COMBINING FULLY CONVOLUTIONAL NETWORKS AND GRAPH-BASED APPROACH FOR AUTOMATED SEGMENTATION OF CERVICAL CELL NUCLEI

Ling Zhang[1]*, Milan Sonka*[2]*, Le Lu*[1]*, Ronald M. Summers*[1]*, Jianhua Yao*[1]

[1]Radiology and Imaging Sciences Department, National Institutes of Health (NIH), Bethesda MD
[2]Iowa Institute for Biomedical Imaging and Department of Electrical and Computer Engineering, The University of Iowa, Iowa City IA

ABSTRACT

Cervical nuclei carry substantial diagnostic information for cervical cancer. Therefore, in automation-assisted reading of cervical cytology, automated and accurate segmentation of nuclei is essential. This paper proposes a novel approach for segmentation of cervical nuclei that combines fully convolutional networks (FCN) and graph-based approach (FCN-G). FCN is trained to learn the nucleus high-level features to generate a nucleus label mask and a nucleus probabilistic map. The mask is used to construct a graph by image transforming. The map is formulated into the graph cost function in addition to the properties of the nucleus border and nucleus region. The prior constraints regarding the context of nucleus-cytoplasm position are also utilized to modify the local cost functions. The globally optimal path in the constructed graph is identified by dynamic programming. Validation of our method was performed on cell nuclei from Herlev Pap smear dataset. Our method shows a Zijdenbos similarity index (ZSI) of 0.92 $\pm$ 0.09, compared to the best state-of-the-art approach of 0.89 $\pm$ 0.15. The nucleus areas measured by our method correlated strongly with the independent standard (r^2 = 0.91).

***Index Terms*—** Deep learning, FCN, graph-based segmentation, Pap smear

1. INTRODUCTION

In cervical cytology diagnosis, nuclear features already include substantial diagnostic information, as all the cytology abnormalities (including atypical squamous cells of undetermined significance [ASC-US], ASC-cannot exclude HSIL [ASC-H], low grade squamous intraepithelial lesion [LSIL], high SIL [HSIL], and squamous cell carcinoma) accompany specific nuclear abnormality [1]. In order to accurately characterize nuclear features, accurate segmentation of nuclei is a necessary step, and is also of utmost importance in automation-assisted cervical screening techniques [2].

This research was supported by the Intramural Research Program of the National Institutes of Health Clinical Center. The authors thank Nvidia for the TITAN Z GPU donation.

Generally, methods for the automated segmentation of cervical nuclei includes two components: initial segmentation and fine segmentation. For initial segmentation, K-means (spatial [3] or fuzzy [4] version), multi-scale watershed [5], and adaptive thresholding [2] are proposed. Besides coarse performance, these methods are based on simple assumption about the nucleus characters (e.g., low intensity, circular borders), which may not be true in all cases (e.g., some nuclei exhibit high intensity and/or irregular shape). Therefore, pre- and post-analysis stage are needed. Recently, a convolutional neural network (CNN) image patch-based classification [6] method is proposed to detect nuclei. This method is able to overcome the aforementioned problems, but does not incorporate spatially dense information. Moreover, the heavy computational burden makes it not preferable in practical use.

For fine segmentation of the nuclei, snake models (specifically a radiating gradient vector flow (RGVF) snake [3]), and graph cuts [2, 6] are the most commonly used techniques. However, these techniques do not involve inherent shape constraints for the segmented boundary, thus may result in irregular nucleus boundary. The graph-based approach [7] embeds the shape prior in the constructed graph, and the nucleus boundary can be detected as a globally optimal solution. However, the success of graph-based approaches relies on two clues: 1) a relatively reliable initial segmentation for constructing the graph and constraining the graph searching range; 2) a robust and task-specific cost function for describing the object. Previously, the first clue is usually obtained based on simple assumptions [7] or low-level features [8]; the second clue is based on hand-crafted/engineered costs [7] and/or machine learning costs based on low-level features [8].

To improve the aforementioned problems, we propose to segment cervical nuclei by combining fully convolutional networks (FCN) [9] and graph-based approaches (denoted by FCN-G). Different from CNN-based segmentations, FCN takes an arbitrary-sized whole image as input and correspondingly-sized ground truth segmentation as output. By building a fully convolutional version of existing CNNs followed by a deconvolution layer which interpolates the coarse output to generate dense pixel output, FCN is able to

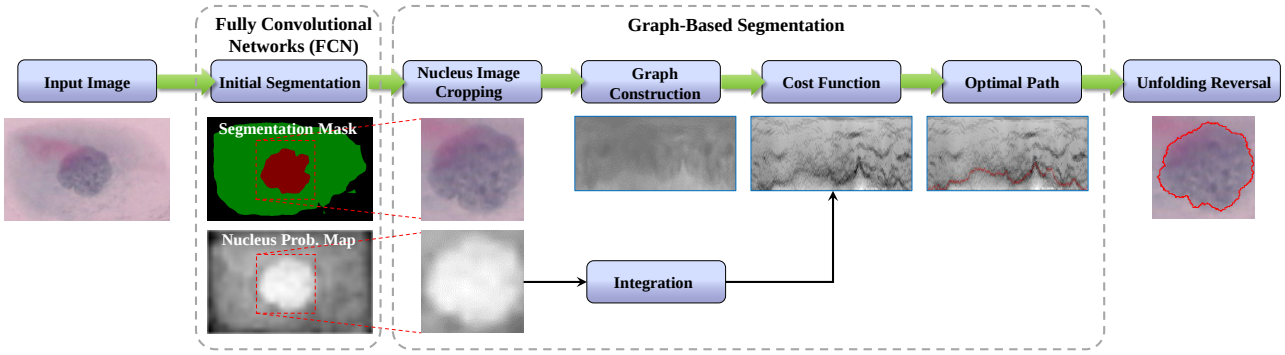

Fig. 1. Framework of combining fully convolutional networks (FCN) and a graph-based approach (denoted by FCN-G).

be trained end-to-end, pixel-to-pixel, and predicts all the pixel labels of a whole image at a time. With these advantages, FCN is efficient, does not require pre- and post-processing, and learns high-level features. For the graph-based segmentation, we employ the basic framework previously proposed in [7], but additionally incorporate FCN-learned nucleus in-region cost and context constraint in a globally optimal solution. Following an initial segmentation of nuclei and cytoplasm by FCN, the graph-based approach yields improved cell nucleus segmentation. We show quantitative comparisons between the FCN-G and the state-of-the-art approaches [3, 5, 4] on the Herlev Pap smear dataset [10].

2. METHODS

Our segmentation framework comprises two stages (as shown in Fig. 1): fully convolutional networks (FCN) initial segmentation and graph-based fine segmentation. (1) FCN is first adopted to segment the whole cell image into background, cytoplasm and nuclei, meanwhile produces a probability map for cell nuclei. FCN allows for capturing implicit and informative deep features of cell nuclei, but usually produce segmentation results that may not precisely localize on the nucleus boundaries. Therefore, (2) the graph-based fine segmentation is exploited, consisting of two sub-steps: (a) a rectangle (sub-image) around the FCN-segmented nucleus is cropped according to an annotation protocol [7] that ensures involving the entire nucleus area; (b) a graph-based approach incorporating the FCN-learned nucleus probability map is used to detect the improved nucleus boundary.

2.1. Fully Convolutional Networks

We adopt FCN [9] to predict the pixel-level labels for background, cytoplasm, and nuclei in cervical cell images. To account for the difference between abnormal and normal nuclei, we separate the nuclei class into abnormal and normal classes, resulting in four classes/labels for FCN training and predicting. The FCN prediction generates binary masks and probabilistic label maps for the four classes. Among them, the nucleus mask is used to crop a sub-image containing nucleus, and the nucleus probability map is used to produce an in-region cost, both are used in the graph-based segmentation.

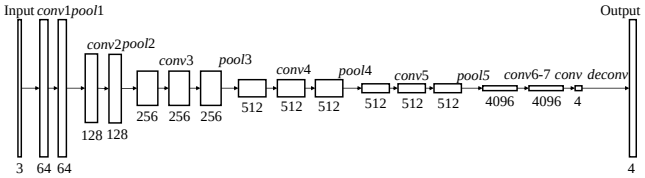

Fig. 2. Schema of the FCN. $conv$: convolution; $pool$: max-pooling; $deconv$: deconvolution.

2.1.1. FCN Training

We adopt the VGGnets architecture [11] as the basis of our FCN. To construct the FCN, VGGnets' last fully connected layer is discarded and the other two fully connected layers are converted to convolution. Then, a 1×1 convolution with 4 channels corresponding to the four cervical cell classes is appended. Finally, a deconvolutional layer is added to upsample the coarse outputs to pixel-level outputs. This architecture is FCN-32s and is shown in Fig. 2.

In our preliminary experiment, we found FCN-16s performing better on our dataset. Specifically, a 1×1 convolutional layer is appended on top of $pool4$ to produce additional predictions, which are then added to a $2\times$ up-sampling version of $conv7$'s predictions. Finally, this sum of predictions is up-sampled back to the image with the 16 pixel stride. For FCN training, transfer learning is used. A pre-trained VGGnets is used to initialize the $conv$ layers and the other layers are initialized with Gaussian distribution. Then all layers of the FCN-16s are fine-tuned by back-propagation algorithm.

2.1.2. FCN Predicting

During the predicting stage, the FCN generates four probability maps, each for one class. The final segmentation mask

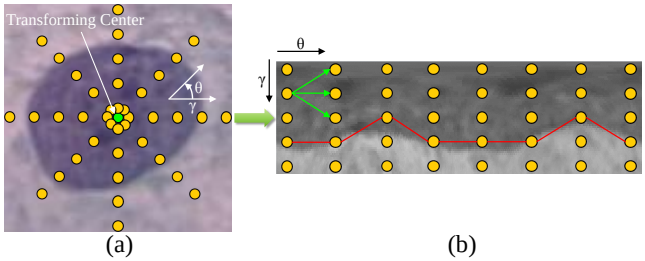

Fig. 3. Schema of construction of a graph in (b) from the image center in (a) using a polar transform.

is obtained by maximum voting of the four probability maps, followed by a combination of abnormal and normal nucleus masks, and then by morphological opening operation. The final nucleus probability map is obtained by summing the abnormal and normal nucleus probability maps.

2.2. Graph-Based Nucleus Segmentation

2.2.1. Graph Construction

Given an cropped image, image transforming is performed by transforming the image coordinates from Cartesian to polar, as shown in Fig. 3. As a result, the ellipse-like border of the nucleus becomes a curve which starts from the first column and ends at the last column in the transformed image. Then a graph with N_g graph columns is constructed and searched for the optimal path, where N_g equals to the number of columns of the transformed image. In this graph, each node (yellow points in Fig. 3(b)) corresponds to a pixel in the transformed image. The successors of a node (pointed by green arrows in Fig. 3(b)) are the three nodes on the subsequent column corresponding to three possible changes of edge direction.

2.2.2. Cost Function Design

The cost assigned to the graph node contains three specific information components related to cervical nuclei.The first two cost terms are detailed in [7]. Briefly, the edge cost c_e is calculated based on the normalized image gradient magnitude and image gradient direction. The region cost c_r is assigned as the sum of the inside and outside variances computed in the graph column.

Given the FCN-learned nucleus probability map, an in-region-based cost is assigned as the cumulative sum of the probability-values p in the graph column as follow,

$$c_{fcn}(i,j) = -\sum_{j'=1}^{j' \le j} p(i,j') \ , \tag{1}$$

Then, the aforementioned three cost terms are combined into a total cost function c to allow more robust segmentation,

$$c(i,j) = \alpha \cdot c_e(i,j) + \beta \cdot c_r(i,j) + \gamma \cdot c_{fcn}(i,j) \ , \tag{2}$$

where α, β and γ are the weights for the edge, region, and FCN term respectively. Each of the three terms is normalized to the range [0,1] before their combination.

2.2.3. Nucleus Context Prior Constraints

The graph-based segmentation might incorrectly identify borders that coincide with large image gradients between cytoplasm and background. To solve this problem, two kinds of constraints including the hard context constraint and the context-prior penalty are used to modify the costs of the graph nodes with Δd pixels above and below the cytoplasm boundary while the costs of all other nodes remain unchanged,

$$c'(i,j) = \begin{cases} f_c(c(i,j)) & if \ -\Delta d \le n(i,j) - b_j \le 0 \\ K & if \ n(i,j) - b_j > 0 \\ c(i,j) & otherwise \end{cases} , \tag{3}$$

where $c(i,j)$ and $c'(i,j)$ denote the costs of node $n(i,j)$ before and after considering the proximity of cytoplasm, respectively. b_j is a cytoplasm boundary point in polar coordinates. The reason for these constraints are: first, the costs of nodes which are immediately above the cytoplasm points need to become less "attractive" by utilizing an update function,

$$f_c(c(i,j)) = c(i,j) \cdot (1 + 0.25e\frac{-d(b_j, n(i,j))^2}{2\Delta d^2}) \ , \tag{4}$$

where $d(.)$ denotes the distance (number of nodes); second, resulting from the hard context constraint, nodes far away from coarsely segmented cytoplasm boundary points become less attractive when modified according to the truncated L_1 distance. K is set to 2.

2.2.4. Boundary Detection

The optimal path in the graph is determined by dynamic programming, and then reversely mapped to the original image in Cartesian coordinate system to get a contour of nucleus.

3. EXPERIMENTS AND RESULTS

The cell data used to train and test the proposed method come from a publicly available dataset (Herlev) [10], which consists of 917 Pap smear images – each containing one cervical cell – with ground truth segmentation and classification. All the cervical images' longer sides are rescaled to a length of 500 pixels, to avoid performance problems on the original resolution we encountered in our preliminary experiments.

The FCN is implemented using MatConvNet package [12], and runs on NVIDIA GeForce GTX TITAN Z GPU with 6 GB of memory. FCN is trained by Stochastic Gradient Descent (SGD), with a mini-batch size of 20, momentum of 0.9, learning rate of 0.0001, and epoch of 50. A dropout of 0.5 is used. The parameters for graph-based approach are set as $Ng = 360$, $\alpha = 1$, $\beta = 0.5$, $\gamma = 0.5$, and $\Delta d = 2$.

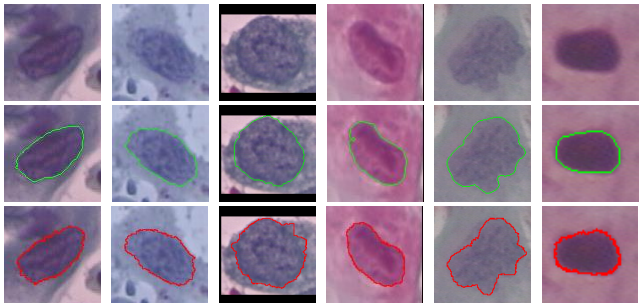

Fig. 4. Examples of our segmentation method on Herlev dataset. First to third rows: original image, ground truth (green boundaries), and our results (red boundaries).

We evaluate our method using five-fold cross validation. The performance evaluation metric is Zijdenbos similarity index (ZSI) ($\frac{2TP}{2TP+FP+FN}$) as in [3, 5, 4]. Fig. 4 shows examples of our segmentation results. Table 1 shows the quantitative comparison of RGVF snake [3], multi-scale watershed [5], fuzzy K-means [4], and our method in terms of mean $\pm$ std. of ZSI for all nuclei from the Herlev dataset. In addition, linear regression analysis is used to compare nucleus areas. Compared with ground truth, FCN-G shows strong correlation of r^2 = 0.91. The whole processing time of FCN-G is 0.36 s per image on average.

Table 1. Performance comparison of our method with previous methods on the Herlev dataset.

Methods	ZSI
Fuzzy K-means [4]	0.80 ± 0.24
RGVF snake [3]	0.87 ± 0.19
Multi-scale watershed [5]	0.89 ± 0.15
FCN-G	$\mathbf{0.92 \pm 0.09}$

4. CONCLUSION

The first combination of fully convolutional networks (FCN) and graph-based approach is proposed for automated and efficient cervical nucleus segmentation. The FCN learning of nucleus region is used to guide the graph construction and generate an in-region cost function. Globally optimal segmentation is guaranteed according to nucleus shape constraint, edge- and region-information, and nucleus context prior. The experimental results on the Herlev dataset demonstrate the superior performance of the proposed method.

5. REFERENCES

[1] D. Solomon and R. Nayar, *The Bethesda System for reporting cervical cytology: definitions, criteria, and explanatory notes*, Springer Science & Business Media, 2004.

[2] L. Zhang, H. Kong, C. T. Chin, S. Liu, Z. Chen, T. Wang, and S. Chen, "Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts," *Comput. Med. Imaging Graph.*, vol. 38, no. 5, pp. 369–380, 2014.

[3] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake," *Pattern Recognit.*, vol. 45, no. 4, pp. 1255–1264, 2012.

[4] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in pap smears," *Comput. Meth. Programs Biomed.*, vol. 113, no. 2, pp. 539–556, 2014.

[5] A. Gençtav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images," *Pattern Recognit.*, vol. 45, no. 12, pp. 4151–4168, 2012.

[6] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multi-scale convolutional network and graph partitioning," *TBME*, vol. 62, no. 10, pp. 2421–2433, 2015.

[7] L. Zhang, S. Liu, T. Wang, S. Chen, and M. Sonka, "Improved segmentation of abnormal cervical nuclei using a graph-search based approach," in *SPIE Medical Imaging*, 2015, pp. 94200W–94200W.

[8] S. Kashyap, I. Oguz, H. Zhang, and M. Sonka, "Automated segmentation of knee MRI using hierarchical classifiers and just enough interaction based learning: Data from osteoarthritis initiative," in *MICCAI*, 2016, pp. 344–351.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[10] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Nature inspired Smart Information Systems (NiSIS 2005)*, pp. 1–9, 2005.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR*, 2015.

[12] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *ICME*, 2015, pp. 689–692.