Brain Tumour Removing and Missing Modality Generation using 3D Wavelet Diffusion Model

André Ferreira*[1,2,3,8,9], Gijs Luijten[3,6], Behrus Puladi[8,9], Jens Kleesiek[3,4,5,10], Victor Alves[1], and Jan Egger[2,3,4,6]

[1]Center Algoritmi / LASI, University of Minho, Braga, 4710-057, Portugal
[2]Computer Algorithms for Medicine Laboratory, Graz, Austria
[3]Institute for AI in Medicine (IKIM), University Medicine Essen, Girardetstraße 2, Essen, 45131, Germany
[4]Cancer Research Center Cologne Essen (CCCE), University Medicine Essen, Hufelandstraße 55, Essen, 45147, Germany
[5]German Cancer Consortium (DKTK), Partner Site Essen, Hufelandstraße 55, Essen, 45147, Germany
[6]Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, Graz, 8010, Austria
[7]Department of Neurosurgery and Spine Surgery, University Hospital Essen, Essen,Germany
[8]Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany
[9]Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Aachen, Germany
[10]Department of Physics, TU Dortmund University, Dortmund, Germany

***id10656@alunos.uminho.pt**

Abstract

Missing data or data with artefacts are very common in medical imaging. This significantly **compromises the effectiveness** of automated brain analysis tools.

Brain lesions, such as tumours, change the morphology of the brain, leading to high variability and **poor performance of prediction models** developed for healthy brains. This problem can be solved with a tool capable of **replacing the lesion with healthy tissue.**

For brain tumour classification and treatment planning, **4 modalities** are usually acquired, namely **T1, T1c, T2** and **FLAIR**, to enable optimal clinical decision making. Several automated tools require the input of all modalities to properly operate. In real clinical practice, it is **difficult to obtain** all MRI modalities, resulting in **missing MRI sequences**. The acquisition of MRI data is very **time consuming**, **costly** and **technically difficult**. Therefore, a tool that can **generate the missing modalities** is of great interest.

Diffusion models are known for their ability to generate realistic synthetic data. However, they require a **lot of computer resources**. The **3D WDM** network has proven to be capable of successfully fulfilling both tasks. By using the **wavelet transformation**, it was possible to generate the scans in **full resolution**, generating the missing modality or realistically inpainting over the unhealthy region.

Conditional versions of the 3D WDM are developed to perform each specific task. These conditions are used as input of the **U-net** in the denoising process, both for **training and inference**.

Methods

The network used for both tasks is based on the 3D Wavelet Diffusion Model (3D WDM) [1]. All models of **BraSyn** were trained for 2.000.000 and **BraInp** for 3.000.000 iterations using the AdamW optimiser with learning rate 1e-5, weight decay of 0, and batch size of 1.

Missing MRI – BraSyn (Global)

The dataset is based on the **BraTS 2023 adult glioma segmentation dataset** [2]. The training set contains all 4 modalities and the respective segmentation (1251 cases). In the validation (219 cases) and test sets (570 cases), one random modality is omitted. Two main solutions were created:

- Default ($\mathbf{D_g}$): 3D WDM with 32 input and output channels (4 modalities*8);
- Known 3 to 1 ($\mathbf{K3T1_g}$): Only the missing modality is generated. The 3 known modalities have no noise in all steps of WDM training and inference, only the modality of interest (MOI). The input shape has 4 more channels to inform the network which modality to generate, as shown in **Figure 1**.

Loss functions:

- **Dg**: defined by Equation 1, where a_0 represents the 4 real modalities, and $\tilde{a}_0$ the generated modalities.
- **K3T1g**: defined by Equation 2. MOI is the modality of interest, and s_3 represents the three tumour regions, tumour core (TC), whole tumour (WT) and enhancing tumour (ET). The loss of each region is calculated and summed. $\lambda_1 = 1$.

$$L_{D_g} = ||\tilde{a}_0 - a_0||_2^2 \quad \textbf{(1)}$$

$$L_{K3T1_g} = ||\tilde{a}_{MOI} - a_{MOI}||_2^2 + \lambda_1 \cdot ||\tilde{a}_{MOI} * s_3 - a_{MOI} * s_3||_2^2 \quad \textbf{(2)}$$

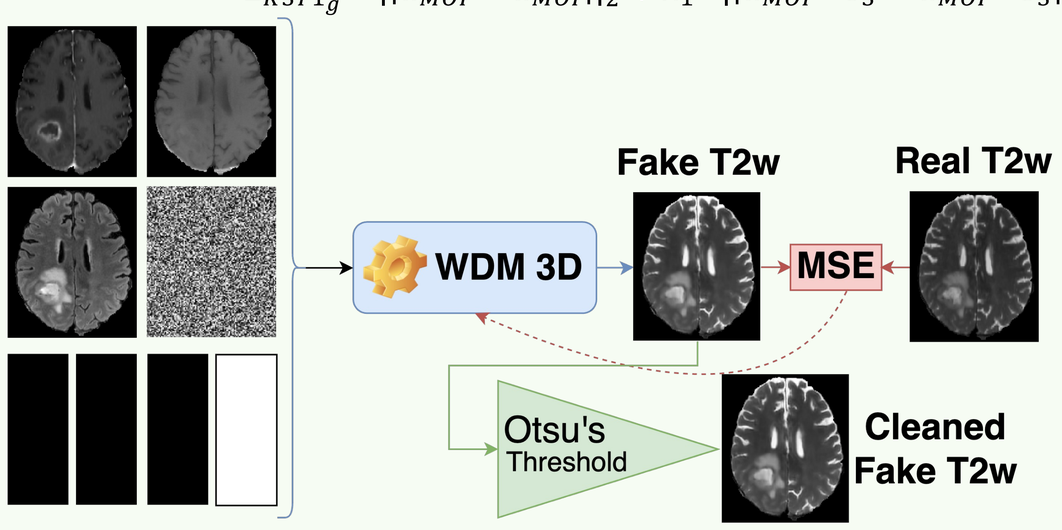

Figure 1: Training and inference pipelines for **BraSyn**. The blue colour means that the steps are shared by both, the red colour means that they are only performed for training and the green colour only for inference.

Inpaint MRI – BraInp (Local)

Same dataset [2], but only contains the **T1 modality**. The **training set** (1251 cases) consists of MRI scans, masks of the tumour (dilated), and masks of a random healthy region. The **validation** (219 cases) and **test** (*several hundred* cases) only contain the MRI scan without the **region to inpaint** and **respective mask**.

Two main solutions were created:

- Default ($\mathbf{D_I}$): 3D WDM without any changes. For **sampling** (**Figure 2**), the known region is replaced by the original scan (at each step *t*), keeping only the generated region to interest (**ROI**), i.e., to inpaint;
- Always known healthy ($\mathbf{AKH_I}$): Unhealthy tissue is removed, even from the ground-truth. The healthy mask is concatenated in the input. Noise is only added to the ROI, the known tissue never has noise, neither during training nor during inference.

Loss functions:

- $\mathbf{D_I}$: defined by Equation 3, x_0 represents the real MRI scan, and $\tilde{x}_0$ the generated scan.
- $\mathbf{AKH_I}$: defined by Equation 4, m_h represents the healthy mask, m_{uh} the unhealthy mask. $\lambda_2 = 10$.

$$L_{D_I} = ||\tilde{x}_0 - x_0||_2^2 \quad \textbf{(3)}$$

$$L_{AKH_I} = ||\tilde{x}_0 \cdot (1 - m_{uh}) - x_0 \cdot (1 - m_{uh})||_2^2 + \lambda_2 \cdot ||\tilde{x}_0 \cdot m_h - x_0 \cdot m_h||_2^2 \quad \textbf{(4)}$$

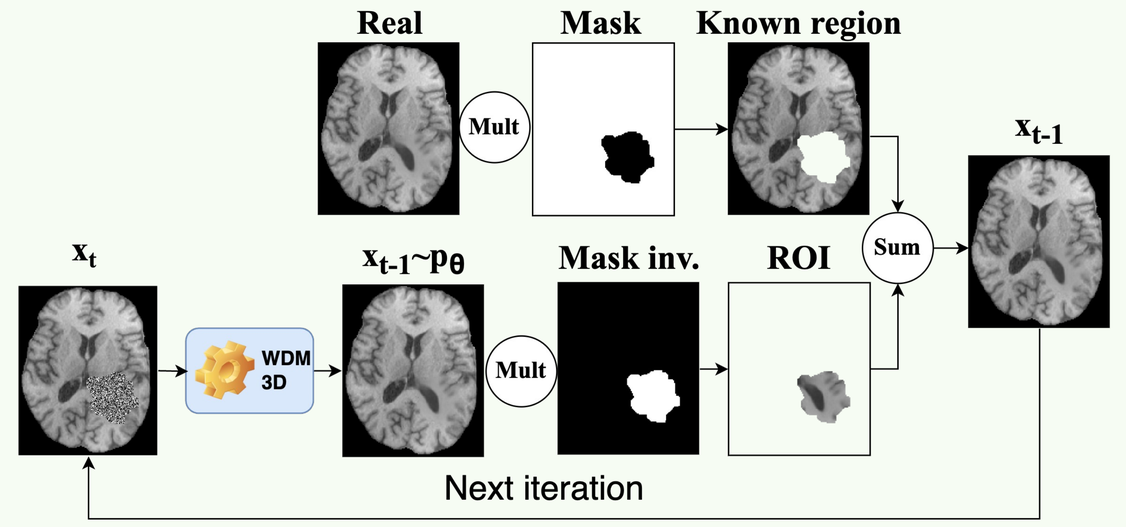

Figure 2: Pipeline for conditional inference in inpainting (**BraInp**) .

Results/Conclusion

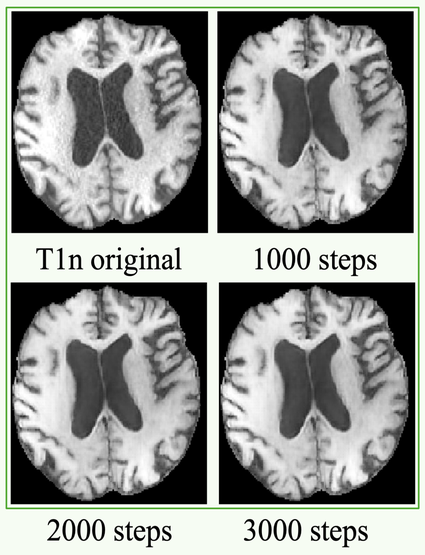

Figure 3: T1n modality sampled using 1000, 2000 or 3000 steps, knowing the other three modalities.

For internal evaluation of **BraSyn**, we only performed **visual inspection**. We tested 1000, 2000, and 3000 sampling steps.

- $\mathbf{D_g}$ - Produced noisy and unrealistic outputs.
- $\mathbf{K3T1_g}$ – Using **3000 sampling steps** improved SSIM and realism of the generations. More sampling steps reduced the noise but increased blur. **Otsu's threshold** was applied to remove background noise and fix the poor segmentations performed by a pre-trained nnUNet model, which improved DSC.

MSE, SSIM, PSNR and visual inspection were used for validation of **BraInp**.

- $\mathbf{D_I}$ was generating unhealthy tissue in some cases, since the ground-truth contained unhealthy tissue.
- $\mathbf{AKH_I}$: with 5000 sampling steps produced the **best results** in the validation:

MSE=0.0109±0.0076; **PSNR**=20.751±3.4009; **SSIM**=0.8078±0.1142

The use of 5000 steps makes the sampling **very slow** (30 min. per case), but it **improves** the objective metrics. However, it increases the **blurring** on the scan. In a context of challenge, this improvement is justifiable, however, this needs to be further investigated to understand if it has an **impact on the performance of the downstream tasks**.

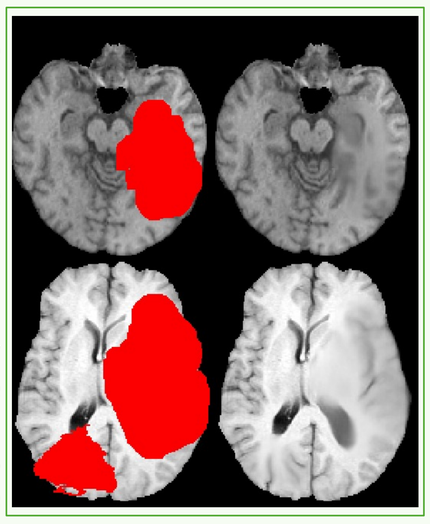

Figure 4: Voided and respective reconstruction. Validation cases.

References

[1] Friedrich, Paul, et al. "WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis." arXiv preprint arXiv:2402.19043 (2024).

[2] Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021).

Acknowledgments: André Ferreira thanks the Fundação para a Ciência e Tecnologia (FCT) Portugal for the grant 2022.11928.BD. This work was supported by FCT within the R\&D Units Project Scope: UIDB/00319/2020. It was also supported by the Advanced Research Opportunities Program (AROP) of RWTH Aachen University and Clinician Scientist Program of the Faculty of Medicine RWTH Aachen University, also some of the computations were performed with computing resources granted by RWTH Aachen University under project rwth1484. This work received funding from enFaced (FWF KLI 678), enFaced 2.0 (FWF KLI 1044) and KITE (Plattform für KI-Translation Essen) from the REACT-EU initiative (EFRE-0801977, \url{https://kite.ikim.nrw/}).