A Comparative Study of CNN and FCN for Histopathology Whole Slide Image Analysis

Shujiao Sun[1,2], Bonan Jiang[3], Yushan Zheng[1,2(✉)], and Fengying Xie[1,2]

[1] Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
sunshujiao@buaa.edu.cn
[2] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China
yszheng@buaa.edu.cn
[3] Beijing-Doblin International College, Beijing University of Technology, Beijing 100124, China

Abstract. Automatic analysis of histopathological whole slide images (WSIs) is a challenging task. In this paper, we designed two deep learning structures based on a fully convolutional network (FCN) and a convolutional neural network (CNN), to achieve the segmentation of carcinoma regions from WSIs. FCN is developed for segmentation problems and CNN focuses on classification. We designed experiments to compare the performances of the two methods. The results demonstrated that CNN performs as well as FCN when applied to WSIs in high resolution. Furthermore, to leverage the advantages of CNN and FCN, we integrate the two methods to obtain a complete framework for lung cancer segmentation. The proposed methods were evaluated on the ACDC-LungHP dataset. The final dice coefficient for cancerous region segmentation is 0.770.

Keywords: Image segmentation · Computational pathology · CNN · FCN · Lung cancer

1 Introduction

Digital pathology has been gradually introduced in clinical practice. However, the manual analysis of whole-slide images (WSIs) is a time-consuming task for pathologists and prone to errors or intra-observer variability. The limited knowledge of pathologists also influences the veracity of diagnosis. As such, automatic analysis of WSIs seems to be particular importance at the background of high incidence of cancer. To relieve the dilemma we are facing [1], a large number of researchers all around the world focus on studying algorithms for the automatic analysis of WSIs.

Y. Zhao et al. (Eds.): ICIG 2019, LNCS 11902, pp. 558–567, 2019.
https://doi.org/10.1007/978-3-030-34110-7_47

In recent years, an increasing number of automatic analysis methods for WSIs have been developed based on machine learning algorithms. Cancerous region segmentation is a popular application among the existing methods based on deep learning networks. However, it's difficult to process a WSI directly using recent deep learning methods because of its high pixel resolution. Therefore, the first step is to divide the WSI into small patches and segment the cancerous regions patch by patch. Some segmentation methods for nature images are usually applied to WSIs analysis. For segmentation, one of the classic networks is Fully convolutional network (FCN) [8]. FCN is to segment images based on pixel-level, the input and output are both 2-D images with the same size. The segmentation results of the region of interest can be directly obtained. There are also many segmentation methods: SegNet [12], CRFs [14], DeepLab [4] and some aimed at WSIs, for instance, U-Net proposed in [9] adopt overlap-tile strategy for seamless segmentation of arbitrary large images and Yang *et al.* [13] combined U-Net and multi-task cascade network to segment WSIs and achieved better accuracy.

Besides FCN, convolutional neural networks (CNNs) including DenseNet [3], ResNet [2], GooLeNet [11] and graph CNN for survival analysis [5] are also widely applied to the segmentation of WSIs. But the WSI should be firstly divided into small patches based on sliding window strategy because of the high resolution. Then, the patches are fed to the network and the labels of the patches are predicted. The segmentation results will be generated from the up-sampled probability maps similar to the result of FCN. This method is not sensitive to boundaries, while, for practical application, the segmentation result is sufficient to analyze WSI for pathologists. In addition, the small blank regions among the tissue would not be excessively segmented and pathologists can get a more integrated results.

In this paper, we conducted a series of experiments to segment cancerous regions from WSIs utilizing FCN and CNN frameworks and verified the feasibility and effectiveness of the two methods (FCN and CNN). Then, we compared the results from the two methods and found that they equally performed. All of our experiments are completed on a public lung histopathology dataset of ACDC-LungHP challenge [6].

Our work addresses the segmentation of lung carcinoma from WSIs. The main contributions of this paper can be summarized as:

(1) We designed two complete strategies for lung carcinoma segmentation based on CNN and FCN, respectively.
(2) We evaluated the accuracy and running time of CNN-based and FCN-based strategies and compared the segmentation performance of the two strategies.
(3) To leverage the advantages of both CNN and FCN, we proposed an integrated flowchart for lung carcinoma segmentation. Our method was evaluated on the ACDC-LungHP challenge and achieved the 4th rank with a dice coefficient 0.770.

The details of our methods and their results comparison are introduced in the following sections.

2 Method

The flowchart of our methods is illustrated in Fig. 1. To analyze histopathology images, pixel-wise segmentation based on FCN will be designed to segment lung carcinoma. Because histopathology WSIs are in high resolution, CNN for patch-level classification can accomplish the segmentation task when the patches are small enough relative to the whole slide. The ensemble of two networks is to leverage both advantages and further improve the segmentation performance.

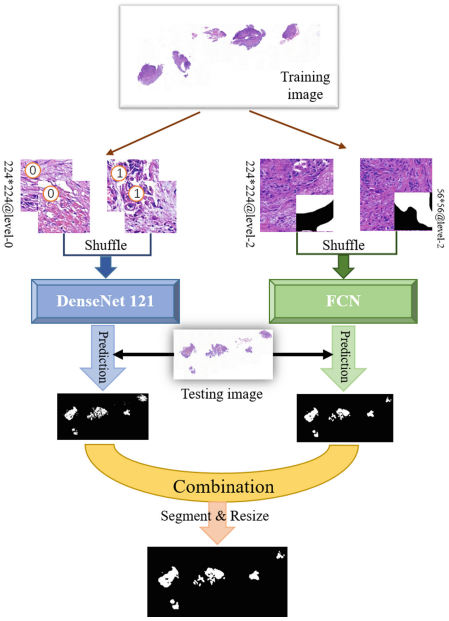

Fig. 1. The flowchart of our methods. It contains 3 sections: CNN, FCN and ensemble.

2.1 Segmentation Based on FCN

FCN is a conventional method for image segmentation. FCN framework can be used to process WSIs the same as nature images by dividing the whole WSI into smaller patches with size of 224×224. The FCN structure was designed based on DenseNet-1 structure [3], as shown in Fig. 2. To limit the computation, the last dense block of DenseNet structure was removed and two transposed convolution layers were connected instead, up-sampling the feature map to 56×56. Namely, the side length of output is $1/4$ to that of the input. Then, dice loss [10] and focal loss [7] were applied to train the networks with label masks. when predicting, the testing WSI was processed by the trained FCN with a sliding window of $1280 \times$

1280 pixels. To relieve the effect of the window border, the window region was padded to 1536×1536 pixels before feeding into the FCN. In corresponding, the output was cropped to remove the padded regions. After prediction, a probability map in high resolution was generated. Then, the segmentation was completed by the threshold on the probability maps.

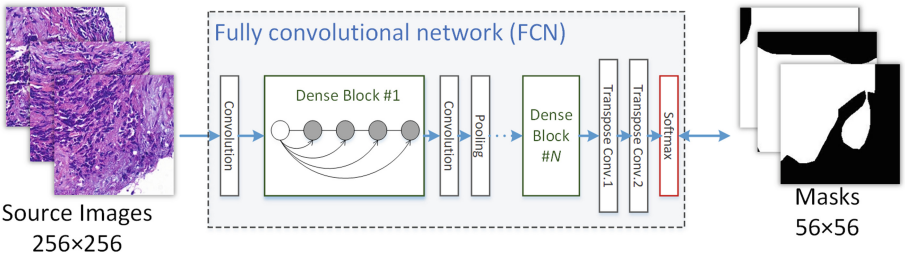

Fig. 2. The structure of FCN based on DenseNet.

2.2 Segmentation Based on CNN

CNNs have been proven effective in image classification and have been successfully introduced into histopathological image analysis because of the high resolution of WSIs. Specifically, to increase comparability, we also employed DenseNet-121 structure [3] with 2 output neurons (as shown in Fig. 3) as the classifier of cancer patches. The network was trained from randomly initial parameters. To relieve overfitting, color noise was randomly added to the patches. As for prediction, the testing WSIs were divided into patches (the same size with the CNN input) following the sliding window paradigm with sliding step set 112 (half of the patch side length) and fed into DenseNet structure. For each window, the output of the positive neural node was recorded and regarded as the probability of cancer. Thereby, a probability mat that indicates the location of cancerous patches was obtained after the sliding window paradigm. Then the mat was up-sampled to fit the original size of WSI. Finally, a threshold was selected to generate the mask for the WSI.

2.3 Ensemble

Finally, we have tried to assemble the probability maps obtained by the two frameworks. Specifically, the maps were averaged and then segmented by a threshold. The CNN can classify a patch into one single category according to the threshold. The small blank regions among the tissue would not be excessively segmented. Thereby, the cancerous regions segmented by CNN are more integrated than those obtained by FCN structure. On the contrary, the FCN can

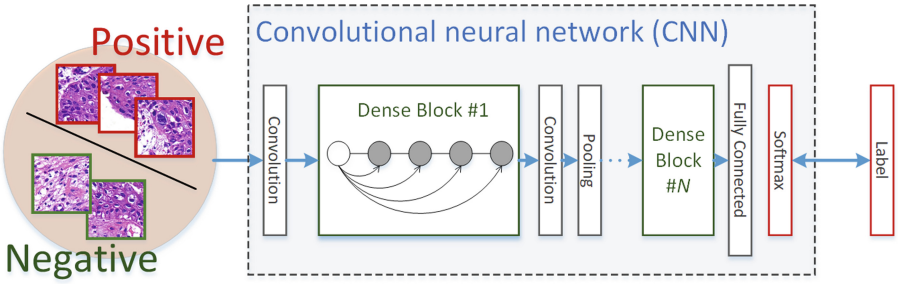

Fig. 3. The structure of CNN based on DenseNet-121.

generate elaborate borders of tissue, since it is designed for pixel-level segmentation. To leverage the both advantages, we fuse the two results and aim at a better accuracy. A diagram as shown in Fig. 4 indicates the process of generating segmentation results.

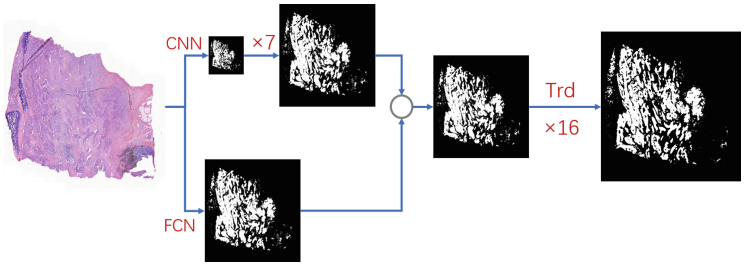

Fig. 4. The flowchart of generating masks from probability maps. The probability matrix obtained from CNN is up-sampled 7 times, to the same size with FCN and averaged with the probability matrix of FCN. A threshold ("Trd" in the figure) is applied to segment the averaged probability matrix. Finally, the segmentation result is resized 16 times to generate the pixel-wise segmentation result of the WSI.

3 Experiments

3.1 Experimental Setting

The proposed method was implemented in python. The experiments that involves CNNs were conducted based on the mxnet framework and the experiments for FCNs were on the tensorflow platform. The training patches along with the labels were transformed to the formats the platforms required.

All the experiments were conducted on a computer with an Intel Core i7-7700k CPU of 4.2 GHz and a GPU of Nvidia GTX 1080Ti.

3.2 Data Preparation

The data used in the experiments are ACDC-LungHP dataset. It concludes a mass of lung cancer biopsy samples stained with hematoxylin and eosin (H&E). To train the neural networks, we designed a flowchart (as shown in Fig. 5) to generate training samples from the WSIs. 150 WSIs with annotations are used to train our networks, among them, 80% are used as training set and the remaining are validation set.

At first, a bounding box was manually annotated to locate the tissue regions for each WSI. To reduce the computation, a threshold was applied to coarsely filter the blank areas (pure white and black pixels). Then, square patches in size of 224×224 pixels for CNN and 256×256 pixels for FCN were randomly sampled from the tissue regions to establish the training datasets. To balance the samples from each WSI, the patches from WSIs with small tissue regions were augmented through randomly flipping & rotating. Correspondingly, the patches from large WSIs were randomly reduced. Overall, about 2000 positive (contain more than 50% cancerous pixels referring to the annotation) and 2000 negative (less than 10% cancerous pixels) patches were generated from each WSI. For the training of FCNs, the mask of cancerous pixels for each patch was simultaneously cropped and used as the ground truth. All the patches and the corresponding labels and masks were shuffled to ensure each batch could contain the general allocation of the WSI data.

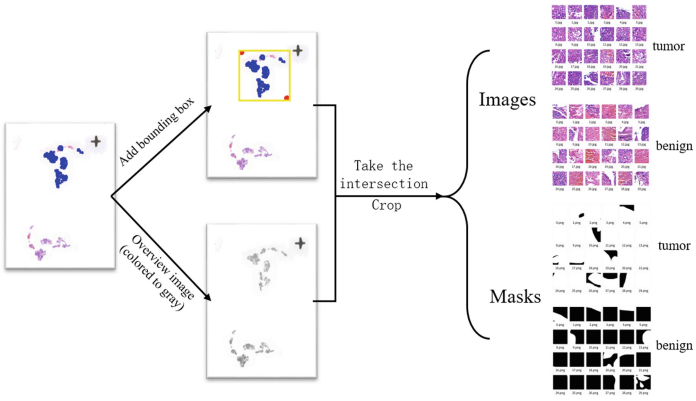

Fig. 5. The process of generating datasets for CNN and FCN.

3.3 Training

The CNN structure used in our experiment is the DenseNet-121 suggested in [3]. The training patches are sampled from WSIs under the resolution of level 0 (defined in ASAP, the resolution is 40× lens) and randomly flipped, rotated and

scaled for augmentation. The cross-entropy with softmax output is used as the loss function. The SGD with momentum is applied as the optimizer.

The FCN consists of three dense blocks where the first block has 5 convolutional layers and the other two have 8 convolutional layers. Before the first dense block, two convolutional layers with 3×3 kernels are applied on the input images and each side of the input tensors is zero-padded by one pixel to keep the feature-map size fixed. Following the dense blocks, two transposed convolutional layers are used to upsample the feature maps to the same size of labels. Focal loss and dice loss were considered in the training stage. We conducted the experiment on WSIs under the resolution of level 2 (the pixel resolution is $1\times$ lens) and tried three kinds of loss combinations: focal loss only, dice loss only and both. Then, the loss type that achieves the best result is chosen for subsequent experiments.

3.4 Results and Discussions

Hyper-Parameter Setting. A number of experiments were conducted on the training and validation sets to determine the settings of our proposed approach. The CNN and FCN frameworks were conducted on resolution of level 0, level 1 and level 2 respectively. The results indicated that, as for CNN, the performance of level 0 was the best and level 2 for FCN achieved best accuracy. Besides, learning rate, batch size and growth rate were determined according to the performance of the validation set.

Loss Functions. One of the important factors is loss function. As for segmentation, focal loss [7] and dice loss [10] are the most commonly used loss functions. Focal loss is aimed at resolving the problem that the proportion of positive and negative samples is seriously out of balance when addressing two-value segmentation and bipartition. Dice loss pays more attention on the object needed be segmented and is mainly used for biomedical images segmentation.

Several experiments were designed, including focal loss [7] employed, only dice loss [10] and combination of focal loss and dice loss. After the prediction mentioned in Sects. 2.1 and 2.2, the results indicated that for FCN, the combination of both loss functions was more appropriate, for CNN, focal loss performed not better than cross-entropy (commonly used for classification).

Segmentation Accuracy. The dice coefficient and running time for different settings of FCN and CNN are presented in Table 1. The FCN achieved a dice score of 0.7525 and the CNN achieved a comparative score of 0.7528. It indicates both the two strategies are adequate for histopathological whole slide image analysis. Actually, the patch-based CNNs can generate a probability map of hundreds by hundreds pixels from high-resolution WSI (Level 0), which is sufficient to help pathologists recognize diagnostically relevant regions from the WSI. The running time of FCN structure is much shorter than CNN. The main reason is that the FCN structure uses a resolution that is much lower than that

of CNN structure. Another reason is that the CNN structure utilizes overlapping patches in the analysis, which has further increased the computation. Furthermore, to exploit the advantages of the two frameworks, we assembled the CNN and FCN structures that achieved the best results separately (using the approach provided in Sect. 2.3). Consequently, the dice coefficient reached to 0.770. But, at the same time, it needs more time including the CNN's and FCN's. The result ranked No.4 in the ACDC-LungHP challenge. The leaderboard is listed in Table 2.

By analysis of each WSI result, several challenging WSIs for our method are displayed in Fig. 6, which are needed to be further improved. For visualization, the segmentation results obtained by our framework can be converted to free curves, which is able to be reloaded in ASAP tools. An instance of the visualization is presented in Fig. 7.

Table 1. The dice coefficients for different setting of the our methods.

Method	Image levels	Loss type	Dice coefficient	Time
CNN	Level-0	Cross entropy loss	**0.7528**	328.8 s
FCN	Level-2	Focal loss	0.7184	7.2 s
		Dice loss	0.7213	
		Focal+Dice loss	**0.7525**	
Combination	Level-0	Cross entropy loss	0.7700	336.0 s
	Level-2	Focal+Dice loss		

Table 2. The leaderboard of ACDC-LungHP challenge 2019.

Rank	Group name	Score (Dice mean)
1	PINGAN Technology	0.8373
2	Lunit Inc	0.8297
3	Turbolag	0.7968
4	**Ours**	**0.7700**
5	BUAA	0.7659
6	Arontier	0.7638
7	Frederick National Laboratory for Cancer Research	0.7552
-	**Ours(CNN)**	**0.7528**
-	**Ours(FCN)**	**0.7525**
8	National Taiwan University of Science and Technology	0.7510
9	Skychain	0.7456
10	University of Maryland	0.7394

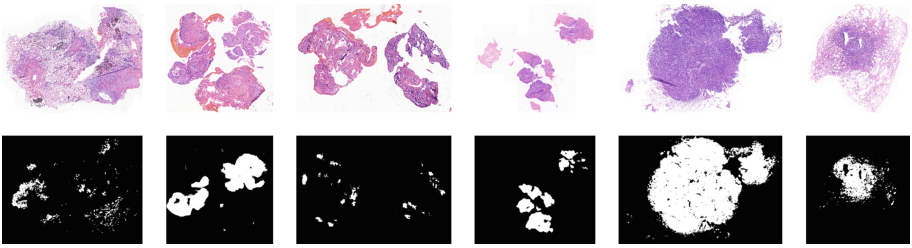

Fig. 6. The challenging WSIs and corresponding segmentation masks.

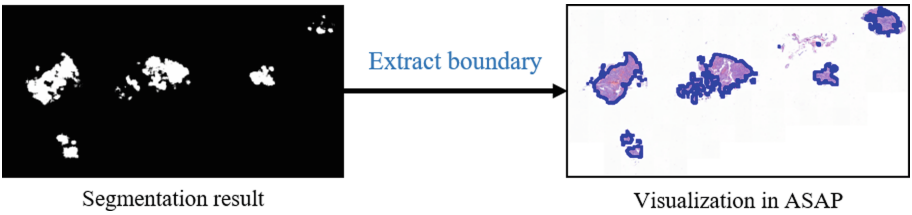

Fig. 7. The flowchart of generating masks from probability maps.

4 Conclusion

According to the results, CNN and FCN both achieved satisfactory performance for histopathological whole slide image analysis. Furthermore, the frameworks based on CNN and FCN achieved comparable segmentation performance. It indicates that the segmentation via patch-wise classification on a high resolution could be equivalent to the segmentation by an FCN under lower resolutions. After a combination of the CNN and FCN results, the metric was further improved. It demonstrates that the information from high and low magnification of WSIs are complementary.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 61771031, 61371134, 61471016, and 61501009), China Postdoctoral Science Foundation (No. 2019M650446) and Motic-BUAA Image Technology Research Center.

References

1. Bejnordi, B.E., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**(22), 2199–2210 (2017)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016

3. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
4. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018). https://doi.org/10.1109/TPAMI.2017.2699184
5. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_20
6. Li, Z., et al.: Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study. CoRR abs/1803.05471 (2018)
7. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: The IEEE International Conference on Computer Vision (ICCV), October 2017
8. 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
9. 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
10. 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/
11. Szegedy, C., et al.: Going deeper with convolutions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015
12. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
13. Yang, Q., et al.: Cervical nuclei segmentation in whole slide histopathology images using convolution neural network. In: Yap, B.W., Mohamed, A.H., Berry, M.W. (eds.) SCDS 2018. CCIS, vol. 937, pp. 99–109. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-3441-2_8
14. Zheng, S., et al.: Conditional random fields as recurrent neural networks. In: The IEEE International Conference on Computer Vision (ICCV), December 2015