FastSurfer-LIT: Lesion Inpainting Tool for Whole Brain MRI Segmentation With Tumors, Cavities and Abnormalities

Clemens Pollak, David Kügler, Tobias Bauer, Theodor Rüber, Martin Reuter
(AI in Medical Imaging, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany)

DOI: doi.org/10.1162/imag_a_00446
Imaging Neuroscience
Published: January 07 2025

https://github.com/Deep-MI/LIT

Limitation of previous neuroimaging analysis pipelines

- **Morphometric analysis of brain MRIs with pathologies:**
  essential to assess the impact of tumors, surgery, chemo/radiotherapy, or drug treatments
- **FreeSurfer** (Fischl, 2012), **FSL** (Jenkinson et al., 2012), or **SPM** (Ashburner, 2009) can segment brain MR images with small lesions (Radwan et al., 2021), but **not with large lesions (Resection cavities, tumors, etc.)**

⇒ registration problem with FreeSurfer (Radwan et al., 2021)

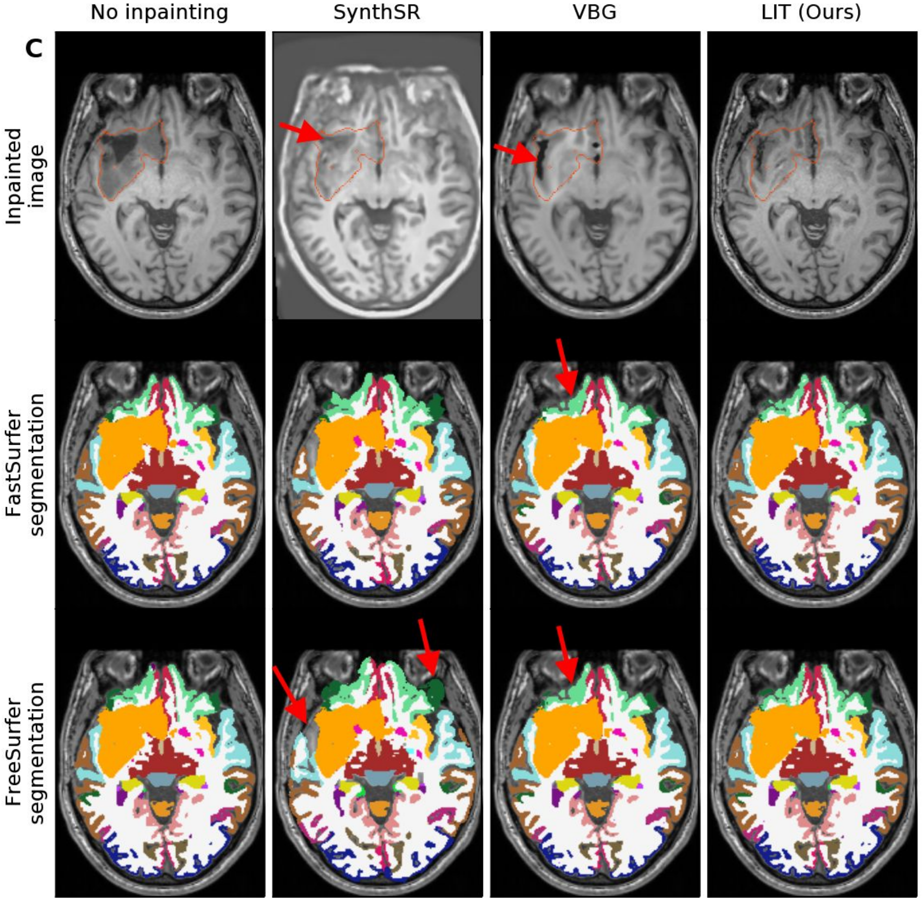
⇒ out-of-distribution data problem with FastSurfer (deep learning method)

Pre-existing pipelines

- **SynthSR** (Iglesias et al., 2023) and **Virtual Brain Grafting** (VBG) (Radwan et al., 2021)
  - inpainting healthy looking tissue in lesion areas
  - sometimes unstable: long runtimes, faulty segmentations, or unreliable estimates of cortical thickness (large lesions in particular)
  - 1mm resolution only (Henschel et al., 2022)

- **VBG** takes 2.5 hours (64% and 192% increase of computation time for FreeSurfer and FastSurfer respectively)
- **SynthSR** computes the inpainting in seconds using a CNN. No mask is used to specify the lesion localisation, meaning changes can occur in unexpected areas.

⇒ need of a **more robust inpainting** method

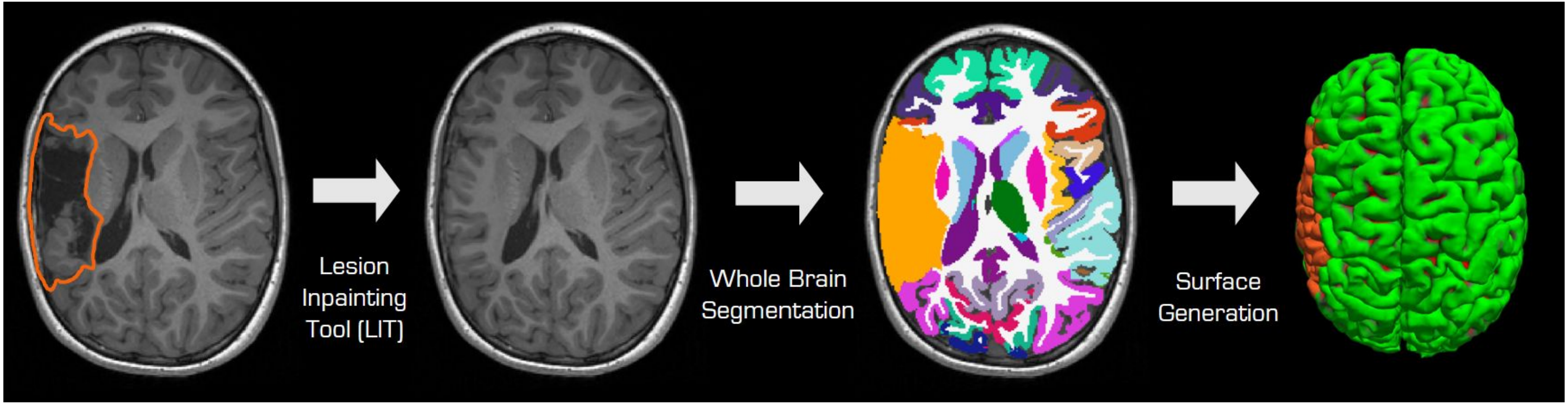


Fastsurfer-LIT pipeline

- **<u>Goals of "FastSurfer Lesion Inpainting Tool" (FastSurfer-LIT)</u>:**

1) **Denoising Diffusion Probabilistic Models (DDPM)**:
   fill lesion areas with healthy tissue that matches and extends the surrounding tissue
   Various resolutions, scanners, and types of lesion
2) Use Freesurfer/Fastsurfer

⇒ reduced selection bias (Guo et al., 2019), increased statistical power and allow group comparison to include cases with abnormalities



Denoising Diffusion Probabilistic Models (DDPM)

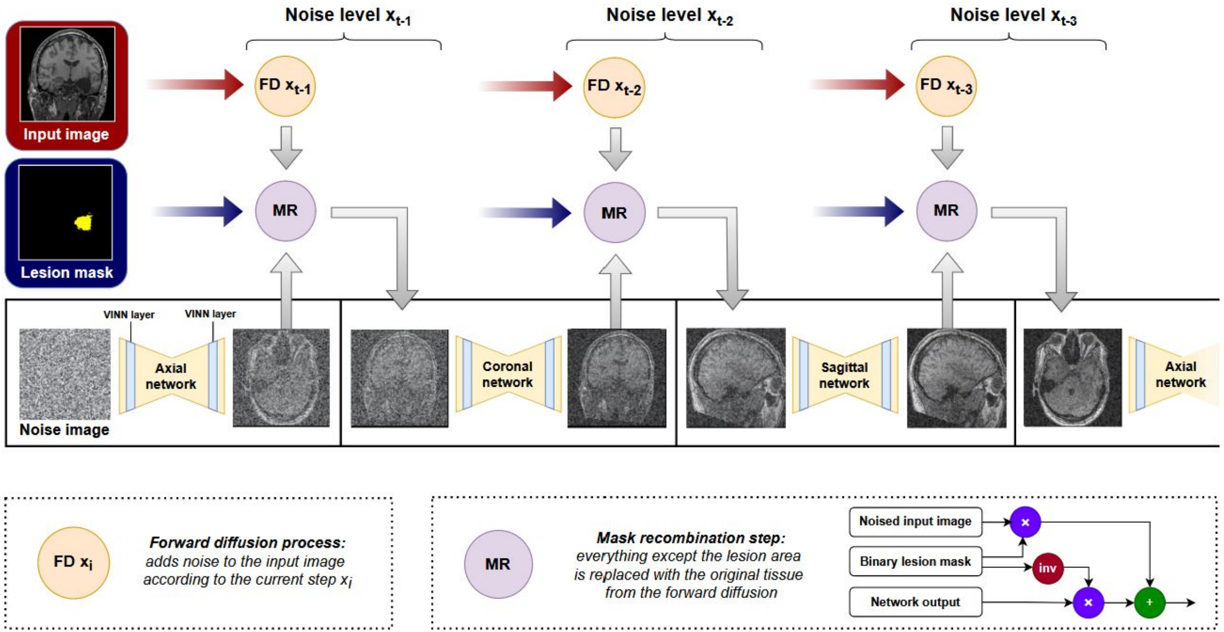


Figure A.1: Overview of the DDPM inpainting process. The architecture for the axial, coronal, and sagittal networks is shared, but weights differ. Mask recombination is used to guide the inpainting process. Denoising per step is exaggerated in the figure for better viewability. Not shown are slab selection and slab-wise denoising.

Datasets

Meta-dataset used for FastSurferVINN (Henschel et al., 2022): training, validation, and test

(a) HCP (Glasser et al., 2013) $(30, 20, 80)$ (b) RS (Breteler et al., 2014) $(30, 20, 80)$ (c) ABIDE-I (Di Martino et al., 2014) $(68, 0, 20)$ (d) ABIDE-II (Di Martino et al., 2017) $(0, 0, 25)$ (e) ADNI (Jack et al., 2008) $(215, 8, 40)$ (f) IXI ("IXI – Information eXtraction from Images", n.d.) $(400, 0, 43)$ (g) LA5C (Poldrack et al., 2016) $(203, 9, 15)$ (h) MBB (Babayan et al., 2019) $(195, 0, 0)$ (i) MIRIAD (Malone et al., 2013) $(30, 7, 0)$ (j) OASIS1 (Marcus et al., 2007) $(79, 11, 35)$ (k) OASIS2 (Marcus et al., 2010) $(65, 5, 17)$

Testing method for FastSurfer-LIT

- very few publicly available datasets with (manual) reference segmentations
  ⇒ **Tumor growth simulation** (Subramanian et al., 2019) in combination with segmentations of lesion free brains
  - 158 generated images of synthetic glioblastoma (Subramanian et al., 2019)
  - 39 generated images of multiple sclerosis (MS) lesions (Commowick et al., 2018; Commowick et 7 al., 2021)

1. **blinded rater comparison** with VBG method
   - 100 cases of hospital patients with 14 different kinds of tumors and surgical cavities
2. **consistency test of cortical thickness estimates** across 14 patients pre and post temporo-mesial resection surgery

Results

Method configuration	PSNR ↑	SSIM ↑	DICE ↑	HD95 ↓
LIT (VINN-DDPM with VA & VSC)	**29.64**	**0.72**	**0.9492**	**0.6217**
VINN-DDPM with VA	29.30	0.70	0.9491	0.6236
DDPM with VA	29.52	**0.72**	0.9484	0.6595
Baseline (2D DDPM)	27.91	0.66	0.9482	0.6640

Table 1: Method ablation results. All scores are calculated on the validation set with synthetic lesions and mass effects. Segmentation differences based on FastSurfer appear small, since segmentations perfectly match on slices unaffected by the lesion.
LIT = Lesion Inpainting Tool (proposed method), VA = view aggregation, VSC = varying spatial context, DDPM = Denoising Diffusion Probabilistic Model, VINN = Voxel Size Independent Neural Network, PSNR = Peak Signal To Noise Ratio, SSIM = Structural Similarity Metric, DICE = Dice similarity coefficient, HD95 = 95th percentile Hausdorff distance in millimeter.

Results

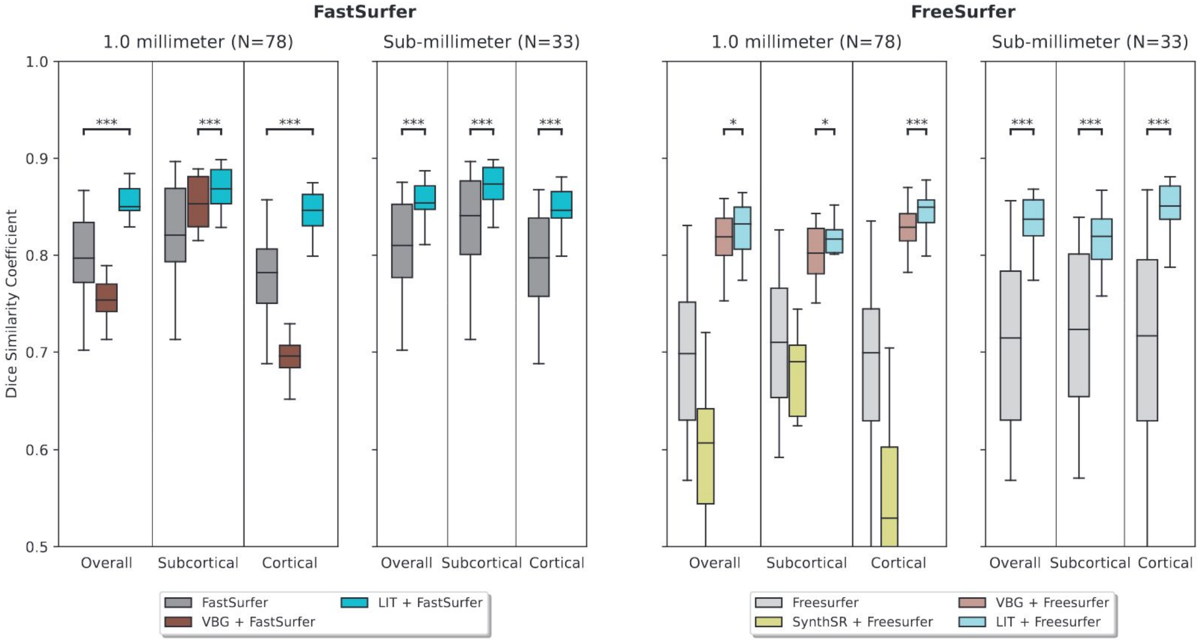


Figure 2: Method comparison on volumes with simulated glioblastoma. Reference segmentations are generated from lesion-free images by FastSurfer (left) and FreeSurfer (right) respectively (see Section 3.2). Our LIT inpainting significantly outperforms the state-of-the-art on all tests, measured by the Wilcoxon rank-sum statistic. p-values are indicated for comparison to the second best method. ($* := p < 0.05$, $** := p < 0.005$, $*** := p < 0.0005$).

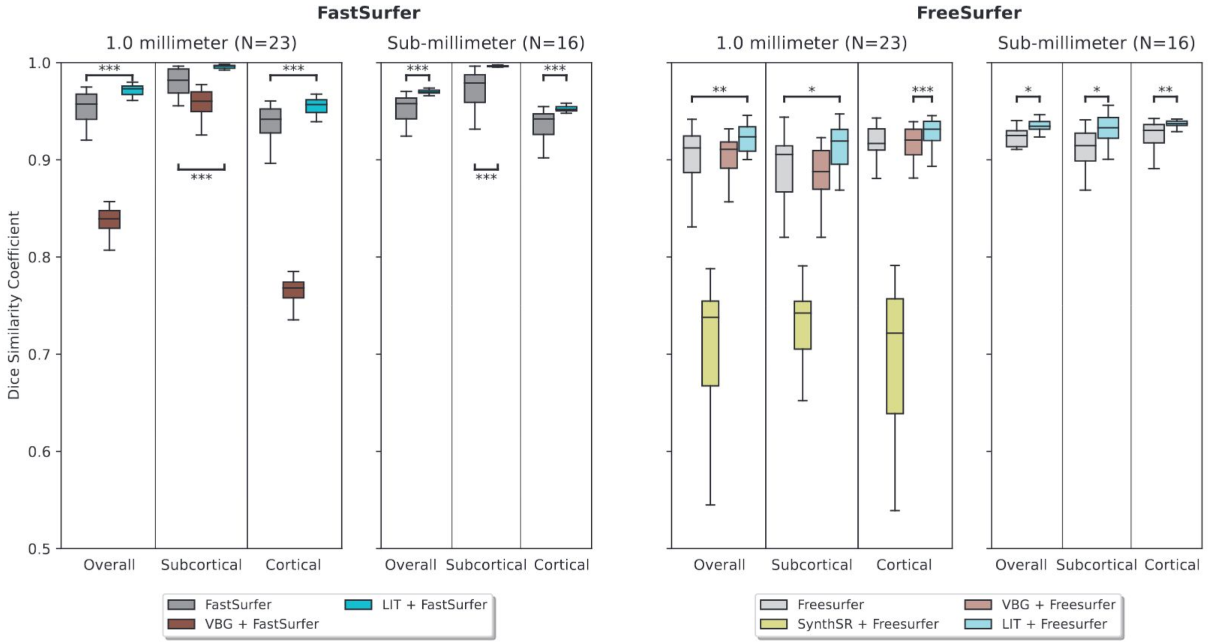


Figure 3: Method comparison on images with synthetic multiple sclerosis lesions. Reference segmentations are generated from lesion-free images by FastSurfer (left) and FreeSurfer (right) respectively (see Section 3.3). Our LIT inpainting significantly outperforms the state-of-the-art on all tests, measured by the Wilcoxon rank-sum statistic. p-values are indicated for comparison to the second best method. ($* := p < 0.05$, $** := p < 0.005$, $*** := p < 0.0005$).

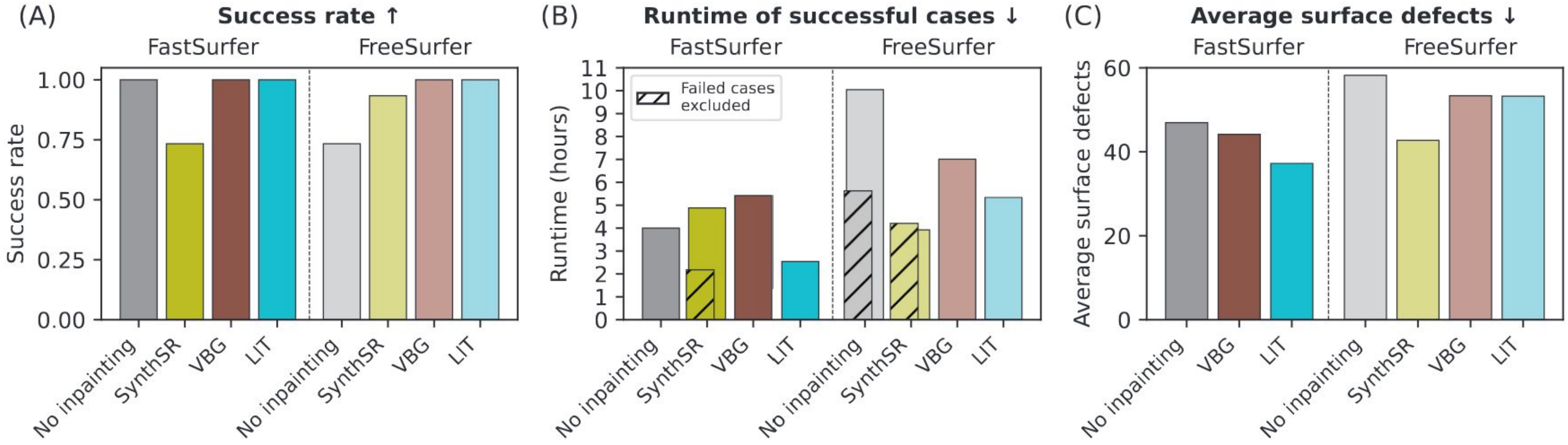


Figure 4: From left to right: The success rate of method (fraction of segmentations and surfaces generated within 24h), runtime on a desktop workstation, number of surface defects before fixing of surface topology. Subfigures A+B are based on 15 cases of the patient dataset, while Subfigure C is based on 100 cases of the synthetic lesion dataset.

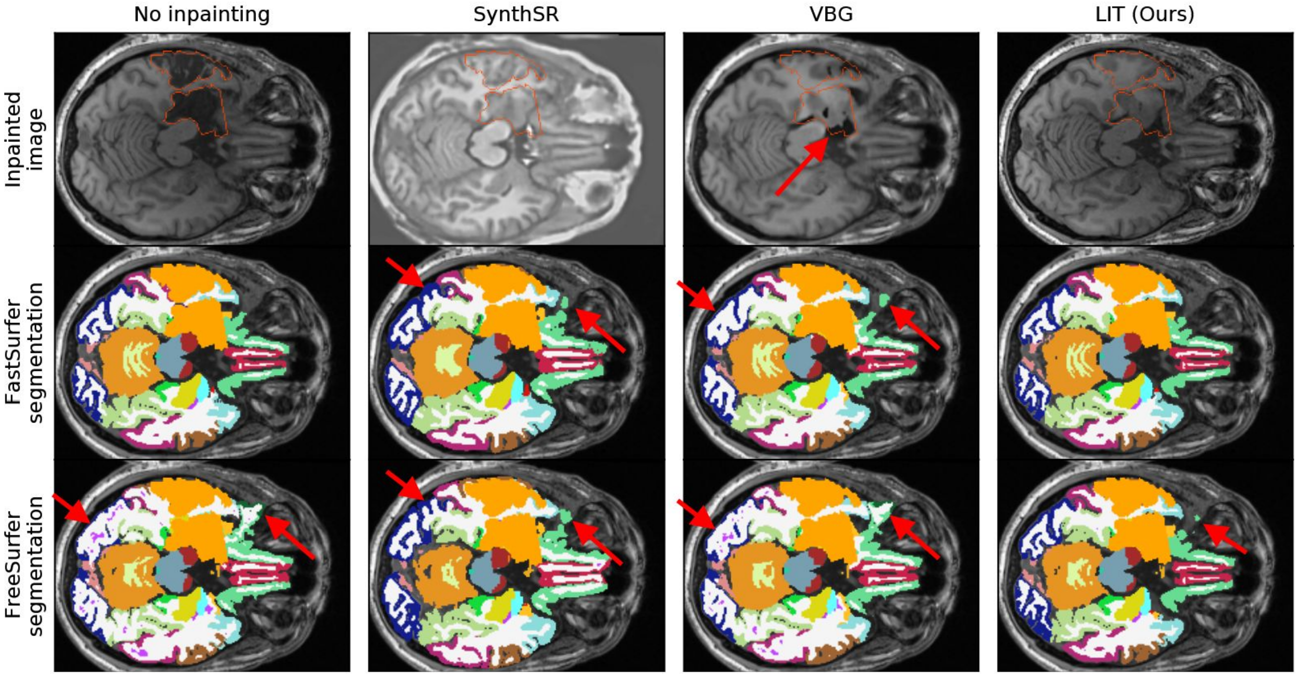


Figure 5: Qualitative comparison of whole brain segmentations for all method combinations on a representative case from the UKB, dataset. For randomly selected and difficult cases see Figure A.2 & A.3 respectively. The shown slices are located at the center of the lesion. Red arrows indicate inpainting and segmentation flaws.

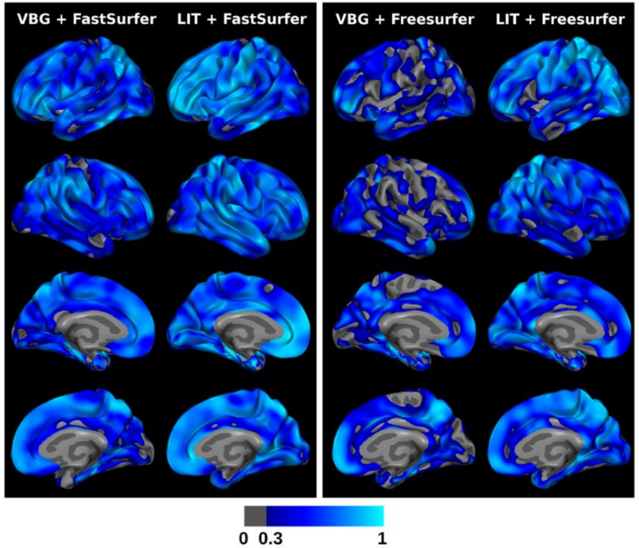


Figure 6: Comparisons of intraclass correlation coefficient (ICC) values for the cortical thickness of patients before and after temporomesial resection. Each column has different views of the semi-inflated template brain surface, where lighter shades of blue indicate a higher ICC and higher reproducibility of thickness estimates, when using inpainting for lesion filling.

Thank you for your attention !