

Swin-UNETR for stroke lesion segmentation

Liam Chalcraft¹ and Ioannis Pappas²

¹ Wellcome Centre for Human Neuroimaging, University College London

² Stevens Institute for Neuroimaging and Informatics, University of Southern California

l.chalcraft@cs.ucl.ac.uk, ipappas@usc.edu

Abstract. We use the hybrid SwinUNETR architecture proposed in [1]. The encoder is constructed from a Swin transformer [2], consisting of 4 stages each with 2 transformer layers containing (3, 6, 12, 24) heads respectively and using an initial feature size of 48. The convolutional decoder is constructed with 5 layers of channel sizes (768, 384, 192, 96, 48). Images are preprocessed using skullstripping, bias correction, reslicing to $1mm$, foreground cropping and z-score normalisation. Training data is augmented using lesion-weighted random crop to 128^3 , random flip, gaussian noise, gaussian blur and intensity shift. Training is performed for 1000 epochs with an Adam optimiser. Final inference is performed across the ensemble using flip-based test-time augmentation. All training was performed using NVIDIA DALI and Auto-Mixed Precision in Pytorch Lightning, and can be trained on new data using the implementation available at <https://github.com/liamchalcraft/MDUNet>.

References

1. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images (2022). <https://doi.org/10.48550/ARXIV.2201.01266>, <https://arxiv.org/abs/2201.01266>
2. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows (2021). <https://doi.org/10.48550/ARXIV.2103.14030>, <https://arxiv.org/abs/2103.14030>