Large Kernel Attention (Encoder) UNet for stroke lesion segmentation

Liam Chalcroft[1] and Ioannis Pappas[2]

[1] Wellcome Centre for Human Neuroimaging, University College London
[2] Stevens Institute for Neuroimaging and Informatics, University of Southern California
l.chalcroft@cs.ucl.ac.uk, ipappas@usc.edu

Abstract. We propose a hybrid U-Net model with a convolution-based transformer encoder consisting of 6 consecutive blocks of decreasing resolution using attention layers equivalent to a 21^3 convolution kernel based on the matrix decomposition method proposed in [1]. The 6 blocks have output channels of (32, 64, 128, 256, 320, 320) respectively. The $5th$ block has 2 transformer layers, whilst all others have 1. The decoder follows the typical CNN layout of nnUNet [2] with symmetrical channels to the encoder. 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/liamchalcroft/MDUNet.

References

1. Guo, M.H., Lu, C.Z., Liu, Z.N., Cheng, M.M., Hu, S.M.: Visual attention network (2022). https://doi.org/10.48550/ARXIV.2202.09741, https://arxiv.org/abs/2202.09741
2. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H.: nnu-net: Self-adapting framework for u-net-based medical image segmentation (2018). https://doi.org/10.48550/ARXIV.1809.10486, https://arxiv.org/abs/1809.10486