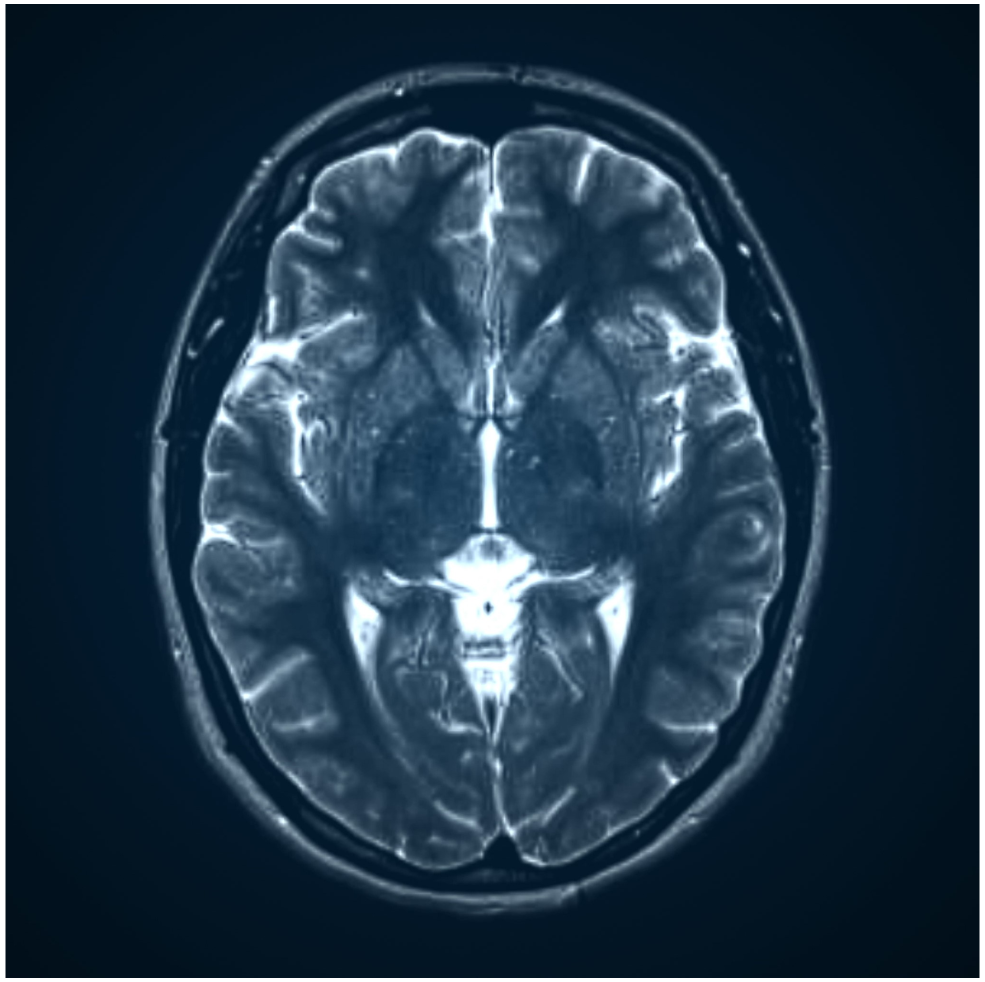
Master Thesis

Counterfactual Diffusion-based Image Editing on Brain MRI

Malek Ben Alaya, 15.10.2024

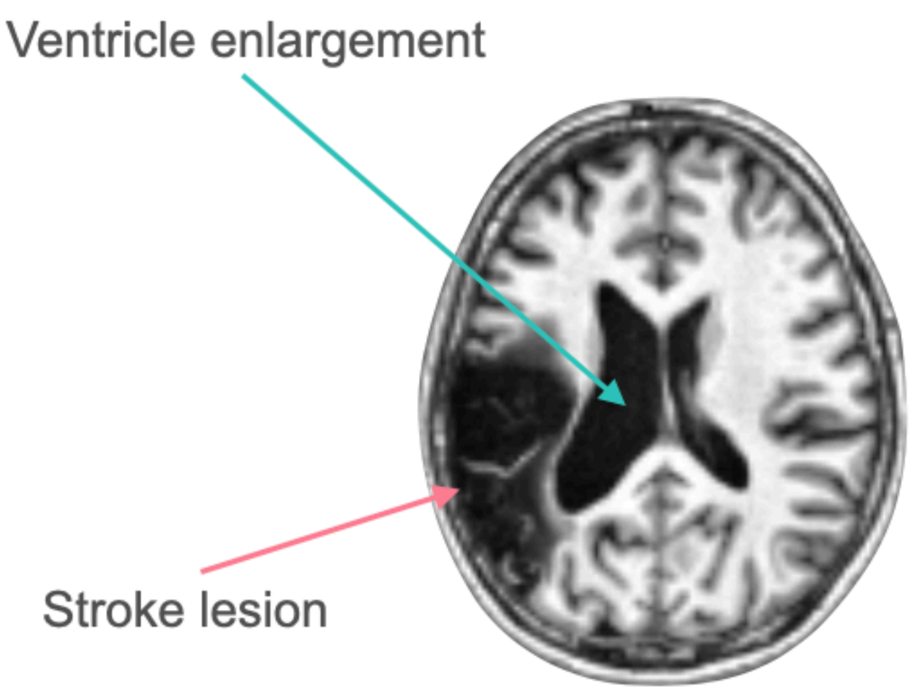
Motivation

- Current anomaly detection benchmark datasets not diverse enough
- Can we create such dataset with counterfactual image editing?

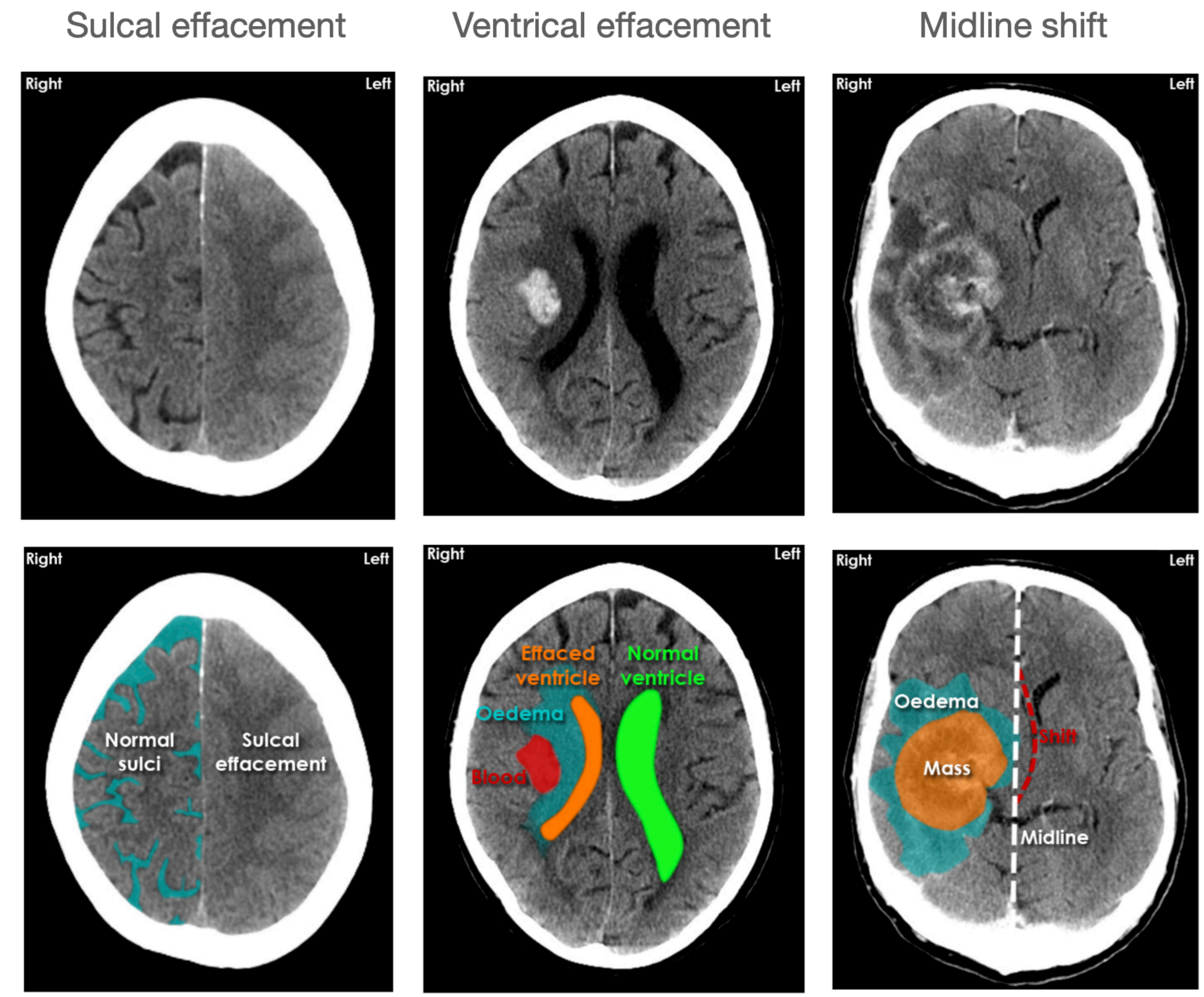


Credits [1]

Problem statement



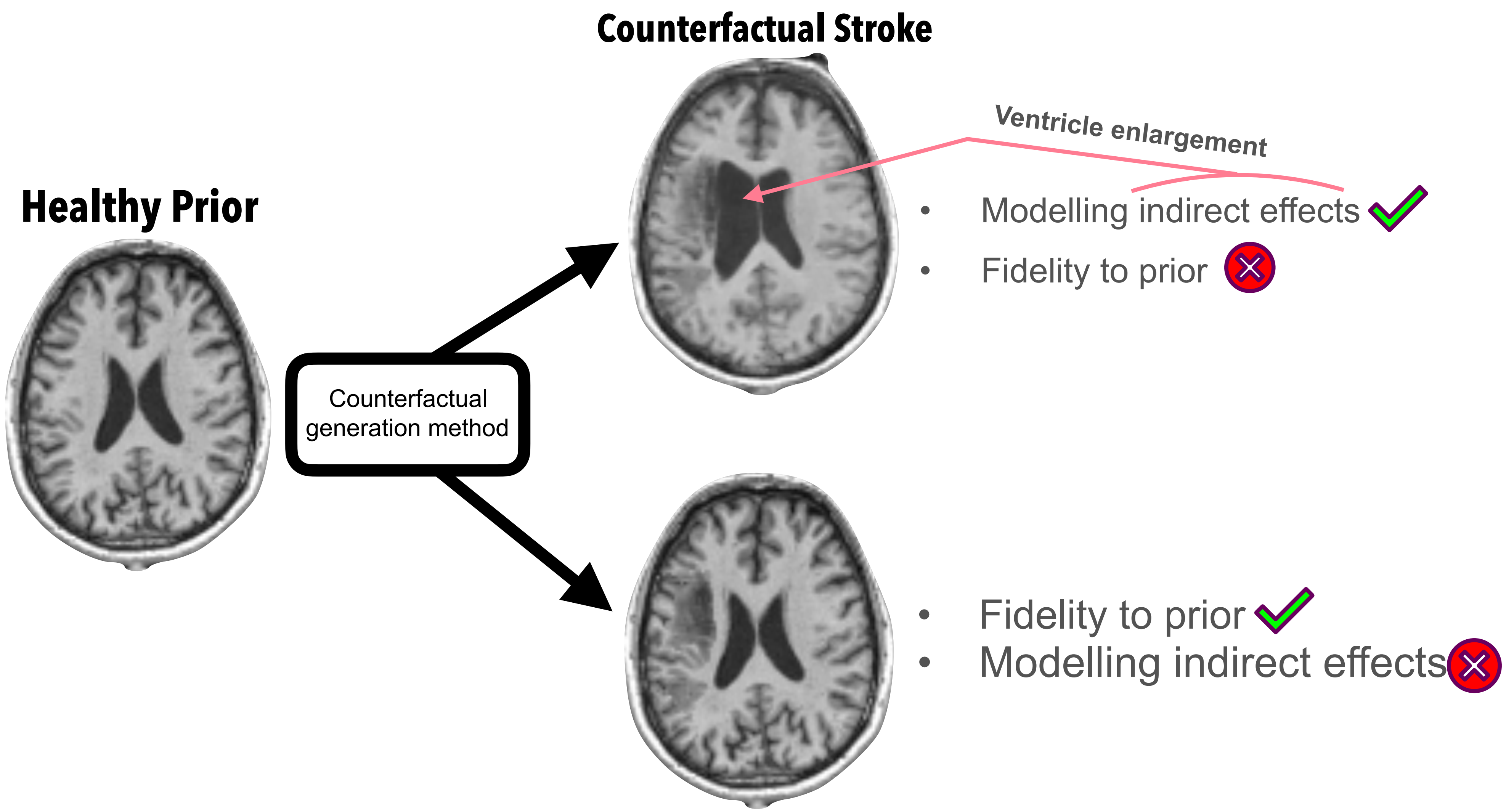
Indirect effects caused by stroke



Indirect effects caused by tumours, taken from [2]

- Indirect pathological effects need to be modelled

Problem statement



Desired to strike balance between:

- Ensuring high fidelity to prior scan
- Modelling indirect effects

Contribution

Contributions:

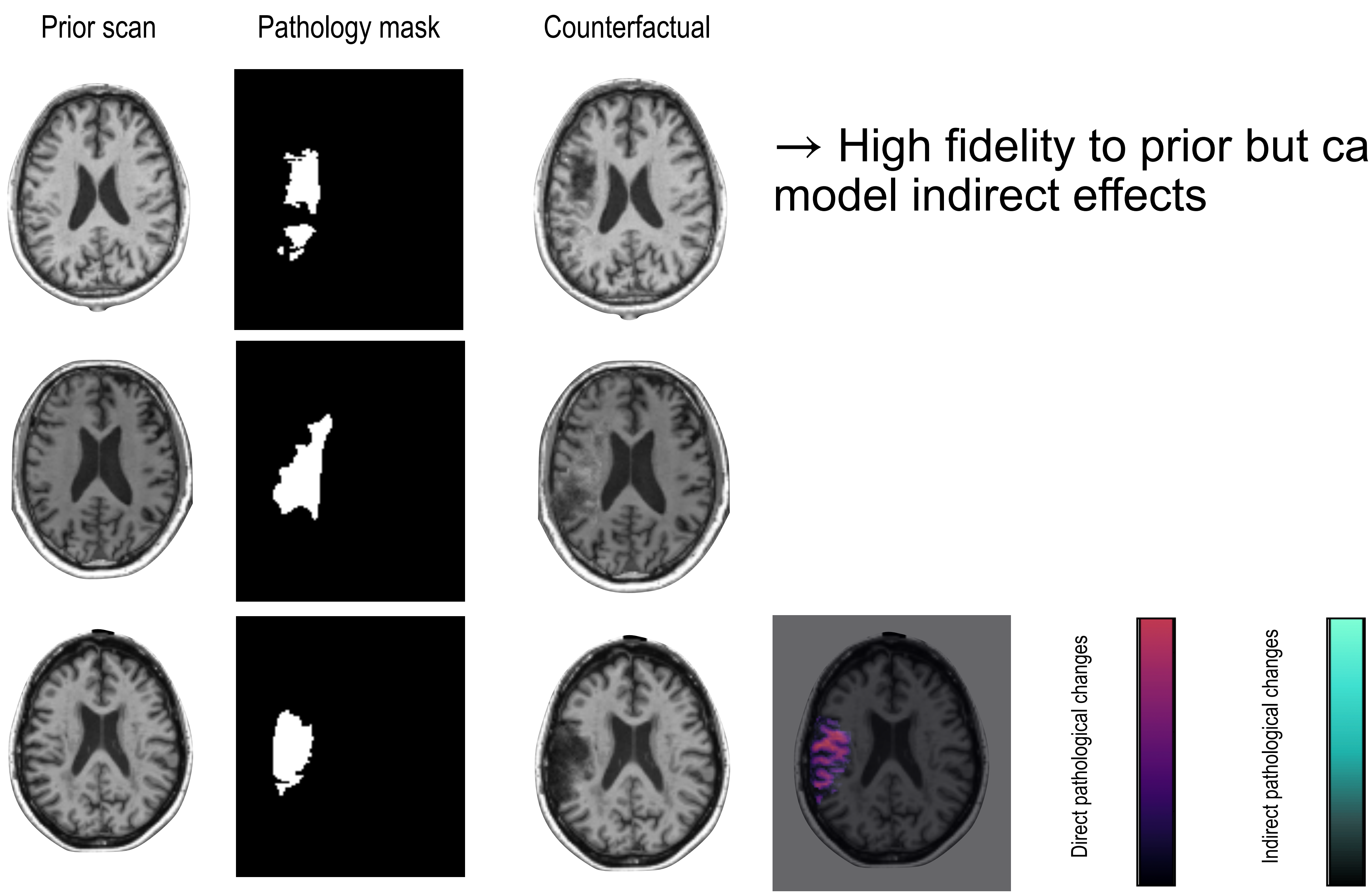
- Evaluate state-of-the-art image editing and inpainting techniques on generating counterfactuals
- Introduce MedEdit, which balances modelling indirect effects and high fidelity to prior
- Validate findings through clinical evaluation
- Improve MedEdit

Datasets:

- Atlas v2.0, BraTS 2023

Palette [3]

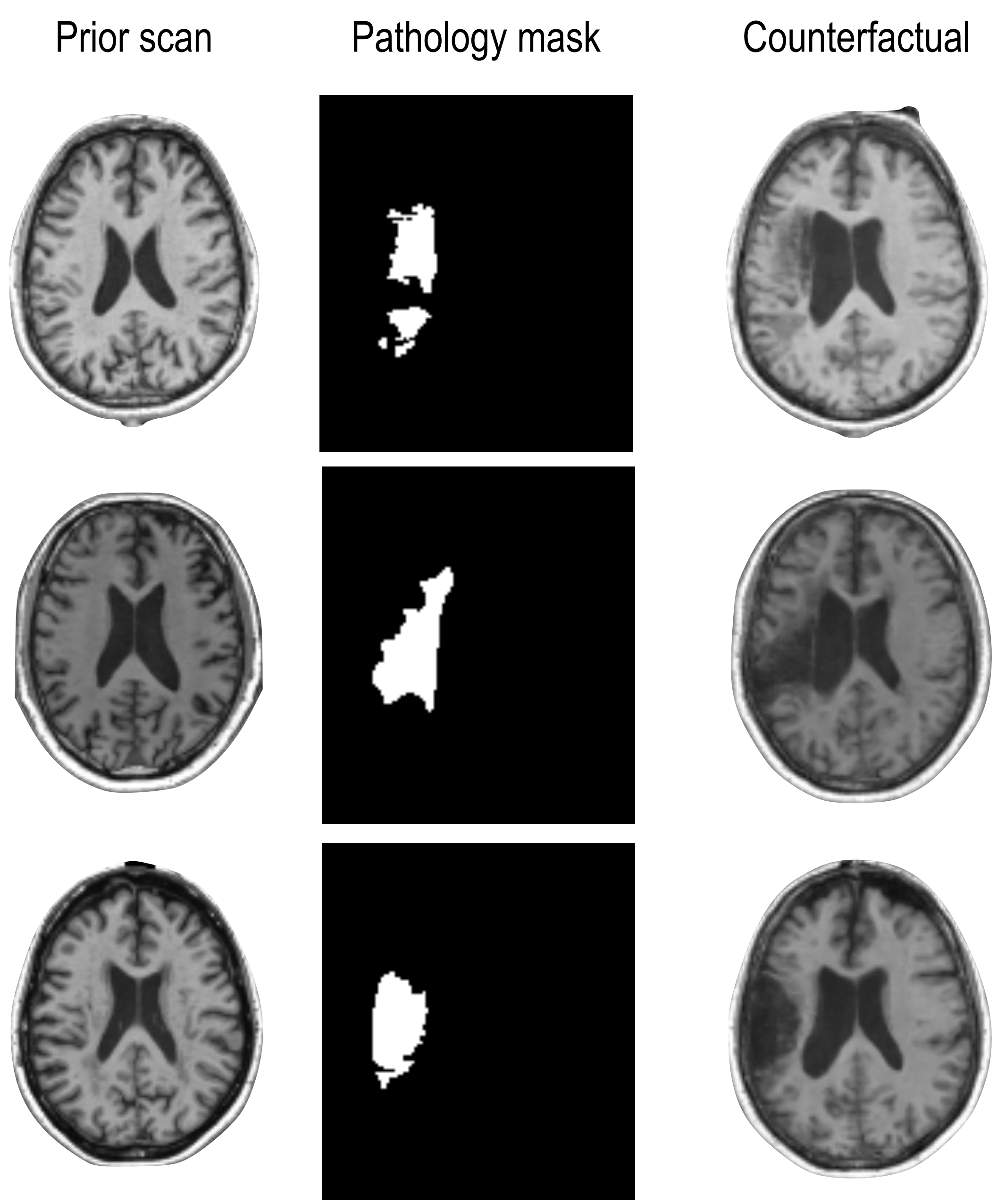
- Trained explicitly to inpaint masked part



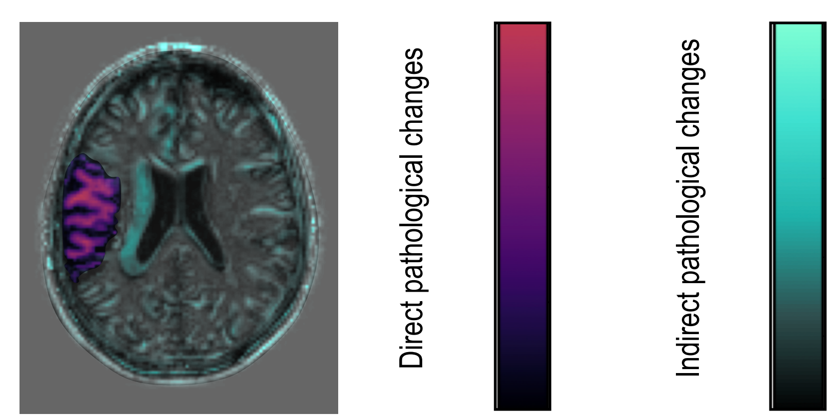
→ High fidelity to prior but cannot model indirect effects

SDEdit [4]

- At training time: train conditional diffusion model on pathological brain scans
- At generation time:
  - noise prior scan up to a certain timestep t
  - denoise conditioned on pathology mask



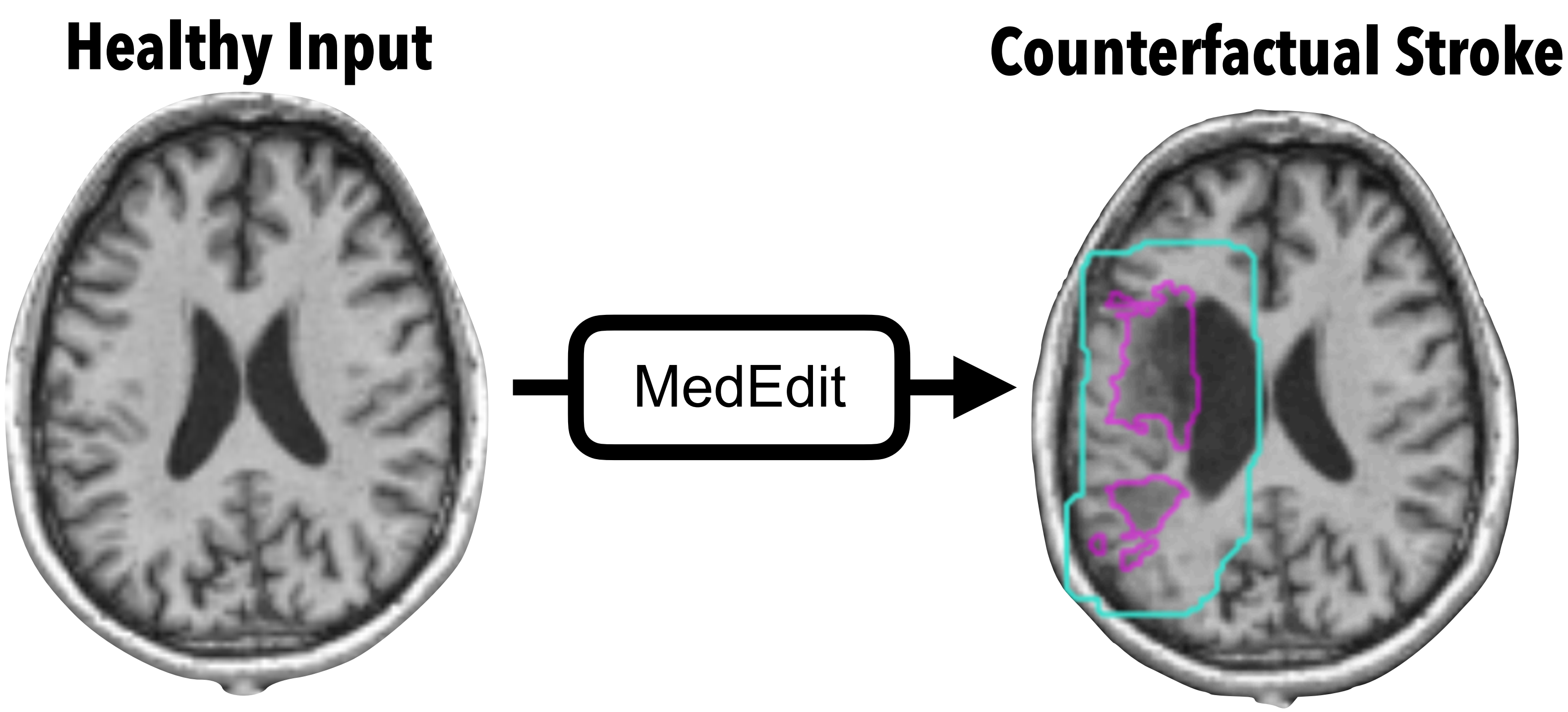
→ Models indirect effects but low fidelity to prior



MedEdit [5]

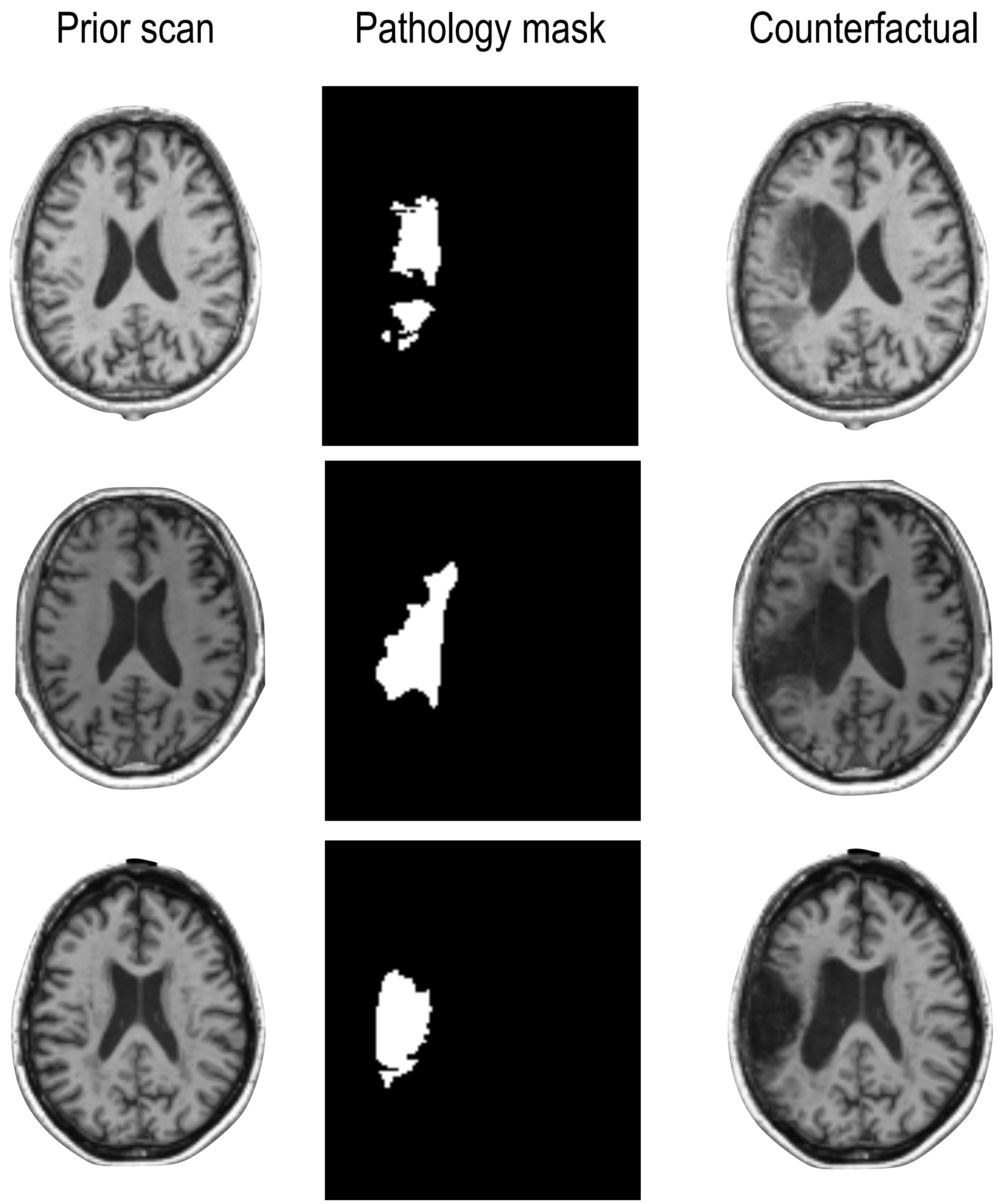
Approach:

- Train a diffusion model to learn the conditional distribution of pathological brain scans
- During inference, **the turquoise area** is edited conditioned on the **pathology mask (pink)** and the brain mask
  - inpainting mask is a dilated version of pathology mask

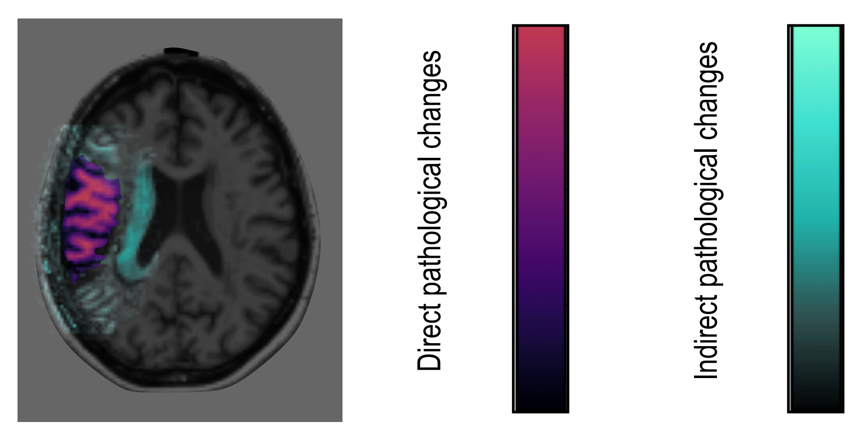


Note: MedEdit heavily bases on RePaint [6], which uses a class-conditional diffusion model and a user-specified inpainting mask

MedEdit



→ Balances modelling indirect effects and high fidelity to prior scan



Quantitative Evaluation

Computational metrics:

- **Realism**: FID [7] between generated counterfactuals and a set of real pathological scans
- **Adherence to desired pathological change:** Dice as identified by a trained nnU-Net [8]

Clinical metrics:

Radiologist's ratings [0 to 4] of:

- Realism,
- Adherence to desired pathological change,
- **Fidelity to prior scans** and
- **Modelling indirect pathological changes**

Quantitative Evaluation: Results

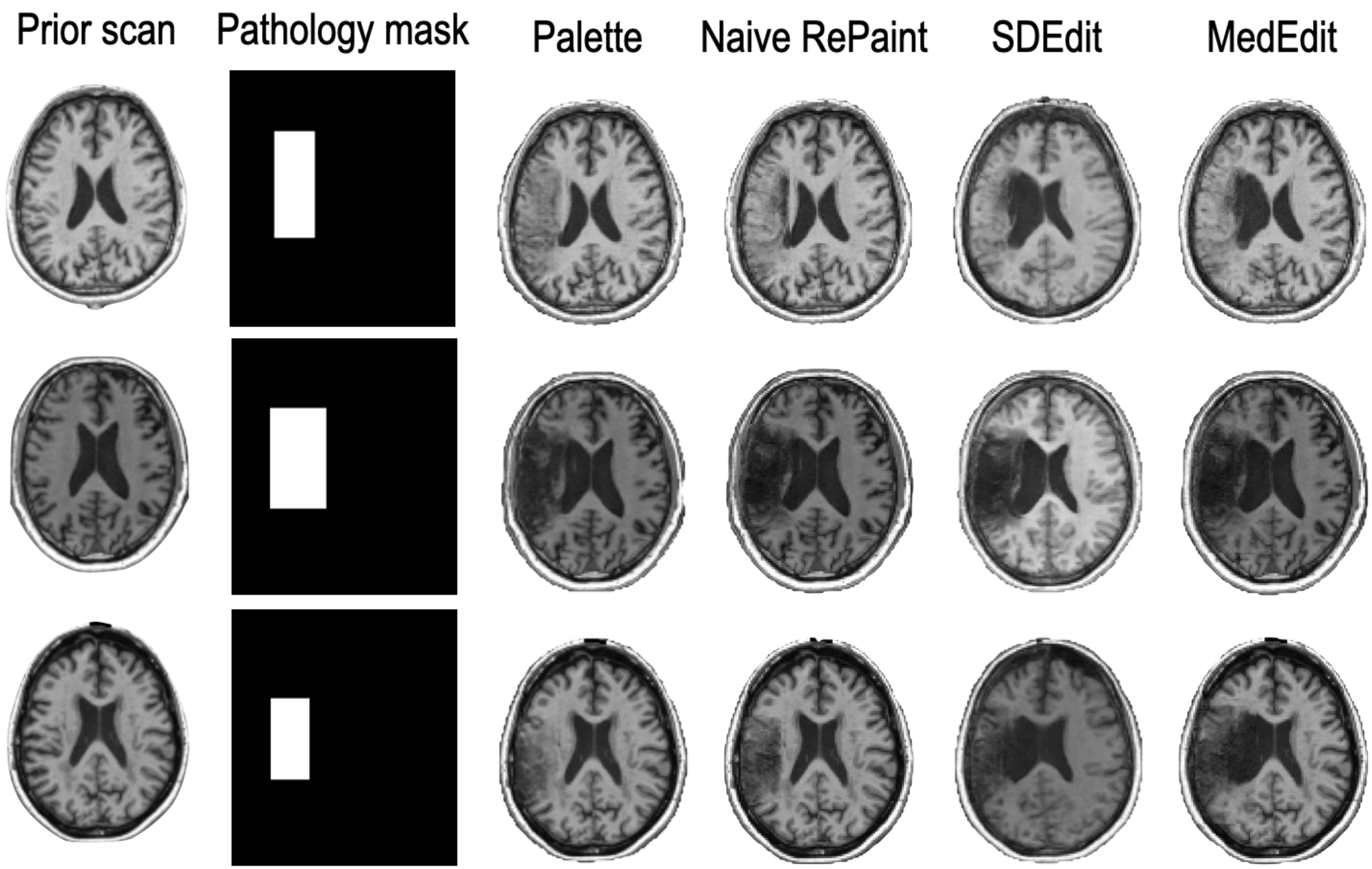
Method	Computational Metrics			Clinical Metrics				
	(1-Dice) * FID $\downarrow$	FID $\downarrow$	Dice $\uparrow$	Realism $\uparrow$	Fidel. $\uparrow$	Path. $\uparrow$	Ind-Path. $\uparrow$	Balance. $\uparrow$
Real samples	-	-	-	**3.20**	-	-	-	-
SDEdit	7.95 ▲ 159%	24.1	**0.67**	2.80	2.10	3.60	3.00	2.55
Palette	5.63 ▲ 83%	9.08	0.38	2.40	3.95	3.65	2.00	2.97
Naïve RePaint	4.24 ▲ 38%	8.31	0.50	2.55	**4.00**	**3.70**	1.85	2.92
MedEdit (ours)	**3.07** ▼ 28%	**8.30**	0.63	**3.20**	3.20	3.45	**3.15**	**3.17**

- Clinical metrics validate balancing high fidelity and indirect effects
- Computational metrics shows MedEdit's superiority

Improving MedEdit

- So far: restricted to strictly delineated masks, which we take from a hold-out set

→ limited generalisability

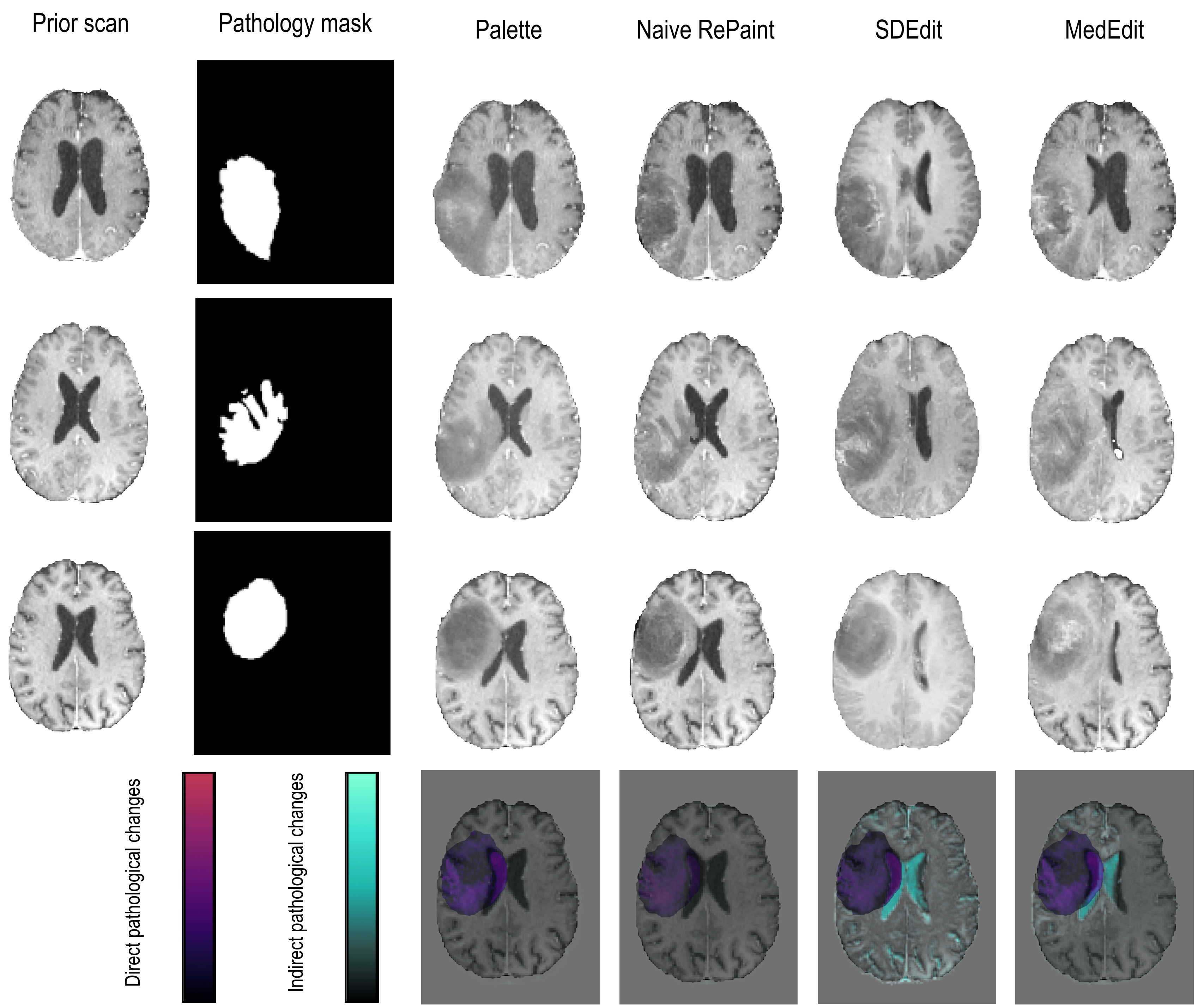


Improving MedEdit: Results

Method	Computational Metrics		
	(1-Dice) * FID ↓	FID ↓	Dice ↑
SDEdit	9.49 ▲ 174%	22.09	<u>0.57</u>
Palette	4.08 ▲ 18%	8.5	0.52
Naïve RePaint	3.73 ▲ 8%	<u>8.48</u>	0.56
MedEdit (ours)	**3.46** ▼ 7%	**8.45**	**0.59**

- MedEdit has better performance

Application to tumors



- Necrotic core not always introduced
- Reason: slicing of 3D volume

Conclusion

- MedEdit balances high fidelity and modelling indirect effects
- Evaluated with clinical and computational metrics
- Improved its generalisability to bounding-box conditioning instead of strictly delineated masks
- Application to tumors falls behind

References

[1] https://www.brainline.org/slideshow/brain-imaging-what-are-different-types [Accessed: 04.12.2023].
[2] R. master class. Mass effect. https://www.radiologymasterclass.co. uk/tutorials/ ct/ct_acute_brain/ct_brain_mass_effect. Accessed: September 25, 2024. 2024.
[3] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi. Palette: Image-to-Image Diffusion Models. 2022. arXiv: 2111. 05826 [cs.CV].
[4] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. 2022. arXiv: 2108.01073 [cs.CV].
[5] M. B. Alaya, D. M. Lang, B. Wiestler, J. A. Schnabel, and C. I. Bercea. MedEdit: Counterfactual Diffusion-based Image Editing on Brain MRI. 2024. arXiv: 2407.15270 [eess.IV].
[7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. “GANs trained by a two time-scale update rule converge to a local nash equilibrium.” In: Advances in Neural Information Processing Systems (2017).
[8] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.” In: Nature methods 18.2 (2021), pp. 203–211.

Appendix

RePaint:

- At training time: train diffusion model
- At generation time: modify reverse diffusion process

At each step of reverse process:

- Stage 1: conditions the inpainting of unknown area on the known regions

$$x_{t-1}^{\text{known}} \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t} x_0, \left(1 - \bar{\alpha}_t\right) \mathbf{I}\right)$$

$$x_{t-1}^{\text{unknown}} \sim \mathcal{N}\left(\mu_\theta\left(x_t, t\right), \sigma_t^2 \mathbf{I}\right)$$

$$x_{t-1} = m \odot x_{t-1}^{\text{known}} + (1 - m) \odot x_{t-1}^{\text{unknown}}$$

- Stage 2: diffuse x_{t-1} back x_t through: $x_t \sim \mathcal{N}\left(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbf{I}\right)$ and repeat the conditioning process for more harmonisation