Method description for the KUBIAC team

Riccardo De Feo, Juhana Kaiponen, Jussi Tohka

A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Finland

Our approach to this segmentation problem (task 2) is to use an ensemble of convolutional neural networks. Our architecture is based on U-Net (Ronneberger et al. 2015) and draws inspiration from QuickNAT (Roy et al. 2018) and the work of Sichtermann et al. (2019) on aneurysm segmentation.

Standardization. For both training and inference, we rescale the pre-processed TOF-MRA and anatomical MRI volumes to a voxel size of 0.357x0.357x0.5 mm. We then normalize the intensities of each volume individually to have zero-mean and unit variance .

Architectures. Each sample includes one high resolution 64x64x64 patch, preserving the full detail of the original volumes, and one low-resolution 128x128x128 context patch, downsampled to 64x64x64. Both patches include two channels, extracted from the TOF-MRA and the anatomical volume.

- The U-Net architecture we implemented presents two different encoding paths, one for the high resolution input patch and one for the low resolution input patch. These are concatenated at the bottleneck and at each skip connection directed to the decoding path, generating our predictions. The Depth of the U-shape is of three levels, with each convolution block including 3 convolution-activation-batchnorm steps, and one residual connection.
- The CNN architecture also includes two input paths of two blocks following the same specifications. Their outputs are concatenated and fed through three more convolution blocks. This architecture does not include any pooling operation.

Training. We extracted 100 negative samples and 121 positive samples (when aneurysms are present) from each volume. We selected more positive samples to balance for those volumes displaying no aneurysms. Positive samples were selected by extracting a random positive voxel, giving more weight to smaller aneurysms, and applying a random shift of up to 32 voxels on all axes. On this dataset we trained 6 networks for each architecture according to the following procedure.

We randomly selected six groups of five subjects. Each network is trained using each group for validation and the remaining data for training. The resulting groups are:

1. 10078F, 10042, 10072F, 10031, 10026

2. 10062B, 10045B, 10071F, 10078F, 10010
3. 10051B, 10070B, 10013, 10057B, 10076B
4. 10061F, 10003, 10076B, 10057B, 10065F
5. 10048F, 10047F, 10015, 10066B, 10016
6. 10070B, 10076B, 10076F, 10042, 10037

We trained our networks using the RAdam optimizer (Liu, Jian, He et al., 2020) with a batch size of 8, shuffling the patches at every epoch. We calculated our loss function by combining three loss functions: weighted categorical cross entropy, Generalized Dice Loss (Sudre et al., 2017), and Boundary Loss (Kervadec et al. 2019). After each epoch we evaluated the results on the test set using the Dice overlap and saved the best model according to this metric. While we optimized to recognize three categories (background, aneurysms, and treated aneurysms) we only select our model based on the Dice overlap for the aneurysm category.

For the U-Net architecture, we further trained 6 networks without using GDL loss, replacing it with non-weighted Dice loss. Our validation results show that a weighted average of the predictions of all three networks provides better results than only using two or one. The entire process resulted in an ensamble of 18 neural networks.

Patch selection. At inference time we select patches according to the following procedure. Because aneurysms can typically be found in the TOF-images as brighter regions, we first select the 0.25% brightest voxels. Next the coordinates are divided into groups based on their location. The algorithm takes the first voxel selected and assigns to the same group all other voxels within a 10-voxel distance on any axis. This is repeated until all selected voxels have been assigned to a group. Finally, for each group the algorithm calculates the average middle point. These middle points are returned as centers for possible aneurysms.

Inference. The input volume is standardized, a list of locations of interest is selected, and sample patches are extracted as described above.

For each neural network, all selected patches are processed to generate an equal number of predictions. A voxel is preliminarly classified as aneurysm if that is predicted as the most likely class. These prediction patches are combined in one single volume, where for each voxel we take the average prediction across all patches including that voxel. Based on validation data, the prediction map intensities are rescaled to shift the optimal classification threshold at 0.5, and assigned a weight based on their validation Dice score. We can thus take a weighted averaged across all neural networks and threshold the result at the value of 0.5. The output is brought back to the original voxel size with a 0-order interpolation and the output file generated.