Method for automatic initialization of trainable active contours for instance segmentation in histological images

Nikita Yakovlev[1], Alexander Khvostikov[1] and Andrey Krylov[1]

[1]*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskie Gory, 1, building 52, Moscow, 119991, Russia*

Abstract

The method of trainable active contour is one of the semi-automatic segmentation methods that can be applied to segment glands in histological images. In this paper, we propose a method for automatic initialization of trainable active contour model, which makes the segmentation method fully automatic. Using a U-Net like architecture, a preprocessed image segmentation masks is predicted for the input image, from which initial approximations of contours are calculated for each gland. The proposed method correctly marks 96.2 % part for the glands on the test set of the PATH-DT-MSU S1-v2 dataset. As a result, we get initial approximations located inside each gland in the image.

Keywords

Glands segmentation, active contours, convolution neural networks, histological images, instance segmentation.

1. Introduction

Today, the analysis of scans of human tissue helps histologists make more accurate diagnoses when examining patients. Glands are a biological structure, by the appearance of which one can draw conclusions about the condition of the organ from which a tissue sample was taken for the study. Therefore, one of the important tasks is to automatically distinguish (segment) these biological structures.

One of the ways to segment glands is to use the active contour method [1], which is a variational method. The active contour is a parametrically defined curve, which changes its position as a result of minimization of the energy functional of this curve, introduced in such a way that its minimum corresponds to the boundary of the object. The functional takes into account the boundaries smoothness and shape of the object being segmented. This is especially useful when segmenting on glands that are often elliptical in shape. Also, an important advantage of the active contours method is that we get an instance segmentation. The main disadvantage of this method is the need for a predefined functional, which needs to be defined separately for each object and image.

This problem can be solved by using a trainable active contours model [2], where the function is determined independently at each point (pixel) of an image using convolutional neural network (CNN). Good segmentation accuracy on the PATH-DT-MSU S1-v1 dataset has been achieved in this paper.

Both of the mentioned above active contours methods are semi-automatic. They require an initial approximation of the contour (see Fig. 1), which will change its position during functional minimization. In most cases, the best results are achieved when the initialization of the contour is marked near the centre of the gland. The quality of the final segmentation depends on its initial position.

Automatic initialization of active contours for glands segmentation requires a separate study, which is the main motivation for the current work.
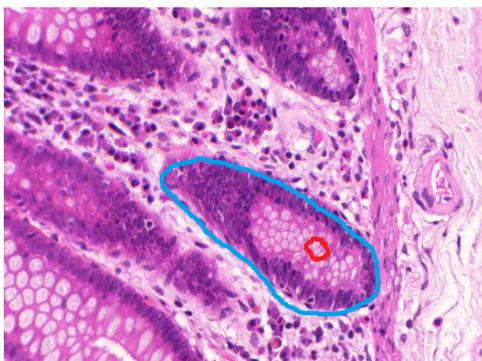
Figure 1: Example of a histological image and segmentation of a single gland using the trainable active contour model. The red contour is the initial approximation, the blue contour is the prediction of the trainable contour model.

One of the options for generating starting points for the method of active contours in the image is the use of neural network SeedNet [3]. The task is formulated as a sequential determination of the starting point position and its type (background or object) using reinforcement learning. In order to speed up and improve the quality of learning, the randomness of the choice is limited to a uniform grid of 20 different positions on the vertical and horizontal axis. The input of the neural network consists of the initial image and several initial points with their type. The output of the neural network is a matrix of size 20×20, each cell of which can contain a point of one of two types: background or object.

In each iteration of the training process the point and its type are selected by the neural network, then the image is segmented using the points set earlier and depending on the quality of the segmentation the algorithm is rewarded or penalized. During the testing phase for more accurate segmentation, an input mask can be fed with the annotated points of the background and/or object.

Another method is proposed in [4]. This method is a combination of classical image processing method with watershed lines, which has a physical justification: segmentation is performed considering that a single-channel image can be represented as a topographic surface, which is filled starting from local minima. Watershed lines are drawn in such a way that the filled water

bodies do not overlap.

As the output of the neural network, the authors propose to obtain a single-channel image, which is a prediction of the class distance of the object pixel from the nearest pixel of the background. The boundaries of a pixel's distance class assignment are common to all objects in the image, which can play a role when working with multi-scale objects.

Another way of defining a starting point was proposed in Object as Points [5]. The authors solved the problem of defining a bounding box for the object. They define the bounding box through the centre point, width and height of the rectangle in which the entire object would be contained. These parameters are determined by a neural network. The loss function takes into account the proximity of the centre point and the sides of the bounding box to the original annotation (ground truth).

2. The proposed method

In this paper, we propose an automatic method of starting point initialization for the active contour method using CNN . Since the segmentation masks of nearby glands very often merge, we proposed to modify the segmentation mask, which takes into account the distance from a pixel inside the gland to the nearest background pixel, and predict it using CNN U-Net. Initial contour approximations are placed in the mask regions, which should correspond to the gland centers. The proposed method correctly marked 96.2% of the glands.

2.1. Description of existing data

In this work we use PATH-DT-MSU S1-v2[1] dataset [6], which contains colon section images. There is an instance segmentation mask for each image in the dataset, i.e., each gland is separately identified. It is worth noting that there are malignant and benign adenocarcinoma in dataset. The former are characterised by a more complex, non-ellipsoidal shape compared to the latter.

It also should be noticed that in PATH-DT-MSU S1-v2 dataset the authors introduce a separate class for open glands. This creates difficulties in processing of such images because the degree of difference in annotation of the same image by different specialists increases.

It should be noted that the images often contain areas that look very similar to glands, but in truth are not. This is most clearly seen in Figure 2.

In Table 1 the different characteristics of the dataset are described. It is easy to see that the images in the dataset have different resolutions.

The test set of the PATH-DT-MSU S1-v2 dataset consists of a total of 40 images on which 1578 glands are annotated, the train set consists of a total of 40 images on which 1362 glands are annotated.

2.2. Preprocessing of image segmentation masks

The merging of closely spaced objects is one of the problems often encountered in segmentation. We propose to solve this problem by preprocessing of the mask.

---

[1]https://imaging.cs.msu.ru/en/research/histology/path-dt-msu
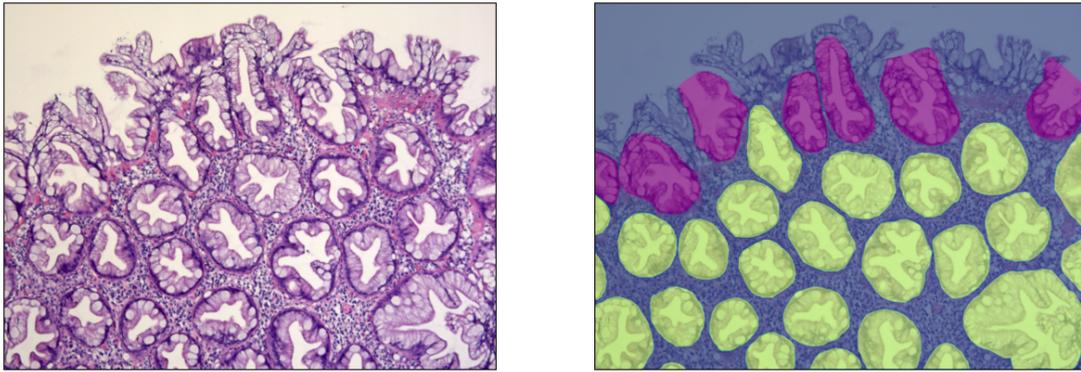
Figure 2: Example of an original image (benign adenocarcinoma) and the same image with superimposed segmentation mask from PATH-DT-MSU S1-v2, glands are marked yellow and open glands are marked in purple.

Table 1

Key characteristics of the used dataset.

Dataset	Training set number of images [resolution, px]		Test set number of images [resolution, px]		Optical magnification
PATH-DT-MSU [6]	40	$\begin{cases} 27 & [1632 \times 2176] \\ 13 & [2442 \times 3263] \end{cases}$	40	$\begin{cases} 31 & [1632 \times 2176] \\ 9 & [2442 \times 3263] \end{cases}$	10 x

Let $I \in \mathbb{R}^{W \times H \times 3}$ - original image, $M \in \mathbb{R}^{W \times H \times 1}$ - preprocessed mask, $\widetilde{M} \in \mathbb{R}^{W \times H \times 1}$ - predicted by used CNN (defined in the following part) mask, where W, H - image width and height.

An euclidean distance transform was applied to the mask of each gland (calculate the distance from the gland pixel to the nearest background pixel), then we normalize the values obtained so that all values are between 0 and 1. To add contrast to the borders, all non-zero values were increased by one, i.e. received a mask where the background corresponds to the value 0, at the border of each gland mask value is 1, at the point of the gland furthest from the background pixels mask value is 2. The result is map M of normalized distances to the nearest background pixel. The normalization was performed for each gland separately. That is, for both small and large glands (see Figure 3), the proposed mask at the gland borders will have a value of 1, at the point of the gland furthest from the background the proposed mask value will be equal to 2.

The values of intensity of preprocessed masks M change continuously in each gland (see Figure 3).

It should be noted that the centre of the object is of particular interest. This is due to the fact that we will achieve good segmentation results using the active contours method if we initialise the initial contour approximately in the centre of the object.

This mask preprocessing will help to separate merged gland mask predictions, since in these cases it will be possible to initialise the starting points at local maxima.
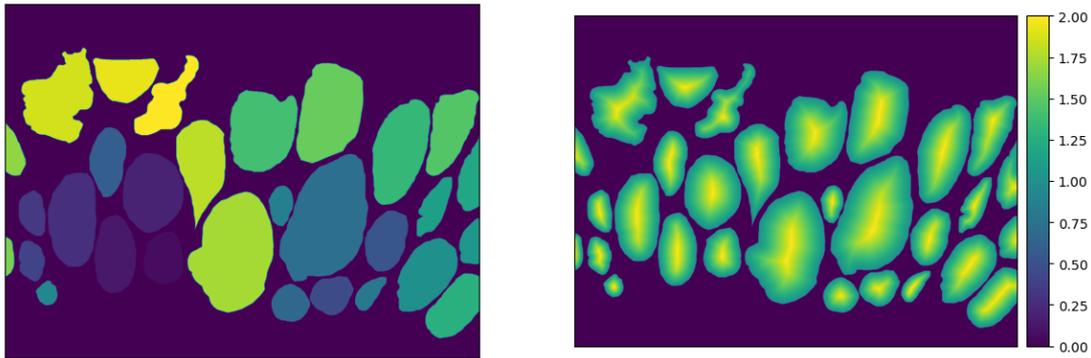
Figure 3: Example of a proposed mask change on an image from PATH-DT-MSU S1-v2. On the left is the object segmentation, on the right the proposed mask preprocessing.

2.3. The architecture of the neural network

A U-Net [7] like architecture was used to process the images *I* and predict the proposed mask *M* for each image. Six pooling layers were used in order to get higher level features (see Fig. 4). Fewer pooling layers gave worse results. The number of neurons in the first layer is 8, then on each successive layer it is doubled (up to 512), as in the original U-Net. This choice of number of parameters allowed finding a balance between performance, available computational power and network capacity. The neural network architecture has a total of 7.79 million parameters, of which 7088 are non-trainable (the variance and mean of the batch normalization layers).

All blocks have sequential convolution, batch normalization, ReLU activation function, convolution, batch normalization, ReLU activation function. The window size at convolution is 3x3. The mean value after batch normalization is 0, the variance of the values is 1. Center block does not have any additional layers. Contractive block last layer has MaxPooling layer with window size 2x2. Expansive block first layer has a transposed convolution with stride 2 (i.e. the input tensor size is doubled in two dimensions).

In the in the original paper, U-Net was used to predict the probability of assigning a pixel to a class, which was the reason for using a probability-predicting activation function (such as *sigmoid* or *softmax*) in the last layer.

Since it was necessary to train the neural network to predict the distance to the nearest background pixel, we used the same ReLU activation function in all convolutions before.

$$ReLU(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } x >= 0, \end{cases}$$

where $x \in \mathbb{R}$ - value coming into the activation function.

The loss function is the Mean Square Error (MSE or L_2 norm):

$$MSE = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} (M_{ij} - \widetilde{M}_{ij})^2$$
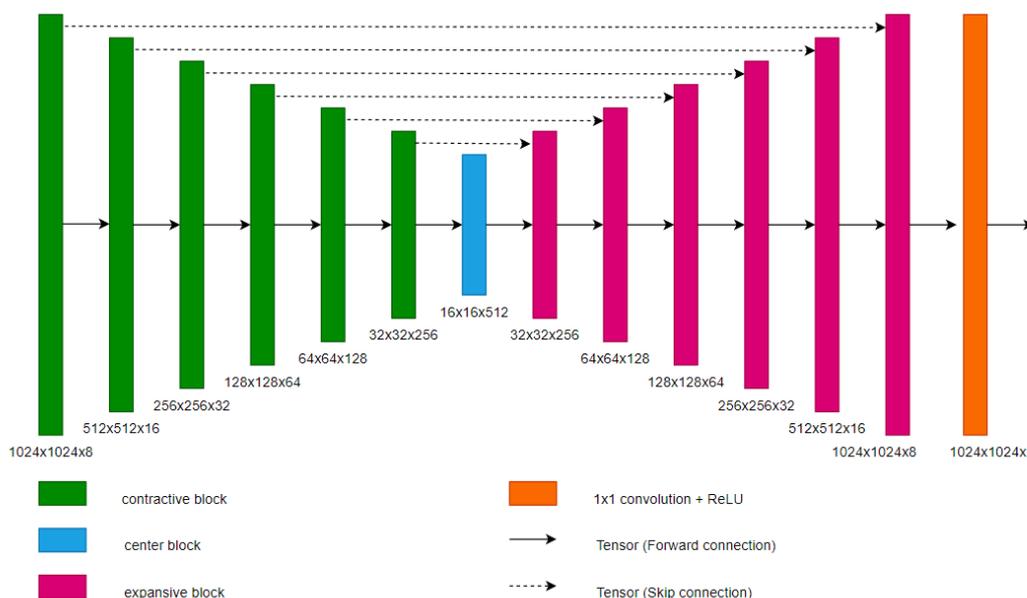
Figure 4: Architecture of the used CNN for distance map prediction ($\widetilde{M}$) from original image (I). The tensor dimensions of each block are indicated underneath the blocks.

where W is image width, H is image height, M - preprocessed mask, $\widetilde{M}$ - predicted by used CNN mask.

2.4. Post-processing the image segmentation masks

A distance map $\widetilde{M}$ obtained using with the proposed neural network, requires further processing. The following algorithm was proposed and applied, because the maximum value of $\widetilde{M}$ should correspond to the furthest point of the gland from the background:

1. All prediction values in $\widetilde{M}$ less than the threshold τ were equated to zero (see fig. 5). τ was chosen in the range between 1 and 2, so that the value corresponds to the areas within the gland (best results were obtained with $\tau = 1.5$). The threshold was adjusted using Hyperopt [8], in the range of 1 to 2.

2. For each connected component slice with the τ prediction $\widetilde{M}$ of the trained neural network, the coordinates of the maximum value was found. Its neighbourhood will be the initial approximation of the contour (see yellow dots on second image in fig. 5). The connectivity components that contain less than 25 pixels are skipped, as they are likely to be an incorrect prediction, since in practice the number of pixels covering every gland is empirically greater than 25.
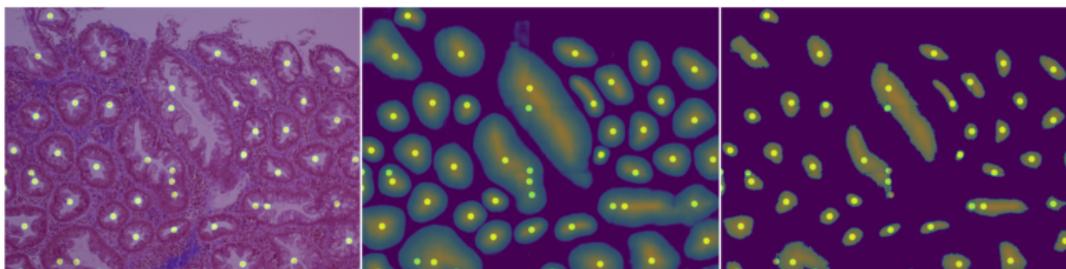
Figure 5: Left to right: the original image (I) of the PATH-DT-MSU S1-v2 test sample, the neural network prediction $\widetilde{M}$, the slice with a 1.5 prediction threshold of the trained neural network. On top of all the images predicted start points are shown in overlay.

3. Experiments

3.1. Neural network training

Due to the variability in resolution of the input images, we trained a neural network on a $1024 \times 1024 \times 3$ patches. The values of each channel were initially scaled into a range [0, 1].

The training set was augmented in the following way. Patches were generated with random rotation up to 45° (if the selected patches fit in the image), random rotations to multiple angles of 90°, reflections around the vertical and horizontal axes, random changes in brightness and contrast parameters, random patch cropping of scale from 0.8 to 1.5 expected and then interpolating it to the desired resolution.

For training we used 36 images from the training set, the 4 remaining images were used for validation.

In this work we used Adam [9] optimiser with standard moment parameters. The initial learning rate (LR) was selected using Hyperopt library [8], using bounds from 5×10^{-5} to 5×10^{-3}. These bounds were chosen because of the assumption that for most problems a initial learning rate between 1×10^{-4} and 1×10^{-3} is appropriate. The selection method used was bayesian optimization, which chooses a balance between finding the best parameter and exploring an as yet untested interval of the parameter to be selected. The best training result was obtained when LR = 10^{-3}.

The number of generated patches per train epoch was chosen so that the number of pixels was 20 times larger than the entire training portion of the sample. Using this number of patches per epoch improved the training results compared to using more epochs when the neural network was fed with the number of patches that matched the number of pixels of the training sample per training epoch. This may be due to the properties of the Adam optimizer, which considers previous epoch steps.

The learning rate was multiplied by 0.3 every five epochs if the resulting minimum loss function value on validation did not change. Also, after the fifth epoch, the learning rate was reduced by 5 percent at each iteration. The batch size was chosen to be 16, due to the limitation of used hardware.

3.2. Neural network training results

Plots of MSE loss function value during training (i.e. used loss function) on PATH-DT-MSU S1-v2 training set and learning rate can be seen in Figure 6. It can be seen that in this case the loss function on validation set has already plateaued after the 20 epoch. The mean squared error on the test set of the PATH-DT-MSU S1-v2 was 0.075, which is approximately equal to the same metric on train, which is 0.069. So, there is no strong overfitting on the training set.
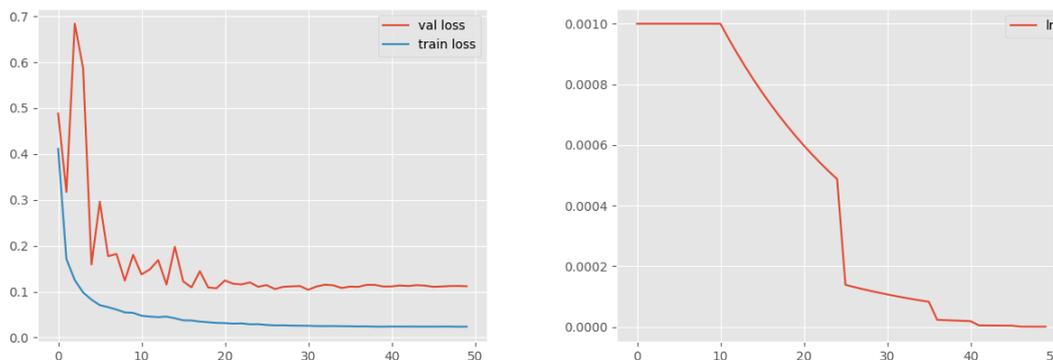
Figure 6: Loss function (MSE) and learning rate (LR) over epoch during training on PATH-DT-MSU S1-v2.

3.3. Post-processing the obtained masks

The test sample of the PATH-DT-MSU S1-v2 dataset consists of a total of 40 images on which 1578 glands are annotated. A total of 2191 starting points was placed on the PATH-DT-MSU S1-v2 test set. The algorithm correctly predicted started point for 1519 glands (0.962 part of the glands) and 2138 starting points fallen into the segmentation mask of any gland (0.975 part of the marks) (i.e. fallen into any gland). Note that multiple start points on the same gland were not considered as an error, since for the active contour method several initial approximations can be specified to further obtain a more accurate estimation of the segmented object.

The average size of uninitialised glands in the image is 11415 pixels and the average size of initialised glands is 115210 pixels. The difference in average size is 10 times.

The smallest image from the dataset has size 1632×2176, i.e. contains 3551232 pixels. On average, the glands in the image covers 3.2 % of the image area. The glands on which the proposed method is wrong cover 0.32 % of the image.

When reviewing the errors, it was found that they were mainly made in the following cases:

1. The gland is located near the border of the image and the algorithm does not mark it (Figure 7);

2. The gland is located on the edge of the open glands and is not marked by the algorithm (Figure 8);
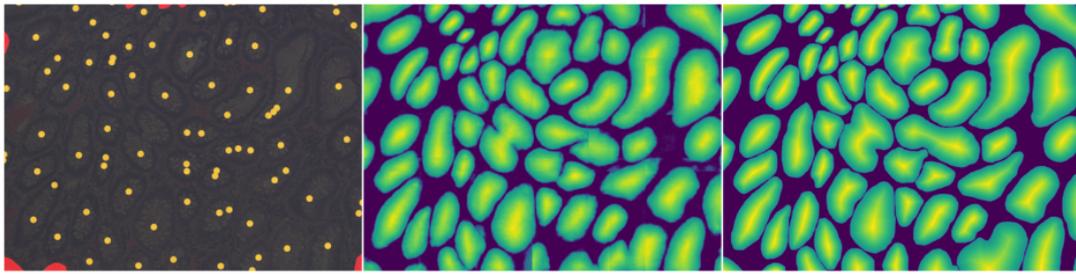
Figure 7: Left to right: the image from the test set of the PATH-DT-MSU S1-v2, with the uninitialized glands (i.e. the predicted contours didn't hit the gland) at the border of the image highlighted in red (all initialized contours highlighted in yellow), the prediction of the trained network from the image $\widetilde{M}$, ground truth M for the image I. 7 glands located near the border of the images are uninitialized.
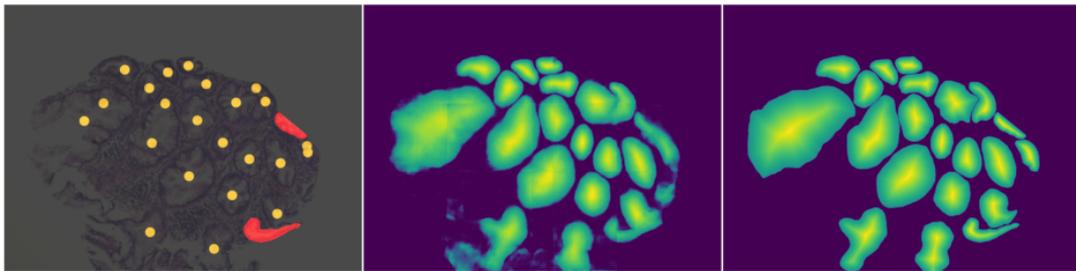
Figure 8: Left to right: the image from the test set of the PATH-DT-MSU S1-v2, with the uninitialized glands (i.e. the predicted contours didn't hit the gland) at the slice border of the image highlighted in red (all initialized contours highlighted in yellow), the prediction of the trained network from the image $\widetilde{M}$, ground truth M for the image I. 2 glands located near the border of the histological slice are uninitialized.

3. Small glands (Fig. 9);
4. Complex image (Fig. 10).

In most cases, such errors are acceptable because the trained active contour method is not supposed to work with glands that are too small and are located near the borders of the image.

4. Conclusion

In this paper we proposed a new method for automatic initialization of active contour model, that can be used to segment individual glands in histological images.

The methods was implemented using TensorFlow framework [10] and Keras library [11] in Python3. The library Albumentation [12] was used to artificially expand the training sample.

The achived value of MSE during training is 0.075 in test set of PATH-DT-MSU S1-v2. When arranging the starting points, 1519 out of 1578 the glands (0.962 part of the gland) were correctly
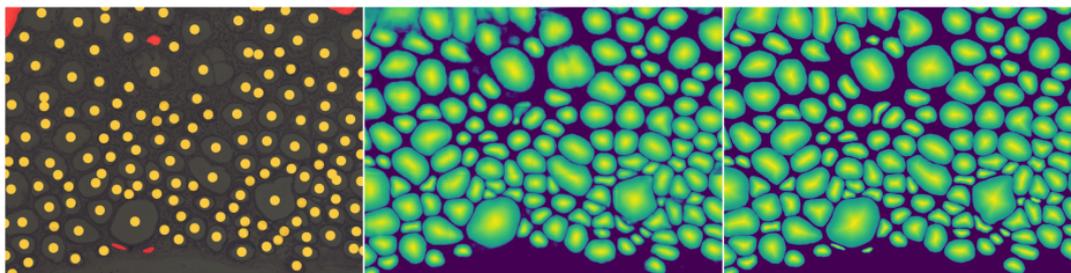
Figure 9: Left to right: the original image with small uninitialized glands (i.e. the predicted contours didn't hit the gland) from the PATH-DT-MSU S1-v2 test set (uninitialized glands are highlighted in red), the prediction of the trained network from the image $\widetilde{M}$, ground truth M for the image I. 5 small glands uninitialized.
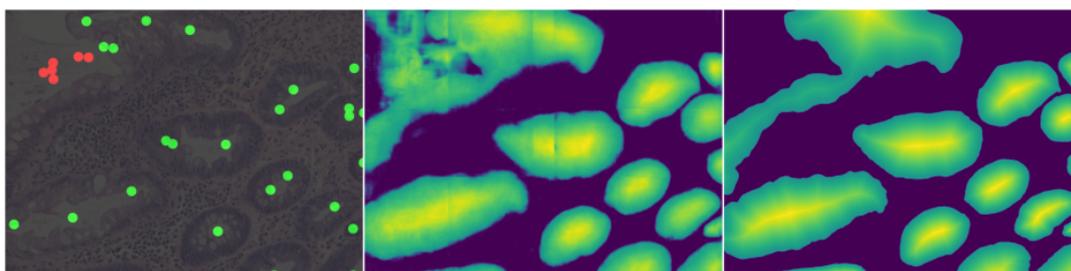
Figure 10: Left to right: the original, difficult to process image from the test set of PATH-DT-MSU S1-v2, with incorrect starting points (correctly initialized starting contours highlighted in green) highlighted in red, the prediction of the trained network by the image $\widetilde{M}$, ground truth M for the image I. 6 starting points are incorrectly set.

marked and 2138 starting points out of 2191 (0.975 part of the predicted points) predicted points hit the glands on the same dataset.

5. Acknowledgments

References

[1] R. Deriche, P. Kornprobst, G. Aubert, Snakes: Active contour models, International Journal of Computer Vision (1987) 321–331. doi:`10.1007/BF00133570`.

[2] I. M. P. M. A.V. Khvostikov, A.S. Krylov, Trainable active contour model for histological image segmentation, Scientific Visualization 11 (2019) 80–91. doi:`10.26583/sv.11.3.06`.

[3] K. M. Lee, H. Myeong, G. Song, Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation (2018) 1760–1768. doi:`10.1109/CVPR.2018.00189`.

[4] M. Bai, R. Urtasun, Deep watershed transform for instance segmentation (2017) 2858–2866. doi:`10.1109/CVPR.2017.305`.

[5] X. Zhou, D. Wang, P. Krähenbühl, Objects as points, arXiv (2019). `arXiv:1904.07850`.

[6] A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Oleynikova, P. G. Mal'kov, Automatic mucous glands segmentation in histological images, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W12 (2019) 103–109. doi:`10.5194/isprs-archives-XLII-2-W12-103-2019`.

[7] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation (2015) 234–241. doi:`10.48550/arXiv.1505.04597`.

[8] J. Bergstra, D. Yamins, D. D. Cox, Making a science of model search (2012). `arXiv:1209.5111`.

[9] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization (2017). `arXiv:1412.6980`.

[10] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/.

[11] F. Chollet, keras, https://github.com/fchollet/keras, 2015.

[12] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Albumentations: Fast and flexible image augmentations, Information 11 (2020) 125–132. doi:`10.3390/info11020125`.