Automatic White Matter Hyperintensity Segmentation via Two-channel U-Net

Hongwei Li[1] Gongfa Jiang[1] Lei Zhao[1] Ruixuan Wang[1] Jianguo Zhang[2] and Wei-Shi Zheng[1]

Sun Yat-sen University, Guangzhou, China,
CVIP, University of Dundee, United Kindom

1 Data Preprocessing

Three datasets named *Utrecht, Singapore, and AmsterdamGE3T* were used as the training set. Each of them contains 20 patients' 3D MR images. Data preparation includes the following three steps.

1.1 Masking and slice selection

To reduce the noise from background, we remove the background information by masking the brain. A brain mask is obtained by simple thresholding and filling the holes inside the initial brain mask. We remove the first and last several slices of each brain since there is little information in such slices. For example, for *Utrecht* the first 5 slices and 5 last ones were removed during the training process.

1.2 Data normalization

Voxel intensity normalization per patient was performed during the training and testing stage. For the scan of each patient, the mean value and standard deviation were calculated based on intensities of all voxels inside the brain. Then each voxel intensity was normalized using the mean and standard deviation. In addition, each slice was cropped or padded such that its size became 200×200.

1.3 Data augmentation

Flipping, rotation, shearing, scaling along horizontal direction (x-scaling), and scaling along vertical direction (y-scaling) were employed for data augmentation. The value range for each augmentation transformation is shown in Table 1.

2 Methodology

2.1 U-net

Our segmentation model follows the basic architecture of U-Net[2], which is an extension of the fully convolutional networks (FCN)[1]. Basically, the U-Net

Methods	Rotation	Shearing	x-scaling & y-scaling
Parameters	[-15, 15]	[-0.1, 0.1]	[0.9, 1.1]

Table 1: Detailed parameters for data augmentation

consists of 27 layers, including 19 convolutional layers, 4 upsampling layers and 4 pooling layers. Unlike the original U-Net architecture, kernels of size 5×5 instead of 3×3 were used in the first two convolutional layers to capture rich local information. The number of kernels in these convolutional layers was modified as well. The input to the U-net consists of two slices, one from the FLAIR MRI and the other from the corresponding slice in the T1 MRI.

2.2 Ensemble model

To improve the robustness of our model, an ensemble method was employed for the final segmentation. Three U-net models with the same architecture were trained with different initializations. Then given a new patient data, each slice will be segmented based on the average scores from the output of the five U-nets, thus resulting a primary segmentation map.

2.3 Post-processing

To improve the performance on the primary segmentation map, we employ a simple strategy: if there exists predicted WMH in the first m slices and last m ones of a brain along the Z-direction, then the WMH regions were considered as false positive and would be removed. m was set to different values according to different datasets.

References

1. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
2. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015), http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a, (available on arXiv:1505.04597 [cs.CV])