MULTI-TASK LEARNING IMPROVES THE BRAIN STROKE LESION SEGMENTATION

Libo Liu [1,2] , *Chengjian Huang* [1,3], *Chunsheng Cai* [4], *Xiaodong Zhang* [1], *Qingmao Hu** [1,2]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3] University of Chinese Academy of Sciences, Beijing, China
[4] Department of Cerebrovascular Disease, Huizhou First People's Hospital, Huizhou, China
qm.hu@siat.ac.cn

ABSTRACT

Fast and accurate segmentation of stroke lesions is highly desirable to help specialists in lesion measurements and making treatment plans. Segmentation of stroke lesions are challenging due to evolvement of stroke, low contrast and highly variable grayscale distributions. In this paper, we propose a novel multi-task learning framework to achieve enhanced segmentation of stroke lesions. In this framework, without increasing the inference time and model parameters in the inference phase, we perform supervised learning on a main task (stroke lesion segmentation) and an auxiliary task (edge segmentation of the brain lesion) to improve the performance of the main task. For validation, a CT brain infarction dataset has been constructed. In terms of the Dice coefficient, the proposed method could enhance at least 4.76% on the constructed CT dataset, and at least 7.97% on the open source dataset Anatomical Tracings of Lesions After Stroke than state-of-the-art models.

Index Terms— Stoke lesion segmentation, deep Learning, multi-task learning, auxiliary task.

1. INTRODUCTION

Stroke, is one of the main causes of mortality and disability in adults [1]. Accurate location and measurement of stroke lesions can provide medical specialists with an accurate basis for diagnosis and treatment plan. Manual delineation is tedious, time consuming and variable among raters, and most importantly may not be feasible due to the need for urgent response to stroke. Therefore, automatic segmentation of stroke lesions is of significance.

Recently, deep learning methods have achieved great success and become the mainstream methods in many medical image analysis tasks. Some of the remarkable progresses include: encoderdecoder architecture based on full convolution neural network such as FCN [2], SegNet [4] and U-Net [3] to improve the performance of segmentation; atrous spatial pyramid pooling [5, 6, 7], [12] based on atrous convolution and pyramid pooling architecture [8] to obtain larger receptive field and obtain multi-scale features without additional computational cost; architecture for multi-scale feature fusion like UNet++ [9]; and attention mechanism borrowed from the field of natural language processing [10] to obtain global context features and focus on important information like Attention U-Net [11]. For the stroke lesion segmentation task, the boundary between the stroke lesion and the background is very blurred, and some of the stroke lesions could be very small. In these existing segmentation methods, the spatial resolution of the feature map is continually reduced to obtain larger receptive fields and global context information. So these methods can not segment well the edge of the stroke lesions and the small stroke lesions because the important low-level feature information is lost. Although some methods like [3, 9] employed a large number of complex structures to obtain multi-scale features and retain low-level feature information, these methods still perform not good enough in segmenting small stroke lesions and stroke lesion edges. In addition, the inference speed of these models has been greatly reduced due to their complex structures. Recently, multi-task learning based image segmentation has been proposed to improve segmentation performance [13, 14, 15], but their inference time is also increased greatly due to the complex structures used to learn features for multi-tasks in these methods.

To address challenges of better segmentation accuracy and shorter inference time, we propose a novel multi-task framework for stroke segmentation in this paper. The main task is stroke lesion segmentation, and the edge segmentation of stroke lesions is the auxiliary task to improve the performance of the main task. For the main task, a single task image segmentation network could be employed. However, how to design model to learn auxiliary task without increasing the number of calculations and model parameters in the inference phase is not trivial. To succeed in the auxiliary task, we need to extract features corresponding to edges of the stroke lesions. The image is a two-dimensional (2D) signal, which is composed of high-frequency signals and a low-frequency signals. As edges are high-frequency signals of the image, they could be attained by subtracting the low-frequency signals from the entire image signals. Likewise, edge features could be obtained by subtracting the low-frequency features from the stroke lesion features. In convolution neural networks, the feature map of the deepest layer often represents low-frequency features. So, the low-frequency features could be extracted from the deepest layer of the stroke lesion segmentation network, which are subtracted from the whole stroke lesion features to attain features corresponding to auxiliary task.

During the training step, the stroke lesion features are supervised by the stroke lesion segmentation task, while the edge features of the stroke lesions are supervised by edges of the stroke lesion task for learning edge prediction. Because the main task branch and the auxiliary task branch can be separated, only parameters of the main task will be retained after training. Therefore, the number of parameters and inference time will not increase compared to a single task brain stroke lesion segmentation network. At the same time, in order to achieve effective supervision of auxiliary task and promote the training of the main task model, a consistency loss function is proposed. In addition to a public dataset (ATLAS [17]), a computed tomog-

*Corresponding author

raphy (CT) dataset has been constructed to verify the effectiveness of this framework. The proposed framework achieved a remarkable improvement as compared to other state-of-the-art 2D models on the two datasets. Here is a summarization of our contributions:

- We proposed a novel multi-task learning framework for stroke lesion segmentation.
- The proposed multi-task learning framework has the same number of parameters and same inference time as the single task stroke lesion segmentation network.
- We proposed a consistency loss for multi-task learning framework to enhance the performance of the stroke lesion segmentation.
- We have carried out experiments on two stroke lesion segmentation datasets and attain the top performance.

2. METHODS

In this section, the proposed method for stroke lesion segmentation is described in detail in 3 sub-sections, i.e., overview of the method, low frequency feature generation module and consistency loss.

2.1. Overview of the Method

The diagram of the proposed method is shown in Fig.1. The input is a 2D brain image which is fed into a stroke lesion segmentation network. The stroke lesion segmentation network can be any kind of image segmentation network using deep learning. In this paper, we use the UNet++ [9] as the stroke lesion segmentation network, which is a classic and powerful medical image segmentation network with U-shaped structure. The output of the stroke lesion segmentation network is then used as the features representing the stroke lesions. For the convenience of description, we use F_{whole} to represent this feature map, which is fed into a convolution layer with kernel size of 1×1 and sigmoid function to predict the mask of the stroke lesions. For the auxiliary task, the deepest layer feature map of the stroke lesion segmentation network is fed into a low frequency feature generation module to get the feature map representing the low frequency features of the stroke lesions. Here, F_{low} and F_{edge} respectively are features maps representing the low frequency features and edges of the stroke lesions, satisfying

$$F_{edge} = F_{whole} - F_{low} \tag{1}$$

Finally, F_{edge} is fed into the edge segmentation head which combines a convolution layer with kernel size of 1×1 and sigmoid function to predict edges of the stroke lesions. In the training step, the main task branch and auxiliary task branch are trained together. The main task branch will calculate L_{mask} loss and L_{lap} loss with the corresponding ground truth to learn a good feature representation. At the same time, the output of the auxiliary task branch will calculate l_{edge} loss with the corresponding ground truth to learn auxiliary tasks to promote the main task. L_{lap} loss, L_{mask} loss and L_{edge} loss will be described in detail in subsection 2.3. After the training step, only the parameters of the main task branch are retained. In the inference phase, we only perform the inference of the main task branch. Therefore, the number of parameters of the proposed method at the inference stage is the same as that of a single task stroke lesion segmentation network. In this way, we have improved the performance of the stroke lesion segmentation network without increasing the number of parameters and slowing down the inference speed.

2.2. Low-Frequency Feature Generation Module

In order to extract better low-frequency features, we designed a simple low-frequency feature generation module (LFFGM). As shown in Fig.1, the module is a skip-connect structure. The input of the LFFGM is the deepest feature map of the stroke lesion segmentation network which is fed into a U-shaped structure to extract better low-frequency features. In the U-shaped structure, two sequential operations starting from one 2×2 Max pooling operation with stride of 2, followed by one 3×3 convolution operation, one batch normalization layer and one ReLU function, are applied to decrease the resolution of the feature map by 4 times. Then two sequential operations starting from one 2×2 up sampling with stride of 2, one 3×3 convolution, one batch normalization layer and one ReLU function are used to recover the resolution of the feature map. Then a concatenation operation is applied to concatenate the feature map generated by U-shaped structure and the input of the LFFGM. Finally, a 1×1 convolution is applied to reduce the dimension of the feature map to derive F_{low} with the same dimension as F_{whole}.

2.3. Consistency Loss

Image segmentation can be regarded as classification of all pixels. The stroke lesion segmentation can be regarded as classification of two classes, i.e., stroke lesion and non-stroke lesion. Here the traditional cross entropy loss and Dice loss [16] are adopted to cope with the extremely unbalanced positive and negative samples, i.e.,

$$Loss_{segment} = \alpha DiceLoss + \beta CE \tag{2}$$

$$CE = \begin{cases} -log(p), & y = 1 \\ -log(1-p), & otherwise \end{cases} \tag{3}$$

$$DiceLoss = 1 - \frac{2|p \cap y|}{|p| + |y|} \tag{4}$$

where p is the prediction of the segmentation model and y is the ground truth segmentation, α and β are weighted coefficients to be set via experiments. In the following experiments, α and β are all set as 1. So for the stroke lesion segmentation task and edges of the stroke lesion task, we use the $Loss_{segment}$ to achieve effective supervision. In addition to measuring the segmentation performance, we also hope that the stroke lesion predicted by the main task branch and the ground truth are as consistent as possible in terms of second derivative. Because the more consistent the second derivative of the two, the better the edge segmentation performance of the model. So we use Laplacian operator to process prediction of the stroke lesions and ground truth to obtain their second derivative and calculate their mean absolute error. The loss can be written as:

$$Loss_{lap} = |f_{laplace}(p) - f_{laplace}(y)| \tag{5}$$

where p is the prediction of the segmentation model, y is the ground truth segmentation, and $f_{laplace}$ is Laplacian operator. So the total loss for our proposed method can be written as:

$$Loss = \gamma Loss_{mask} + \lambda Loss_{edge} + \delta Loss_{lap} \tag{6}$$

where $Loss_{mask}$ is the $Loss_{segment}$ loss of the stroke lesion segmentation, $Loss_{edge}$ is the loss of the edges of stroke lesion segmentation, and $Loss_{lap}$ is the loss which is described in formula 5, γ, λ and δ weighted coefficients to be set via experiments. In the following experiments, γ, λ and δ are set as 1, 1 and 10.

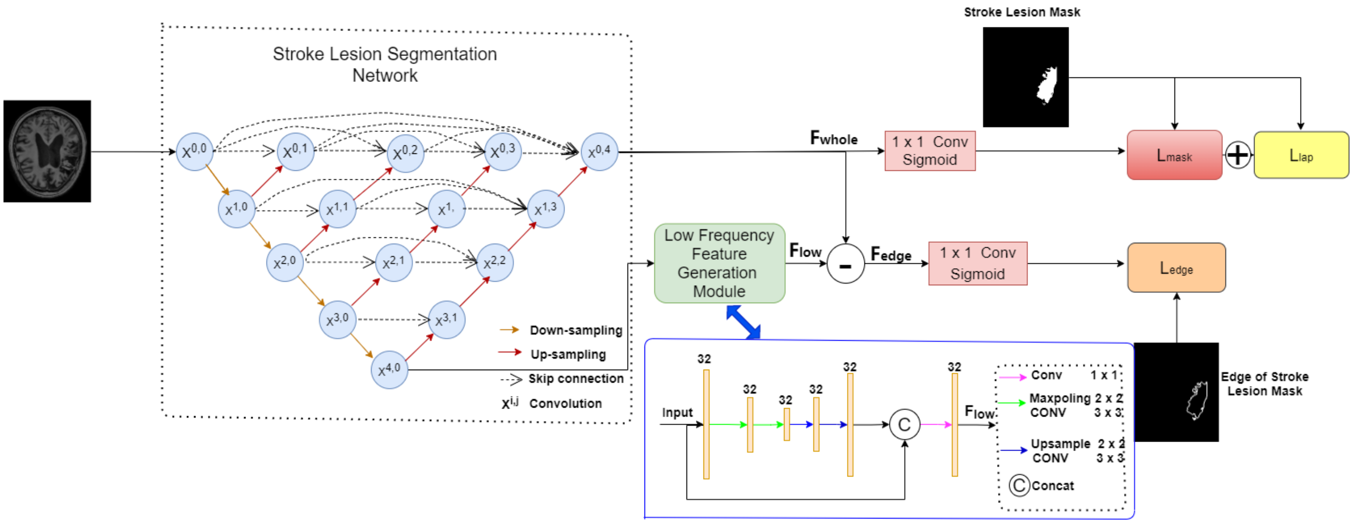

Fig. 1. Overview of the proposed multi-task framework. The stroke lesion segmentation network is exactly UNet++ [9]. In the low-frequency feature generation module, the input features are down-sampled sequentially through 2×2 max pooling which are then up-sampled by 2×2 linear interpolation to the same resolution as the input for concatenation. 32 is the number of channels in each feature map.

Table 1. Comparison of stroke lesion segmentation on private CT dataset.

Method	Dice	IoU
U-Net [3]	0.8464	0.8119
SegNet [4]	0.8338	0.8012
Attention U-Net [11]	0.8795	0.8420
DeepLab v3+ [12]	0.8712	0.8372
UNet++ [9]	0.8946	0.8568
TransUNet [19]	0.8693	0.8337
Swin-Unet [20]	0.8546	0.8214
Proposed	**0.9372**	**0.9049**

3. EXPERIMENTS

3.1. Dataset

In this paper, we train and validate our method on an open-source magnetic resonance imaging (MRI) dataset ATLAS [17] and our CT brain infarct segmentation dataset to evaluate the performance of the proposed method.

1) **Private CT Brain Infarct Segmentation Dataset**
In this dataset, 45 acute ischemic stroke patients with an onset time less than 6 hours were included. The infarcts were marked on the axial slices by experienced clinicians using the MITK Workbench software (https://www.mitk.org). The size of these non-enhanced CT images are $512 \times 512 \times (24 - 32)$ and the voxel spacing is $0.43 \times 0.43 \times 4.1mm^3$ to $0.43 \times 0.43 \times 4.8mm^3$. 40 patients were randomly picked for training and the rest 5 patients were used for testing.

2) **ATLAS Dataset [17]**
The dataset is collected from 11 cohorts and consists of 304 T1-weighted MRI scans, with the stroke lesions being labeled by trained students and research fellows. In the subsequent experiments, a subset of the dataset called MNI-152 is chosen according to [17], consisting of 229 T1-weighted normalized 3D MRI volumes and corresponding labels of stroke lesions. Each 3D volume has a dimension of $197 \times 233 \times 189mm^3$ and the voxel sizes are isotropic being 1mm in X, Y, and Z directions. The stroke lesion size ranges from $10mm^3$ to $2.8 \times 105mm^3$. Altogether there are 189×229 axial slices each having 233×197 voxels in the ATLAS dataset to be processed. In this paper, we randomly selected 120 subjects for training, 40 for validation, and 60 for testing. The stroke lesions could be embolic and hemorrhagic with slightly lower intensities and blurred boundaries.

3.2. Implementation Details and Evaluation Metrics

The proposed method is implemented in Keras and optimized with Adam [18]. The learning rate is initially set as 0.001. Other details are: batch size being 8, the maximum number of epochs being 100, and the loss being defined in formula 6. The model is trained using the NVIDIA TESLA P100 with the 16GB memory. Moreover, to evaluate the performance of relevant methods, two evaluation metrics are adopted to quantify the performance, namely, Dice score, and Intersection over Union (IoU).

3.3. Ablation Study

To show if the proposed multi-task learning, LFFGM and consistency loss can enhance stroke lesion segmentation, we have carried out comparative experiments. As shown in Table 2, auxiliary task implemented with respectively, a simple segmentation head, LFFGM, and LFFGM + consistency loss to predict edges of stroke lesions, could enhance both Dice and IoU in an ascending order, with the proposed LFFGM + consistency loss multi-task framework the most efficient to attain an increase of 2.76% for Dice and 2.75% for IoU. The performance of the stroke lesion segmentation of different multi-task learning strategies are summarized in Table 2.

Table 2. The segmentation performance on private CT dataset of different multi-task learning strategies.

Method	Dice	IoU
UNet++ [9]	0.8946	0.8568
UNet++ [9]+simple edge segmantation head	0.8996	0.8659
UNet++ [9]+LFFGM	0.9096	0.8774
UNet++ [9]+LFFGM +consistency loss (proposed)	**0.9372**	**0.9049**

Table 3. Comparison of stroke lesion segmentation on ATLAS dataset.

Method	Dice	IoU
U-Net [3]	0.5688	0.4416
Attention U-Net [11]	0.5404	0.4169
DeepLab v3+ [12]	0.5170	0.3888
X-Net [21]	0.5429	0.4218
CLCI-Net [22]	0.5826	0.4549
TransUNet [19]	0.5324	0.4053
Swin-Unet [20]	0.5620	0.4347
Proposed	**0.6290**	**0.5017**

3.4. Comparison with State-of-the-art Methods

In this sub-section, we are going to compare the proposed method with state-of-the-art methods. For this comparison, we have searched literature and chosen seven representative 2D methods to implement for the private dataset, including U-Net [3], SegNet [4], Attention U-Net [11], DeepLab V3+ [12], UNet++ [9], TransUNet [19] and Swin-Unet [20], where TransUNet [19] and Swin-Unet [20] are representative of the latest transformer methods. The performances of these methods are summarized in Table 1. For these seven 2D methods, their Dices and IoUs are all below 0.90, while the proposed method could yield a Dice of 0.9372 and IoU of 0.9049, an increase of at least 4.76% for Dice and 5.61% for IoU!

For the ATLAS [17] dataset, we also compare the proposed method with seven state-of-the-art methods, including U-Net [3], Attention U-Net [11], DeepLab V3+ [12], X-Net [21], CLCI-Net [22], TransUNet [19] and Swin-Unet [20] (Table 3), where X-Net [21] and CLCINet [22] are the two methods achieving state-of-the-art performance on ATLAS [17]. The proposed method could enhance these existing methods with an increase of Dice greater than 7.97% (from 0.5826 to 0.6290) and an increase in IoU greater than 10.29% (from 0.4549 to 0.5017). Typical segmentation results are shown in Fig.2 for the UNet++ [9] and the proposed method on the private dataset. Here are some comments regarding the rationale of the enhanced performance of the proposed method. Because the auxiliary task is the segmentation of edges of the stroke lesion, which is the complementary information of the main task (i.e., stroke lesion regions), the introduction of the auxiliary could add more supervisions of the UNet++ [9] due to the supervision of edges in the auxiliary branch on the one hand (Fig.1), and strengthen the shared features for edge and region segmentation of stroke lesions to learn better feature representations of the UNet++ [9] on the other hand. Specifically, it can be seen from Fig.2 that the proposed two task learning could segment both the regions and edges of stroke lesion much better than UNet++ [9].

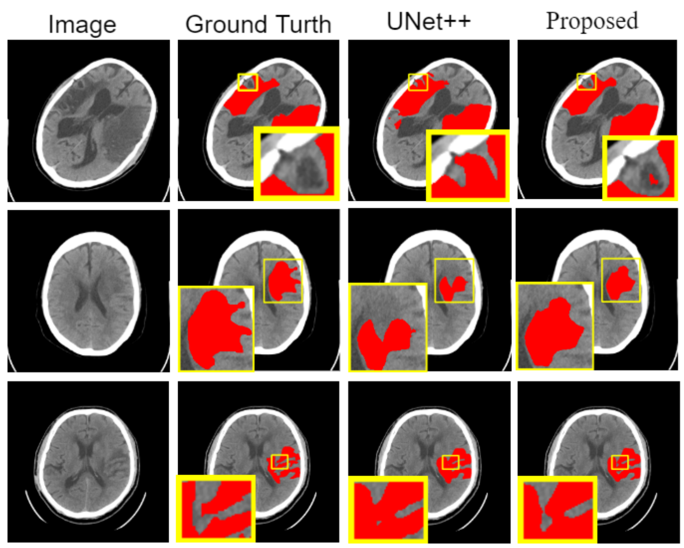

Fig. 2. Visual comparison of UNet++ [9] and the proposed method with corresponding original images and ground truths on the private dataset.

4. CONCLUSION

In this paper, a novel multi-task learning framework for stroke lesion segmentation has been proposed and validated. In the framework, a low-frequency feature generation module and a consistency loss are proposed to enhance the performance of the stroke lesion segmentation. Comprehensive experiments have been carried out on two stroke lesion segmentation datasets including private CT brain infarct segmentation dataset and ATLAS [17] dataset. The proposed method could enhance existing methods on the private CT dataset by at least 4.76% for Dice and 5.61% for IoU, and could enhance existing methods on the public ATLAS [17] dataset by at least 7.97% for Dice and 10.29% for IoU. Furthermore, the proposed method has exactly the same inference time as its single task counterpart (here the UNet++ [9]) as only the main task branch is needed at the inference stage.

5. ACKNOWLEDGEMENT

This work has been supported in part by the National Natural Foundation of China (No. 62001462), and Shenzhen Key Basic Research Project (No. JCYJ20200109114816594).

6. REFERENCES

[1] Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2016 update: a report from the American Heart Association. Circulation 2016; 133(4): e38-e360.

[2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015; 3431-3440.

[3] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 2015; 234-241.

[4] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017; 39(12): 2481-2495.

[5] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017; 40(4): 834-848.

[6] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.

[7] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:1412.7062, 2014.

[8] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2015; 37(9): 1904-1916.

[9] Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. UNet++: a nested UNet architecture for medical image segmentation. International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support 2018; 3-11.

[10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017.

[11] Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.

[12] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision 2018; 801-818.

[13] Takikawa T, Acuna D, Jampani V, et al. Gated-scnn: Gated shape cnns for semantic segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision 2019; 5229-5238.

[14] Li X, Li X, Zhang L, et al. Improving semantic segmentation via decoupled body and edge supervision. Proceedings of the European Conference on Computer Vision 2020; 435-452.

[15] Zhen M, Wang J, Zhou L, et al. Joint semantic segmentation and boundary detection using iterative pyramid contexts. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020; 13666-13675.

[16] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision; 565-571

[17] Liew S L, Anglin JM, Banks NW, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific data 2018; 5(1): 1-11.

[18] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[19] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

[20] Cao H, Wang Y, Chen J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537, 2021.

[21] Qi K, Yang H, Li C, et al. X-Net: brain stroke lesion segmentation based on depth-wise separable convolution and long-range dependencies. International Conference on Medical Image Computing and Computer-Assisted Intervention 2019; 247-255.

[22] Yang H, Huang W, Qi K, et al. CLCI-Net: cross-level fusion and context inference networks for lesion segmentation of chronic stroke. International Conference on Medical Image Computing and Computer-Assisted Intervention 2019; 266-274.