2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

AUTO-CONTEXT FULLY CONVOLUTIONAL NETWORK FOR LEVATOR HIATUS SEGMENTATION IN ULTRASOUDN IMAGES

Na Wang[1], *Yi Wang*[1], *Huifang Wang*[2], *Baiying Lei*[1], *Tianfu Wang*[1], *Dong Ni*[1]*

[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[2]Department of Ultrasound, the First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, Shenzhen 518035, China

ABSTRACT

The accurate boundary segmentation of levator hiatus (LH) in pelvic ultrasound (PUS) images benefits the diagnosis of female pelvic floor dysfunction (FPFD). Since manual boundary delineation is tedious, time-consuming, and of large observer variability, automatic method is highly demanded to improve the efficacy and objectiveness of LH segmentation. In this paper, we propose an auto-context fully convolutional network (AC-FCN), in which a customized FCN is integrated with the auto-context model for iteratively refining the prediction map of LH. Benefited from the progressive refinement, the final prediction map possesses both the local spatial consistency and boundary details. Hence the LH boundary can be readily segmented by applying an auxiliary active shape model (ASM) on this prediction map. The efficacy of the proposed method is validated on extensive challenging LH PUS images. Experimental results demonstrate that the proposed method provides more accurate segmentation than state-of-the-art methods do.

Index Terms— Levator hiatus, segmentation, fully convolutional networks, auto-context, active shape model

1. INTRODUCTION

Female pelvic floor dysfunction (FPFD) may present a group of clinical conditions that include stress urinary incontinence (SUI), pelvic organ prolapse (POP), fecal incontinence (FI), etc. These conditions affect up to 50% of women [1]. In general, levator trauma (LT) caused by pregnancy and childbirth is the major reason of FPFD. Thus the quantitative measurement of levator hiatus (LH) can be used for the diagnosis of FPFD [2].

Pelvic ultrasound (PUS) imaging is widely used for the examination and diagnosis of FPFD because it is radiation-free, real-time and inexpensive. In clinical practice, manual delineation (Fig. 1) is still the conventional way to segment the LH boundary from PUS images for the further diagnosis of FPFD. However, because of a low signal-to-noise ratio, PUS usually falls short of clearly depicting the LH, leading to fuzzy and incomplete LH boundaries (yellow arrows in Fig. 1). Moreover, the unilateral or bilateral injury of the levator ani further breaks the completeness of LH boundaries (blue arrows in Fig. 1). Thus the manual segmentation of LH remains a challenging and time-consuming task often leading to large observer variability. Developing automatic methods for the LH segmentation is a necessary and crucial image-processing issue that needs to be resolved to achieve more objective and efficient FPFD examinations.

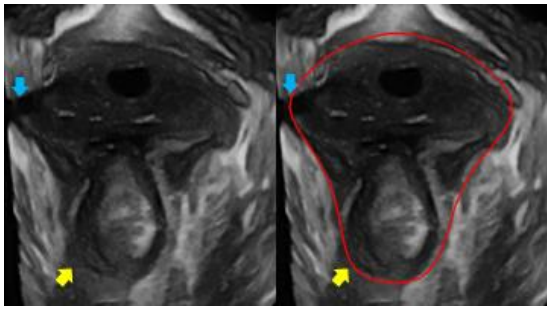

Fig. 1. One PUS image (left) and its segmentation (right). The red contour indicates the manually delineated LH boundary. Yellow and blue arrows denote boundary deficiency caused by US characteristic and injury of the levator ani, respectively.

To address the problem of LH segmentation, Sindhwani *et al*. [3] proposed a semi-automatic method using level set to outline the LH. However, this method requires a user-assisted labeling process to initialize the outlining, hence making the segmentation process tedious and subjective. In last decade, the advanced deep learning technique has shown its promising performance in image processing and understanding. The Convolutional Neural Networks (CNNs) [4] can achieve satisfactory performance in the task of foreground classification. However, CNNs cannot achieve pixel-level segmentation effectively because the fully connected layer in CNNs generally causes loss of spatial information. In contrast, with the end-to-end manner, the Fully Convolutional Networks (FCNs) [5] can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation, thus becoming the

state-of-the-art semantic segmentation method. Nevertheless, when encountering the issue of boundary deficiency, such as the fuzzy and incomplete LH boundaries in PUS images, a single conventional FCN tends to be incompetent to provide accurate prediction of boundary details. Thus additional boundary refinement strategy is demanded to improve the segmentation of the LH boundary.

In this study, we propose an auto-context fully convolutional network (AC-FCN) for the automatic LH segmentation in PUS images. To the best of our knowledge, this is the first fully automatic method to address the problem of LH segmentation. In the proposed network, a customized FCN is integrated with the auto-context model for iteratively refining the prediction map of LH. Specifically, a customized FCN based on FCN-8s [5] is tailored to the problem of LH segmentation for more efficient and accurate LH ROI prediction. Then the auto-context model [6] is integrated with the customized FCN to progressively refine the local spatial consistency and boundary details of the prediction map. After obtaining the refined prediction map, an auxiliary ASM [7] is applied for realizing LH shape constraint to further well segment LH from the prediction map. Experimental results demonstrate that the proposed method provides more accurate LH segmentation than state-of-the-art methods do.

2. METHODOLOGY

The proposed LH segmentation framework is illustrated in Fig. 2. Our customized FCN model is first trained to estimate LH region from PUS image by generating the dense prediction map. Then the generated prediction map is used iteratively as context information along with the original PUS image at each auto-context level, to progressively refine itself. The final segmentation is obtained by applying an auxiliary ASM on the prediction map generated from the last auto-context level.

2.1. Customized Fully Convolution Networks

Fully Convolution Networks [5] can take arbitrary-sized input and predict correspondingly-sized dense output with efficient inference and learning. Both inference and learning are performed whole-image-at-a-time. With the end-to-end, pixels-to-pixels training manner, FCNs possess great advantages in accurate semantic segmentation.

We choose the most successful fully convolutional model FCN-8s [5], which originated from VGG16 net [8], as our backbone architecture. The customization for the task of LH segmentation can be realized in two aspects. Firstly, because we are only dealing with two-class segmentation problem to extract LH region from the whole PUS image, there is no need to employ a completed FCN-8s with a large amount of parameters. Thus we discard the 6*th* and 7*th* convolutional layers of FCN-8s to reduce the model complexity meanwhile avoiding potential overfitting. This

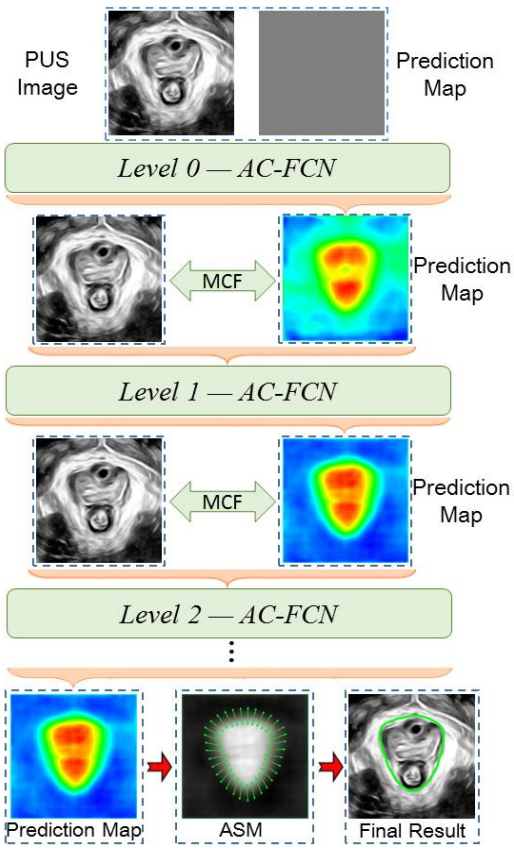

Fig. 2. Illustration of the proposed framework. The auto-context fully convolutional network (AC-FCN) can progressively refine the prediction map. The final segmentation is obtained by applying an ASM on the estimated prediction map.

customization is also beneficial for shortening the training time, as well as reducing the memory occupancy rate. Due to the discard of the 6*th* and 7*th* convolutional layers of FCN-8s, the 4*th* and 5*th* layers become the last two convolutional layers in our model. Thus we further append a fusion layer for fusing feature maps from 4*th* and 5*th* pooling layers to strengthen feature learning. Secondly, in this study, the LH target is always located in the central region of PUS image far away from the image edge. Therefore, instead of using pad value of 100 as conventional FCN-8s, we set pad value of 1 for the 1*st* convolutional layer. With such modification, we can further reduce the complexity of our model. Besides, according to the relationship between input and output dimension of convolutional layer:

$$F_o = \frac{F_i - K + 2 \times P}{S} + 1, \quad (1)$$

the output dimension F_o can be always equal to the input dimension F_i when kernel size K, stride S, and pad parameter P is set to 3, 1, 1, respectively. By applying this setting, we further discard all the crop layers in front of loss layers to simplify the model.

2.2. Refinement with Auto-Context Model

When facing the problem of boundary deficiency, such as the fuzzy and incomplete LH boundaries in PUS images, a customized FCN may still be incompetent to provide accurate prediction of boundary details. Thus additional boundary refinement strategy is demanded to improve the prediction of the LH boundary. Therefore the auto-context model [6] is integrated with the customized FCN to progressively refine the details of the prediction map.

Context information has shown to be useful in image segmentation tasks. The main concept of auto-context model is that the input of the k*th* level model contains not only the appearance features from original image but also the context features obtained from the output of the (k-1)*th* level model. In such a way, the generated prediction map is used iteratively as context information along with the original PUS image at each auto-context level, to progressively refine itself. The general iterative process can be formulated as:

$$y^k = F^k(P(x, y^{k-1})), \tag{2}$$

where F^k represents the model mapping function at level k, x and y^{k-1} mean PUS image and prediction map generated by (k-1)*th* level model, respectively. $P(\cdot)$ indicates the parallel concatenation combining information from x and y^{k-1}.

2.3. Improvement with Active Shape Model

Although the proposed AC-FCN model can be robust in predicting LH region regardless of common boundary deficiency caused by ultrasound imaging characteristic, it is still challenging when facing boundary incompleteness caused by unpredictable levator trauma. Thus after obtaining the refined prediction map, we apply an auxiliary ASM [9] to constrain the LH shape for the further well LH segmentation. This auxiliary ASM is constructed based on 248 annotated LH prediction maps. These prediction maps are the output of our AC-FCN model on 248 PUS images in training dataset. Each prediction map is annotated with 12 main and 60 secondary landmarks. Because the refined prediction map obtained by our AC-FCN model already possesses fine shape information, the auxiliary ASM can effectively fit to the accurate LH boundary without the influence of little ambiguous in the prediction map.

3. EXPERIMENTAL RESULTS

Experiments were carried on the datasets obtained from 284 patients at Shenzhen Second People's Hospital. From the 284 patients, we collected 372 PUS images using Mindray-Resona7 color Doppler ultrasound diagnostic instrument. Given that the difficulty of clinical data acquisition, we adopted data augmentation strategy to enrich training set by random cropping, translation, rotation, scaling or any combination of these actions. Of 372 PUS images, 248 images were augmented to 1488 images as training set, and the other 124 images were taken as testing set. An experienced physician first manually segmented all images. Then another experienced physician made necessary refinement and confirmed the ground truth.

Table 1. Comparison of different segmentation methods

Method	Dice	Jaccard	CC	ADB
CNN	0.8903	0.8026	0.7531	13.2707
SegNet	0.9319	0.8728	0.8536	9.7832
U-net	0.9230	0.8578	0.8320	11.0753
FCN-8s	0.9523	0.9094	0.8994	6.8874
AC-FCN-*L0*	0.9615	0.9262	0.9195	5.3720
AC-FCN-*L1*	0.9638	0.9303	0.9245	5.0556
AC-FCN-*L2*	0.9642	0.9312	0.9255	4.9853
AC-FCN-*L2*-ASM	**0.9646**	**0.9318**	**0.9261**	**4.9607**

The metrics employed to quantitatively evaluate segmentation results included region-based metrics such as Dice, Jaccard, and Conformity Coefficient (CC), and Average Distance of Boundaries (ADB) as distance-based metric [10]. The three region-based metrics can be calculated as:

$$Dice = \frac{2 \times A(G \cap S)}{A(G)+A(S)}, \tag{3}$$

$$Jaccard = \frac{A(G \cap S)}{A(G \cup S)}, \tag{4}$$

$$Conformity\ Coefficient = 2 - \frac{A(G \cup S)}{A(G \cap S)}, \tag{5}$$

where G and S denote the ground truth and automatic segmented boundary, respectively, and $A(\cdot)$ represents the area operator. The ADB is the average over the shortest pixel distances between the boundary points of the shapes and is defined as:

$$ADB = \frac{\sum_{p_G \subset G} d_{min}(p_G, S) + \sum_{p_S \subset S} d_{min}(p_S, G)}{\sigma_G + \sigma_S}, \tag{6}$$

where $d_{min}(p_G, S)$ and $d_{min}(p_S, G)$ represent the distance from point p_G on G to the nearest point on S and point p_S on S to the nearest point on G, respectively. σ_G and σ_S are the number of points on G and S, respectively.

We compared our method with state-of-the-art methods, including CNN [4], SegNet [11], U-net [12], and FCN-8s [5]. All of the compared networks were pre-trained and fine-tuned properly. Table 1 demonstrates the comparison between our method and state-of-the-art methods, as well as the comparison between different levels (AC-FCN-*L0~2*). As shown in Table 1, our AC-FCN-*L0* already outperforms the other four state-of-the-art networks, which demonstrates the efficacy of our specially tailored FCN. By further observing the results from different levels, as context information propagates from level 0 to level 2, our AC-FCN progressively achieves improvement. However, this improvement contributed by auto-context degrades exponentially when context level increases, and too much context levels may lead to overfitting problem, thus we only apply three levels and the AC-FCN-*L2* achieves satisfactory performance. Finally, by comparing AC-FCN-*L2* and AC-FCN-*L2*-ASM, we observe that the segmentation can still be improved by using ASM to

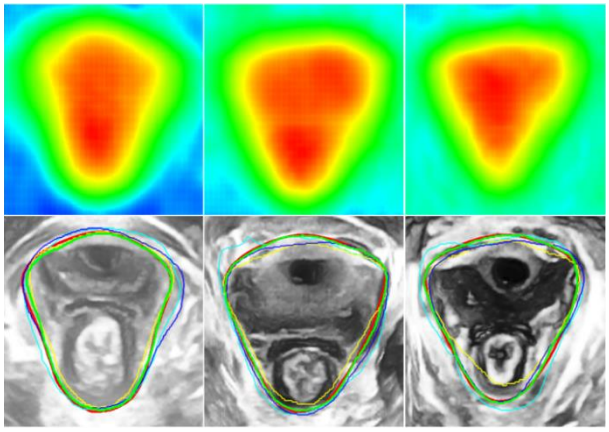

Fig. 3. Qualitative segmentation comparison of different methods. First row: the refined prediction maps obtained by our AC-FCN-*L2*. Second row: segmented boundaries of SegNet (yellow), FCN-8s (cyan), AC-FCN-*L2* (blue), AC-FCN-*L2*-ASM (green) and the ground truth (red).

constrain the LH shape. Fig. 3 visualizes the segmentation results obtained by different methods, as well as the refined prediction maps obtained by our AC-FCN-*L2*. Note that to clearly show the difference of boundaries obtained by different methods and also considering the inferior performance of CNN and U-net in Table 1, we only illustrate five different boundaries in Fig. 3 for comparison. The five are segmented boundaries of SegNet, FCN-8s, AC-FCN-*L2*, AC-FCN-*L2*-ASM and the ground truth. As shown in Fig. 3, the AC-FCN-*L2*-ASM method achieves the most similar LH boundaries to the ground truth. Compared to other methods, the AC-FCN-*L2*-ASM can successfully approach the real LH boundary due to the consideration of both context information and shape constraint.

4. CONCLUSION

This paper presents an auto-context fully convolutional network (AC-FCN) for the automatic LH segmentation in PUS images. To our best knowledge, this is the first fully automatic method to address the problem of LH segmentation. In the proposed network, a customized FCN is integrated with the auto-context model for iteratively refining the prediction map of LH. Benefited from the iterative refinement, the final prediction map can possess both the local spatial consistency and boundary details. Hence the LH boundary can be readily segmented by applying an auxiliary ASM on this prediction map. The experimental results indicate that by considering both context information and shape constraint, the proposed method provides more accurate segmentation than state-of-the-art methods do. In general, we integrate the current popular deep learning technique with classical shape modeling method to build a clinically practical framework for the automatic LH segmentation. The proposed framework can be general and easily extended to other medical image segmentation tasks.

5. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Funds of China (Nos. 61571304, 61701312, 81571758 and 81771922) and National Key Research and Develop Program (No. 2016YFC0104703).

6. REFERENCES

[1] S. Hagen and D. Stark, "Conservative prevention and management of pelvic organ prolapse in women," Cochrane Database of Systematic Reviews, vol. 8, p. CD003882, 2011.

[2] H. P. Dietz, "Quantification of major morphological abnormalities of the levator ani," Ultrasound in Obstetrics & Gynecology, vol. 29, pp. 329–334, 2007.

[3] N. Sindhwani, D. Barbosa, M. Alessandrini, B. Heyde, H. P. Dietz, J. D'Hooge, et al., "Semi-automatic outlining of levator hiatus," Ultrasound in Obstetrics & Gynecology the Official Journal of the International Society of Ultrasound in Obstetrics & Gynecology, vol. 48, p. 98, 2016.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in International Conference on Neural Information Processing Systems, 2012, pp. 1097-1105.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.

[6] Z. Tu, "Auto-context and its application to high-level vision tasks," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," Computer vision and image understanding, vol. 61, pp. 38-59, 1995.

[8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Computer Science, 2014.

[9] B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," IEEE transactions on medical imaging, vol. 21, pp. 924-933, 2002.

[10] H. H. Chang, A. H. Zhuang, D. J. Valentino, and W. C. Chu, "Performance measure characterization for evaluating neuroimage segmentation algorithms," Neuroimage, vol. 47, pp. 122-135, 2009.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," arXiv preprint arXiv:1511.00561, 2015.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234-241.