A new hybrid method for gland segmentation in histology images

Li Yang Wang, Yu Zhou, and Bogdan J. Matuszewski

Computer Vision and Machine Learning Research Group, School of Engineering
University of Central Lancashire, UK
{lywang, yzhou11, bmatuszewski1}@uclan.ac.uk

Abstract. Gland segmentation has become an important task in biomedical image analysis. An accurate gland segmentation could be instrumental in designing of personalised treatments, potentially leading to improved patient survival rate. Different gland instance segmentation architectures have been tested in the work reported here. A hybrid method that combines two-level classification has been described. The proposed method achieved very good image-level classification results with 100% classification accuracy on the available test data. Therefore, the overall performance of the proposed hybrid method highly depends on the results of the pixel-level classification. Diverse image features reflecting various morphological gland structures visible in histology images have been tested in order to improve the performance of the gland instance segmentation. Based on the reported experimental results, the hybrid approach, which combines two-level classification, achieved overall the best results among the tested methods.

Keywords: Gland segmentation, Two-level classification, Deep learning, Random forest

1 Introduction

Colorectal cancer is one of the most commonly diagnosed cancers, which affects both women and men. Accurate cancer grading is essential for individual cancer treatment planning and resulting patient survival rate. Different morphological structures of the gland objects can be used for grading. Therefore, accurate segmentation of gland structures in histology images is important in order to support assessment of the cancer. One of the reasons why gland segmentation is challenging is that the structure and appearance of histology images, even for the same tissue, can look significantly different. Furthermore, the gland size, shape, texture and appearance could vary significantly even within the same gland category. Due to these reasons, gland segmentation is a challenging problem. The proposed hybrid method deals with these challenges by dividing the images into tissue type categories first and solving the pixel-wise classification separately for each predefined category of the histology image.

2 Related work

A large number of published papers demonstrate that gland segmentation has become an important problem in biomedical image analysis [3-7]. This section reviews some of the approaches previously proposed for gland segmentation.

Wu [3] proposed a region growing method where featureless areas are used as initial seeds for region growing. The chain of epithelial nuclei is used for termination of the region growing. The drawback of this method is that it only achieves good performance for images displaying benign tissue, and is not effective for images showing malignant tissue, due to the deformation of morphological structures of gland objects.

Gunduz-Demir [4] introduced an approach that has employed graph connectivity to classify initial seeds for region growing. This method is different from the method proposed by Wu as it uses pixel information to represent each tissue type. The edges between different gland objects are employed as the stopping criterion. Again, this method performs well only on images representing benign tissue.

Recently, deep learning methods have been used and achieved excellent performance in gland segmentation. Kainz [5] proposed an approach, which used two convolutional networks as pixel level classifier. The input for these two networks was processed by using the red channel of original images. A total weight variation of global segmentation was used to determine the final output.

Chen [6] introduced a method based on fully convolutional network (FCN). This method takes advantage of multi-level feature representations. The network uses generic encoder-decoder architecture. The down-sampling encoder generates the multi-level features and the up-sampling decoder is used to restore the original image size and provides the gland occurrence probability maps.

Li [7] proposed an approach, which combines deep learning and handcrafted features to train the SVM classifier. Different sizes of the patches for both handcrafted and deep learning features have been tested in [7] in order to improve the performance.

3 Method

Segmentation approaches considered in this paper could be divided into segmentation with and without pre-classification. Segmentation with the pre-classification could be further divided into pre-classification at the image and feature levels. Fig. 1 shows these three options adopted for the gland segmentation problem. Method 1 is a simplest approach. Both Method 2 and Method 3 use the pre-classification and are variants of the Method 1. The difference between Methods 2 and 3 is that for the Method 2 the image features are learned and extracted separately for the benign and malignant gland images, whereas for the Method 3 the features are learned for all training images and the separation between benign/malignant images is performed after features are extracted (i.e. at the feature level). Segmentation with the pre-classification is in effect a two-level classification and consists of image-level classification and pixel-level classification. The image-level classification part is to separate the histology images into benign and malignant cases, and the pixel-level classification is to perform the actual gland

segmentation. The final segmentation results for method 2 and 3 are superposition of the segmentation results for benign and malignant cases.

The morphological structure of glands in benign and malignant cases is significantly different (see Fig. 3 for typical histology images showing benign and malignant tissue). To obtain good segmentation results, it is not only needed to separate the images into benign or malignant cases but also develop a way to describe the morphological structure of gland objects for these cases. In this work, two and three different target classes have been tested in order to find the best way to describe the local discriminative patterns for benign and malignant tissue. For the two-target classes, the gland and non-gland parts of the image were considered, whereas for a three-target classes case, gland inside, gland boundary, and gland outside image regions were taken into consideration.

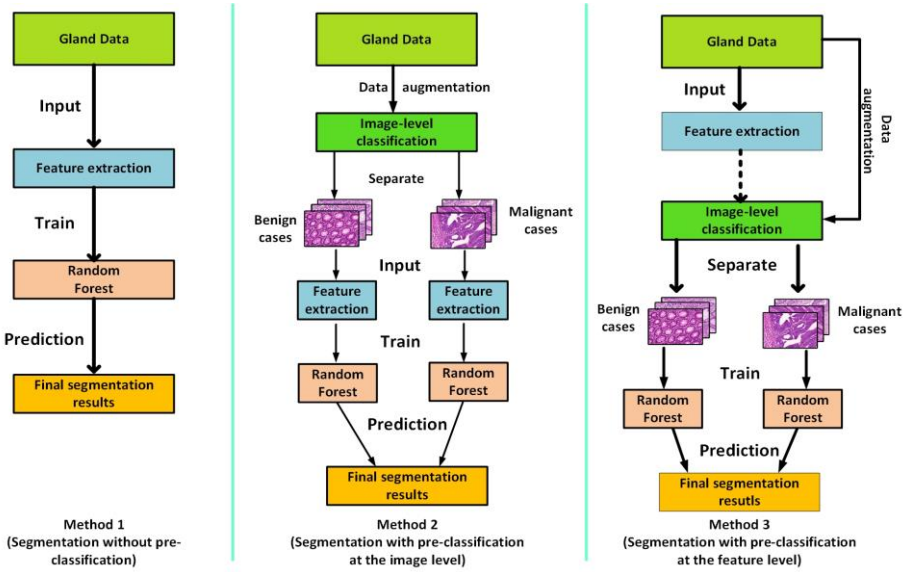

Fig 1. Different segmentation architectures adopted for gland segmentation problem. From left to right: Method 1, represents the segmentation without pre-classification; Method 2, is segmentation with the pre-classification at the image level; and Method 3, uses segmentation with pre-classification at the feature level.

3.1 Image-level classification

As mentioned previously, image-level classification aims to separate the histology images into benign and malignant cases. Recently, deep learning techniques have achieved excellent performance in image classification tasks. Three different deep learning architectures, AlexNet [8], GoogleNet [9] and ResNet-50 [10] have been tested on the gland image classification problem.

There are only 85 training images and 80 test images in the gland dataset [1, 2, 18] used in the reported experiments. It is therefore important to increase the number of training images in order to avoid overfitting. Data augmentation methods, adopted to increase the number of training images use local image deformations and colour jitter.

Colour jitter changes image appearance by modifying image tones without changing morphological structure. The local image deformations are used to modify shape of structures and textures without changing the colour. In the adopted implementation, a 2D thin-plate spline [11] has been used to augment the training data. The deformation model uses fixed 10x10 grid and a random displacement of each grid point with the maximum displacement of 9 pixels. Fig. 2 shows an example of original images and images after using data augmentation.

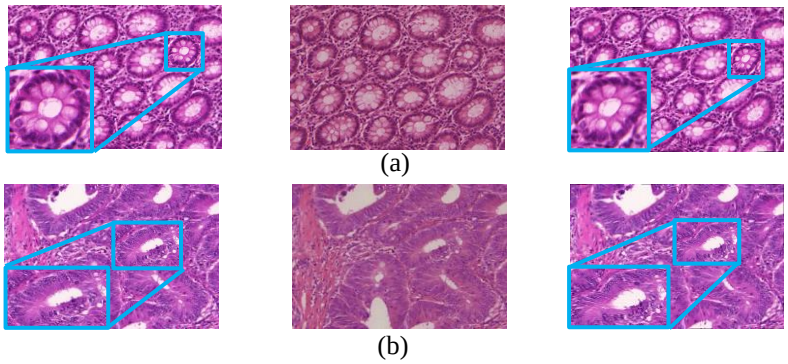

Fig 2. Original images and images after applying the adopted data augmentations. (a) From left to right: an original image with benign tissue, corresponding image after colour jitter and image after local image deformation (b) From left to right: an original image with malignant tissue, corresponding image after colour jitter and image after local image deformation.

For evaluation of different image classification architectures, 80% of training images from each category are used to generate additional training images using described image augmentation process. As result, there are 6392 training images in total. The remaining 17 histology images are used for validation. The original 80 test images from the gland dataset are used for testing. All the tested networks employ the Adam optimisation method [12].

3.2 Pixel level classification

Pixel level classification deals with the gland instance segmentation. There are many types of feature extraction methods, which have been widely used in classification tasks. In gland segmentation, both handcrafted features and deep learning features have been applied. In this work number of handcrafted features were tested, including ring histograms [13], rotation-invariant local uniform patterns (LBP) [14] and circular Fourier HOG features [15], as well as deep features, including LeNet-5 and GoogleNet.

As described in Section 3, the pixel level classification is to solve gland segmentation for each image category. The details of two and three target classes adopted for the experiments are given below.

Two target classes. For images with benign or malignant tissue, the two target classes are defined as gland and non-gland (background) image areas. The labels for gland and

background are provided in the gland database, with a sample ground truth shown in Fig. 3. The images and the provided two-class ground truth have been used to train feature extraction methods with the random forest used as pixel level classifier. The results are determined by using a set of morphological post-processing steps on probability maps generated by the random forest.

Three target classes. For images with benign or malignant tissue, the three target classes are defined as 'gland inside', 'gland boundary' and 'gland outside'. The 'gland boundary' labels are generated by applying the erosion to the original ground truth images with two labels, and subsequently performing the XOR operation between the original and the eroded images. Fig.3 shows the labels for these three target classes.

Pixel-level classifier. Random forest technique has been used as a pixel-level classifier to learn the local patterns in the histology images. The forest applied for the gland segmentation uses Breiman model [16]. The Gini impurity is used to split samples in each tree node. The mathematical expression of Gini impurity is given as:

$$Gini\ impurity = 1 - \sum_{i=1}^{c} P_i^2 \tag{1}$$

where c is the total number of classes in each splitting node, and P_i is the probability of the class i.

The class output of the forest model is defined by the majority vote collected from all decision trees in the forest.

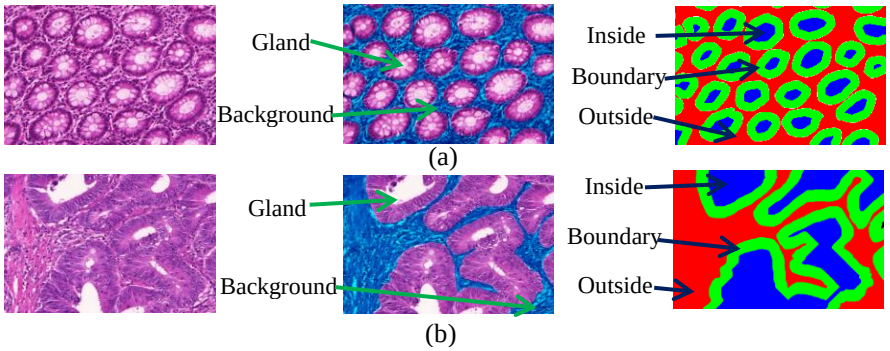

Fig 3. Example images and corresponding labels for two and three target classes. (a) From left to right: image with benign tissue, the ground truth for two target classes and the ground truth for three target classes (b) From left to right: image with malignant tissue, the ground truth for two target classes and the ground truth for three target classes.

4 Evaluation

The performance of segmentation results was evaluated by using four evaluation measures [18, 20]: (1) detection accuracy of individual glands; (2) object-level segmentation accuracy; (3) object-level shape similarity using Hausdorff distance, and (4) object-level shape similarity using Boundary Jaccard index.

Detection accuracy. F1 score has been used in this work in order to estimate the detection accuracy for individual glands [18]. If a segmented gland object overlaps at least 50% with the corresponding ground truth it is treated as true positive (**TP**); otherwise, it treated as false positive (**FP**). The difference between the number of ground truth and the number of true positive has been treated as the number of false negative (**FN**). The F1 score is defined as:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{2}$$

where

$$Precision = \frac{TP}{TP+FP}\ ,\ Recall = \frac{TP}{TP+FN} \tag{3}$$

Segmentation accuracy. Dice index [17] is a metric to measure similarity between two sets. The range of Dice index is between 0 and 1, where the higher the value, the better the segmentation result. However, in this work, object-level Dice index has been used to evaluate the Dice index for individual glands. The definition of object-level Dice index is as follows [18]:

$$Dice_{\text{obj}}(\boldsymbol{g}, \boldsymbol{s}) = \frac{1}{2}\left[\sum_{i=1}^{n_g} \omega_i Dice\left(G_i, \boldsymbol{S}_*(G_i)\right) + \sum_{j=1}^{n_s} \widetilde{\omega_j}\, Dice\left(\mathbf{G}_*(S_j), S_j\right)\right] \tag{4}$$

where

$$\omega_i = \left|G_i\right| \Big/ \sum_{p=1}^{n_g} \left|G_p\right|,\quad \widetilde{\omega_j} = \left|S_j\right| \Big/ \sum_{p=1}^{n_s} \left|S_p\right| \tag{5}$$

n_g and n_s are the total number of ground truth objects and segmented objects. $Dice\left(G_i, \boldsymbol{S}_*(G_i)\right)$ is to estimate the overlapping area between ground truth and corresponding segmented objects, and $Dice\left(\mathbf{G}_*(S_j), S_j\right)$ is to evaluate the overlapping area between segmented objects and corresponding ground truth.

Shape similarity. Hausdorff distance [19] is used to estimate the shape similarity between segmented object and corresponding ground truth. Object-level Hausdorff distance is to measure the shape similarity of individual gland objects for gland instance segmentation, and it is defined as [18]:

$$\mathbf{H}_{\text{obj}}(\boldsymbol{g}, \boldsymbol{s}) = \frac{1}{2}\left[\sum_{i=1}^{n_g} \omega_i \mathbf{H}\left(G_i, \boldsymbol{S}_*(G_i)\right) + \sum_{j=1}^{n_s} \widetilde{\omega_j}\, \mathbf{H}\left(\boldsymbol{G}_*(S_j), S_j\right)\right] \tag{6}$$

Boundary Jaccard index [20] is another measure used to estimate the similarity between contours. It is sensitive to the infra-segmentation and over-segmentation, but contrary

to the Hausdorff distance, this measure is not sensitive to boundary outliers and its value is bounded between 0 and 1. Object-level boundary Jaccard index is to evaluate boundary Jaccard index for individual glands, and is defined as:

$$\mathbf{BJ}_{\text{obj}}(\boldsymbol{g},\boldsymbol{s}) = \frac{1}{2}[\sum_{i=1}^{n_g} \omega_i \mathbf{BJ}\big(G_i, \boldsymbol{S}_*(G_i)\big) + \sum_{j=1}^{n_s} \widetilde{\omega_j}\, \mathbf{BJ}\big(\boldsymbol{G}_*(S_j), S_j\big)] \tag{7}$$

where

$$\boldsymbol{BJ} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP+FN}} \tag{8}$$

The first term measure Boundary Jaccard index between the ground truth and corresponding segmented results, and the second term measures Boundary Jaccard index between the segmentation objects and corresponding ground truth. $\boldsymbol{BJ}$ is the Boundary Jaccard index, see [20] for details.

5 Results

5.1 Results for image-level classification

The results of image-level classification problem are shown in Table 1. **TP** refers to identifying correctly images with benign tissue, **FP** refers to images with malignant tissue predicted as benign tissue; **FN** refers to images with benign tissue predicted as malignant tissue, and **TN** refers to correctly identifying images showing malignant tissue.

Table 1. Results for image-level classification using three different networks on test images

Network name	**TP**	**FP**	**FN**	**TN**
AlexNet	23	14	13	30
GoogleNet	34	3	8	35
ResNet-50	37	0	0	43

Image-level classification results show that the more up-to-date deep learning techniques are better in separating images in the gland dataset. Using the proposed data augmentation methods, the ResNet-50 performed with 100% classification accuracy on the available test data.

5.2 Results for pixel-level classification

Tables 2, 3 and 4 show the quantitative results for pixel level classification using three different segmentation architectures for benign and malignant cases as well as the overall results. The number in bold in each column in Tables 2, 3 and 4 presents the best results using a corresponding evaluation measure. The best overall performance is achieved using Method 3 with two target classes. For the benign case, the best results were obtained with the LeNet-5 deep feature, whereas for the malignant and overall results the GoogleNet deep features turned out to perform the best.

Table 2. Segmentation results for benign cases with the LeNet-5 deep features

Method name	Number of target classes	F1 score	Object-level Dice index	Object-level Hausdorff distance	Object-level Boundary Jaccard index
Method 1	2	0.63	0.68	194.16	0.69
Method 2	2	0.55	0.64	126.34	0.66
Method 3	2	0.62	0.70	**105.67**	0.72
Method 2	3	**0.74**	**0.77**	108.18	**0.78**
Method 3	3	0.70	0.70	128.95	0.72

Table 3. Segmentation results for malignant cases with the GoogleNet deep features

Method name	Number of target classes	F1 score	Object-level Dice index	Object-level Hausdorff distance	Object-level Boundary Jaccard Index
Method 1	2	0.64	0.67	210.65	0.68
Method 2	2	0.60	0.62	215.46	0.64
Method 3	2	**0.61**	**0.70**	**164.69**	**0.71**
Method 2	3	0.57	0.56	254.08	0.57
Method 3	3	0.52	0.53	230.68	0.55

Table 4. Overall Segmentation results with the GoogleNet features

Method name	Number of target classes	F1 score	Object-level Dice index	Object-level Hausdorff distance	Object-level Boundary Jaccard index
Method 1	2	0.67	0.70	150.38	0.71
Method 2	2	0.63	0.65	173.76	0.67
Method 3	2	0.66	**0.73**	**136.49**	**0.75**
Method 2	3	**0.68**	0.69	202.48	0.70
Method 3	3	0.61	0.67	215.57	0.68

For the benign tissue, three-target class model provides better performance than using two-target classes. This is because the morphological structure of benign tissue is better described by a three-class target model. However, the morphological structure of malignant tissue, it is better represented by two-target class model. Fig. 4 shows a sample of qualitative segmentation results for pixel-level classification. From visual inspection, the segmentation results for benign cases are better than those for malignant cases.

5.3 Methods comparison

As already mentioned in the related work section, a number of different segmentation techniques have been proposed for gland segmentation. For example, segmentation results reported for the fully convolutional neural networks are particularly impressive [6, 21, 22]. However, interpretation and understanding of these state of the art results is somewhat difficult. Although the overall performance of these networks is very good it is not easy to associate this with any particular characteristics of the images or indeed specific parts of the network. The work reported in this paper has been focused on using random forests, as these techniques facilitate: a simple approach for using different features, changes in number of target classes, and adjustment of the classifier to work in a

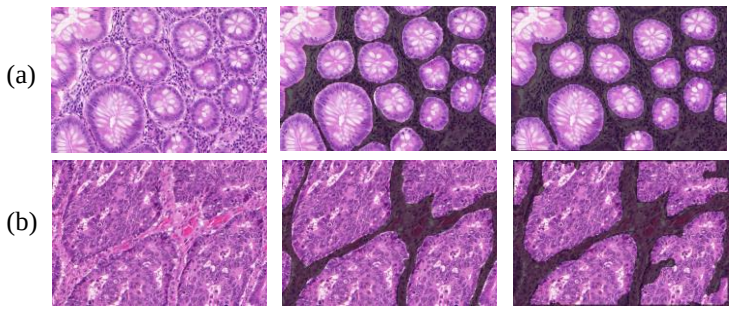

Fig 4. A sample of the segmentation results using Method 3 with 2 target classes. (a) From left to right: test image showing benign tissue, the ground truth (gland part is highlighted) and the segmentation result. (b) From left to right: test image showing malignant tissue, ground truth and segmentation result.

semi-supervised fashion to ease the burden for manual segmentation when much large datasets are available. Furthermore, operation of the classifier could be better understood, by using standard random forest analysis techniques, e.g to find the important features, describe features interactions, or indeed relate specific features to the quality of the segmentation results.

Random forest have been used for gland segmentation before. A recent work on applying random forest for gland segmentation has been reported in [23]. Table 5 provides the quantitative comparison between results reported in [23] and the results ofreported in this paper (Method 3). It is evident that the proposed method achieves a better segmentation performance

Table 5. Comparison results for the methods using random forest

Method name	F1 score		Object-level Dice index		Object-level Hausdorff distance	
	A	B	A	B	A	B
Proposed method	**0.66**	**0.68**	**0.75**	**0.68**	**107**	**223**
Method using random forest [23]	0.54	0.52	0.65	0.57	126	262

6 Conclusion

The paper describes three methods developed for gland segmentation in histology images. The proposed methods have been assessed using number of different metrics evaluating detection accuracy as well as region and contour segmentation accuracy. Two of the proposed methods use image pre-classification assigning each image to two possible categories: benign and malignant. The adopted image pre-classification method together with the training data augmentation achieves 100% classification accuracy on the available test data. Overall, the best results are obtained based on using segmentation with pre-classification at the feature level. This outperforms the method with the pre-classification at the image level as the former enables more data for the feature

extraction training, what is particularly important in cases of limited training data availability. Furthermore, it has been shown that the best results for segmentation of benign glands are obtained with three-class setting, whereas the malignant glands when two-class setting is used. This could be explained by noticing that for majority of malignant glands they lack a distinctive “inside” pattern.

References

1. https://warwick.ac.uk/fac/sci/dcs/research/tia/glascontest/download/, last accessed 2019/06/15
2. Sirinukunwattana, K., Snead, D.R. and Rajpoot, N.M.: A stochastic polygons model for glandular structures in colon histology images. IEEE transactions on medical imaging, 34(11), 2366-2378 (2015)
3. Wu, H.S, Xu, R., Harpaz, N., Burstein, D., Gil, J.: Segmentation of intestinal gland images with iterative growing. Journal of Microscopy, 220(3), 190-204 (2005).
4. Gunduz-Demir, C., Kandemir, M., Tosun, A.B., Sokmensuer, C.: Automatic segmentation of colon glands using object-graphs. Medical image analysis, 14(1), 1-12 (2010).
5. Kainz, P., Pfeiffer, M., Urschler, M.: Semantic segmentation of colon glands with deep convolutional networks and total variation segmentation. arXiv preprint arXiv:1511.06919 (2015).
6. Chen, H., Qi, X., Yu, L., Heng, P. A.: DCAN: deep contour-aware networks for accurate gland segmentation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2487-2496 (2016).
7. Li, W., Manivannan, S., Akbar, S., Zhang, J., Trucco, E., McKenna, S.J.: Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 1405-1408 (2016)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097-1105 (2012)
9. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9 (2015).
10. He, K., Zhang, X., Ren, S., Sun, J: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778 (2016).
11. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on pattern analysis and machine intelligence, 11(6). 567-585 (1989)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
13. Xiaoling, W.: A novel circular ring histogram for content-based image retrieval. 2009 First International Workshop on Education Technology and Computer Science, 2, 785-788 (2009)
14. Zhao, G., Ahonen, T., Matas, J. and Pietikainen, M.: Rotation-invariant image and video description with local binary pattern features. IEEE transactions on image processing, 21(4), 1465-1477 (2011).
15. Skibbe, H. and Reisert, M: Circular Fourier-HOG features for rotation invariant object detection in biomedical images. ISBI, 450-453 (2012).
16. Breiman, L.: Classification and Regression Trees. Routledge. (2017).
17. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology, 26(3), 297-302 (1945).

18. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., Böhm, A: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis, 35, 489-502 (2017).
19. Beachemin, M., Thomson, K.P.B., Edwards, G.: On the Hausdorff distance used for evaluation of segmentation results. Can. J. Remote Sens. 24(1), 3–8 (1998)
20. Fernandez-Moral, E., Martins, R., Wolf, D., Rives, P.: A new metric for evaluating semantic segmentation: leveraging global and contour accuracy. 2018 IEEE Intelligent Vehicles Symposium (IV), 1051-1056 (2018).
21. Kainz, P., Pfeiffer, M., Urschler, M: Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. arXiv preprint arXiv:1511.06919. (2015).
22. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Snead, D., Tsang, Y.W., Rajpoot, N.: MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. Medical image analysis, 52. 199-211 (2019).
23. AP, R., Khan, S.S., Anubhav, K., Paul, A.: Gland Segmentation in Histopathology Images Using Random Forest Guided Boundary Construction. arXiv preprint arXiv:1705.04924 (2017).