Cancer Sensitive Cascaded Networks (CSC-Net) for Efficient Histopathology Whole Slide Image Segmentation

Shujiao Sun[1,2], Huining Yuan[1], Yushan Zheng[2,1*], Haopeng Zhang[1,2], and Zhiguo Jiang[1,2]

[1] Image Processing Center, School of Astronautics, Beihang University, Beijing, 100191, China

[2] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, 100191, China

Abstract. Automatic segmentation of histopathological whole slide images (WSIs) is challenging due to the high resolution and large scale. In this paper, we proposed a cascade strategy for fast segmentation of WSIs based on convolutional neural networks. Our segmentation framework consists of two U-Net structures which are trained with samples from different magnifications. Meanwhile, we designed a novel cancer sensitive loss (CSL), which is effective in improving the sensitivity of cancer segmentation of the first network and reducing the false positive rate of the second network. We conducted experiments on ACDC-LungHP dataset and compared our method with 2 state-of-the-art segmentation methods improved from U-Net. The experimental results have demonstrated that the proposed method can improve the segmentation accuracy and meanwhile reduce the amount of computation. The dice score coefficient and precision of lung cancer segmentation are 0.694 and 0.947, respectively, which are superior to the compared methods.

Keywords: image segmentation, digital pathology, whole slide image analysis, computer-aided-diagnosis, cascaded convolutional neural network

1 Introduction

The manual analysis of histopathological whole slide images (WSIs) is a time-consuming task for pathologists and often suffers from errors and intra-observer variability because of the diversity of cancerous organization [4]. Currently, the number of pathologists cannot meet the requirement of cancer diagnosis, especially in remote regions of developing countries [3]. Therefore, it is significant to develop automatic analysis of histopathological WSIs based on artificial intelligence. Owing to the accessibility of large amounts of WSIs, the computer-aided diagnosis methods based on histopathological WSIs have become popular [2, **?**].

Tumor region segmentation is a popular topic in the domain of histopathological WSIs analysis. Generally, the results are desired to be a segmentation

map which indicates the precise location of cancerous regions. To resolve the segmentation tasks, some classical segmentation networks, such as, fully convolutional networks (FCN) [7], U-Net [10], SegNet [12], were developed. However, all the types of objects in a WSI are equally regarded in both the training and predicting stages of these networks. The challenging area related to cancer diagnosis are insufficiently concerned, which limited the performance of segmentation. Moreover, the widespread background and negative regions that are easy to segment consumes redundant computation. In recent years, zoom-in-net [13], cascade-net [14] and scan-net [6] have been proposed to reduce the computation amount raised by large scale images. Besides, a novel attention gate [9] was applied to the common U-Net to highlight the useful features for various tasks and reduce computation while increasing the segmentation accuracy. Nevertheless, they did not design specific mechanism for the segmentation of challenging cancerous regions in the histopathological WSIs.

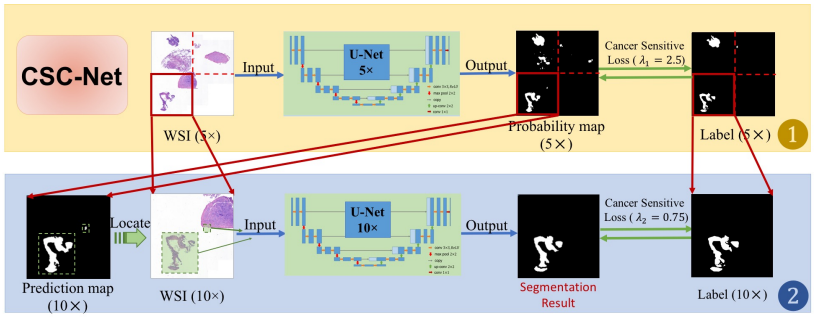

Fig. 1. The Cancer Sensitive Cascaded Networks proposed in this paper, where the architecture consists of two stages based on two U-Nets. Stage 1 is trained with samples from a low magnification to obtain the segmentation outline. Stage 2 is trained with a higher magnification to refine the segmentation results according to the probability map output by the first stage.

In this paper, we propose a novel Cancer Sensitive Cascaded Network (CSC-Net) for the segmentation of histopathological WSIs. The CSC-Net consists of two U-Net structures [10]. The first network is used to comprehensively segment cancerous regions and filter blank background and negative area to reduce computation for the next high-resolution network. The second network specializes in further refining cancerous regions segmented in the first stage and meanwhile filtering the false positive regions that are mis-segmented. Aiming at our CSC-Net, we designed a specific Cancer Sensitive Loss (CSL) to guarantee a high recall of cancerous regions in the first stage and ensure the accuracy of segmentation in the second stage. We conducted experiments on the ACDC-LungHP dataset and the results have demonstrated that our method performed better than the

existing methods in regards of accuracy and running time. The contribution in this paper includes 1) a novel cascade framework for cancerous regions segmentation from histopathological WSIs and 2) a novel cancer sensitive loss function for the cascaded framework.

2 Methodology

The pipeline of the proposed CSC-Net is shown in Fig. 1. The structure of cascaded networks and cancer sensitive loss functions are two essential components in our method, which are detailed in this section.

2.1 Cascaded Networks

A WSI contains a number of cancer-free regions that can be easily recognized by automatic analysis algorithms even at low magnifications. Therefore, it is unnecessary to segment the entire WSI at high magnifications where the amount of computation is times higher than that at lower magnifications. While, the tiny and boundary cancerous regions (as shown in Fig. 2) are challenging to segment. And they need to be processed in a higher resolution to ensure an accurate segmentation result.

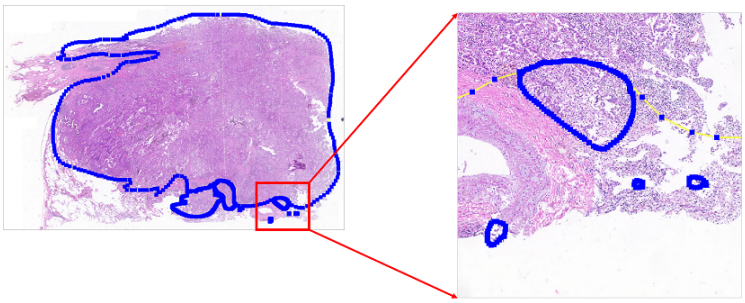

Fig. 2. A WSI in our dataset, where the regions encircled with blue curves are cancer tissues and the zoom-in patch shows the tiny cancerous spots and borders that are difficult to segment.

The proposed cascaded framework aims at the challenge of histopathological WSIs segmentation including accuracy and computational efficiency. The cascaded networks consist of two stages. For the first stage (as illustrated in the upper part of Fig. 1), we train the U-Net with samples at low magnification. It is used to filter the pixels which are easy to recognize (such as most of the negative area) and obtain the glancing regions of probable cancerous tissues. Then, we train the second network with the probable cancerous tissues on higher magnification to optimize the cancerous regions. The regions fed to the second network

are sampled based on the probability map output by the first network. Specifically, a sliding-window in size of $n \times n$ is applied to the probability map and the probable regions are determined by a threshold t referring to the equation

$$\frac{1}{n^2}\sum_{k=1}^{n^2} p_k > t, \tag{1}$$

where p_k is the cancerous probability of the k-th pixel that output by the first network. As the samples to train the second network do not involve easy-recognized cancer-free patches, the trained network is potential to focus on distinguishing the challenging regions and thus is able to improve the segmentation performance.

When predicting, the WSI is first fed to the low-resolution network. Then, the patches filtered by Eq. 1 are extracted at high magnification and fed to the high-resolution network to refine the segmentation results. Compared with the segmentation methods that directly process WSIs at a high magnification, the cascade strategy proposed in this paper can improve the segmentation accuracy and meanwhile reduce the amount of computation.

2.2 Cancer Sensitive Loss

In the field of medical images segmentation, Dice score coefficient (DSC) [11] is a most frequently used metric to evaluate the segmentation results. DSC is used to calculate the overlapping rate between prediction region and ground truth. The 2-class DSC formula adopted in this paper is illustrated in Eq. 2:

$$DSC_c = \frac{2\sum_{i=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} + \sum_{i=1}^{N} g_{ic} + \epsilon}, \tag{2}$$

where $p_{ic} \in [0, 1]$ is the predicted probability of the i-th pixel to the c-th class and g_{ic} is the label with $g_{ic} = 1$ representing cancerous pixel and $g_{ic} = 0$, otherwise. N is the total number of pixels in WSIs and ϵ is used to protect the division operation.

Dice Loss (DL) [8] is a widely used loss function in medical image segmentation, which is based on DSC and defined as

$$DL = \sum_{c} \left(1 - DSC_c\right). \tag{3}$$

The limitation of the DL is that false positive (FP) and false negative (FN) get equal attentions. As for our cascaded networks, the requirements of two stages are different. The first stage is expected to be sensitive to the cancerous area and achieve a high recall and the second is supposed to get more accurate segmentation results. Aiming at the motivation of the cascaded network, we proposed a novel Cancer Sensitive Loss (CSL) function based on DL. The CSL is defined as

$$CSL = \sum_{c} \left(1 - DSC_c\right)^{\lambda}, \tag{4}$$

where $\lambda \in (1, \infty)$ is a coefficient to balance the penalty on FP and FN. Supposing a round ground truth with an area of 1, we tune the area of a predicted foreground region that has its center within the ground truth and observe the change of CSL to the area of the region. The curves as functions of the foreground regions are illustrated in Fig. 3, where a line indicating a constant loss (e.g. 0.2) with black arrows is used to point out the effect of λ to the original DL. Obviously, the λ with a value above 1 can loose penalty of the predictions that are already correct to the ground truth and meanwhile assigns a lower loss for the segmentation with large false positive pixels. Specifically for histopathological image segmentation, the CSL encourages the network has a high sensitivity for cancerous region segmentation. Therefore, we train the the first network with the proposed CSL in our cascaded structure to ensure the cancerous regions can successfully go through the first network and will be further considered at higher magnification by the second network. We tuned the λ to assess the CSL in the training of the network and found that $\lambda = 2.5$ is the most appropriate to the first cascaded network. As for the second stage, $\lambda \in (0, 1)$ become our consideration because the loss value is bigger than dice loss when the segmentation result is not accurate whether with high FP or FN (as the vertical line shown in Fig. 3).

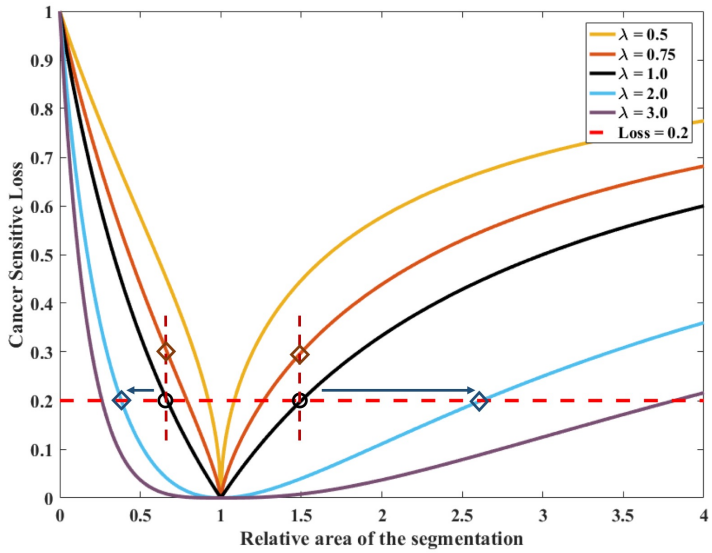

Fig. 3. The curves of the proposed cancer sensitive loss as functions of the cancerous area segmented by the networks, where the area of the ground truth is supposed as 1. The horizontal indicates $loss = 0.2$.

Table 1. Quantitative comparison between the CSC-Net and other existing methods.

Model	DSC	Recall	Precision	Time (min)
Attention U-Net (10×) [9]	0.667	0.699	0.945	50.17
Multi-scale-input Attention U-Net (10×) [1]	0.635	0.647	0.941	70
U-Net + Cross Entropy (5×)	0.487	0.434	0.923	11
U-Net + Cross Entropy (10×)	0.615	0.592	0.939	57.83
U-Net + Cross Entropy (cascade)	0.610	0.574	0.939	28.17
U-Net + DL (5×)	0.617	0.640	0.937	11.67
U-Net + DL (10×)	0.675	0.706	0.946	56.5
U-Net + DL (cascade)	0.674	0.710	0.946	28.5
U-Net + FTL (5×)	0.628	0.656	0.939	11.33
U-Net + FTL (10×)	0.664	0.772	0.943	57.33
U-Net + FTL (cascade)	0.664	0.764	0.944	27.67
U-Net + CSL (5×)	0.674	**0.801**	0.944	11.5
U-Net + CSL (10×)	0.690	0.788	0.945	57.33
U-Net + CSL (cascade)	**0.694**	0.792	**0.947**	**27.33**

3 Experiments

We verified our CSC-Net on ACDC-LungHP dataset [5] which consists of 150 lung cancer WSIs in high resolution up to $170,000 \times 80,000$ pixels (by 40× object lens). This dataset includes WSIs with different cancerous regions proportion. We randomly chose 100 WSIs with a 80-20 train-validation split and divided them into smaller patches with size of 512×512 to train the first network. For the second network, the patch size was set to 1024×1024 to ensure a patch contains the same amount of information as that in the first stage. The remainder 50 WSIs were used to validate our network by sliding-window setup.

To demonstrate the effectiveness of our cascade strategy, we adopted U-Net [10] as the basic network to construct our cascaded networks. The U-Net in each stage includes 4 pooling layers, 4 up-sampling layers as shown in Fig. 1. We first evaluated our cascade strategy on existing loss functions, including cross entropy, dice loss and focal Tversky loss. The thresholds in the Eq. 1 are set to $t = 0.05$ according to the best performance of validation data. Finally, we compared the CSC-Net with two improved U-Net structures, attention U-Net [9] and multi-scale-input attention U-Net [1].

All the experiments were conducted on a computer with an Intel Core i7-7700k CPU of 4.2 GHz. To improve the training speed, the parallel training model was adopted in the U-Net structure using two GPUs of Nvidia GTX 1080Ti.

4 Results and Discussions

Table 1 quantitatively shows the performance of the proposed CSC-Net and several comparision methods. Dice score coefficient (DSC), recall, precision and the

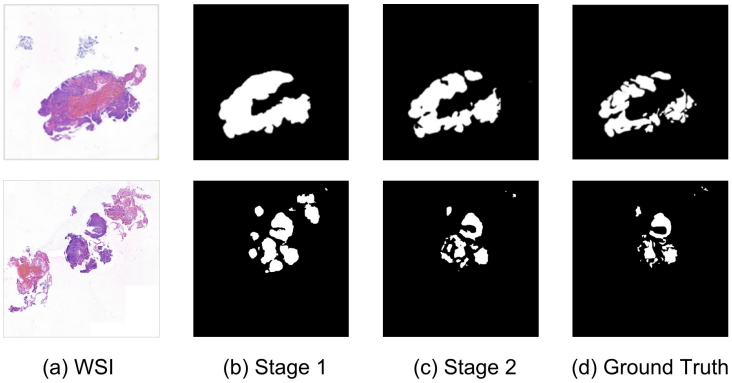

Fig. 4. Visualization of the segmentation results at different stages, where (a) is the original WSI, (b) is the intermediate results output by the low-resolution network, (c) is the final segmentation results output by the high-resolution network and (d) is the ground truth.

running time were used to evaluate our method. The running time is the total time of processing the whole 50 testing WSIs. And the time of cascaded networks consists of the time consumed at the two stages. Overall, our CSC-Net performed the best with a DSC 0.694 and a precision 0.947. Specifically, for the DSC, our method outperformed the attention U-Net [9] by 4.05% and the multi-scale-input attention U-net [1] by 9.29%. The two compared methods were designed to segment WSIs in a single high magnification and did not focus on the challenging or easy-recognized regions related to cancer diagnosis. In contrast, our framework filtered the easy-recognized regions through the low-magnification network and make the high-magnification network focus on distinguishing cancer regions and hard negative regions, and therefore achieved better performance for cancer segmentation. Benefited by the cascaded segmentation structure, our framework consumed about only half of the running time compared to the methods using single high magnification of WSIs.

Table. 1 also presents the comparisons for different loss functions. It is obvious that the proposed CSL function achieved the best performance. The method U-Net-CSL achieved a recall of 0.801, which was much higher than those based on other loss functions and trained under 5× lenses. The high recall has ensured more cancerous candidates could pass the first network and thereby delivered a better segmentation result. Furthermore, The results within each common loss function also demonstrated the effectiveness of the proposed cascaded strategy.

Several representational WSIs segmentation results is displayed in Fig. 4. It shows that our CSC-Net segmented the outline of the cancerous regions in the first stage and refined the detail within the outline, the qualitative result is consistent with the quantitative analysis.

5 Conclusion

In this paper, we proposed a cancer sensitive cascaded network (CSC-Net), including the cascade strategy and a cancer sensitive loss (CSL) function. The cascade strategy is effective in improving the accuracy of segmentation result while reducing about half of the running time consumed by common methods. The cancer sensitive loss function designed for our cascaded structure can improve the sensitivity of the first network and finally deliver a more precise segmentation result. The experimental results indicted that the cascaded network trained with CSL outperformed the state-of-the-art methods. The future work will focus on designing an end-to-end cascaded structure to further improve accuracy and efficiency of our algorithm.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61771031, 61901018, 61471016, and 61501009), China Postdoctoral Science Foundation (No. 2019M650446) and Motic-BUAA Image Technology Research Center. Asterisk indicates the corresponding author. E-mail: yszheng@buaa.edu.cn

References

1. Abraham, N., Khan, N.M.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 683–687. IEEE (2019)
2. Gurcan, M.N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Histopathological image analysis: A review. IEEE reviews in biomedical engineering **2**, 147 (2009)
3. Kelleher, M.J.: The professional ideology of social pathologists transformed: The new political orthodoxy in sociology. The American Sociologist **32**(4), 70–88 (2001)
4. Kumar, V., Abbas, A., Fausto, N., Aster, J.: Robbins and Cotran Pathologic Basis of Disease, Professional Edition E-Book. Robbins Pathology, Elsevier Health Sciences (2014), https://books.google.co.jp/books?id=jJllBAAAQBAJ
5. Li, Z., Hu, Z., Xu, J., Tan, T., Chen, H., Duan, Z., Liu, P., Tang, J., Cai, G., Ouyang, Q., Tang, Y., Litjens, G.J.S., Li, Q.: Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study. CoRR **abs/1803.05471** (2018)
6. Lin, H., Chen, H., Graham, S., Dou, Q., Rajpoot, N., Heng, P.A.: Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. IEEE transactions on medical imaging **38**(8), 1948–1958 (2019)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
9. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)

10. 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.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
11. SORENSEN, T.A.: A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. Biol. Skar. **5**, 1–34 (1948), https://ci.nii.ac.jp/naid/10008878962/en/
12. V.Badrinarayanan, A.Kendall, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (Dec 2017). https://doi.org/10.1109/TPAMI.2016.2644615
13. Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H., Wang, X.: Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 267–275. Springer (2017)
14. Yang, Q., Wu, K., Cheng, H., Gu, C., Liu, Y., Casey, S.P., Guan, X.: Cervical nuclei segmentation in whole slide histopathology images using convolution neural network. In: International Conference on Soft Computing in Data Science. pp. 99–109. Springer (2018)