Integrated Extractor, Generator and Segmentor for Ischemic Stroke Lesion Segmentation

Tao Song(✉) and Ning Huang

SenseTime Inc., Shanghai, China
{songtao, huangning}@sensetime.com

Abstract. The challenge of Ischemic Stroke Lesion Segmentation 2018 asks for methods that allow the segmentation of stroke lesion based on acute CT perfusion data, and provided a data set of 103 stroke patients and matching expert segmentations. In this paper, a novel deep learning framework with extractor, generator and segmentor for ischemic stroke lesion segmentation has been proposed. Firstly, the extractor is to extract the feature map from processed perfusion weighted imaging (PWI). Secondly, the output of extractor, cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) and time of peak of the residue function (Tmax), etc. as the input of the generator to generated the Diffusion weighted imaging (DWI) modality. Finally, the segmentor is to precisely segment the ischemic stroke lesion using the generated data. In order to overcome the over-fitting, the data augmentation (e.g. random rotations, random crop and radial distortion) is used in training phase. Therefore, generalized dice combined with cross entropy were used as loss function to handle unbalanced data. All networks are trained end-to-end from scratch using the 2018 Ischemic Stroke Lesion Challenge dataset which contains training set of 63 patients and testing set of 40 patients. Our method achieves state-of-the-art segmentation accuracy in the testing set.

Keywords: Extractor · Generator · Segmentor · Generalized dice

1 Introduction

CT perfusion (CTP) [1] is an important diagnostic method in ischemic stroke. It enables differentiation of salvageable ischemic brain tissue (penumbra) from irrevocably damaged infarcted brain (infarct core). This is useful when assessing a patient for treatment (clot retrieval or thrombolysis). Compared with CTP, magnetic resonance images (MRI) is more sensitive to the early parenchymal changes of infarction. But its clinical application has been limited due to difficulties in timely access of MRI in many hospitals. For ischemic stroke patients, rapid imaging is especially important in the clinical treatment workflow.

The quantitative perfusion parameters of CTP (included CBV, CBF, MTT, Tmax, etc.) are usually used to identify the ischemic penumbra and the infarct core. The infarct core is defined as an area with prolonged MTT or Tmax, with markedly decreased CBF and CBV. The ischemic penumbra, which in most cases surrounds the

A. Crimi et al. (Eds.): BrainLes 2018, LNCS 11383, pp. 310–318, 2019.
https://doi.org/10.1007/978-3-030-11723-8_31

infarct core, also has prolonged MTT or Tmax (typically >6 s), but in contrast has only moderately decreased CBF and, importantly, near normal or even increased CBV [16]. Although the parameters of CTP can provide abundantly information in treatment of ischemic stroke, the accuracy is affected by factors such as the placement of arterial input function (AIF), and the deconvolution method used. In this work, we are also trying the extract the perfusion information directly from the CTP data, instead of indirectly from the perfusion parameters extracted.

In recent years, deep convolutional neural networks (CNNs) [2], have achieved great successes in image classification, segmentation and detection tasks and rapidly become the most popular technique in the medical imaging analysis. Image segmentation plays an important role in medical imaging, so the CNNs were firstly applied on medical image segmentation using patch-wise pixel classification. Later on, the global and local information are considered in the fully convolutional network (FCN) [3], which have encoder, decoder and show state-of-the-art performance in segmentation. The U-Net [4] is developed based on the FCN framework and uses the skip-connections to combine the low-level feature maps with higher-level feature maps, which has achieved better result in breast cancer segmentation in pathology. Cross-entropy (CE) was commonly used as loss function in segmentation networks of medical imaging, where background pixels are in majority which could cause serious class-imbalance problem, therefore the Dice loss function was proposed in [5] to alleviate this problem.

In this paper, we proposed an integrated network of ischemic stroke lesion segmentation for CTP data, which consists of an extractor, a generator and a segmentor. The FCN-like framework, with encoder and decoder, is used in extractor, generator and segmentor, respectively. This network is trained and tested on 2018 Ischemic Stroke Lesion Challenge [11], and achieved the first place of this challenge. Figure 1 shows the overall pipeline of our method, which contains three networks: (1) the extractor is to learn the representative image or the most important information from CT perfusion images; (2) the generator is to generate the DWI data using the output of the extractor

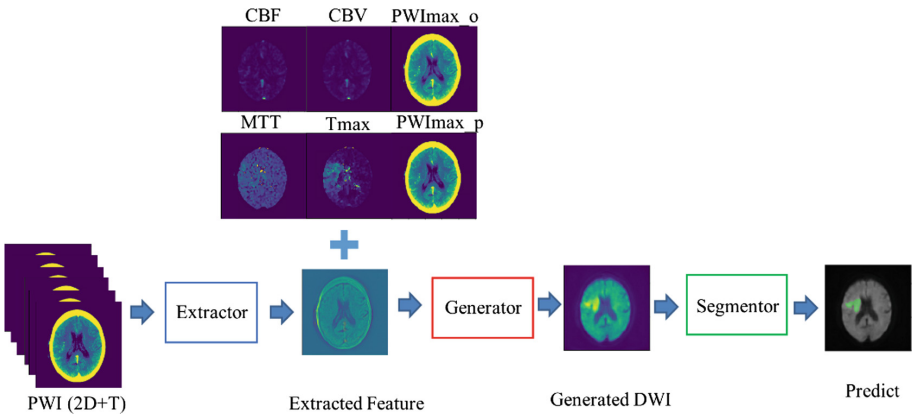

Fig. 1. The overall pipeline of our integrated segmentation framework. Here, *PWImax_o* and *PWImax_p* represents the max value in the time dimension of original PWI and preprocessed PWI, respectively.

and the perfusion parameters, which provides a better input for the segmentor; (3) the segmentor is to precisely segment the ischemic stroke lesion using the generated data. All networks are trained end-to-end, and the infarct core will be automatically predicted in the inference phase.

2 Method

2.1 Dataset and Preprocessing

The integrated segmentation framework is trained and tested using the 2018 Ischemic Stroke Lesion Challenge dataset. Imaging data from acute stroke patients in two centers who underwent CTP within 8 h of stroke onset, and were further scanned with MRI with DWI sequence within 3 h after CTP were included. Infarcted brain tissue is hyperintense on the DWI images. The training set consist of 63 patients, which contains, plain CT, DWI and CTP as well as perfusion maps of CBF, CBV, MTT and Tmax. Ground-truth segmentations were also provided. 40 patients with perfusion maps (CBF, CBV, MTT, Tmax), CT and CTP are provided in the testing phase, with no DWI or ground truth. The perfusion maps are calculated from original PWI using deconvolution method. The preprocessing of CTP data contains three stages: (1) firstly, the pixel values of a certain CTP frame are summed at any given time point, which forms a single time intensity curve for the whole CTP data of any case; (2) secondly, this time intensity curve is smoothed using a Gaussian smoothing filter, with a kernel size of 5. (3) finally, the 11 frames of CTP are selected, which are sampled between the onset of contrast injection and the end of the first pass. For normalization of the input data, the Batch Normal layer with no parameters is chosen.

2.2 Network Architecture

Our integrated segmentation framework consists of extractor, generator and segmentor, with all networks adapted from U-Net, which is a fully convolutional neural network and uses skip connections to combine low-level feature maps with higher-level ones. The U-Net consists of four blocks in the downsampling stage, each block has two 3×3 convolutions, each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with a stride of 2. At each downsampling step, the number of feature channels doubles. Every step in the expansive path consists of an upsampling of the feature map followed by a 2×2 convolution ("deconvolution") that halves the number of feature channels, a concatenation with the correspondingly feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. The network of extractor is a small U-Net that halves number of feature channels of U-Net in the downsampling and upsampling stage. The network of generator is an original U-Net without modification, except in the first convolution to adapt to the input data.

With regard to segmentor, the network is an attention U-Net as illustrated in Fig. 2. Compared with the original U-Net, it has an inserted network block, called squeeze-and-excitation networks (SE Block) [6], and with a switchable normalization (SN) [7] layer. The SE Block adaptively recalibrates channel-wise feature responses by

explicitly modelling interdependencies between channels using the attention mechanism. And the SN layer is to learn the weights of batch-wise, channel-wise and layer-wise normalizers for normalization, it is robust to a wide range of batch sizes, maintaining high performance even when the batch size is small.

In the training phase, the same two small networks are used to help train the generator following the generator, each network consists of five 3×3 convolutions with stride 2, each followed a ReLU, and two adaptive average pooling layers in the end of network with size 7×1 and 1×7, respectively.

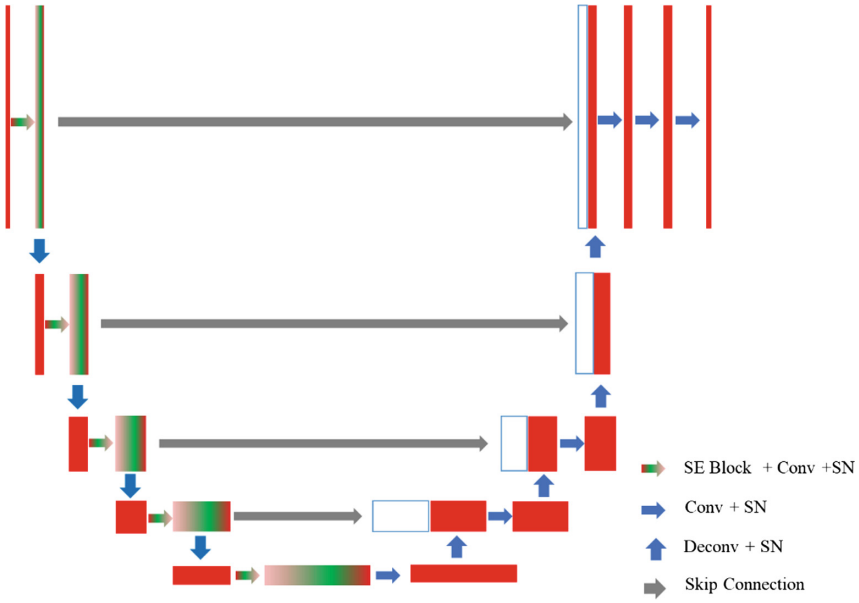

Fig. 2. The architecture of our segmentation network (Attention U-Net) adapted from 2D U-Net.

2.3 Loss Function

The output of extractor is a single channel feature map with the same size of input, which is calculated before sigmoid activation in the final layer. Through the sigmoid activation, this single channel feature map indicated the confidence probability of each pixel to be foreground. The L1 loss function is used for training in the extractor to regress the confidence probability and be expressed as

$$L_e = \lambda_1 * \|p - y\|_1 \tag{1}$$

where p and y represent the confidence probability of predicted and ground-truth respectively.

The loss function of generator has two parts: one part is the L2 loss to constrain the distance between the generated DWI and real DWI; the other part is to calculate the L2

distance of feature maps extracted from the generated DWI and the real DWI. Thus the whole loss function of generator can be written as

$$L_g = \lambda_2 * W * \left\| DWI_g - DWI_r \right\|_2 + \lambda_3 * \left(\left\| F_A - F_B \right\|_2 \right) \tag{2}$$

where the DWI_g and DWI_r is the generated DWI and real DWI, respectively. F_A is high-level feature map, extracted from generated DWI using model A, and F_B is the high-level feature map, extracted from real DWI using model B, the model A and B have the same network to extract high-level information, similar as perceptual loss. Here, W is the heat map of ground truth, calculated by signed distance function (SDF) [9], as is shown in Fig. 3.

As for the segmentor, its network predicts two output channels of the same size as the input, which indicate the probability of each pixel to be foreground or background after the pixel-wise softmax activation. In the medical image segmentation task, the background pixels are the majority, so the balance of the sample gradient must be considered. In this work, generalized dice [8] combined with cross entropy were used as loss function to handle the class-imbalance problem. The loss function is defined as the following

$$L_s = \lambda_4 * \{ W * CE - \log(generalized\ dice) \} \tag{3}$$

In this loss function, we consider the pixel-wise classification, and the similarity between the foreground prediction and the given ground truth, it is robust to a wide variance of input data. In order to balance the gradient size of cross entropy and generalized dice, the log operate is used.

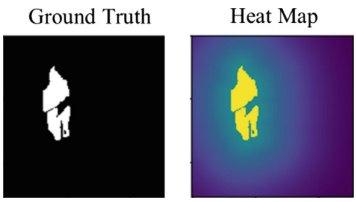

Fig. 3. The heat map of ground truth.

2.4 Training and Testing

As mentioned above, the extractor, generator and segmentor are the main components of our integrated segmentation framework, also the same two small networks (model A and B) need to be trained in the training phase. The five networks are trained end-to-end from scratch using the 2018 Ischemic Stroke Lesion Challenge dataset which contains training set with 63 patients and testing set with 40 patients. The training set is divided to four subsets to validate the trained models using the cross-validation method. Firstly, the extractor with input size 256 * 256 * 11 is used to extract the feature map using the regression of confidence probability. Then a concatenation with the extractor's feature map and other perfusion maps with a size of 256 * 256 * 7 as input of generator. In the end, the segmentor is used to predict the foreground region

(probability) using generated DWI with a size of 256 * 256 * 1. In training, the weights of all networks are initialized using Xavier initialization [12] and updated using RMSprop [14] optimizer with a batch size of 5 samples. The strategy of warm-up [13] and step-by-step learning rate decay are used, and the learning rate is initialized at 0.002 and reduced by factor 0.2 after 180, 300 epochs, and the λ_1, λ_2,λ_3 and λ_4 we set at 1.0, 0.002, 1.2 and 1.0, respectively.

In the testing phase, we preprocess a testing case to get the 11 slices of PWI as the input of extractor, and concatenate the extractor's feature map, CBF, CBV, MTT, Tmax, *PWImax_o* and *PWImax_p* to feed into the network of generator, then the Segmentor to predict the probability of infarct core using the generated DWI. The final segmentation region is an ensemble result of the four cross-validation models by computing their mean probability, and post-processing method of connected-component analysis is used to ensure the continuity of predicted area in space.

3 Result

The integrated segmentation framework is implemented by PyTorch [10] with cuDNN, and all experiments are performed on a workstation with 32 GB of memory, Intel Core i7 6700 k @ 4.0 GHz, and four Nvidia GTX 1080Ti 11G GPUs. In the training stage of 2018 ischemic stroke lesion challenge, the training datasets were divided into four subsets to validate the trained models by cross-validation, so all analysis of the trained model is performed on the cross-validation dataset. The detailed statistics are listed in Table 1.

Perfusion maps without DWI is provided in the testing phase, so we firstly attempt to feed the perfusion maps (CBV, CBF, MTT, Tmax) into the network of U-Net and use cross entropy after softmax activation for training, which caused a lot of false positives with a Dice score of only 0.53. In order to reduce the gradient of background to balance the gradient in the training phase, a novel loss function was designed using a combination of cross entropy and generalized dice, it improved the result with a Dice score of 0.55 and made the training more stable. Meanwhile, we designed a novel network (Attention U-Net), adapted form U-Net, which used an attention block (SE block) to get better performance with a Dice score of 0.56. An initial experiment using DWI as input achieves a dice score of 0.83, which inspired us to propose a two stage segmentation framework, which contains a generator with U-Net and a segmentor with Attention U-Net. In this framework, the *PWImax_p*, CBF, CBV, MTT and Tmax are used as the input of the generator to generate a pseudo-DWI, then the generated DWI is used to predict the region of infarct core. It increased the dice score by two percentage points on the original result. The heat map, calculated by sign distance function, is used as loss function for the generator and the segmentor, which makes the network focus more on the region of infarct core, so a Dice score of 0.59 was achieved. With the network becoming larger and larger, smaller batch number is needed in the training phase, so we replace Batch Norm (BN) [15] with Switch Norm (SN) in the networks. The SN is to learn the probabilities of different normalizers for normalization, it is robust to a wide range of batch sizes, maintaining high performance even when the batch size is small. Therefore, a Dice score of 0.60 is obtained by the SN.

Next, we try to extract the useful information from original CTP data to generate a more realistic DWI, firstly we extract 6 slices from PWI in time dimension, then a small U-Net was used to extract the representative images to feed into the generator's network. Compared with previous result, using the extractor makes the result increased by one percentage point. Later on, the two same small networks (model A and model B) are used following the generator, it makes the gradients descent faster and the training more stable. To calculate a more reasonable mean and standard deviation, the batch normal layer without learned parameters, designed before first convolution layer, is used to normalize the input data. Finally, the Dice score 0.62 is obtained in the validation set.

As shown as Fig. 4, the predictions are compared with ground truths, and the generated DWI is compared with the original DWI, respectively. It demonstrates the effectiveness of our integrated segmentation framework.

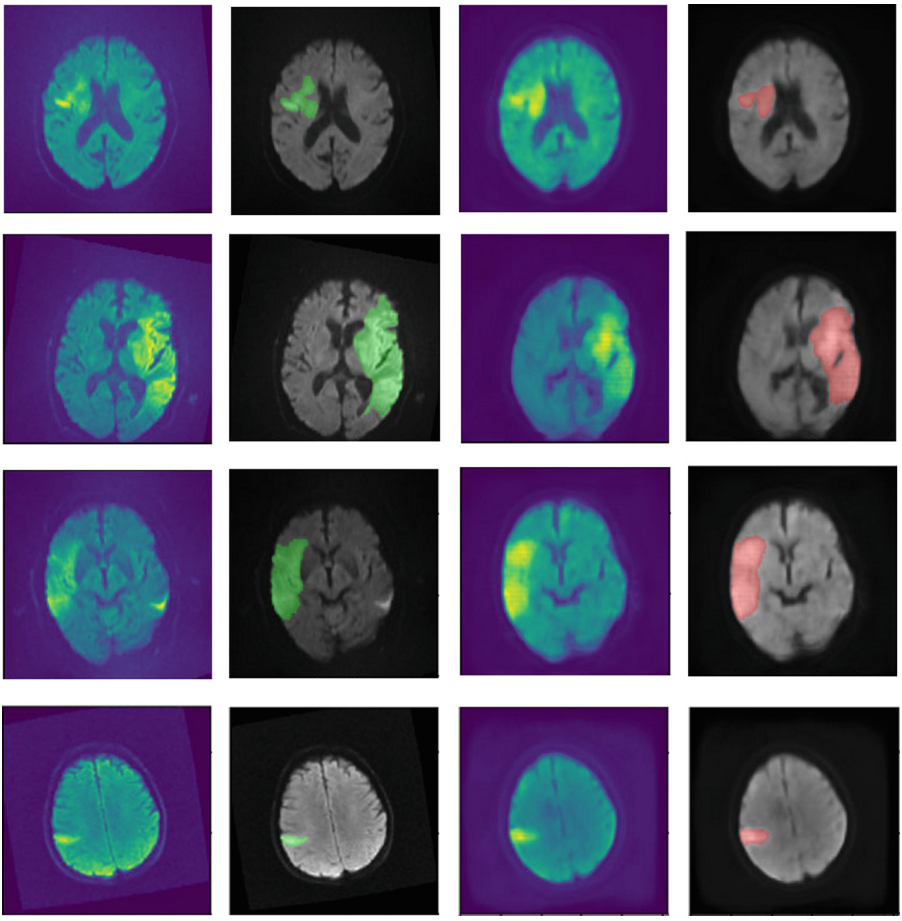

Fig. 4. Segmentation results of 4 cases in the validation set compared with the ground truth. The ground truths and predictions are given in green and red, respectively. From left to right, the original DWI, ground truth superimposed on original DWI, the generated DWI and the predictions superimposed on the generated DWI. (Color figure online)

Table 1. Ablation study of segmentation results using different methods. Here, AU-Net is the attention U-Net, and GD is generalized dice. And the E and G is the extractor and generator, respectively.

Stage	Method	Dice
One	U-Net + CE	0.53
One	U-Net + CE + GD	0.55
One	AU-Net + CE + GD	0.56
Two	AU-Net + CE + GD + G	0.58
Two	AU-Net + CE + GD + G + SDF	0.59
Two	AU-Net + CE + GD + G + SDF + SN	0.60
Three	AU-Net + CE + + GD + G + SDF + SN + E	0.61
Three	AU-Net + CE + GD + G + SDF + SN + E + BN	0.62

4 Conclusion

This paper detailed an integrated segmentation framework, which consists of extractor, generator and segmentor, for ischemic stroke lesion segmentation. First, the extractor is to extract the representative image from the original CTP data. Secondly, the generator is to generate the DWI from extractor's output and perfusion maps. Finally, the segmentor is to precisely segment the infarct core using the generated data. The network achieved a dice coefficient of 0.62 in cross validation stage and won the first place in the 2018 ischemic stroke lesion challenge in the test stage.

References

1. Eastwood, J.D., Lev, M.H., Azhari, T.: CT perfusion scanning with deconvolution analysis: pilot study in patients with acute middle cerebral artery stroke. Radiology **222**(1), 227–236 (2002)
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1097–1105 (2012)
3. 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)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
5. Zhang, J., Shen, X., Zhuo, T., et al.: Brain tumor segmentation based on refined fully convolutional neural networks with a hierarchical dice loss. arXiv preprint arXiv:1712.09093 (2018)
6. Luo, P., Ren, J., Peng, Z., Zhang, R., Li, J.: Differentiable learning-to-normalize via switchable normalization. arXiv preprint arXiv:1806.10779 (2018)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507 (2017)

8. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS - 2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
9. https://en.wikipedia.org/wiki/Signed_distance_function
10. Pytorch. http://pytorch.org/
11. http://www.isles-challenge.org/
12. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
14. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Netw. Mach. Learn. **4**(2), 26–31 (2012)
15. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
16. https://radiopaedia.org/articles/ct-perfusion-in-ischaemic-stroke