Class-balanced Vascular Lesion Segmentation with Deeply Supervised 2D UNet

Shuai Chen,

Erasmus MC, Biomedical Imaging Group Rotterdam, Departments of Radiology and Nuclear Medicine, The Netherlands

1 Data

We use only the MRI scans provided by *Valdo 2021 Challenge* to train each task (e.g. train Task 1 without using the data from Task 2 and 3). For task 1 (Segmentation of Enlarged Perivascular Spaces (EPVS)) 40 subjects (T1-weighted, T2-weighted, and FLAIR) are available and the EPVS labels are provided over the selected region masked (only available for a fraction of the dataset); For task 2 (Segmentation of Cerebral Microbleeds (CMB)) 72 subjects (T1-weighted, T2-weighted, and T2S) are available and cerebral microbleed labels are provided for all subjects; For task 3 (Segmentation of Lacunes) 40 subjects (T1-weighted, T2-weighted, and FLAIR) are available and the lacune labels are provided for all subjects.

2 Preprocessing

2.1 2D patches extraction

There are severe class-imbalance problems in all three segmentation tasks where the target objects are relatively small to detect from the background area. Therefore, we extract only 2D slices that contain the target object for training in all tasks. The total number of 2D slices are 701, 364, and 421 for each task respectively. Around 20% are used as validation set for each task. 2D Patches with the resolution 192×160 (padded to 225×225) for PVS, 512×512 for CMB, and 384×320 (padded to 451×451) for LAC are further extract for the actual training.

2.2 Data augmentation

Several 2D data augmentation strategies were applied on the training patches, including 2D random rotation (180°), zooming ($0.7 \sim 1.4$), shifting (10 voxels), as well as random flipping in both x and y axis.

3 Method

Our method is a deeply supervised 2D UNet where the hyperparameters are defined by nnUNet [1]. Both the encoder and decoder paths have a 5-layer depth, where each layer consists of (3×3 convolution, instance normalization, leaky ReLU)×2 with zero padding. There is 2×2 downsampling or upsampling between each layer. The number of feature channels in the first layer is 32 and doubles in each deeper layer. We use sigmoid as activation function for the Dice loss in all segmentation tasks. Deep supervision is applied in all 5 depth layers. We use batch size 26 and learning rate 0.01 (SGD) and train the network for 1000 epochs. The runtime of each model is 1.5 hours on a GTX 1080 GPU. Early stopping is applied when there is no improvements on the validation performance for 10 epochs to avoid overfitting to the validation sets. An empirically selected threshold (0.5) is applied to the prediction from all models.

4 Initial Results

The validation Dice performance of the selected models for the three tasks are: 0.51 (PEVS), 0.76 (CMB), and 0.77 (Lacune). Note that the evaluation is only performed on the extracted slices that contain positives.

References

1. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Köhler, G., Norajitra, T., Wirkert, S.J., Maier-Hein, K.H.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. CoRR **abs/1809.10486** (2018), `http://arxiv.org/abs/1809.10486`