Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation

Conference Paper · April 2017

DOI: 10.1109/ISBI.2017.7950607

6 authors, including:

Wu Lingyun
Shenzhen University

Xin Yang
Shenzhen University

Shengli Li
Shenzhen Maternity & Child Healthcare Hospital, Southern Medical University

Pheng-Ann Heng
The Chinese University of Hong Kong

CASCADED FULLY CONVOLUTIONAL NETWORKS FOR AUTOMATIC PRENATAL ULTRASOUND IMAGE SEGMENTATION

Lingyun Wu[1], *Xin Yang*[2], *Shengli Li*[3], *Tianfu Wang*[1], *Pheng-Ann Heng*[2], *Dong Ni*[1,⋆]

[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, China

[2]Department of Computer Science and Engineering, The Chinese University of Hong Kong

[3]Department of Ultrasound, Affiliated Shenzhen Maternal and Child Healthcare Hospital of Nanfang Medical University, China

ABSTRACT

Computerized prenatal ultrasound (US) image segmentation methods can greatly improve the efficiency and objectiveness of the biometry interpretation. However, the boundary incompleteness and ambiguity in US images hinder the automatic solutions severely. In this paper, we propose a cascaded framework for fully automatic US image segmentation. A customized Fully Convolutional Network (FCN) was utilized to exploit feature extractions from multiple visual scales and distinguish the anatomy with a dense prediction map. To enhance the local spatial consistency of the prediction map, we further implant the FCN core into an Auto-Context scheme. By modifying the join operator in traditional Auto-Context scheme from parallelization to summation, our framework gains extra considerable improvement. We demonstrate the efficacy of our method on two challenging datasets: fetal head and abdomen US images. Extensive experimental results show that our method can bridge severe boundary incompleteness and achieves the best segmentation accuracy when compared with state-of-the-art methods.

Index Terms— Prenatal ultrasound, image segmentation, fully convolutional networks, auto-context, join operator

1. INTRODUCTION

Ultrasound (US) imaging is a dominant modality in prenatal examinations due to its superior advantages, such as real-time imaging, low-cost and free of radiation. Manually delineating the fetal anatomy boundary is generally a prerequisite for fetal biometric interpretation from US images. Fig. 1 illustrates two typical manual boundary delineations from an expert for fetal head and abdomen.

In clinical practice, manual boundary delineation is tedious, time-consuming and often leads to large inter- and intra-observer variability. In this situation, automatic solution for boundary segmentation is highly desired for improving clinical work-flow and diagnosis objectiveness. However, automatically segmenting target fetal anatomy from US images remains a challenging task for several reasons. Firstly, US images often present various intensity distributions due to different imaging conditions. Secondly, a number of factors including acoustic shadows, speckle noise and low contrast between objects and surrounding tissues cause the typical boundary ambiguities and long-span occlusion [1], as arrows shown in Fig. 1. Thirdly, different press sources tend to make the fetal anatomical structures deform.

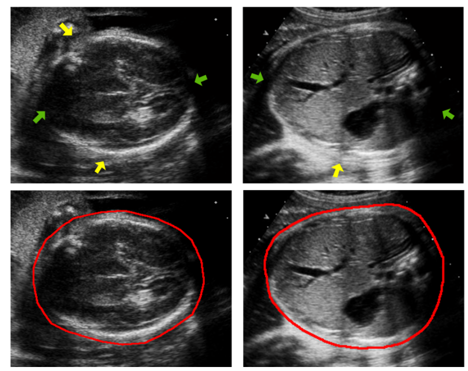

Fig. 1. Boundary delineation (red line) of feta head (left) and abdomen (right) ultrasound images. Green and yellow arrows denote boundary deficiency and ambiguity, respectively.

Accordingly, intensive researches for automatic fetal anatomy segmentation in US have been conducted. Previous studies [2, 3] made use of discriminative learning methods to segment fetal key anatomies. Promising as they are, the handcrafted features and limited size of training data [2, 3] may degrade their performance. The recently relighted Convolutional Neural Networks (CNNs) [4] perform well for foreground classification, but they often need auxiliary object detectors to provide region of interest [5]. By integrating cues from multiple visual scales, Fully Convolutional Networks (FCNs) [6] with pre-training become dominant in the image segmentation field. However, the popular network designs for

⋆ Corresponding author: nidong@szu.edu.cn

natural image analysis usually bear the model complexities which exceed the required amount in medical image field. Also, when faced with the boundary deficiency and occlusion in US images, a single FCN tends to be insufficient in predicting local boundary details well. Compared to the computational intensive refinement strategy in [7], Auto-Context [8] is a more straightforward solution for boundary refinement. By revisiting fetal US image in one channel and context cues of prediction results in another parallel channel, Auto-Context scheme can significantly enrich the local boundary prediction details. However, this parallelization manner not only increases the model complexity with more parameters to learn but also may underestimate the importance of the prediction map, which is often more informative than the US image.

To address the aforementioned issues, in this paper, we propose a cascaded FCNs (denoted as *casFCN*) for automatic prenatal US image segmentation. We characterize our framework with three features. Firstly, we specifically tailored the popular FCN-8s network [6] into a lite FCN to conduct efficient anatomy region prediction. This customization fits the model complexity well with our segmentation tasks as it provides better segmentation accuracies than FCN-8s. Secondly, we implanted the lite FCN into an Auto-Context model to refine the prediction map details successively. Thirdly, instead of using parallelization, we investigate the usage of a summation operator on fetal US image and prediction map in the Auto-Context scheme. It may be worthy of exploration to see its efficacy with the FCN model on our segmentation tasks.

2. METHODOLOGY

Our proposed casFCN framework is illustrated in Fig. 2. The tailored FCN models are trained to recognize fetal anatomy region from input by outputting dense boundary maps. The generated boundary map then goes through the following Auto-Context levels. At each level, the input is the summation of fetal US image with prediction map from previous level. The boundary map in level 0 is initialized as void. The prediction map is gradually refined in local as it is revisited by Auto-Context scheme from level to level. The final segmentation mask is obtained from the last Auto-Context level.

2.1. Fully Convolutional Networks Architecture

FCNs possess great advantages in accurate and dense classification, since it learns to combine coarse abstractions from deep layers with fine details from shallow layers. Composed of downsampling and upsampling path, FCNs can be trained in a promising end-to-end manner [6].

The most successful FCN architecture is FCN-8s, which originates from VGG 16-layer net [9] but discards the final classification layer and replaces all fully connected layers with convolution layers. The 6^{th} and 7^{th} convolutional layers

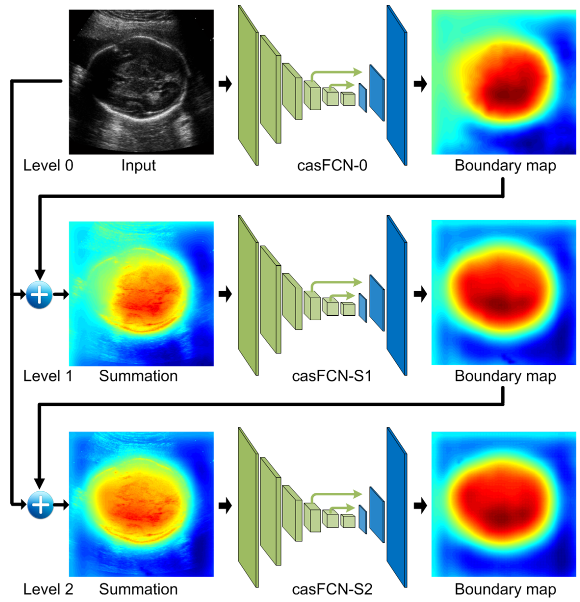

Fig. 2. Intuition of our proposed cascaded Fully Convolutional Networks.

of FCN-8s are both designed to generate 4096 feature maps for segmentation of 21 classes, while our task is a two-class only segmentation problem. So we further discard the 6^{th} and 7^{th} convolutional layers of FCN-8s to decrease model complexity and thus avoid possible overfitting. The incremental improvement of this customization can be observed in section 3. Then we append a 1×1 convolution layer to generate two score maps for fetal anatomy (M_{fg}) and background (M_{bk}) at each coarse output location, followed by a deconvolution layer to upsample the coarse maps to dense pixel-wise outputs ($M^{'}_{fg}$ and $M^{'}_{bk}$). We take the foreground score map $M^{'}_{fg}$ as the boundary prediction map. Our tailored FCN retains the weights from VGG16 model and is then fine-tuned on our target US images to address the possible overfitting problem on small training dataset. Let W be the weights of FCN, then the cross-entropy based loss function used in this paper for training is

$$\mathcal{L}(\mathcal{X}; W) = \sum_{x_i \in \mathcal{X}} -\log p(y_i = \ell(x_i)|x_i; W) \qquad (1)$$

where $\mathcal{X}$ represents the training samples and $p(y_i = \ell(x_i)|x_i ; W)$ is the probability of target class label $\ell(x_i)$ corresponding to the pixel $x_i \in \mathcal{X}$.

2.2. Refinement with Auto-Context Scheme

Faced with the boundary deficiency in US, a single FCN presents limited ability in estimating the occluded parts of fetal anatomical structures. We therefore combine FCN with the iterative refinement framework, Auto-Context, to effectively explore local contextual information and thus remove the ambiguity when predicting on current position [10].

The core idea of Auto-Context is to stack a series of models in a way that, the model at level k not only utilizes the appearance features in intensity image, but also the contextual features extracted from the prediction map generated by the model at level $k-1$. The revisiting of previous prediction map contributes greatly to the successive refinement on the prediction map. Eq. 2 illustrates the general iterative process of an Auto-Context scheme, where $\mathcal{H}^k$ is the mapping function of the model at level k, x and y^k are the fetal US image and the prediction map from level $k-1$, respectively. $\mathcal{J}$ is a join operator to combine information from x and y^k. Classic Auto-Context scheme adopts parallelization as join operator, that is, parallel channels for extracting features from different image sources. Also, empirically designed and fixed structure are needed for models to collect contextual knowledge, such as the star shape used in [8]. However, no specific structure is required for our casFCN framework, because FCN can gather contextual information flexibly from hierarchical receptive fields. Usually, the refinement with Auto-Context Scheme diminishes exponentially as context level increases and the overfitting or performance drop might be caused by stacking too many context levels. So, three context levels were finally adopted in this paper for time efficiency. Taking class labels of the maximum values between $M^{'}_{fg}$ and $M^{'}_{bk}$, last context level outputs the final segmentation mask. Notably, our Auto-Context level k is pre-trained with parameters from level $k-1$ to facilitate the training process of system.

$$y^k = \mathcal{H}^k(\mathcal{J}(x, y^{k-1})) \qquad (2)$$

As validated in section 3, different join operators $\mathcal{J}$ in Eq. 2 have different impact on refinement performance. Classical choice is a *parallelization* manner which parallelizes intensity image with previous prediction map and then conduct feature extraction on them with two independent channels. A casFCN using this manner is denoted as *casFCN-Pk*, where *k* is context level. However, an extra channel in parallelization not only involves more parameters to learn, but also underestimates the importance of prediction map in enhancing the anatomy region in intensity image. In this paper, we choose to use *summation* as the join operator $\mathcal{J}$. Apart from parameter sharing, summarizing the intensity image with prediction map is instinctive and explicit to highlight the fetal anatomical structures and also suppress the irrelevant strong background noise. A casFCN using this manner is denoted as *casFCN-Sk*, where *k* is context level. Detailed investigation on the coefficients in summation is elaborated in section 3.

3. EXPERIMENTAL RESULTS

Our method is validated on fetal head and abdomen US image segmentation tasks. Our training dataset contains 900 and 688 images for fetal head and abdomen, respectively. Our testing dataset contains 236 fetal head and 505 fetal abdomen images, with size 768×576. These two datasets cover the gestational age (GA) from 19 to 40 week. An experienced expert provided ground truth for all images.

Table 1. Comparison of fetal head segmentation methods

Method	Dice	Adb	Conform	Jaccard
CNN	0.9723	3.6686	0.9428	0.9464
FCN-8s	0.9659	4.3620	0.9293	0.9343
U-Net	0.9694	3.9201	0.9368	0.9408
casFCN-0	*0.9774*	*2.9581*	*0.9536*	*0.9558*
casFCN-P1	0.9809	2.4981	0.9610	0.9626
casFCN-P2	0.9811	2.4804	0.9615	0.9630
casFCN-S1	0.9839	2.1006	0.9672	0.9683
casFCN-S2	**0.9843**	**2.0459**	**0.9680**	**0.9690**

Table 2. Comparison of fetal abdomen segmentation methods

Method	Dice	Adb	Conform	Jaccard
CNN	0.9349	10.2482	0.8589	0.8789
FCN-8s	0.9452	8.1213	0.8832	0.8965
U-Net	0.9271	10.7047	0.8414	0.8650
casFCN-0	*0.9541*	*6.9491*	*0.9024*	*0.9133*
casFCN-P1	0.9560	6.9340	0.9068	0.9166
casFCN-P2	0.9561	6.8360	0.9070	0.9168
casFCN-S1	0.9604	6.3327	0.9169	0.9244
casFCN-S2	**0.9618**	**5.7775**	**0.9200**	**0.9628**

Our strict evaluation criteria include Dice coefficient (Dice), Average Distance of Boundaries (Adb), Conformity and Jaccard index [11]. We extensively compared our casFCN with well established methods, including CNN, FCN-8s and U-net [12]. All of the compared networks were pre-trained and fine-tuned properly. The detailed quantitative evaluation results are shown in Table 1 and 2. Without cascaded refinement, our casFCN-0 already outperforms other 3 networks. As cascade level propagates, our casFCN tends to gain improvement successively from level to level. The casFCN-S consistently outperforms the casFCN-P, which provides a convincing evidence for our operator design. We use Dice and Adb as reference to balance the appearance and contextual information in summation. Shown as Fig. 3, best performance is achieved when about 20% intensity image and 80% prediction map were involved, respectively.

Fig. 4 presents the qualitative evaluation of our proposed casFCN method. As observed, large variations of GA, size, appearance and shape in anatomies are well addressed, especially on fetal abdomen boundary delineation tasks.

4. CONCLUSIONS

In this paper, we present a casFCN framework for fully automatic prenatal US image segmentation. Promising segmen-

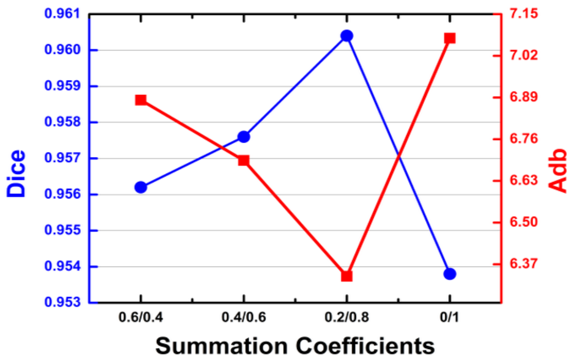

Fig. 3. Performance of casFCN-S1 on fetal abdomen segmentation with different summation coefficient combination.

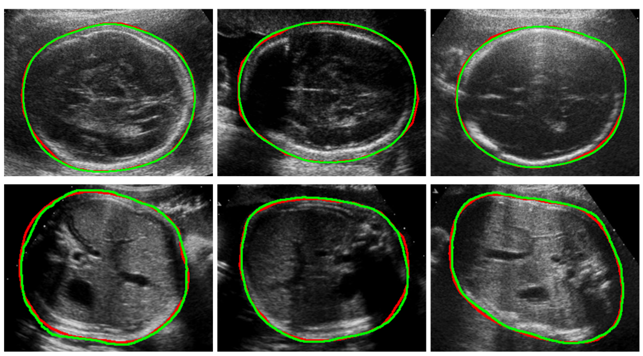

Fig. 4. Segmentation exemplars for fetal head (top) and abdomen (bottom) images. Red and green lines denote ground truth and our casFCN-S2 segmentation results, respectively.

tation accuracies are achieved in both two challenging tasks. We get a good starting point by tailoring the FCN-8s for considering both low model complexity and high segmentation accuracy. Implanting our tailored FCN into an Auto-Context scheme enhances the boundary details by dynamically exploring local context, and thus gains remarkable improvement from level to level. Additionally, we provide a new perspective to the classical Auto-Context scheme with the investigation on two types of join operators. We also conduct thorough experiments to support our parameter setting. The proposed casFCN framework is general and can be easily extended to other US image segmentation tasks.

5. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Funds of China (No. 61571304 and 81571758), and Hong Kong Research Grants Council General Research Fund (No. CUHK14203115).

6. REFERENCES

[1] J Alison Noble, "Reflections on ultrasound image analysis," *Medical Image Analysis*, vol. 33, pp. 33–37, 2016.

[2] Gustavo Carneiro, Sara Good, and et al., "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE transactions on medical imaging*, vol. 27, no. 9, pp. 1342–1355, 2008.

[3] Mohammad Yaqub, J Alison Noble, and et al., "Improving the classification accuracy of the classic rf method by intelligent feature selection and weighted voting of trees with application to medical image segmentation," in *MLMI*. Springer, 2011, pp. 184–192.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[5] Hariharan Ravishankar, Sahana M Prabhu, and et al., "Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning," in *ISBI*. IEEE, 2016, pp. 779–782.

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[7] Hao Chen, Yefeng Zheng, Pheng-Ann Heng, and S Kevin Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," in *MICCAI*. Springer, 2016, pp. 487–495.

[8] Zhuowen Tu and Xiang Bai, "Auto-context and its application to high-level vision tasks and 3d brain image segmentation," *IEEE TPAMI*, vol. 32, no. 10, pp. 1744–1757, 2010.

[9] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] Yaozong Gao, Li Wang, Yeqin Shao, and Dinggang Shen, "Learning distance transform for boundary detection and deformable segmentation in ct prostate images," in *MLMI*. Springer, 2014, pp. 93–100.

[11] Herng-Hua Chang, Audrey H Zhuang, and et al., "Performance measure characterization for evaluating neuroimage segmentation algorithms," *Neuroimage*, vol. 47, no. 1, pp. 122–135, 2009.

[12] Olaf Ronneberger, Thomas Brox, and et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.