of 150 cases (val-150) for evaluation (Section 2.6) and normalize intensities per case.

2.3 Healthy-brain augmentation

Beyond the provided glioma cases, we expand the training data in three ways; the added lesion-free brains also serve as donors for the retrieval prior (Section 2.4).

Mask augmentation. We sample multiple tumor-shaped decoys per training brain from a pool of real tumor shapes, placed in normal-appearing tissue at least 5 voxels from the tumor, which multiplies the number of distinct training targets and exposes the model to a wide range of void sizes and locations.

Cross-cohort data (GoAT). To widen the anatomical and appearance variation beyond adult glioma, we add 100 non-glioma brains from the BraTS Generalizability Across Tumors (GoAT) collection, which spans tumor types including meningioma [25], metastasis [29], and pediatric tumors [22]. Because GoAT repackages some of the glioma cases used here under anonymized identifiers, we deduplicate by image content against val-150 before pooling, so no held-out subjects leak into training.

Healthy brains (HCP). We incorporate 808 T1-weighted volumes from the Human Connectome Project (HCP) [36,16], using the FreeSurfer-processed, skull-stripped images [13]. Each brain is rigidly registered to the SRI24 atlas with a mutual-information criterion [15] and regridded onto the standard BraTS resolution, after which we insert synthetic tumor-shaped voids on the fly during training. Because these brains contain no tumor, the entire void carries ground-truth healthy tissue and is supervised directly, unlike glioma cases where the tumor region has no reference.

2.4 Retrieval prior

The dominant residual error is in the interior of large voids, which is underdetermined by the surrounding tissue. We supply the missing structure with a retrieved-donor prior, an estimate of the healthy tissue built, for each case, from the real brains most similar to the patient (Fig. 1). Figure 2 shows example retrievals across void sizes. The prior is computed from the voided image and the mask, and never sees the tissue under the void.

Donor pool. The pool comprises 1909 real brains in the common reference space: the 1101 glioma training cases outside the held-out split, together with the 808 lesion-free HCP brains of Section 2.3. Each donor contributes its full T1 volume, but only tissue *outside* the query's void is ever used to compare donors, so retrieval cannot see the tissue it is reconstructing.

Coarse-to-fine retrieval. Comparing full-resolution volumes against the whole pool for every case is expensive, so retrieval is two-stage. First, each brain is summarized once by a compact structural embedding computed without a learned encoder: the volume is normalized by its 99.5th-percentile intensity, trilinearly

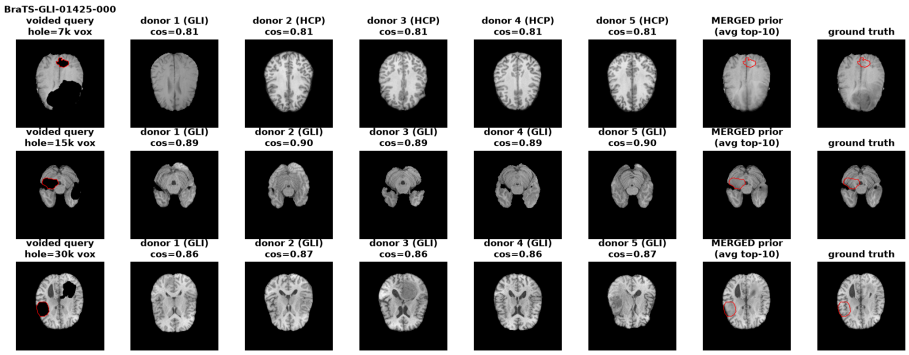

Fig. 2: Example retrievals for small, medium, and large voids (rows). Each shows the voided query, its five nearest donors with cosine similarity, and the merged top-10 prior, with the scored healthy region outlined in red.

downsampled to a fixed $24 \times 24 \times 16$ grid, flattened, and z-scored, which captures gross anatomy while discarding the intensity scale. The query's visible tissue is embedded identically, and the closest donors are shortlisted by cosine similarity between these embeddings. Second, the donors are re-ranked at full resolution by mean squared error with the query over the visible region only. Because the embeddings are precomputed and the costly full-resolution comparison touches only the shortlist, a query takes about two seconds. We average the $K = 10$ closest re-ranked donors into the merged prior, where K is a hyperparameter trading prior fidelity against inference cost.

Prior selection. We hypothesize that a useful prior must produce errors sufficiently decorrelated from those of the backbone model. The backbone already predicts the conditional mean from context, so a prior helps only where its errors are decorrelated from the backbone's. If the errors are similar in the same voxels, it is redundant, but if the errors are in different locations it is potentially accurate where the backbone makes mistakes. For two unbiased estimates the variance of their optimal combination is smallest when their errors are uncorrelated, which motivates this ranking. We rank candidate priors, a contralateral mirror, a population template, donor retrieval, and a tissue atlas, by the voxel-wise correlation between their reconstruction error and the base model's error inside the scored region. A low correlation marks complementary information. This separates a redundant prior (a fill sampled from the diffusion model itself) from a complementary one (donor retrieval), and selects the merged retrieval prior.

Conditioning. The merged prior is transformed to the wavelet domain and concatenated as an additional conditioning input, expanding the network from 24 to 32 input channels. The K donors are merged by a voxel-wise mean in image space, then mapped through the same Haar transform as the other inputs.

Starting from the trained unconditioned model, we zero-initialize the input-convolution weights of the new channel so that fine-tuning begins exactly at the base model [40] and learns to exploit the prior as the whole network is fine-tuned. For the HCP donors, whose voids are generated on the fly, the prior is recomputed each step. To test whether the effect is specific to the generative model, we condition a 3D SegResNet [30] regression baseline on the same prior in the same way.

2.5 Loss and optimization

Reconstruction. Every prediction is composited so that only the void is synthesized,

$$\hat{x} = \mathrm{IDWT}(x_0) \odot m + v \odot (1 - m), \tag{1}$$

with v the voided input and m the mask. Because everything outside the mask is copied from the input, the network receives gradient only inside the void during training.

Objective. The model is supervised in the wavelet domain and the image space. A wavelet-space term is the diffusion objective, the mean squared error between the predicted and true target coefficients (x_0). An image-space term applies an L_1 and squared-error reconstruction loss to the composited volume of Eq. 1 inside the scored healthy region. The total loss weights the two terms equally. A structural-similarity term [41], $(1 - \mathrm{SSIM})$ over the crop, is available and is used for the regression baseline but disabled for the diffusion model, which it did not improve. Wavelet-coefficient clipping is disabled, as the coefficients are not bounded to $[-1, 1]$.

Training parameters. We first train the unconditioned base to convergence, then warm-start the retrieval-conditioned model from it and fine-tune, which we found more stable than training with the prior from scratch. All models are trained on mask-centered 128^3 crops (batch size four) with random axis flips and on-the-fly resampling of the healthy decoy, so each brain yields many distinct voids across epochs, using 1000 noising steps with a linear-β schedule. We optimize with Adam and an exponential moving average of the weights (decay 0.999): the base at learning rate 10^{-4} on a cosine schedule, and the conditioned model fine-tuned for 40 epochs at 5×10^{-5}. Training runs in mixed precision on a single GPU with the full configuration released with our code at https://github.com/minnelab/retrieval-augmented-inpainting-brats-2026.

2.6 Evaluation

Because the public validation labels are not released, all evaluation is carried out on a fixed held-out split of 150 of the 1251 training cases (val-150). Each case is scored with the official inpainting metrics on a single frozen healthy decoy, computed only inside the healthy region after percentile-normalizing by

the voided image; predictions are always composited, leaving everything outside the mask equal to the input. We report SSIM, PSNR, and MSE, together with the per-case rank-sum over the three that defines the challenge objective. Our challenge submission is the wavelet diffusion model conditioned on the merged top-10 retrieval prior (Table 1, bold row; final row of Table 2), with inference by $n_{\text{avg}} = 8$ sample averaging and 8-flip test-time augmentation.

Stratification by decoy size. Reconstruction difficulty is dominated by the size of the region to be synthesized: large voids are harder because their centers lie beyond the receptive field of the surrounding tissue and are filled less sharply, whereas small voids are recovered almost exactly from local context. For this size-stratified analysis we sample five decoys per case rather than the single decoy used elsewhere (750 decoys over val-150), so that larger voids are better represented.

3 Results

All results use the held-out val-150 split (Section 2.6) and the official metrics; unless stated otherwise, scores use $n_{\text{avg}} = 8$ sample averaging.

3.1 Reconstruction performance

Table 1 compares our model against a SegResNet regression baseline and the unconditioned diffusion baseline. The diffusion model with the merged retrieval prior attains the best SSIM, the lowest MSE, and the best per-case rank-sum, ahead of both the unconditioned diffusion baseline and the regression baseline. An oracle prior built from a blurred copy of the ground truth (an isotropic Gaussian blur, $\sigma = 4$ voxels) is included as an upper bound. The two diffusion models have the same architecture, training data, and augmentations where "Diffusion, no prior" is the fully augmented base without the retrieval-prior channel. The SegResNet baseline is a regression model trained on the same voided image and masks. On a 150-case sample of the BraTS training split, with each case excluded as its own retrieval donor, the submitted model attains SSIM 0.8793, PSNR 24.32, and MSE 0.00410, which is within noise of the val-150 result (Table 1) and indicates no meaningful overfitting.

3.2 Ablation

In Table 2, we report an ablation of each training step in the retrieval-augmented model. Its final row is the submitted model of Table 1. The largest SSIM gains come from extended training with an exponential moving average and mask-shape augmentation, and from pooling additional brains through the healthy-brain augmentation. The merged retrieval prior then adds a smaller SSIM increment, but its effect on the per-case rank-sum, the metric the challenge ranks on, is comparable to the earlier training refinements (from 8.76 to 6.67), as it also improves PSNR and MSE. Its benefit is larger still on harder voids (Fig. 3).

Table 1: Held-out val-150 (single frozen decoy per case, $n_{\text{avg}} = 8$, with test-time augmentation). Rank-sum: mean per-case rank over {SSIM, PSNR, MSE} (lower better) across the three methods; the oracle is a non-participating upper bound (–).

Method	SSIM ↑	PSNR ↑	MSE ↓	Rank-sum ↓
SegResNet regression baseline	0.8583	23.14	0.00489	8.45
Diffusion, no prior	0.8755	23.77	0.00425	5.99
Diffusion + merged retrieval prior (ours)	**0.8786**	**23.99**	**0.00408**	**3.57**
Oracle: blurred-GT prior (upper bound)	0.9245	25.24	0.00260	–

Table 2: Ablation of the retrieval-augmented model (val-150, $n_{\text{avg}} = 8$; final row adds test-time augmentation). Rank-sum over {SSIM, PSNR, MSE} (lower better) across the six rows.

Step	SSIM ↑	PSNR ↑	MSE ↓	ΔSSIM	Rank-sum ↓
Conditional wavelet diffusion	0.8021	20.32	0.00944	–	17.70
+ long training, EMA, mask-shape aug.	0.8413	21.85	0.00614	+0.039	15.08
+ healthy-brain augmentation	0.8626	22.85	0.00502	+0.021	11.46
+ diversity, longer training	0.8699	23.34	0.00462	+0.007	8.76
+ merged retrieval prior (ours)	0.8737	23.59	0.00443	+0.004	6.67
+ test-time augmentation	**0.8786**	**23.99**	**0.00408**	+0.005	**3.33**

3.3 Structural priors

To isolate the effect of the prior itself, Table 3 compares the candidate priors of Section 2.4 under identical inference and a matched training budget, so the rows differ only in which prior the model receives. The benefit tracks the prior's fidelity, its correlation with the true tissue inside the void. Low-fidelity priors (a tissue atlas, or a single or blurred donor, fidelity ≤ 0.51) are neutral to slightly harmful. Only the merged prior, which raises fidelity from 0.51 for a single donor to 0.67, meaningfully improves over the no-prior base; the blurred ground-truth oracle (fidelity 0.81) would gain far more (+0.055), which bounds the achievable headroom.

In the submitted configuration, the merged prior improves SSIM by +0.0031 over the no-prior diffusion model (Table 1; paired Wilcoxon signed-rank over the 150 cases, improving 88% of them, $p \approx 6 \times 10^{-21}$). The prior also transfers across architectures: it improves a SegResNet regression baseline, relative to a matched no-prior control, by a significant margin (+0.0005 SSIM, 69% of cases, paired Wilcoxon over the 150 cases, $p < 10^{-4}$), so the retrieved signal is useful independent of the model family.

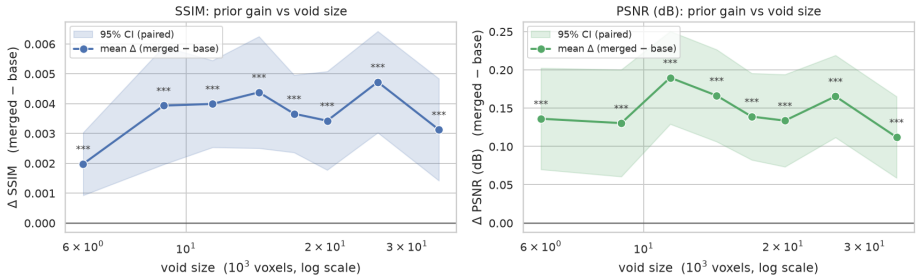

Fig. 3: No-prior base versus merged prior as a function of void size (five decoys per case); the prior's advantage grows with larger voids. *** indicates $p < 0.001$.

Table 3: Prior comparison at comparable training budget (val-150, $n_{\text{avg}} = 8$, no test-time augmentation). Fidelity: correlation of the prior with ground truth inside the void.

Prior	Type	Injection	Fidelity	SSIM (Δ)
No prior (base)	–	–	–	0.8699
Tissue atlas	tissue	concat	0.50	0.8682 (-0.0017)
Blurred retrieval	single donor	concat	0.51	0.8695 (-0.0004)
Single retrieval	single donor	concat	0.51	0.8700 ($+0.0001$)
Merged top-10 (ours)	merged K=10	concat	0.67	**0.8708** (**+0.0009**)
Oracle blurred-GT	GT (σ=4)	concat	0.81	0.9245 ($+0.0546$)

A broader evaluation with five decoys per case, where larger voids are better represented, raises the prior's advantage to +0.0045 SSIM (74% of the 750 decoy-samples, $p < 10^{-40}$), concentrated in the largest, hardest voids (Fig. 3).

3.4 Downstream parcellation

The pixel metrics presented above do not necessarily correspond to downstream performance on relevant applications for the synthesized tissue. We therefore parcellated every reconstruction and its ground truth and measured label agreement inside the decoy (Table 4). Because the validation data is skull-stripped, we report two segmenters: FastSurfer [17] (DKT-atlas cortical parcellation with subcortical segmentation) and SynthSeg [8], which is trained to be robust to contrast and skull-stripping. The compared reconstructions are a matched pair differing only in the retrieval-prior channel (same architecture and training), evaluated with test-time augmentation, for which the prior adds +0.0011 SSIM. Across both segmenters the prior yields significantly higher agreement on all three measures (paired Wilcoxon, $p < 10^{-2}$), improving a majority of cases, with the effect larger under FastSurfer than under SynthSeg.

Table 4: Downstream parcellation agreement inside the decoy (val-150, matched pair). Concordance is the fraction of decoy voxels labeled identically to the ground-truth parcellation. "Improved" counts cases with $\Delta > 0$ (paired Wilcoxon p).

Metric	Base	+ prior	Δ	Improved (p)
FastSurfer (DKT + aseg)				
Voxel concordance	0.8691	0.8712	+0.0021	116/150 ($p<10^{-9}$)
Foreground concordance	0.8736	0.8758	+0.0022	115/150 ($p<10^{-11}$)
Macro-Dice	0.7384	0.7442	+0.0058	105/150 ($p<10^{-5}$)
SynthSeg (aseg, skull-strip-robust)				
Voxel concordance	0.8827	0.8830	+0.0003	87/150 (p=0.005)
Foreground concordance	0.8826	0.8828	+0.0002	88/150 (p=0.002)
Macro-Dice	0.8569	0.8575	+0.0006	96/150 (p=0.007)

4 Discussion

Our results show that a retrieved-donor prior improves healthy-tissue inpainting beyond a strong unconditioned diffusion model, but only when the prior carries information the model cannot already infer from context.

The prior's advantage grows with void size (Fig. 3), largest in the interior of big voids that the visible anatomy under-determines. An oracle prior built from blurred ground truth would gain far more, so this residual is not irreducible in principle; it is bounded instead by the fidelity a cross-subject donor can reach under the void, which no realistic prior matches. The prior also makes the fill modestly more anatomically usable: it raises downstream parcellation agreement inside the decoy by a small but significant margin that survives a robustness check with a skull-strip-robust segmenter, again largest on bigger voids. The larger effect under FastSurfer than under this robust segmenter suggests part of the FastSurfer gain reflects its sensitivity to skull-stripped cortex. Retrieval adds about two seconds per case, and our results show the prior is mainly beneficial for larger voids.

4.1 Limitations and future directions

There are two main limitations to the proposed approach. The absolute gain is modest, because the base model already sits close to the achievable ceiling on most voids, and the prior helps most where the task is hardest. Furthermore, most retrieval donors are themselves tumor patients, so a retrieved donor can carry a lesion near the filled region and inject pathological rather than healthy tissue into the prior. Averaging over several donors and the lesion-free HCP brains decreases this effect but does not remove it, and masking each donor's lesion with its tumor segmentation so that only healthy tissue is averaged would address it more directly. Promising directions include merging more donors and a

cheaper embedding-space merge that avoids loading donor volumes at inference. The framework does not depend on the HCP dataset specifically: the glioma cases already form a self-contained donor pool, with HCP adding lesion-free donors and directly supervised voids. We did not isolate the HCP and GoAT contributions, which the ablation folds into one healthy-brain augmentation step worth +0.021 SSIM (Table 2).

Acknowledgments. Challenge data were obtained through Synapse ID (syn74274097). Data were further provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Computations used Alvis and Arrhenius via NAISS (Swedish Research Council grant No. 2022-06725) and the Berzelius resource at the National Supercomputer Centre (Knut and Alice Wallenberg Foundation), with visualization and processing on the MAIA infrastructure at KTH [7]. The project is partially funded by the Swedish Research Council (grant 2022-03389), Digital Futures, MedTechLabs, the Swedish Childhood Cancer Foundation (Barncancerfonden MT2022-0008), Vinnova through AIDA (project ID 2319), and Hjärt-Lungfonden (grant No. 2022-0492).

References

1. Baid, U., Ghodasara, S., Mohan, S. et al.: The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification (Sep 2021). https://doi.org/10.48550/arXiv.2107.02314, http://arxiv.org/abs/2107.02314, arXiv:2107.02314
2. Bakas, S., Akbari, H., Sotiras, A. et al.: Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection (2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q, https://www.cancerimagingarchive.net/analysis-result/brats-tcga-gbm/
3. Bakas, S., Akbari, H., Sotiras, A. et al.: Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection (2017). https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF, https://www.cancerimagingarchive.net/analysis-result/brats-tcga-lgg/
4. Bakas, S., Akbari, H., Sotiras, A. et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data **4**(1), 170117 (Sep 2017). https://doi.org/10.1038/sdata.2017.117, https://www.nature.com/articles/sdata2017117
5. Bakas, S., Reyes, M., Jakab, A. et al.: Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge (Apr 2019). https://doi.org/10.48550/arXiv.1811.02629, http://arxiv.org/abs/1811.02629, arXiv:1811.02629
6. Bakas, S., Thakur, S.P., Faghani, S. et al.: BraTS-Path Challenge: Assessing Heterogeneous Histopathologic Brain Tumor Sub-regions (May 2024). https://doi.org/10.48550/arXiv.2405.10871, http://arxiv.org/abs/2405.10871, arXiv:2405.10871
7. Bendazzoli, S., Persson, S., Astaraki, M. et al.: MAIA: a collaborative medical AI platform for integrated healthcare innovation. npj Artificial Intelligence **1**(1), 45 (Dec 2025). https://doi.org/10.1038/s44387-025-00042-6, https://www.nature.com/articles/s44387-025-00042-6
8. Billot, B., Greve, D.N., Puonti, O. et al.: SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (May 2023). https://doi.org/10.1016/j.media.2023.102789, https://linkinghub.elsevier.com/retrieve/pii/S1361841523000506
9. Blattmann, A., Rombach, R., Oktay, K. et al.: Retrieval-Augmented Diffusion Models. In: Advances in Neural Information Processing Systems 35. pp. 15309–15324. Neural Information Processing Systems Foundation, Inc. (NeurIPS), New Orleans, Louisiana, USA (2022). https://doi.org/10.52202/068431-1114, http://www.proceedings.com/068431-1114.html
10. Cardoso, M.J., Li, W., Brown, R. et al.: MONAI: An open-source framework for deep learning in healthcare (Nov 2022). https://doi.org/10.48550/arXiv.2211.02701, http://arxiv.org/abs/2211.02701, arXiv:2211.02701
11. Deng, F.: RANO criteria for brain metastases (RANO-BM). In: Radiopaedia.org. Radiopaedia.org (Jul 2021). https://doi.org/10.53347/rID-91171, https://radiopaedia.org/articles/91171
12. Durrer, A., Bieder, F., Friedrich, P. et al.: fastWDM3D: Fast and Accurate 3D Healthy Tissue Inpainting. In: Mukhopadhyay, A., Oksuz, I., Engelhardt, S. et al. (eds.) Deep Generative Models, vol. 16128, pp. 171–181. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-05472-2_17, https://link.springer.com/10.1007/978-3-032-05472-2_17

13. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (Aug 2012). https://doi.org/10.1016/j.neuroimage.2012.01.021, https://linkinghub.elsevier.com/retrieve/pii/S1053811912000389
14. Friedrich, P., Wolleb, J., Bieder, F. et al.: WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis. In: Mukhopadhyay, A., Oksuz, I., Engelhardt, S. et al. (eds.) Deep Generative Models, vol. 15224, pp. 11–21. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-72744-3_2, https://link.springer.com/10.1007/978-3-031-72744-3_2
15. Garyfallidis, E., Brett, M., Amirbekian, B. et al.: Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics **8** (Feb 2014). https://doi.org/10.3389/fninf.2014.00008, http://journal.frontiersin.org/article/10.3389/fninf.2014.00008/abstract
16. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A. et al.: The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage **80**, 105–124 (Oct 2013). https://doi.org/10.1016/j.neuroimage.2013.04.127
17. Henschel, L., Conjeti, S., Estrada, S. et al.: FastSurfer - A fast and accurate deep learning based neuroimaging pipeline. NeuroImage **219**, 117012 (Oct 2020). https://doi.org/10.1016/j.neuroimage.2020.117012, https://linkinghub.elsevier.com/retrieve/pii/S1053811920304985
18. Ho, J., Jain, A., Abbeel, P.: Denoising Diffusion Probabilistic Models (Dec 2020). https://doi.org/10.48550/arXiv.2006.11239, http://arxiv.org/abs/2006.11239, arXiv:2006.11239
19. Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z, https://www.nature.com/articles/s41592-020-01008-z
20. Karargyris, A., Umeton, R., Sheller, M.J. et al.: Federated benchmarking of medical artificial intelligence with MedPerf. Nature Machine Intelligence **5**(7), 799–810 (Jul 2023). https://doi.org/10.1038/s42256-023-00652-2, https://www.nature.com/articles/s42256-023-00652-2
21. Kazerooni, A.F., Khalili, N., Liu, X. et al.: The Brain Tumor Segmentation in Pediatrics (BraTS-PEDs) Challenge: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs) (2024). https://doi.org/10.48550/ARXIV.2404.15009, https://arxiv.org/abs/2404.15009
22. Kazerooni, A.F., Khalili, N., Liu, X. et al.: The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs) (2023). https://doi.org/10.48550/ARXIV.2305.17033, https://arxiv.org/abs/2305.17033
23. Kofler, F., Meissen, F., Steinbauer, F. et al.: The Brain Tumor Segmentation (BraTS) Challenge: Local Synthesis of Healthy Brain Tissue via Inpainting (Sep 2024). https://doi.org/10.48550/arXiv.2305.08992, http://arxiv.org/abs/2305.08992, arXiv:2305.08992
24. LaBella, D., Abramova, V., Astaraki, M. et al.: Analysis of the 2024 BraTS Meningioma Radiotherapy Planning Automated Segmentation Challenge (Jul 2025). https://doi.org/10.48550/arXiv.2405.18383, http://arxiv.org/abs/2405.18383, arXiv:2405.18383
25. LaBella, D., Adewole, M., Alonso-Basanta, M. et al.: The ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2023: Intracranial Meningioma (2023). https://doi.org/10.48550/ARXIV.2305.07642, https://arxiv.org/abs/2305.07642

26. Li, H.B., Conte, G.M., Hu, Q. et al.: The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn) (2023). https://doi.org/10.48550/ARXIV.2305.09011, https://arxiv.org/abs/2305.09011
27. Mazumder, M., Banbury, C., Yao, X. et al.: DataPerf: Benchmarks for Data-Centric AI Development (Oct 2023). https://doi.org/10.48550/arXiv.2207.10062, http://arxiv.org/abs/2207.10062, arXiv:2207.10062
28. Menze, B.H., Jakab, A., Bauer, S. et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (Oct 2015). https://doi.org/10.1109/TMI.2014.2377694, http://ieeexplore.ieee.org/document/6975210/
29. Moawad, A.W., Janas, A., Baid, U. et al.: The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI (2023). https://doi.org/10.48550/ARXIV.2306.00838, https://arxiv.org/abs/2306.00838
30. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization (2018). https://doi.org/10.48550/ARXIV.1810.11654, https://arxiv.org/abs/1810.11654
31. Phung, H., Dao, Q., Tran, A.: Wavelet Diffusion Models are fast and scalable Image Generators. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10199–10208. IEEE, Vancouver, BC, Canada (Jun 2023). https://doi.org/10.1109/CVPR52729.2023.00983, https://ieeexplore.ieee.org/document/10204642/
32. Rohlfing, T., Zahr, N.M., Sullivan, E.V. et al.: The SRI24 multichannel atlas of normal adult human brain structure. Human Brain Mapping **31**(5), 798–819 (May 2010). https://doi.org/10.1002/hbm.20906, https://onlinelibrary.wiley.com/doi/10.1002/hbm.20906
33. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M. et al. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28, http://link.springer.com/10.1007/978-3-319-24574-4_28
34. Saharia, C., Chan, W., Chang, H. et al.: Palette: Image-to-Image Diffusion Models (May 2022). https://doi.org/10.48550/arXiv.2111.05826, http://arxiv.org/abs/2111.05826, arXiv:2111.05826
35. Song, J., Meng, C., Ermon, S.: Denoising Diffusion Implicit Models (Oct 2022). https://doi.org/10.48550/arXiv.2010.02502, http://arxiv.org/abs/2010.02502, arXiv:2010.02502
36. Van Essen, D.C., Smith, S.M., Barch, D.M. et al.: The WU-Minn Human Connectome Project: An overview. NeuroImage **80**, 62–79 (Oct 2013). https://doi.org/10.1016/j.neuroimage.2013.05.041, https://linkinghub.elsevier.com/retrieve/pii/S1053811913005351
37. de Verdier, M.C., Saluja, R., Gagnon, L. et al.: The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI (2024). https://doi.org/10.48550/ARXIV.2405.18368, https://arxiv.org/abs/2405.18368
38. Zhang, J., Weng, Y., Chen, K.: U-Net Based Healthy 3D Brain Tissue Inpainting (Jul 2025). https://doi.org/10.48550/arXiv.2507.18126, http://arxiv.org/abs/2507.18126, arXiv:2507.18126

39. Zhang, J., Weng, Y., Chen, K.: Robust 3D Brain MRI Inpainting with Random Masking Augmentation. In: Bakas, S., Dennis, E., Astaraki, M. et al. (eds.) Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries, vol. 16377, pp. 102–109. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-16370-7_9, https://link.springer.com/10.1007/978-3-032-16370-7_9
40. Zhang, L., Rao, A., Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3813–3824. IEEE, Paris, France (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.00355, https://ieeexplore.ieee.org/document/10377881/
41. Zhou Wang, Bovik, A., Sheikh, H. et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (Apr 2004). https://doi.org/10.1109/TIP.2003.819861, https://ieeexplore.ieee.org/document/1284395/