

Loss Ensembles for Extremely Imbalanced Segmentation

Jun Ma

Department of Mathematics, Nanjing University of Science and Technology, China

Abstract

This short paper briefly presents our methodology details of automatic intracranial aneurysms segmentation from brain MR scans. We use ensembles of multiple models trained from different loss functions. Our method ranked first place in the ADAM challenge segmentation task. The code and trained models are publicly available at <https://github.com/JunMa11/ADAM2020>.

1. Introduction: Task and Dataset

Early detection of intracranial aneurysms, as well as accurate measurement and assessment of shape, plays a critical role in clinical routine, which can allow to monitor of the growth and rupture risk of aneurysms and make proper treatment decisions to be made [6]. In the MICCAI 2020, ADAM (Aneurysm Detection And segMentation, <http://adam.isi.uu.nl/>) challenge was held to compare different methods of intracranial aneurysms detection and segmentation from Time of Flight Magnetic Resonance Angiographs (TOF-MRAs). The purpose of this short paper is to present our method details for the intracranial aneurysms segmentation task.

The challenge organizers provide 113 cases for training and 141 cases for testing. Each case includes one Time of Flight MRA (TOF-MRA) and one structural MR image (either T1, T2, or FLAIR). In the training set, 20 cases do not contain a diagnosed intracranial aneurysm. For the remaining 93 cases, manual annotations of unruptured intracranial aneurysms are provided in the form of binary masks. All manual annotations were made by experts in the field of medical annotations and checked by an experienced radiologist. In the testing set, 117 cases have at least one untreated, unruptured intracranial aneurysm and 26 cases do not have intracranial aneurysms. It should be noted that the testing set is not publicly released and participants are required to submit their method with a Docker container.

2. Method

The main difficulty in the aneurysms segmentation task is the extremely imbalanced problem. Specifically, the median image size and aneurysm voxel size are $512 \times 512 \times 140$ and 238, respectively. The foreground ratio is 6.5×10^{-6} where the aneurysm occupies very small partition in the whole image.

To handle the the extremely imbalanced problem, our key idea is to introduce ‘anti-imbalance’ loss functions to training neural networks and then fuses different models with ensembles. Our method is based on the well-known nnU-Net [3] and details are presented in the following subsections.

2.1. Preprocessing

We only use the official preprocessed TOF images to train 3D nnU-Net models. The preprocessing includes foreground (non-zero regions) cropping, rasampling, and Z-Score normalization, which are default settings in nnU-Net.

2.2. Network architecture

We use the default 3D full-resolution nnU-Net [3, 1] and set the maximum channel number as 360 (default 320).

2.3. Training protocols

We train two groups five-fold cross-validation models by different loss function combinations [4]: Dice[5, 2] + Cross-Entropy (CE) loss, and Dice + TopK loss [7]. We train all the models on NVIDIA TITAN V100 GPU with a patch size of $256 \times 224 \times 56$ and a batch size of 2. Each fold costs about 5 days.

2.4. Testing protocols

Five best models during the cross-validation are select for ensembles. To speed up the inference time, we disable the default test-time augmentation during inference.

2.5. Post-processing

We remove the predicted aneurysms less than 11 voxels from the segmentation results.

Table 1. Quantitative segmentation results (DSC) of different loss functions. The bold numbers denotes the best results.

Fold	Dice + CE loss	DSC + TopK loss
0	0.4370	0.4921
1	0.5476	0.4888
2	0.5108	0.4926
3	0.6173	0.5998
4	0.4404	0.5240

3. Experimental Results

3.1. Cross-validation results

Table 1 presents the cross-validation results in terms of DSC scores. It can be found that none of the loss function can achieve the best DSC score among all five folds. Thus, we select the five best-fold models for final ensembles.

3.2. Testing set results

Table 2 shows the quantitative results of top-2 participant teams on ADAM Challenge Leaderboard¹. The team ‘junma’ achieved the best DSC and Volumetric Similarity and the team ‘joker’ achieved the best HD95. However, it should be noted that the differences between them are marginal.

Table 2. Quantitative results of top-2 participant teams on ADAM Challenge Leaderboard.

Team	DSC	HD95	Volumetric Similarity	Rank
junma	0.41	8.96	0.50	1
joker	0.40	8.67	0.48	2

4. Conclusion

This short paper briefly presents our solution to the aneurysms segmentation in the ADAM challenge. The main idea is to use ensembles of multiple models trained with different ‘anti-imbalance’ loss functions. Our method ranked first place in the ADAM challenge segmentation task. However, the DSC score is still low, which has a large room for improvement. In the future, we will extend this work and evaluate our method on more extremely imbalanced segmentation tasks.

References

[1] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432, 2016.

[2] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer, 2016.

[3] F. Isensee, P. F. Jäger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 2020.

[4] M. Jun. Segmentation loss odyssey. *arXiv preprint arXiv:2005.13449*, 2020.

[5] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571, 2016.

[6] J. Wardlaw and P. White. The detection and management of unruptured intracranial aneurysms. *Brain*, 123(2):205–221, 2000.

[7] Z. Wu, C. Shen, and A. v. d. Hengel. Bridging category-level and instance-level semantic image segmentation. *arXiv preprint arXiv:1605.06885*, 2016.

¹<http://adam.isi.uu.nl/results/results-miccai-2020/>