Deep Attentional Features for Prostate Segmentation in Ultrasound

Yi Wang[1,2], Zijun Deng[3], Xiaowei Hu[4], Lei Zhu[4,5(✉)], Xin Yang[4], Xuemiao Xu[3], Pheng-Ann Heng[4], and Dong Ni[1,2]

[1] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[2] Medical UltraSound Image Computing (MUSIC) Lab, Shenzhen, China
[3] School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[4] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
lzhu@cse.cuhk.edu.hk
[5] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract. Automatic prostate segmentation in transrectal ultrasound (TRUS) is of essential importance for image-guided prostate biopsy and treatment planning. However, developing such automatic solutions remains very challenging due to the ambiguous boundary and inhomogeneous intensity distribution of the prostate in TRUS. This paper develops a novel deep neural network equipped with deep attentional feature (DAF) modules for better prostate segmentation in TRUS by fully exploiting the complementary information encoded in different layers of the convolutional neural network (CNN). Our DAF utilizes the attention mechanism to selectively leverage the multi-level features integrated from different layers to refine the features at each individual layer, suppressing the non-prostate noise at shallow layers of the CNN and increasing more prostate details into features at deep layers. We evaluate the efficacy of the proposed network on challenging prostate TRUS images, and the experimental results demonstrate that our network outperforms state-of-the-art methods by a large margin.

1 Introduction

Prostate cancer is the most common noncutaneous cancer and the second leading cause of cancer-related deaths in men [9]. Transrectal ultrasound (TRUS) is the routine imaging modality for image-guided biopsy and therapy of prostate cancer. Segmenting prostate from TRUS is of essential importance for the treatment

Y. Wang and Z. Deng contributed equally to this work.

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11073, pp. 523–530, 2018.
https://doi.org/10.1007/978-3-030-00937-3_60

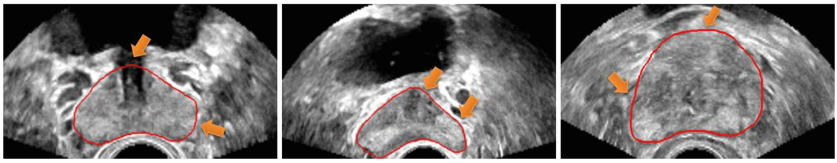

Fig. 1. Example TRUS images. Red contour denotes the prostate boundary. There are large prostate shape variations, and the prostate tissues present inhomogeneous intensity distributions. Orange arrows indicate missing/ambiguous boundaries.

planning [10], and can help surface-based registration between TRUS and preoperative MRI during image-guided interventions [11]. However, accurate prostate segmentation in TRUS remains very challenging due to the missing/ambiguous boundary and inhomogeneous intensity distribution of the prostate in TRUS, as well as the large shape variations of different prostates (see Fig. 1).

The problem of automatic prostate segmentation in TRUS has been extensively exploited in the literature. One main methodological stream utilizes shape statistics for the prostate segmentation. Shen *et al.* [8] presented a statistical shape model for prostate segmentation. Yan *et al.* [14] developed a partial active shape model to address the missing boundary issue in ultrasound shadow area. Another direction is to formulate the prostate segmentation as a foreground classification task. Ghose *et al.* [3] performed supervised soft classification with random forest to identify prostate. In general, all above methods used handcrafted features for segmentations, which are ineffective to capture the high-level semantic knowledge, and thus tend to fail in generating high-quality segmentations when there are ambiguous boundaries in TRUS. Recently, deep neural networks are demonstrated to be a very powerful tool to learn deep features for object segmentation. For TRUS segmentation, Yang *et al.* [15] proposed to learn the shape prior with recurrent neural networks and achieved state-of-the-art segmentation performance.

One of the main advantages of deep neural networks is to generate well-organized features consisting of abundant semantic and fine information. However, directly using these features at individual layers to conduct prostate segmentation cannot guarantee satisfactory results. It is essential to leverage the complementary advantages of features at multiple levels and to learn more discriminative features targeting for accurate and robust segmentation. To this end, we propose to fully exploit the complementary information encoded in multi-layer features (MLF) generated by a convolutional neural network (CNN) for better prostate segmentation in TRUS images. Specifically, we develop a novel prostate segmentation network with deep attentional features (DAFs). The DAF is generated at each individual layer by learning the complementary information of the low-level detail and high-level semantics in MLF, thus is more powerful for the better representation of prostate characteristics. Our DAFs at shallow layers can learn highly semantic information encoded in the MLF to suppress its non-prostate regions, while our DAFs at deep layers are able to select the

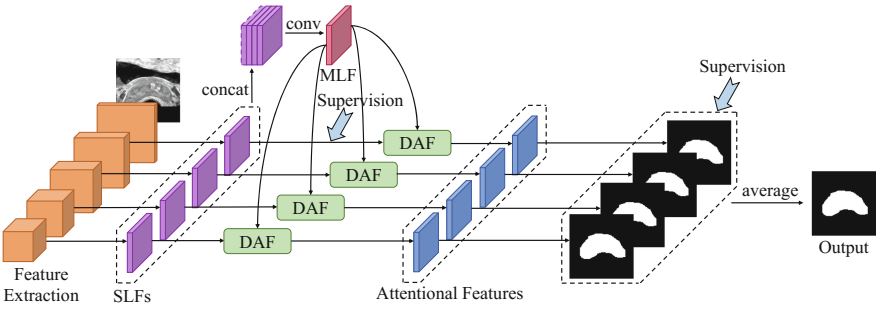

Fig. 2. The schematic illustration of our prostate segmentation network with deep attentional features (DAF). SLF: single-layer features; MLF: multi-layer features.

fine detail features from the MLF to refine prostate boundaries. Experiments on TRUS images demonstrate that our segmentation using deep attentional features outperforms state-of-the-art methods. The code is publicly available at https://github.com/zijundeng/DAF.

2 Deep Attentional Features for Segmentation

Segmenting prostate from TRUS images is a challenging task especially due to the ambiguous boundary and inhomogeneous intensity distribution of the prostate in TRUS. Directly using low-level or high-level features, or even their combinations to conduct prostate segmentation may often fail to get satisfactory results. Therefore, leveraging various factors such as multi-scale contextual information, region semantics and boundary details to learn more discriminative prostate features is essential for accurate and robust prostate segmentation.

To address above issues, we present a deep neural network with deep attentional features (DAFs). The following subsections present the details of the proposed method and elaborate the novel DAF module.

2.1 Method Overview

Figure 2 illustrates the proposed prostate segmentation network with deep attentional features. Our network takes the TRUS image as the input and outputs the segmentation result in an end-to-end manner. It first produces a set of feature maps with different resolutions by using the CNN. The feature maps at shallow layers have high resolutions but with fruitful detail information while the feature maps at deep layers have low resolutions but with high-level semantic information. The highly semantic features can help to identify the position of prostate and the fine detail is able to indicate the fine boundary of the prostate.

After obtaining the feature maps with different levels of information, we enlarge these feature maps with different resolutions to a quarter of the size of original input image by linear interpolation (the feature maps at the first layer

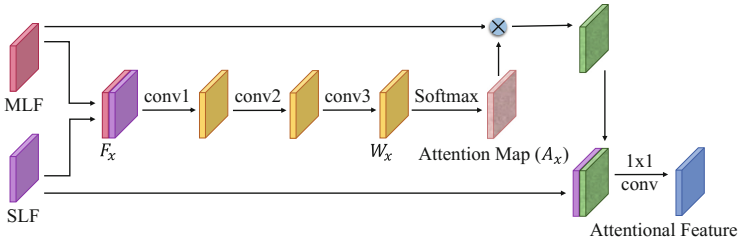

Fig. 3. The schematic illustration of the *deep attentional feature* (*DAF*) module.

are ignored due to the memory limitation). The enlarged feature maps at each individual layer are denoted as "single-layer features (SLF)", and the multiple SLFs are combined together, followed by convolution operations, to generate the "multi-layer features (MLF)". Although the MLF encodes the low-level detail information as well as the high-level semantic information of the prostate, it also inevitably incorporates noise from the shallow layers and losses some subtle parts of the prostate due to the coarse features at deep layers. Hence, the straightforward segmentation result from the MLF tends to contain lots of non-prostate regions and lose parts of prostate tissues.

In order to refine the features of the prostate ultrasound image, we present a DAF module to generate deep attentional features at each layer in the principle of the attention mechanism. The DAF module leverages the MLF and the SLF as the inputs and produces the refined feature maps; please refer to Sect. 2.2 for the details of our DAF module. Then, we obtain the segmentation maps from the deep attentional features at each layer by using the deeply supervised mechanism [4,13] that imposes the supervision signals to multiple layers. Finally, we get the prostate segmentation result by averaging the segmentation maps at each individual layer.

2.2 Deep Attentional Features

As presented in Sect. 2.1, the feature maps at shallow layers contain the detail information of prostate but also include non-prostate regions, while the feature maps at deep layers are able to capture the highly semantic information to indicate the location of the prostate but may lose the fine details of the prostate's boundaries. To refine the features at each layer, we present a DAF module (see Fig. 3) to generate the deep attentional features by utilizing the attention mechanism to selectively leverage the features at MLF to refine features at the individual layer.

Specifically, given the single-layer feature maps at each layer, we concatenate them with the multi-layer feature maps as F_x, and then produce the unnormalized attention weights W_x (see Fig. 3):

$$W_x = f_a(F_x; \theta), \tag{1}$$

where θ represents the parameters learned by f_a which contains three convolutional layers. The first two convolutional layers use 3×3 kernels, and the last convolutional layer applies 1×1 kernels.

After that, our DAF module computes the attention map A_x by normalizing W_x across the channel dimension with a Softmax function:

$$a_{i,j}^k = \frac{exp(w_{i,j}^k)}{\sum_k exp(w_{i,j}^k)}, \tag{2}$$

where $w_{i,j}^k$ denotes the value at spatial location (i, j) position and k-th channel on W_x, while $a_{i,j}^k$ denotes the normalized attention weight at spatial location (i, j) and k-th channel on A_x. After obtaining the attention map, we multiply it with the MLF in a element-by-element manner to generate a new refined feature map. The new features are concatenated with the SLF and then we apply a 1×1 convolution operation to produce the final attentional features for the given layer (see Fig. 3).

We apply the DAF module on each layer to refine its feature map. During this process, the attention mechanism is used to generate a set of weights to indicate how much attention should be paid to the MLF for each individual layer. Hence, our DAF enables the features at shallow layers to select the highly semantic features from the MLF in order to suppress the non-prostate regions, while the features at deep layers are able to select the fine detail features from the MLF to refine the prostate boundaries.

3 Experiments

3.1 Materials

Experiments were carried on TRUS images obtained using Mindray DC-8 ultrasound system in the First Affiliate Hospital of Sun Yat-Sen University. Informed consent was obtained from all patients. In total, we collected 530 TRUS images from 17 TRUS volumes which were acquired from 17 patients. The size of each TRUS image is 214×125 with a pixel size of 0.5×0.5 mm. We augmented (i.e., rotated, horizontally flipped) 400 images of 10 patients to 2400 as training dataset, and taken the remaining 130 images from 7 patients as testing dataset. All the TRUS images were manually segmented by an experienced clinician.

3.2 Training and Testing Strategies

Our proposed framework was implemented on PyTorch and used the ResNeXt101 [12] as the feature extraction layers (the orange parts in the left of Fig. 2).

Loss Function. Cross-entropy loss was used for each output of this network. The total loss $\mathcal{L}_t$ was defined as the summation of loss on all predicted score maps:

$$\mathcal{L}_t = \sum_{i=1}^{n} w_i \mathcal{L}_i + \sum_{j=1}^{n} w_j \mathcal{L}_j + w_f \mathcal{L}_f, \tag{3}$$

where w_i and $\mathcal{L}_i$ represent the weight and loss of i-th layer; while w_j and $\mathcal{L}_j$ represent the weight and loss of j-th layer after refining features using our DAF; n is the number of layers of our network; w_f and $\mathcal{L}_f$ are the weight and loss for the output layer. We empirically set all the weights (w_i, w_j and w_f) as 1.

Training Parameters. In order to reduce the risk of overfitting and accelerate the convergence of training, we used the weights trained on ImageNet [2] to initialize the feature extraction layers and other parts were initialized by random noise. The framework was trained on the augmented training set which contained 2400 samples. Stochastic gradient descent (SGD) with the momentum of 0.9 and weight decay of 0.01 was used to train the whole framework. We set the learning rate as 0.005 and it reduced to 0.0001 at 600 iterations. Learning stopped after 1200 iterations. The framework was trained on a single GPU with a mini-batch size of 4, only taking about 20 min.

Inference. In testing, for each input TRUS image, our network produced several output prostate segmentation maps since we added the supervision signals to all layers. We computed the final prediction map (see the last column of Fig. 2) by averaging the segmentation maps at each layer. After getting the final prediction map, we applied the fully connected conditional random field (CRF) [5] to improve the spatial coherence of the prostate segmentation map by considering the relationships of neighborhood pixels.

3.3 Segmentation Performance

We compared results of our method with several advanced methods, including Fully Convolutional Network (FCN) [6], Boundary Completion Recurrent Neural Network (BCRNN) [15], and U-Net [7]. For a fair comparison, we obtain the results of our competitors by using either the segmentation maps provided by corresponding authors, or re-training their models using the public implementations and adjusting training parameters to obtain best segmentation results.

The metrics employed to quantitatively evaluate segmentation included Dice Similarity Coefficient (Dice), Average Distance of Boundaries (ADB, in pixel), Conformity Coefficient (CC), Jaccard Index, Precision, and Recall [1]. A better segmentation shall have smaller ADB, and larger values of all other metrics.

Table 1 lists the metric results of different methods. It can be observed that our method consistently outperforms others on almost all the metrics. Figure 4 visualizes some segmentation results. Apparently, our method obtains the most similar segmented boundaries to the ground truth. Furthermore, as shown in Fig. 4, our method can successfully infer the missing/ambiguous boundaries, and it demonstrates the proposed deep attentional features can efficiently encode complementary information for accurate representation of the prostate tissues.

Table 1. Metric results of different methods (best results are highlighted in bold)

Method	Dice	ADB	CC	Jaccard	Precision	Recall
FCN [6]	0.9188	12.6720	0.8207	0.8513	0.9334	0.9080
BCRNN [15]	0.9239	11.5903	0.8322	0.8602	**0.9446**	0.9051
U-Net [7]	0.9303	7.4750	0.8485	0.8708	0.8985	0.9675
Ours	**0.9527**	**4.5734**	**0.9000**	**0.9101**	0.9369	**0.9698**

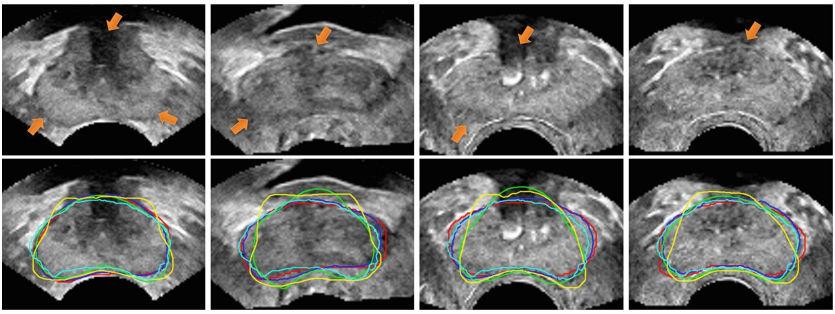

Fig. 4. Visual comparison of prostate segmentation results. Top row: prostate TRUS images with orange arrows indicating missing/ambiguous boundaries; bottom row: corresponding segmentations from our method (blue), U-Net (cyan), BCRNN (green) and FCN (yellow), respectively. Red contours are ground truths. Our method has the most similar segmented boundaries to the ground truth.

4 Conclusion

This paper develops a novel deep neural network for prostate segmentation in ultrasound images by harnessing the deep attentional features. Our key idea is to select the useful complementary information from the multi-level features to refine the features at each individual layer. We achieve this by developing a DAF module, which can automatically learn a set of weights to indicate the importance of the features in MLF for each individual layer by using an attention mechanism. Furthermore, we apply multiple DAF modules in a convolutional neural network to predict the prostate segmentation maps in different layers. Experiments on challenging TRUS prostate images demonstrate that our segmentation using deep attentional features outperforms state-of-the-art methods. In addition, the proposed method is a general solution and has the potential to be used for other medical image segmentation tasks.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (61701312; 61571304; 61772206), in part by the Natural Science Foundation of SZU (No. 2018010), in part by the Shenzhen Peacock Plan (KQTD2016053112051497), in part by Hong Kong Research Grants Council (No. 14202514) and Innovation and Technology Commission under TCFS (No. GHP/002/13SZ), and in part by the Guangdong Natural Science Foundation (No. 2017A030311027). Xiaowei Hu is funded by the Hong Kong Ph.D. Fellowship.

References

1. Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C.: Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage **47**(1), 122–135 (2009)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR (2009)
3. Ghose, S., et al.: A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images. Med. Image Anal. **17**(6), 587–600 (2013)
4. Hu, X., Zhu, L., Qin, J., Fu, C.W., Heng, P.A.: Recurrently aggregating deep features for salient object detection. In: AAAI (2018)
5. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: NIPS (2011)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
7. 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, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Shen, D., Zhan, Y., Davatzikos, C.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans. Med. Imaging **22**(4), 539–551 (2003)
9. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. CA: Cancer J. Clin. **68**(1), 7–30 (2018)
10. Wang, Y., et al.: Towards personalized statistical deformable model and hybrid point matching for robust MR-TRUS registration. IEEE Trans. Med. Imaging **35**(2), 589–604 (2016)
11. Wang, Y., Zheng, Q., Heng, P.A.: Online robust projective dictionary learning: shape modeling for MR-TRUS registration. IEEE Trans. Med. Imaging **37**(4), 1067–1078 (2018)
12. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: CVPR (2017)
13. Xie, S., Tu, Z.: Holistically-nested edge detection. In: ICCV (2015)
14. Yan, P., Xu, S., Turkbey, B., Kruecker, J.: Discrete deformable model guided by partial active shape model for TRUS image segmentation. IEEE Trans. Biomed. Eng. **57**(5), 1158–1166 (2010)
15. Yang, X., et al.: Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: AAAI (2017)