Cervical Nuclei Segmentation in Whole Slide Histopathology Images Using Convolution Neural Network

Qiuju Yang[1], Kaijie Wu[1(✉)], Hao Cheng[1], Chaochen Gu[1], Yuan Liu[2], Shawn Patrick Casey[1], and Xinping Guan[1]

[1] Department of Automation, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai Jiao Tong University, Shanghai 200240, China
{napolun279, kaijiewu, jiaodachenghao, jacygu, shawncasey, xpguan}@sjtu.edu.cn

[2] Pathology Department, International Peace Maternity and Child Health Hospital of China Welfare Institute, Shanghai 200030, China
sean_han@163.com

Abstract. Pathologists generally diagnose whether or not cervical cancer cells have the potential to spread to other organs and assess the malignancy of cancer through whole slide histopathology images using virtual microscopy. In this process, the morphology of nuclei is one of the significant diagnostic indices, including the size, the orientation and arrangement of the nuclei. Therefore, accurate segmentation of nuclei is a crucial step in clinical diagnosis. However, several challenges exist, namely a single whole slide image (WSI) often occupies a large amount of memory, making it difficult to manipulate. More than that, due to the extremely high density and variant shapes, sizes and overlapping nuclei, as well as low contrast, weakly defined boundaries, different staining methods and image acquisition techniques, it is difficult to achieve accurate segmentation. A method is proposed, comprised of two main parts to achieve lesion localization and automatic segmentation of nuclei. Initially, a U-Net model was used to localize and segment lesions. Then, a multi-task cascade network was proposed to combine nuclei foreground and edge information to obtain instance segmentation results. Evaluation of the proposed method for lesion localization and nuclei segmentation using a dataset comprised of cervical tissue sections collected by experienced pathologists along with comparative experiments, demonstrates the outstanding performance of this method.

Keywords: Nuclei segmentation · Whole slide histopathology image
Deep learning · Convolutional neural networks · Cervical cancer

1 Introduction

Worldwide, cervical cancer is both the fourth-most common cause of cancer and cause of death from cancer in women, and about 70% of cervical cancers occur in low and middle-income countries [1]. Its development is a long-term process, from precancerous

B. W. Yap et al. (Eds.): SCDS 2018, CCIS 937, pp. 99–109, 2019.
https://doi.org/10.1007/978-981-13-3441-2_8

changes to cervical cancer, which typically takes 10 to 20 years [1]. In recent years, with the widespread use of cervical cancer screening programs which allows for early detection and intervention, as well as helping to standardize treatment, mortality has been dramatically reduced [2]. With the development of digital pathology, clinicians routinely diagnose disease through histopathological images obtained using whole slide scanners and displayed using virtual microscopy. In this approach, the morphology of nuclei is one of the significant diagnostic indices for assessing the degree of malignancy of cervical cancer. It is of great significance to make accurate nuclei segmentation in order to provide essential reference information for pathologists. Currently, many hospitals, particularly primary medical institutions lack experienced experts, which influences diagnostic efficiency and accuracy. Therefore achieving automatic segmentation of nuclei is necessary to reduce the workload on pathologists and help improve efficiency, as well as to assist in the determination of treatment plans and recovery prognosis.

Whole slide images (WSI) with high resolution usually occupies large amounts of memory. Therefore, it is difficult to achieve high efficiency and throughput if WSI are directly processed. Due to overlapping, variant shape and sizes, extremely high density of nuclei, as well as factors such as low contrast, weakly defined boundaries, and the use of different staining methods and image acquisition techniques, accurate segmentation of nuclei remains a significant challenge.

In recent years, with the application of deep learning methods for image segmentation, a significant amount of research has been devoted to the development of algorithms and frameworks to improve accuracy, especially in areas of non-biomedical images. Broadly speaking, image segmentation includes two categories; semantic and instance segmentation methods. The semantic method achieves pixel-level classification, which transforms traditional CNN [3] models into end-to-end models [4] such as existing frameworks including FCN [5], SegNet [6], CRFs [7], DeepLab [8], U-Net [9], and DCAN [10]. Based upon semantic segmentation, the instance segmentation method identifies different instances, and includes MNC [11], FCIS [12], Mask RCNN [13], R-FCN [14], and similar implementations. Although these methods achieved considerable results, their application in the field of biomedical images with complex background is relatively poor, with the exception of U-Net [9]. U-Net [9] is a caffe-based convolutional neural network which is often used for biomedical image segmentation and obtains more than acceptable results in many practical applications.

In the case of whole slide images of cervical tissue sections, recommendation of a pathologists' clinical diagnostic process was followed, localizing lesions and segmenting nuclei for diagnosing diseases. The method relies upon two steps with the first being localization and segmentation of lesions in WSI using the U-Net [9] model (Fig. 1, Part1). The second step, nuclei segmentation, builds a multi-task cascade network to segment the nuclei from lesions areas, hereinafter referred to as MTC-Net (Fig. 1, Part2). Similar to DCAN [10], MTC-Net leverages end-to-end training which reduces the number of parameters in the fully connected layer and improves computational efficiency. MTC-Net combines nuclei foreground and edge information for accurate instance segmentation results. However it differs from DCAN [10] in that an intermediate learning process, a noise reduction network of nuclei foreground and a distance transformation learning network, are added. A nuclei segmentation dataset of

stained cervical sections was used for comparative study, and the results show that segmentation accuracy has been improved by using this method, especially in the case of severely overlapping nuclei.

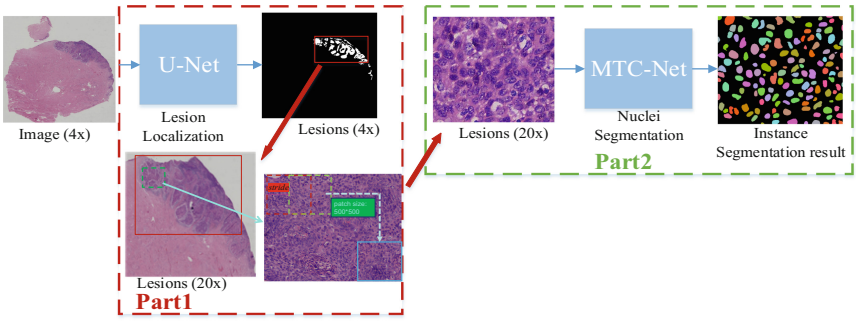

Fig. 1. The overview of the proposed method. Part1 is lesion localization using U-Net [9], the input is a cervical cell image at 4x magnification. The output is a probability map of the input. The lesion region with its coordinates, are chosen and mapped to the same image at 20x magnification. In Part2, a randomly cropped nuclei image from the lesion localized in Part1 is used as the input image of MTC-Net, finally obtaining the instance segmentation result.

2 Experiments

In this section, we describe in detail the preparation of our dataset, detailed explanation of the network structure and loss function of every stage.

2.1 Dataset and Pre-processing

All of the cervical tissue section images in our WSI dataset were collected from the pathology department of International Peace Maternity & Child Health Hospital of China welfare institute (IPMCH) in Shanghai. The dataset contains 138 WSI of variant size, with each sample imaged at 4x and 20x magnification and all ground truth annotations labeled by two experienced pathologists.

Images at 4x magnification were chosen for the initial portion of the algorithm using U-Net [9]; ninety for training/validation and 48 images for testing. Pathologists labeled lesions present in all images in white with the rest of image, viewed as the background region, masked in black. All training/validation images were resized to *512 * 512* in order to reduce computational and memory overhead.

Taking into account the time-consuming nature of labeling nuclei, while implementing the second step MTC-Net, fifty randomly cropped images from the lesions of the WSI dataset were prepared as our nuclei segmentation dataset, with a size of *500 * 500* pixels at 20x magnification. Then pathologists marked nuclei in every image with different colors in order to distinguish between different instances. Ground truth instance and boundary labels of nuclei were generated from pathologists' labels in preparation for model training. We chose 35 images for the training/validation and 15

images for the testing portion. Given the limited number of images, the training/validation dataset was enlarged using a sliding window with a size of *320 * 320* pixels, cropping in increments of 50 pixels. After obtaining small tiles using the sliding window, each tile was processed with data augmentation strategies including vertical/horizontal flip and rotation (0º, 90º, 180º, 270º). Finally, there were 3124 training images in total.

2.2 Lesion Localization

A fully convolutional neural network, U-Net [9], was used as the semantic segmentation model to separate the lesions from the whole slide images (Fig. 2). The input is an *RGB* image at 4x magnification, and the output of this network is a probability map of grayscale pixel values varying from 0 to 1, with a threshold set to 0.6 in order to obtain final segmentation result which is binary. When comparing with the binary ground truth label with pixel values are 0 (background) and 1 (lesions), the semantic segmentation loss function L_l is defined as:

$$L_l(\theta_l) = L_{bce}(output, label) \tag{1}$$

L_{bce} is the binary cross entropy loss function, θ_l denotes the parameters of the semantic segmentation network U-Net [9].

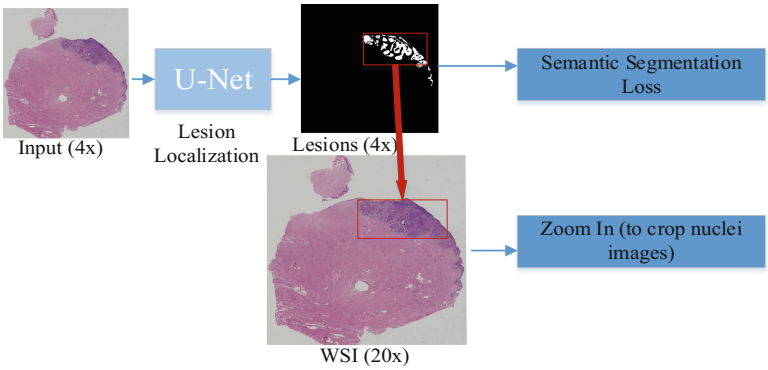

Fig. 2. Procedure of lesion localization. Input is an *RGB* image and the output is a probability map with grayscale pixel values varying from 0 to 1.

2.3 Nuclei Segmentation

Loss Function

The training details of this network (Fig. 3) is divided into four stages, where UNET1 and UNET2 are both U-Net [11] models. The whole loss function L_{seg} is defined as:

$$L_{seg} = \begin{cases} L_1 & stage1 \\ L_1 + L_2 & stage2 \\ L_1 + L_2 + L_3 & stage3 \\ L_1 + L_2 + L_3 + L_4 & stage4 \end{cases} \quad (2)$$

L_1 is the binary cross entropy loss of UNET1, L_2 is the mean squared error loss of stack Denoising Convolutional Auto-Encoder (sDCAE) [15], L_3 is the mean squared error loss of UNET2, L_4 is the binary cross entropy loss of Encoder-Decoder (ED) [16].

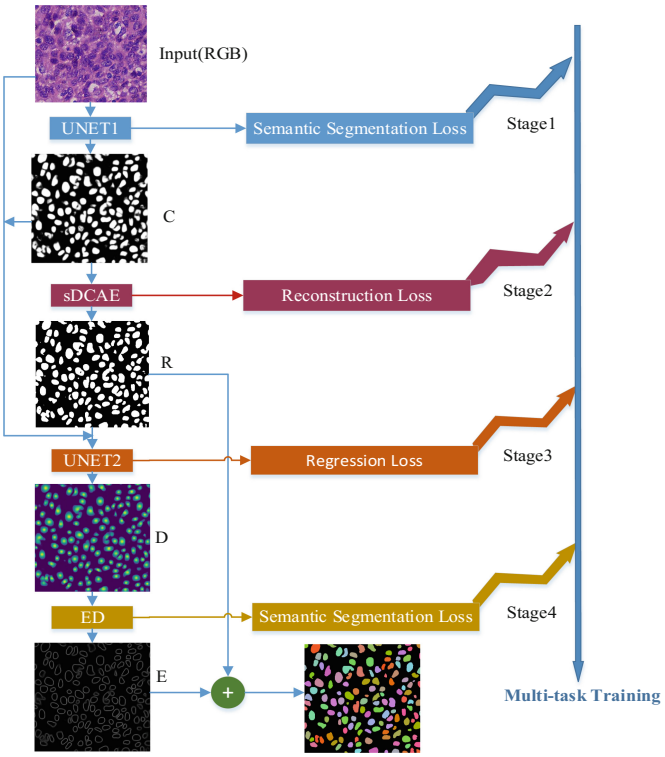

Fig. 3. The procedure of Cervical nuclei segmentation using a multi-task cascaded network (MTC-Net).

Training and Implementation Details

During training stages, the network in each stage focuses on the learning of a sub-task and relies upon the previous output. Therefore, the whole training process is a multi-task cascaded network (MTC-Net). The first stage implements UNET1 for foreground extraction network to isolate the nuclei from the complex background, as much as possible. The input is an *RGB* image, and the semantic output C is the preliminary segmentation image, with semantic segmentation loss L_1 defined as:

$$L_1(\theta_1) = L_{bce}(C, input(RGB)) \tag{3}$$

L_{bce} is the binary cross entropy loss function, θ_1 denotes the parameters of UNET1.

The second stage implements sDCAE [15] as the noise reduction network to reconstruct nuclei foreground and segments edges from the semantic output C. As an end-to-end training, fully convolutional network, sDCAE [15] is not sensitive to the size of input images and more efficient with less parameters when compared to fully connected layers. The input is semantic output C and the output R is the reconstruction image after noise reduction, semantic reconstruction loss is defined as:

$$L_2(\theta_2) = L_{mse}(R, C) \tag{4}$$

L_{mse} is the mean squared error loss function, θ_2 denotes the parameters of sDCAE [15].

The third stage is using UNET2 as the distance transformation learning network of the nuclei. Inputs are the *RGB* image, C and R, with the output D is a distance transformation image. At the same time, distance transformation is used to convert the ground truth instance labels into distance transformation labels (*DT*). Then making a regression on *DT* and D, so regression loss L_3 is defined as:

$$L_3(\theta_3) = L_{mse}(D, DT) \tag{5}$$

L_{mse} is the mean squared error loss function, θ_3 denotes the parameters of UNET2.

The last stage uses ED [16] as the edge learning network of the nuclei. The construction of ED [16] uses conventional convolution, deconvolution and pooling layers. The input is D and output is the prediction segmentation mask E of nuclei. According to ground truth boundary label B, the semantic segmentation loss L_4 is defined as:

$$L_4(\theta_4) = L_{bce}(E, B) \tag{6}$$

L_{bce} is the binary cross entropy loss function, θ_4 denotes the parameters of ED [16].

When generating the final instance result of the input image, the predicted probability maps of R and E were fused, and the final segmentation mask *seg* is defined as:

$$seg(i,j) = \begin{cases} 1 & E(i,j) \geq \lambda \text{ and } R(i,j) \geq \omega \\ 0 & otherwise \end{cases} \tag{7}$$

where *seg(i, j)* is one of the pixel of *seg*, *E(i, j)* and *R(i, j)* are the pixels at coordinate (i, j) of the nuclei segmentation prediction mask E and the predicted probability maps R respectively, λ and ω are thresholds, set to 0.5 empirically. Then each connected domain in *seg* is filled with different values to show the instance segmentation result of nuclei.

The whole framework is implemented under the open-source deep learning network Torch. Every stages' weights were initially set as 0, the learning rate was set as $1e^{-4}$ initially and multiplied by 0.1 every 50 epochs.

3 Evaluation and Discussion

To illustrate the superiority and provide effective evaluation metrics for our model, the winning model of the Gland Segmentation Challenge Contest in MICCAI 2015–DCAN [10] was chosen as a baseline to perform a comparative experiment.

3.1 Evaluation Metric

In the initial step (Lesion Localization), U-Net [9] used the common metric *IoU* to evaluate the effect of localization. *IoU* is defined as:

$$IoU(G_w, S_w) = (|G_w \cap S_w|)/(|G_w| \cup |S_w|) \tag{8}$$

where $|G_w|$ and $|S_w|$ are the total number of pixels belonging to the ground truth lesions and the semantic segmentation result of lesions respectively.

In second step (Nuclei Segmentation), the evaluation criteria include traditional dice coefficient D_1 and ensemble dice D_2. D_1 measures the overall overlapping between the ground truth and the predicted segmentation results. D_2 captures mismatch in the way the segmentation regions are split, while the overall region may be very similar. The two dice coefficients will be computed for each image tile in the test dataset. The *Score* for the image tile will be the average of the two dice coefficients. The score for the entire test dataset will be the average of the scores for the image tiles. D_1 and D_2 are defined as:

$$\begin{cases} D_1(G_n, S_n) = (|G_n \cap S_n|)/(|G_n| \cup |S_n|) \\ D_2 = 1 - \frac{|G_n - S_n|}{|G_n| \cup |S_n|} \\ Score = \frac{D_1 + D_2}{2} \end{cases} \tag{9}$$

Where $|G_n|$ and $|S_n|$ are the total number of pixels belonging to the nuclei ground truth annotations and the nuclei instance segmentation results respectively, *Score* is the final comprehensive metric of the method.

3.2 Results and Discussion

Some semantic segmentation results of testing data in lesion localization, and the visualization of the comparative instance segmentation results in nuclei segmentation, were analyzed.

The architecture of U-Net [9] combines low-level features to ensure the resolution and precision of the output and high-level features used to earn different and complex features for accurate segmentation at the same time. Another advantage is that U-Net [9] utilizes the auto-encoder framework to strengthen the boundary recognition capabilities by adding or removing noise automatically.

U-Net [9] in Part1 can accurately localize and segment the lesions from WSI (Fig. 4). The semantic segmentation results of the network with the threshold set to 0.6 are almost the same as the ground truth, and the results achieved the *IoU* above 97%, which laid the foundation for the subsequent work of nuclei instance segmentation to obtain good results.

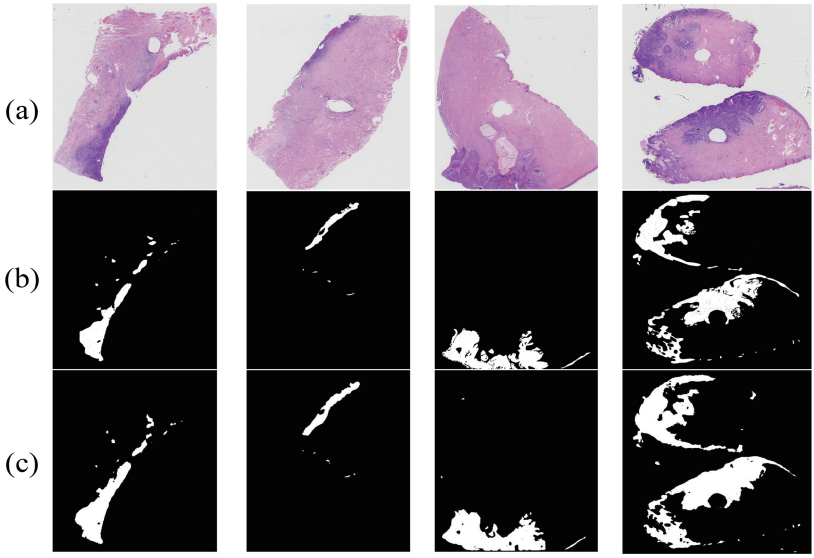

Fig. 4. Semantic segmentation results of testing data in lesion localization. (a): WSI at 4x magnification. (b): ground truth masks of WSI. (c): segmented images.

Nuclei instance segmentation results compared with DCAN [10] (Fig. 5), with MTC-Net exhibiting higher sensitivity for nuclei with severe overlap or blurred boundaries. The application of UNET2 enhanced the segmentation edges and improved the model sensitivity of nuclei edges, and then improved the accuracy of this model.

Quantitative comparative results between DCAN [10] and MTC-Net on the nuclei segmentation dataset were obtained (Table 1), with thresholds λ and ω both set to 0.5. In order to account for possible errors from edge segmentation in nuclei foreground, both segmentation results of DCAN [10] and MTC-Net were operated by morphological expansion. MTC-Net achieves better performance, with the final score about 3% higher than DCAN [10]. The comparative results demonstrate MTC-Net is more effective than DCAN [10] in the field of nuclei segmentation.

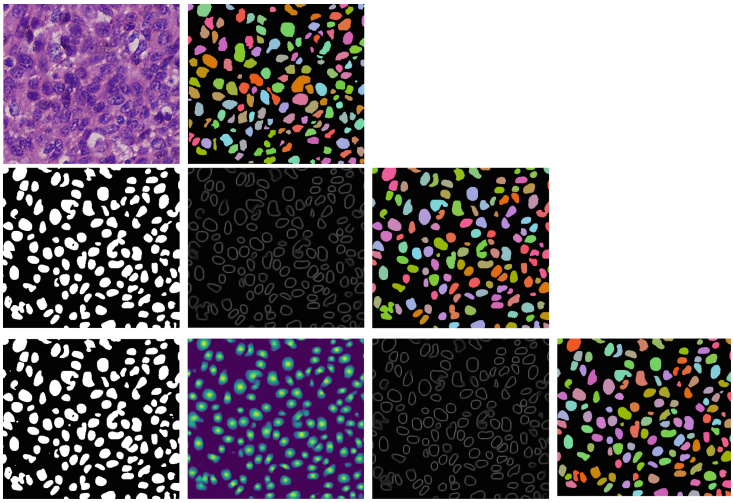

Fig. 5. The comparative nuclei segmentation results using DCAN [10] and MTC-Net. The first row are original image and the ground truth segmentation of this image (left to right). The second row are segmentation results of nuclei foreground, nuclei edges and instance segmentation results (left to right) using model DCAN [10]. The third row are nuclei foreground noise reduction segmentation results, the distance transformation results, nuclei edges segmentation results and the instance segmentation results (left to right) using MTC-Net.

Table 1. The quantitative comparative results between DCAN [10] and MTC-Net on our nuclei segmentation dataset.

Method	Performance		
	D1	D2	Score
DCAN [10]	0.7828	0.7021	0.7424
MTC-Net	**0.8246**	**0.7338**	**0.7792**

4 Conclusions

A two-part method for lesion localization and automatic nuclei segmentation of WSI images of stained cervical tissue sections was introduced. A U-Net [9] model to localize and segment lesions was implemented. A multi-task cascaded network, named MTC-Net, was proposed to segment nuclei from lesions, which is potentially a crucial step for clinical diagnosis of cervical cancer. Similar to DCAN [10], MTC-Net combines nuclei foreground and edge information to obtain instance segmentation results, but the difference is that MTC-Net adds intermediate learning process in the form of a noise reduction network of nuclei foreground and a distance transformation learning network of nuclei. Comparative results were obtained based on our nuclei segmentation dataset, which demonstrated better performance of MTC-Net. After practical application, it was found to some extent that this work provides essential reference information

for pathologists in assessing the degree of malignancy of cervical cancer, which can reduce the workload on pathologists and help improve efficiency. Future work will continue to optimize MTC-Net and focus on training with a larger dataset to achieve higher segmentation accuracy.

Acknowledgements. This work is supported by National Key Scientific Instruments and Equipment Development Program of China (2013YQ03065101) and partially supported by National Natural Science Foundation (NNSF) of China under Grant 61503243 and National Science Foundation (NSF) of China under the Grant 61521063.

References

1. Mcguire, S.: World cancer report 2014. Geneva, Switzerland: world health organization, international agency for research on cancer, WHO Press, 2015. Adv. Nutr. **7**(2), 418 (2016)
2. Canavan, T.P., Doshi, N.R.: Cervical cancer. Am. Fam. Physician **61**(5), 1369 (2000)
3. LeCun, Y.: http://yann.lecun.com/exdb/lenet/. Accessed 16 Oct 2013
4. Saltzer, J.H.: End-to-end arguments in system design. ACM Trans. Comput. Syst. (TOCS) **2** (4), 277–288 (1984)
5. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2014)
6. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **12**(39), 2481–2495 (2017)
7. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., et al.: Conditional random fields as recurrent neural networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 1529–1537 (2015)
8. Chen, L.C., Papandreou, G., Kokkinos, I., et al.: 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)
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. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2487–2496 (2016)
11. Dai, J., He, K., Sun, J.: Instance-aware semantic segmentation via multi-task network cascades. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3150–3158 (2015)
12. Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y.: Fully convolutional instance-aware semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4438–4446 (2017)
13. He, K., Gkioxari, G., Dollár, P., et al.: Mask R-CNN. In: IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988 (2017)
14. Dai, J., Li, Y., He, K., et al.: R-FCN: Object detection via region-based fully convolutional networks. Advances in Neural Information Processing Systems 29 (NIPS) (2016)

15. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**(12), 3371–3408 (2010)
16. Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. Computer Science (2014)