DA-RefineNet: Dual-inputs Attention RefineNet for Whole Slide Image Segmentation

Ziqiang Li[1], Rentuo Tao[1], Qianrun Wu[2], Bin Li[1]

[1]University of Science and Technology of China
[2]Hefei University of Technology

iceli@mail.ustc.edu.cn

Abstract

Automatic medical image segmentation has wide applications for disease diagnosing. However, it is much more challenging than natural optical image segmentation due to the high-resolution of medical images and the corresponding huge computation cost. The sliding window is a commonly used technique for whole slide image (WSI) segmentation, however, for these methods based on the sliding window, the main drawback is lacking global contextual information for supervision. In this paper, we propose a dual-inputs attention network (denoted as DA-RefineNet) for WSI segmentation, where both local fine-grained information and global coarse information can be efficiently utilized. Sufficient comparative experiments are conducted to evaluate the effectiveness of the proposed method, the results prove that the proposed method can achieve better performance on WSI segmentation compared to methods relying on single-input.

Motivation

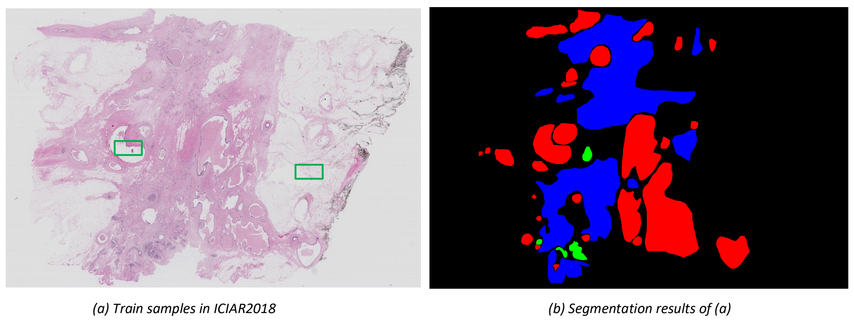

Figure 1: Demonstration of training medical images and its corresponding segmentation results. (a) training samples in ICIAR2018 dataset, where green rectangles denote sliced regions; (b) segmentation results of (a), color pixels represent different region types (normal, benign, in situ, invasive).

It's easy to find that sliced patches with similar texture sometimes were labeled differently due to the context discrepancy, which can be seen in the left image of Fig.1. Intuitively, this may caused by the discrepancy between surrounding regions of different slice images. For a specific region, more informative context information can be derived by simply increasing the size of slide window, however, it will also bring more computation costs. In order to balance slice image size and global context information, we propose a dual-inputs attention network named DA-Refinenet to combine fine texture features and coarse spatial features together, which allows a larger receptive field with little computation cost increasing.

Method: Dual-input Attention RefineNet

Framework

The architecture of proposed DA-RefineNet was demonstrated in Fig. **??**, where we can see the model mainly composed of three parts: two encoders (ENC_{slice}, ENC_{full}) for slice image and full image processing, and a refine decoder (DEC_{refine}) for producing segmentation results. Color arrows in green, blue, orange and gray represent down-sampling, skip-connection, up-sampling and feature fusion operations respectively.

Attn-Refine Block and Feature Fusion

Based on the intuition that global coarse images along with fine-grained local images features can help improve model performance, we adopt attention mechanism and proposed the attention refinement block(Attn-Refine). Detailed architecture of the proposed Attn-Refine block can be seen in the top part of Fig. 2, where color arrows represent different type of input features. Attn-Refine block contains three individual components: Attention block, RCU (residual convolution unit) and CRP (chained residual pooling). The proposed attention block is designed for feature fusion, which can be seen in the bottom part of Fig. 2. To make comparison, we keep RCU and CRP module the same as RefineNet [1].

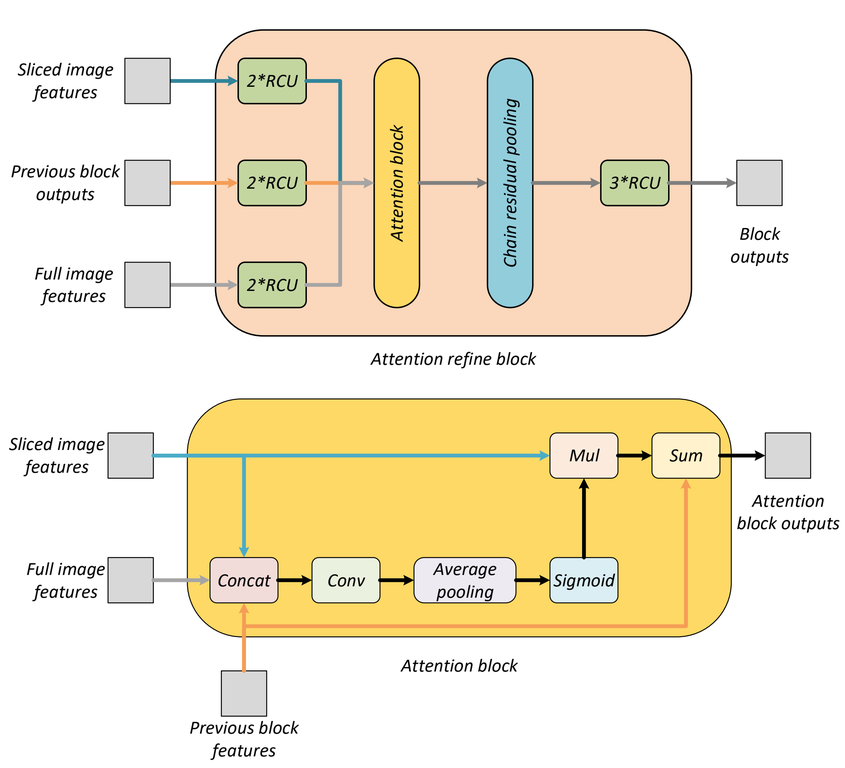

Figure 2: Attn-Refine Block: the top part denote the architecture of a attention refine block while the bottom part denote the attention block structure.

Experiments

Dual-inputs vs Single-input

Encoder	Decoder	Input	$MIoU$	IoU_0	IoU_1	IoU_2	IoU_3	$Accuracy$	$Score$	$Params$
U-net	U-net	Single-input	31.2	45.5	25.0	20.1	50.2	57.2	49.1	150M
Resnet-50	Add-Refine	Single-input	36.2	60.5	28.0	21.8	55.0	67.1	58.8	334M
		Multi-size dual-inputs	44.8	**60.9**	24.2	31.1	**63.1**	74.1	71.2	441 M
		Our dual-inputs	**45.9**	55.3	**32.9**	**37.1**	58.6	**75.1**	**71.1**	441M
Resnet-101	Add-Refine	Single-input	42.2	58.5	22.3	29.8	58.1	73.8	69.0	417M
		Multi-size dual-inputs	44.4	**59.1**	28.4	33.0	**62.9**	73.7	69.9	517M
		Our dual-inputs	**46.5**	58.0	**34.5**	**38.9**	61.4	**75.1**	**71.6**	517M
Resnet-152	Add-Refine	Single-input	39.7	59.0	27.7	28.3	52.5	72.7	69.7	480M
		Multi-size dual-inputs	44.7	**60.4**	25.3	31.2	**61.8**	74.9	71.1	580M
		Our dual-inputs	**46.8**	59.5	**38.1**	**28.8**	60.9	**75.5**	**71.5**	580M

Table 1: Quantitative Comparison of Dual-inputs vs Single-input

Feature Fusion Strategies

Dual-inputs Encoder	Fusion strategy	$MIoU$	IoU_0	IoU_1	IoU_2	IoU_3	$Accuracy$	$Score$
ResNet50_50	Concat	48.9	60.1	38.9	39.2	58.3	75.7	71.5
	Add	45.9	55.3	32.9	37.1	58.6	75.1	71.1
	Attention	**51.0**	**61.1**	**40.9**	**39.4**	**62.3**	**76.4**	**72.0**
ResNet101_50	Concat	47.8	58.3	36.6	42.0	60.0	74.4	71.9
	Add	46.5	58.0	34.5	38.9	**61.4**	**75.1**	71.6
	Attention	**49.9**	**59.3**	**36.6**	**44.0**	59.7	74.8	**72.1**

Table 2: Quantitative Comparison of Feature Fusion Strategies

Qualitative Evaluation Results

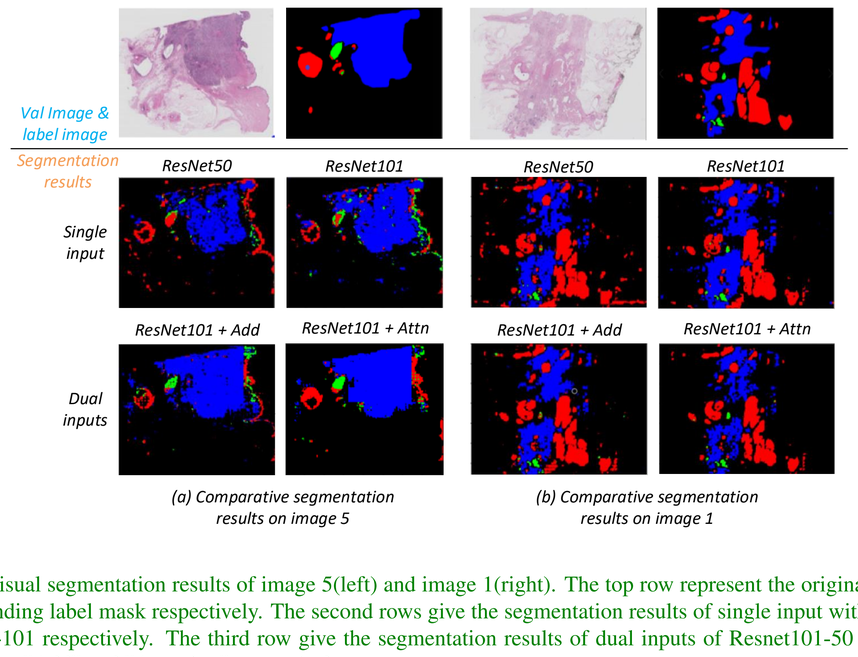

Figure 3: Visual segmentation results of image 5(left) and image 1(right). The top row represent the original image and its corresponding label mask respectively. The second rows give the segmentation results of single input with ResNet-50 and ResNet-101 respectively. The third row give the segmentation results of dual inputs of Resnet101-50 with feature fusion scheme "Add" and "Attention" respectively.

Conclusions

- We proposed a new attention based dual-inputs framework for whole slide image segmentation, which can incorporate global image information and obtain larger receptive fields. It proves that the proposed method can achieve better performance compared to methods that rely on single-input.
- We explore several feature fusion strategies and propose a simple but effective fusion strategy based on attention mechanism under the intuition that coarse global features can help reorganize fine-grained local features.

References

[1] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grand No.U19B2044 and No.61836011. We also want to thank the data provider organizer of the ICIAR2018 Grand Challenge.