Gland segmentation in pancreas histopathology images based on selective multi-scale attention

Yang, Changxing, Xiang, Dehui, Bian, Yun, Lu, Jianping, Jiang, Hui, et al.

Changxing Yang, Dehui Xiang, Yun Bian, Jianping Lu, Hui Jiang, Jianming Zheng, "Gland segmentation in pancreas histopathology images based on selective multi-scale attention," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115962M (15 February 2021); doi: 10.1117/12.2581039

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

Gland Segmentation in Pancreas Histopathology Images Based on Selective Multi-scale Attention

Changxing Yang[1], Dehui Xiang[*1], Yun Bian[2], Jianping Lu[2], Hui Jiang[3], Jianming Zheng[3]

[1]School of Electronic and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China, 20184228037@stu.suda.edu.cn

[2]Department of Radiology, Changhai Hospital, The Navy Military Medical University, Shanghai, China, bianyun2012@foxmail.com

[3]Department of Pathology, Changhai Hospital, The Navy Military Medical University, Shanghai, China, jianghui5131@163.com

ABSTRACT

Pathology is an important subject in the treatment of pancreatic cancer. The tumor presented in the pathological images includes not only the tumor cells, but also the surrounding background structures. Automatic and accurate gland segmentation in histopathology images plays a significant role for cancer diagnosis and clinical application, which assist pathologists to diagnose the malignancy degree of pancreas caner. Due to the large variability of size and shape in glandular appearance and the heterogeneity between different cells, it is a challenging task to accurately segment glands in histopathology images. In this paper, a selective multi-scale attention (SMA) block is proposed for gland segmentation. First, a selection unit is used between the encoder and decoder to select features by amplifying effective information and suppressing redundant information according to a factor obtained during training. Second, we propose a multi-scale attention module to fuse feature maps at different scales. Our method is validated on a dataset of 200 images of size 512×512 from 24 H&E stained pancreas histological images. Experimental results show that our method achieves more accurate segmentation results than that of state-of-the-art approaches.

Keywords: pancreatic histopathology image, gland segmentation, selective multi-scale attention

1. INTRODUCTION

Pancreatic cancer is a tumor of digestive tract with high degree of malignancy. Its 5-year survival rate is less than 10% [11], and it is one of the malignant tumors with the worst prognosis. Histopathology image analysis is often one of the first steps in diagnosis of cancer. The tumor presented on histopathology images includes tumor cells and surrounding tissues such as blood vessels, nerves, etc. [1, 2]. The size and shape of glands are highly related to the presence and severity of diseases [8]. The area and proportion of these glands are crucial to the diagnosis, treatment and prognosis of tumors. Therefore, it is of great significance to realize the segmentation of tumor cells and other surrounding tissues. As is shown in Fig.1, it is still a challenging task to segment glands in histopathology images. First, the boundaries of tumor cells are often unclear (Fig.1 (a)). Second, the color differences are very small between the nerve cells and the surrounding stroma (Fig.1 (b)). Third, blood vessels and nerves are similar in color (Fig.1 (c)).

Deep convolutional neural networks (DCNNs) have been wildly used in image segmentation, such as UNet [3], SegNet [4] and FCN [5], and inspire many researchers to apply them in gland segmentation in histopathology images. Most gland segmentation methods extracted features using an encoder and recover spatial dimension from low encoder layers gradually. Skip connections were employed to help decoder layers output a more accurate segmentation result by fusing features from encoder layers. S. Graham [6] presented a minimal information loss dilated network by re-introducing the original image at multiple points to retain maximum information during feature extraction. Shan E Ahmed Raza [9] proposed a convolutional neural network, which bypassed the max-pooling operation through extra layers to retain information, aiming to improve the segmentation of glands. Qu [10] developed a full resolution network that maintained full resolution feature maps and proposed a variance constrained cross entropy loss in order to learn the spatial relationship between pixels in the same instance.

* Corresponding author: Dehui Xiang, E-mail: xiangdehui@suda.edu.cn

Medical Imaging 2021: Image Processing, edited by Ivana Išgum, Bennett A. Landman, Proc. of SPIE Vol. 11596, 115962M · © 2021 SPIE · CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2581039

Although DCNNs with U-structure have achieved remarkable performances in gland segmentation, the capability of information extraction in each single stage is insufficient. On one hand, there is no effective extraction of multi-scale information when dealing with glands that vary greatly in size and shape. On the other hand, the number of the filters at each layer is usually set to be large in order to capture as many features as possible, but it is effective to make a selection of the features. Therefore, a selective multi-scale attention (SMA) block is proposed in our method. First, we focus on extracting the multi-scale information in encoder and designing a selection unit to amplify effective information and suppress redundant information. Second, we fuse the selected feature maps at different scales from the encoder and decoder with a multi-scale attention module.

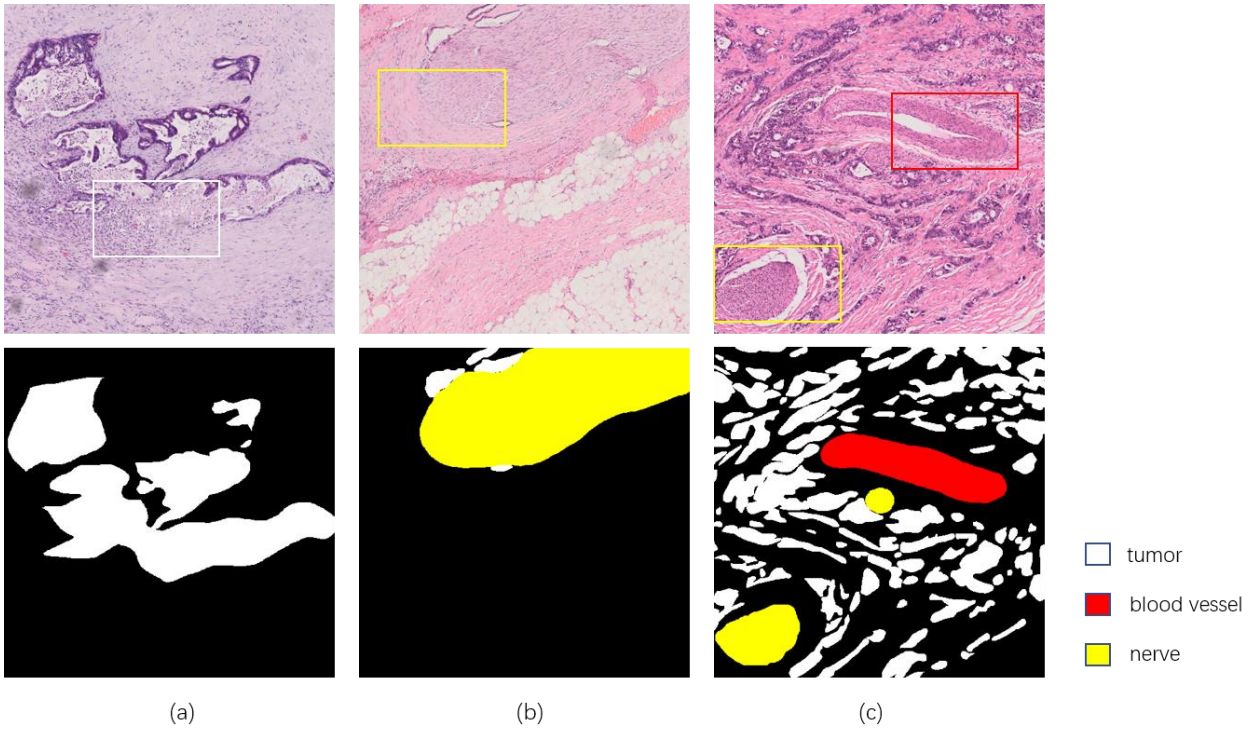

Fig.1. Illustration of the challenges in gland segmentation. (a) White box denotes the boundaries of tumor cells are often unclear. (b) Yellow box represents the color differences are very small between the nerve and the surrounding stroma. (c) Blood vessel in red box are of large similarity of nerve in yellow box

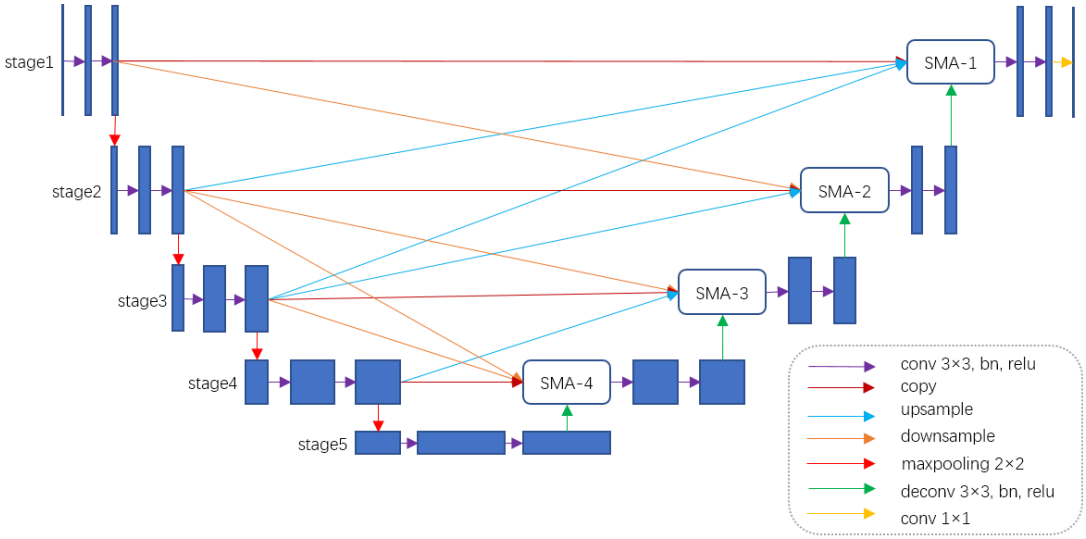

Fig.2 Framework of proposed method for pancreas gland segmentation.

Our contributions are summarized as follows:

•A selective multi-scale attention block is proposed for feature selection and fusion between the encoder and decoder.

•A selection unit is used to select features by amplifying effective information and suppressing redundant information according to a factor obtained during training.

•A multi-scale attention module is designed to fuse feature maps at different scales to deal with glands that vary greatly in size and shape.

2. METHODS

2.1 Network Architecture

The overview of our proposed network is shown in Fig.2. The input of our network is a RGB image and its size is 512×512. We use the encoder-decoder structure based on UNet and propose a selective multi-scale attention block between the encoder and decoder for feature selection and fusion. It consists of two parts: selection unit and multi-scale attention module. The selection unit is used for selecting features by amplifying effective information and suppressing redundant information according to a factor learned during training. The multi-scale attention module is designed to fuse feature maps of different scales from the encoder and decoder. The output of our network has four channels, representing background, tumor cell, blood vessel and nerve.

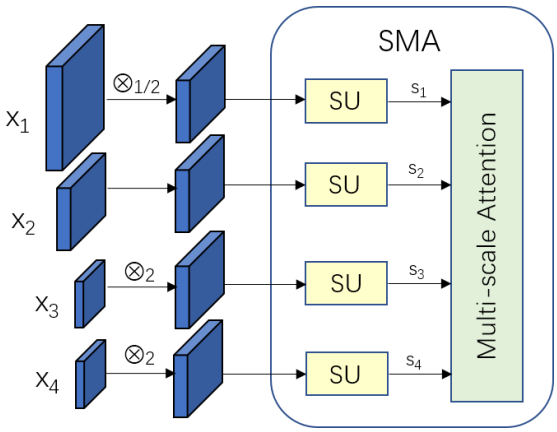

Fig.3 Selective multi-scale attention (SMA) block.

2.2 Selective Multi-scale Attention Block

The selective multi-scale attention block is proposed for feature selection and fusion. In SMA block, the skip-connection is revised by combining the feature maps of this stage with the feature maps of adjacent stages from the encoder and the high-level feature maps in the decoder. For example, Fig.3 shows the SMA-2 block. First, feature maps x_1 from stage1 are down-sampled by a factor of 1/2, x_3 from stage3 and x_4 from decoder are up-sampled by a factor of 2 to the same size as x_2. The down-sample operations are implemented by convolution with different strides and up-sample operations at different scales are implemented by deconvolution. Next, in order to select the information, the feature maps are sent to a select unit in parallel. Finally, the selected feature maps are fused and weighted in the multi-scale attention module.

The operation of selection unit (Fig.4 (a)) can be formulated as below:

$$s = x + cbt(x) \otimes x = x \otimes (cbt(x) + 1) \tag{1}$$

where $\otimes$ means the pixel multiplication, and $\mathrm{cbt(x)} = \tanh(\mathrm{bn}(\mathrm{conv(x)}))$. The value of each pixel in the selected feature map s is obtained by multiplying the input feature map x with a factor learned during training. The output of tanh activation function ranges from -1 to 1, so the factor $cbt(x) + 1$ ranges from 0 to 2. When the factor is less than 1 or more than 1, the value of pixel is suppressed or amplified. We use it to select features at different scales to focus on the important information rather than using all available information.

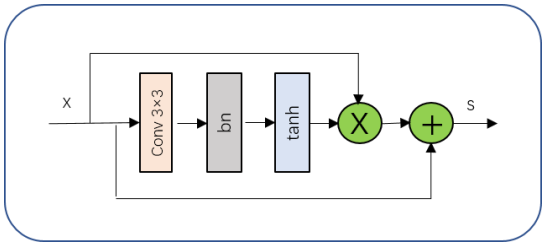

Fig.4 (a) Selection unit (SU).

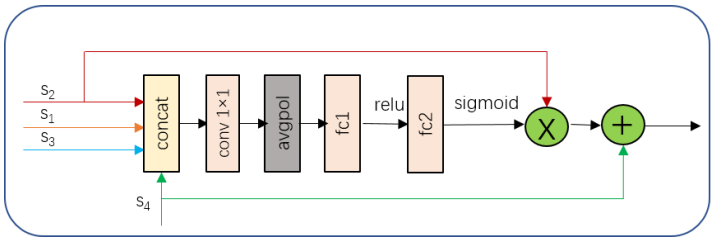

Fig.4 (b) Multi-scale attention (MA) module.

Multi-scale attention module (Fig.4 (b)) is designed for feature fusion. Since the size and shape of glands vary greatly, the capability of information extraction in each single stage is insufficient; we fuse feature maps from three adjacent stages in encoder and feature maps from decoder to learn different structure information. The fused features are used to weight feature from the encoder. The working scheme of the multi-scale attention module can be formulated as:

$$c = concat(s_1, s_2, s_3, s_4) \tag{2}$$

$$a = attention(\mathrm{c}) \tag{3}$$

$$\mathrm{y} = s_2 * a + s_4 \tag{4}$$

where s_1, s_2, s_3 and s_4 are the results of x_1, x_2, x_3, and x_4 passing through the select unit. We first use a 1×1 convolution to reduce the channels of concatenated feature maps to the same number as s_2, and then use a global average pooling layer, two fully-connected layers and sigmoid activation function inspired by SE-block [7] to generate a weight vector to make a weighted attention on s_2. The output of multi-scale attention module is acquired by adding the weighted feature maps s_2 in encoder and the feature maps s_4 in decoder.

2.3 Loss Function

Cross entropy loss is widely used in segmentation tasks. The cross entropy loss is defined as:

$$L_{ce} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} y_n^c \, log \, x_n^c \tag{5}$$

where N is the number of all pixels, C is the number of classes, x_n^c is the prediction of pixel n belonging to class c, and y_n^c is the groundtruth of pixel n belonging to class c.

We also use the multi-class dice loss during the process of training. The dice loss can be expressed as:

$$L_{dice} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{n=1}^{N} y_n^c x_n^c + \varepsilon}{\sum_{n=1}^{N}(y_n^c + x_n^c) + \varepsilon} \tag{6}$$

where ε denotes a smoothing factor 1e-6. The total loss can be denoted as:

$$L_{total} = L_{ce} + \lambda \cdot L_{dice} \tag{7}$$

where λ is the balance weight of L_{dice} and we set $\lambda = 1$ in our experiments.

3. RESULTS

3.1 Datasets

Our data is extracted from 24 H&E stained pancreas histological WSIs, scanned with a NanoZoomer Scanner with a pixel resolution of 0.221 μm/pixel (equivalent to 40× objective magnification). After scanning, the WSIs are down-sampled 16 times and then 200 image tiles are extracted. These images are all malignant and of size 512×512 pixels. Clinical experts manually label them as tumor cell, blood vessel and nerve. In our experiments, we randomly split the 200 images into four portions, and each part contains 50 images for the four-fold cross validation.

3.2 Data Augmentation

Considering that our data is limited, it is necessary to generate a large number of images to alleviate the overfitting of deep neural network. Thus, we horizontally flip, vertically flip and randomly rotate the images to make data augmentation during training. The corresponding scales are {True, True, 30}.

3.3 Implementation Details

Our model is trained on a workstation equipped with one NVIDIA Tesla K40m GPU with 12G memory for 200 epochs. We use the 'poly' learning rate policy, where $\mathrm{lr} = \mathrm{baselr} \times (1 - \frac{iter}{total_{iter}})^{power}$, the basic learning rate is set to 0.05 and power is set to 0.9. The batch size is 4 and stochastic gradient descent (SGD) is adopted in our experiments, in which momentum and weight decay are set to 0.9 and 0.0001, respectively.

3.4 Segmentation Results

To quantitatively assess the performance of our proposed method, we compare the segmentation results with the ground truth according to the following three metrics: dice similarity coefficient (DSC), recall and precision. The DSC calculates the similarity between the segmentation results and ground truth, and it is defined as

$$\mathrm{DSC} = \frac{2TP}{FP+2TP+FN} \tag{8}$$

where TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives. Recall and precision metrics are computed as:

$$\mathrm{Recall} = \frac{TP}{TP+FN} \tag{9}$$

$$\mathrm{Precision} = \frac{TP}{FP+TP} \tag{10}$$

We compare the proposed method against UNet [3], SegNet [4], FCN-8 [5] and MILD-Net [6] on our dataset. Segmentation results of different methods are shown in Fig.5. We compare the quantitative segmentation results to different methods in Table 1. It shows that our method achieves the highest dice similarity coefficient in all three categories. Specifically, our method improves the dice similarity coefficient of tumor cell by 0.8%, blood vessel by 1.7% and nerve by 1.6% over UNet.

In order to prove the effectiveness of the modules, we choose UNet as our baseline and the corresponding ablation experiments results are shown in the last two lines in Table 1. UNet + MA means that feature maps at different scales are directly sent to the multi-scale attention module without feature selection by the selection unit. UNet + SU + MA represents our method, which adds selection unit and multi-scale attention module to UNet.

Table 1. Comparison of quantitative segmentation results on our dataset using different methods.

Methods	tumor			blood vessel			nerve		
	Dice	Precision	Recall	Dice	Precision	Recall	Dice	Precision	Recall
SegNet [4]	0.8099	0.8707	0.7795	0.7051	0.8403	0.6681	0.7396	0.8484	0.7197
FCN-8 [5]	0.8245	**0.8927**	0.7912	0.6584	0.8192	0.6218	0.7084	0.8036	0.6919
MILD-Net [6]	0.8271	0.8724	0.8126	0.7048	0.8587	0.6628	0.7261	**0.8874**	0.6914
UNet [3]	0.8261	0.8843	0.7969	0.7379	0.8415	0.7107	0.7411	0.8517	0.7271
UNet + MA	0.8303	0.8909	0.8006	0.7503	**0.8593**	0.7181	0.7512	0.8797	0.7187
UNet + SU + MA	**0.8347**	0.8649	**0.8216**	**0.7548**	0.8547	**0.7301**	**0.7576**	0.8442	**0.7492**

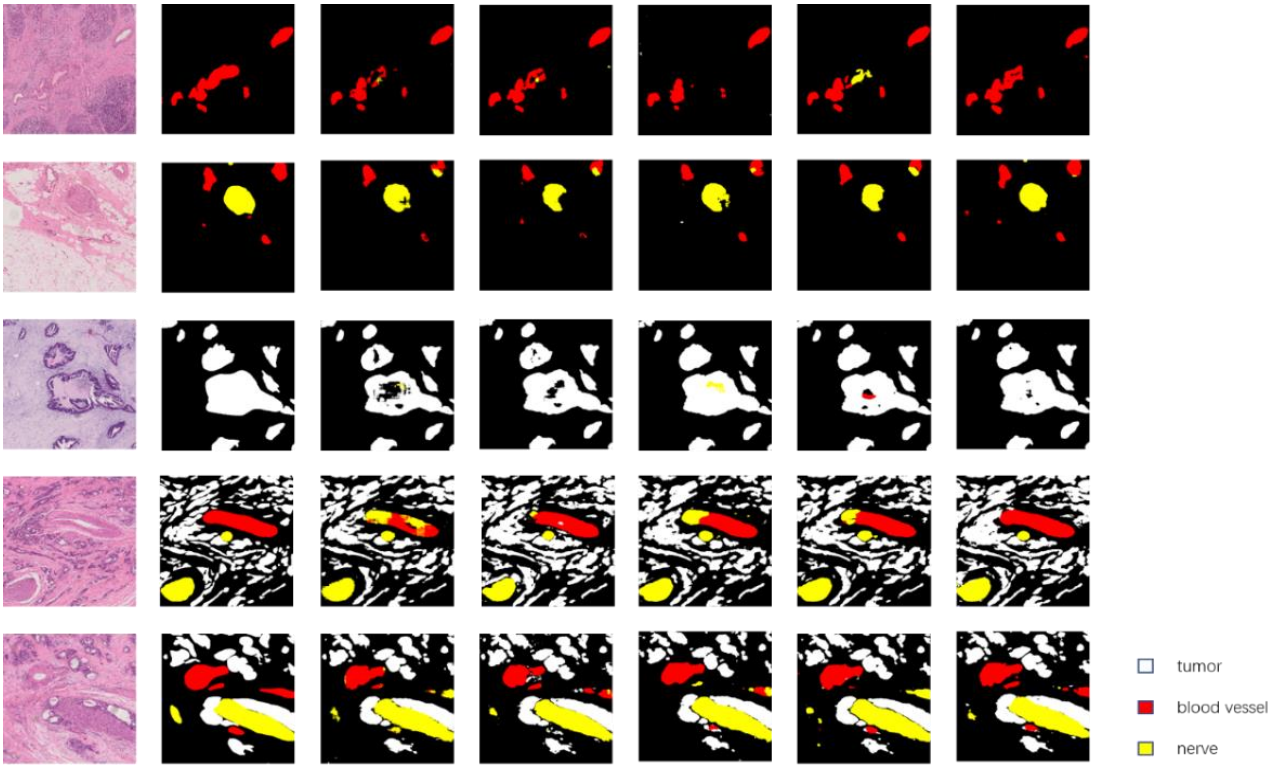

Fig.5 Gland segmentation results of different methods. The first column is the original image; the second column is the ground truth; the next columns are the results of SegNet, FCN-8, MILD-Net and UNet; the last column is our segmentation results.

4. CONCLUSIONS

In this paper, a selective multi-scale attention block is introduced for gland segmentation in pancreas histopathology images. First, we use a selection unit in order to select features by amplifying effective information and suppressing redundant information according to a factor obtained during training. Second, a multi-scale attention module is proposed to fuse feature maps at different scales from the encoder and decoder.

5. ACKNOWLEDGEMENTS

This work has been supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61971298, and in part by the National Key R&D Program of China under Grant 2018YFA0701700.

6. REFERENCE

[1] QuailDF, JoyceJA. Microenvironmental regulation of tumor progression and metastasis [J]. Nat Med, 2013, 19(11): 1423-1437.
[2] FeigC, GopinathanA, NeesseA, et al. The pancreas cancer microenvironment [J]. Clin Cancer Res, 2012, 18(16): 4266-4276.
[3] 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 9351, pp. 234–241, 2015
[4] 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)
[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] S. Graham, H. Chen and J. Gamper et al. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. Medical Image Analysis 52 (2019) 199–211

[7] Hu Jie, L. Shen, and G. Sun. Squeeze-and-Excitation Networks. Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2018
[8] Gurcan, M.N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.:Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. 2, 147 (2009)
[9] Shan E Ahmed Raza, Linda Cheung et al. Micro-Net: A unified model for segmentation of various objects in microscopy images. Medical Image Analysis 52 (2019) 160–173
[10] Qu, H., Yan, Z., Riedlinger, G.M., De, S., Metaxas, D.N.: Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 378–386. Springer(2019)
[11] Chu LC, Park S, Kawamoto S, et al. Utility of CT Radiomics Features in Differentiation of Pancreatic Ductal Adenocarcinoma From Normal Pancreatic Tissue[J]. AJR Am J Roentgenol, 2019, 213(2): 349-357.