Ensembling Convolutional Neural Networks for Ischemic Stroke Lesion Segmentation Based on CT Images

Mikhail Goncharov[1,2], Maxim Pisov[1,2], and Mikhail Belyaev[3,1]

[1] IITP RAS, Moscow, Russia
[2] MIPT, Moscow, Russia
[3] Skoltech, Moscow, Russia
goncharov.myu@phystech.edu, maksim.pisov@phystech.edu,
m.belyaev@skoltech.ru

Abstract. The problem of segmentation of lesion tissue in acute stroke based on CT images is critically important for clinicians. In this paper we describe a method of solving this problem exploiting state-of-art convolutional neural networks, which are now widely used to segment medical images. The final prediction is obtained as a result of averaging the predictions of 3 different U-Net-based models. The method was developed within the framework of the ISLES-18 challenge and showed quite good results on challenge's dataset.

Keywords: Convolutional Networks, Stroke Lesion Segmentation

1 Introduction

Ischemic stroke occurs when the blood supply to an area of the brain is disrupted, resulting in cells' death. Defining the location of the affected area is a key part of the decision-making process in acute stroke. At present, actually, clinicians manually outline core lesions on MRI images. Comparing with MRI, CT is more preferable due to its speed, availability and lack of contraindications. However, now there are no accurate methods, either manual or automated, defining regions of damaged tissue based on CT images. That is why *Ischemic Stroke Lesion Segmentation Challenge* (ISLES-18) [1] asks for advanced data analysis techniques that could help to define these regions on perfusion CT images. In recent years deep learning methods performed well in solving a wide variety of image processing tasks [5,6], including medical ones [8,3]. In this work we propose a method which combines inferences of several convolutional neural networks in order to decrease variability of the result of the segmentation.

2 Data and problem formulation

ISLES-18 dataset contains image data for 63 stroke patients. For some patients 2 slabs are provided and in total there are 94 instances. For each instance CT brain image and binary segmentation mask of the stroke lesion are provided. CT images have shape $5 \times 256 \times 256 \times s$ where 5 stands for a number of perfusion CT modalities and s is a number of slices, which varies from 2 up to 22. The provided ground-truth segmentation masks were manually drawn on DW-MRI images and have corresponding spatial shapes.

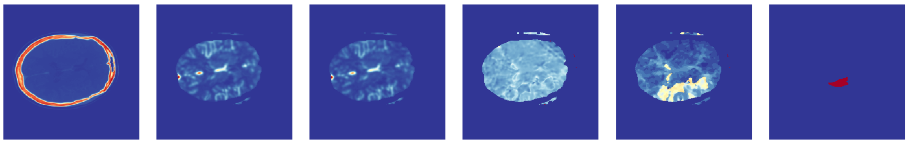

Fig. 1: An example of one slice of CT brain image and the corresponding segmentation mask.

Our task was to develop an automated method for segmentation stroke lesion based on CT images, which is quite challenging, taking into account the relatively small amount of data, its anisotropy, different ways of obtaining input data and target data (CT and MRI respectively).

3 Method

Since most of the images consist of 2 or 4 slices the use of 3D convolutional networks is unfeasible. That is why in our method we use 2D convolutional networks (the training and inference are performed on axial slices).

U-Net [8] is a fully-convolutional neural network architecture for segmentation, which is very popular in medical imaging. U-Net consists of a downsampling branch which alternates convolutional and pooling layers and a symmetric upsampling branch which alternates convolutional and upsampling layers. Outputs from layers of the downsampling branch are stacked with inputs of layers of the upsampling branch. Thus, the network is able to combine patterns from different scales in order to yield a more precise segmentation.

T-Net [7] is a variation of the U-Net architecture with convolutional layers inserted to the connections between the branches, which slightly increase the amount of processing at each scale.

In our method we ensemble both these architectures as well as a modification of T-Net aimed at yielding a better result according to one of three quality metrics considered in the challenge. In each network we replace simple convolutional layers with residual blocks [2] and add initial convolutional layers, which is a standard approach to improve learnability of deep networks.

Preprocessing As it turned out, there are slices in the data, on which there is almost no brain tissue, and therefore, no stroke. In order not to learn on such untypical edge cases we leave out the slices with background area greater than a fixed threshold and predict an empty mask for them; the threshold was chosen

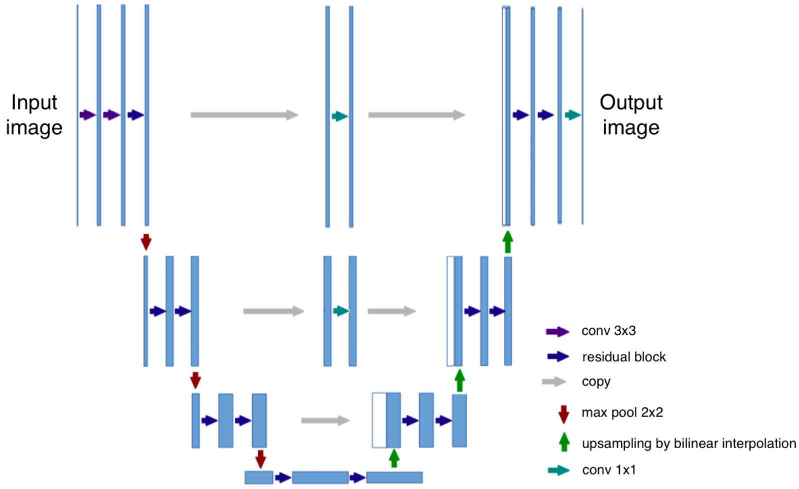

Fig. 2: T-Net architecture

as the maximal background area among the slices with lesions in the train set. Then we crop all images to their 3-dimensional bounding boxes and rescale them to the shape 256×256 in the axial plane.

Setup In our experiments we used Adam optimizer [4] with a constant learning rate of 10^{-3} or 10^{-4}.

To test the models' performance we used group 5-fold cross validation. In order to avoid overfitting the splits were made so that all images belonging to the same patient were presented in the same fold. We used the following quality metrics: Dice coefficient, Hausdorff distance and average symmetric surface distance (from the ground truth), which were proposed by the organizers of the challenge and also precision and recall coefficients. Each metric highlights different aspects of the segmentation quality, e.g. Hausdorff metric is very sensitive to false positives that are located far from the ground truth, whereas Dice metric is much more robust and estimates the fraction of the intersection of our prediction and the ground truth.

As loss functions we used simple binary cross entropy (BCE-loss) and also weighted binary cross entropy (WCE-loss)

$$WCE(x, y) = -\sum_i y_i \log(\sigma(x_i)) + w_i(1 - y_i)\log(1 - \sigma(x_i)),$$

where x is the network's output, y is the target, σ is the sigmoid function, which transforms network's outputs into probabilities; $w_i = 1 + 0.1d(i, y)$ – weights for background pixels, $d(i, y)$ is the distance from the i-th pixel to the lesioned region, i.e. this loss strongly penalizes distance from the ground truth false positives. We exploited this fact in order to train model that would yield a good result according to the Hausdorff metric.

Also, in order to overcome the dataset size limitations we use data augmentation techniques: random flips with respect to the sagittal plane and random rotations in the axial plane.

Models In our method we use 3 models

- U-Net with residual blocks operating on images downsampled by a factor of 4 along each dimension. During training BCE-loss was optimized.
- T-Net (Fig. 2). This model differs from U-Net only by the presence of additional convolutional layers between the branches, however, this change significantly influences the output, as can be seen in Fig. 3.
  Intuitively, both of the above models make relatively coarse predictions.
- A model similar to T-Net, but deeper and operating on images in the original scale (T-Net HD). Due to high resolution of the input and network's depth this model makes more localized predictions. During training WCE-loss was optimized, which, as expected, results in improvement of the Hausdorff score by 10% (comparing to the same model but trained with simple BCE-loss, see Tab. 1).

The final segmentation, which is equivalent to a pixelwise binary classification, is performed according to the averaging decision rule

$$z_i = \mathbb{1}\left(\frac{\sigma(x_i^1)+\sigma(x_i^2)+\sigma(x_i^3)}{3} \geq 0.5\right)$$

where x^1, x^2, x^3 are the outputs of the U-Net, T-Net and T-Net HD respectively.

4 Results

Table 1 shows the models' performances according to considered metrics. As seen, averaging improves both Dice score and Hausdorff distance, which is expected, as averaging unites confident predictions and effectively filters out random outliers.

Model	Mean Dice	Mean HD	Mean ASSD	Mean Precision	Mean Recall
U-Net	0.52 ± 0.01	29.65 ± 0.49	3.19 ± 0.26	0.60 ± 0.01	0.53 ± 0.01
T-Net	0.51 ± 0.01	28.45 ± 0.34	3.36 ± 0.27	0.59 ± 0.01	0.52 ± 0.01
T-Net HD	0.51 ± 0.01	24.75 ± 0.01	2.51 ± 0.27	0.64 ± 0.01	0.50 ± 0.01
T-Net HD-BCE	0.51 ± 0.01	27.58 ± 1.32	2.80 ± 0.38	0.63 ± 0.01	0.49 ± 0.01
Averaging	0.53 ± 0.01	23.29 ± 0.02	2.44 ± 0.32	0.65 ± 0.01	0.52 ± 0.01

Table 1: Segmentation results. T-Net HD-BCE denotes the same model as T-Net HD but trained with simple BCE.

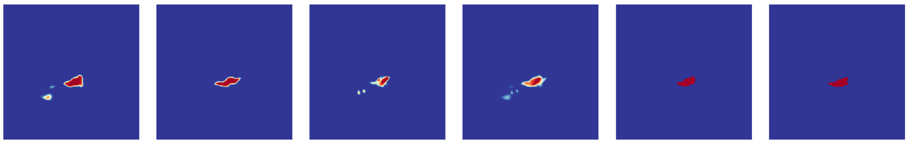

Fig. 3: An example of the final segmentation. From left to right: predicted by U-Net, T-Net, T-Net HD and averaged probability maps, binary segmentation mask and the ground truth.

5 Conclusion

From the very beginning of the challenge to the moment of writing this paper, we tested a large number of models with different setups and, finally, averaged predictions of the best of them. In most cases this method makes visually adequate predictions and yields quite good quality (average values of the quality metrics suffer due to several complex cases, our method can not cope with). Now we can not say with certainty whether the method is applicable in practice, since it could be possible to identify a particular group of people for whom the method yields acceptable quality or it can be useful in another formulation of the problem, for example, in determining the volume of the affected area. Unfortunately now the competition is not over yet and we can not compare our results with other participants, but it may well be impossible to achieve a significantly better result using deep learning methods. In the future work, in order to improve the method one can try to get a final prediction by ensembling more models and applying more sophisticated postprocessing techniques.

References

1. Web page of the challenge: http://www.isles-challenge.org/
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
3. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic for brain tumor segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 138–149. Springer (2016)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
6. Long, J., E., S., Darrell, T.: Fully convolutional networks for semantic segmentation arXiv:1411.4038 (2014)
7. Makarchuk, G., Kondratenko, V., Pisov, M., Pimkin, A., Krivov, E., Belyaev, M.: Ensembling neural networks for digital pathology images classification and segmentation. arXiv preprint arXiv:1802.00947 (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)