Quality Assessment of Fetal Head Ultrasound Images Based on Faster R-CNN

Zehui Lin[1,2,3], Minh Hung Le[1,2,3], Dong Ni[1,2,3], Siping Chen[1,2,3], Shengli Li[4], Tianfu Wang[1,2,3(✉)], and Baiying Lei[1,2,3(✉)]

[1] School of Biomedical Engineering, Shenzhen University, Shenzhen, China
{tfwang, leiby}@szu.edu.cn
[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen, China
[3] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China
[4] Department of Ultrasound, Affiliated Shenzhen Maternal and Child Healthcare, Hospital of Nanfang Medical University, Shenzhen, People's Republic of China

Abstract. Clinically, the transthalamic plane of the fetal head is manually examined by sonographers to identify whether it is a standard plane. This examination routine is subjective, time-consuming and requires comprehensive understanding of fetal anatomy. An automatic and effective computer aided diagnosis method to determine the standard plane in ultrasound images is highly desirable. This study presents a novel method for the quality assessment of fetal head in ultrasound images based on Faster Region-based Convolutional Neural Networks (Faster R-CNN). Faster R-CNN is able to learn and extract features from the training data. During the training, Fast R-CNN and Region Proposal Network (RPN) share the same feature layer through joint training and alternate optimization. The RPN generates more accurate region proposals, which are used as the inputs for the Fast R-CNN module to perform target detection. The network then outputs the detected categories and scores. Finally, the quality of the transthalamic plane is determined via the scores obtained from the numbers of detected anatomical structures. These scores detect the standard plane as well. Experimental results demonstrated that our method could accurately locate five specific anatomical structures of the transthalamic plane with an average accuracy of 80.18%, which takes only an approximately 0.27 s running time per image.

Keywords: Fetal head · Quality assessment · Ultrasound images
Faster R-CNN · Anatomical structure detection

1 Introduction

Ultrasound image has been preferred as an imaging modality for prenatal screening due to its noninvasive, real-time tracking, and low-cost. In prenatal diagnosis, it is important to obtain standard planes (e.g., the transthalamic plane) for prenatal ultrasound diagnosis. With the standard plane, doctors can measure the fetal physiological parameters to assess the growth and development of the fetus. Moreover, the weight of the fetus also can be obtained by measuring the parameters of biparietal diameter and

D. Stoyanov et al. (Eds.): POCUS 2018/BIVPCS 2018/CuRIOUS 2018/CPM 2018,
LNCS 11042, pp. 38–46, 2018.
https://doi.org/10.1007/978-3-030-01045-4_5

head circumference. This clinical practice is challenging for novices since it requires high-level clinical expertise and comprehensive understanding of fetal anatomy. Normally, ultrasound images scanned by novices are evaluated by experienced ultrasound doctors in the clinical practice, which is time-consuming and unappealing. To assist junior doctors by tracking the quality of the scanned image, automatic computer aided diagnosis for the quality assessment of ultrasound image is highly demanded. Accordingly, "intelligent ultrasound" [1] has become an inevitable trend due to the rapid development of image processing techniques. Powered by the machine learning and deep learning techniques, many dedicated research works have been proposed for this interesting topic, which mainly focus on the quality assessment of fetal ultrasound images to locate and identify the specific anatomical structures. For instance, Li *et al.* [2] combined Random Forests and medicine prior knowledge to detect the region of interest (ROI) of the fetal head circumference. Vaanathi *et al.* [3] utilized FCN architecture to detect the fetal heart in ultrasound video frames. Each frame is classified into three common standard views, e.g. four chamber view (4C), left ventricular outflow tract view (LVOT) and three vessel view (3V) captured in a typical ultrasound screening. Dong *et al.* [4] found the standard plane by fetal abdominal region localization in ultrasound using radial component model and selective search. Chen *et al.* [5] proposed an automatic framework based on deep learning to detect standard planes. The automatic framework achieved competitive performance and showed the potential and feasibility of deep learning for regions localization in ultrasound images. However, there are still lack of existing methods proposed under the clinical quality control criteria for quality assessment of fetal transthalamic plane in ultrasound images [6].

For quality control under the clinical criteria, the quality evaluation of the ultrasound images is scored via the number of the detected regions of important anatomical structures. The scores are given by comparing the detected region results with the bounding boxes annotated by doctors. Specifically, a standard transthalamic plane of fetal consists of 5 specific anatomical parts which can be clearly visualized, including lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP) and third ventricle (TV). The ultrasound map and the specific pattern of the fetal head plane including transthalamic plane, transventricular plane, transcerebellar plane are shown in Fig. 1. However, the ultrasound images of these three planes are very similar and the doctors are confusing. In addition, there are remaining challenges for quality assessment of the ultrasound images due to the following limitations: (1) The quality of ultrasound images is often affected by noise; (2) The anatomical structure's area is scanned in different magnification levels; (3) The scanning angle and the fetal location are unstable due to the rotation of the anatomical structure; (4) There are high variations in shapes and sizes of the anatomical structures among the patients.

To solve the above-mentioned challenges, we propose a deep learning based method for quality assessment of the fetal transthalamic plane. Specifically, our proposed method is based on the popular faster region-based convolutional network (Faster R-CNN [7]) technique. The remarkable ability of Faster R-CNN has been demonstrated in effectively learning and extracting discriminative features from the training images. Faster R-CNN is able to simultaneously perform classification and detection tasks. First, the images and the annotated ground-truth boxes are fed into Faster R-CNN. Then, Faster R-CNN generates the bounding boxes and the scores to

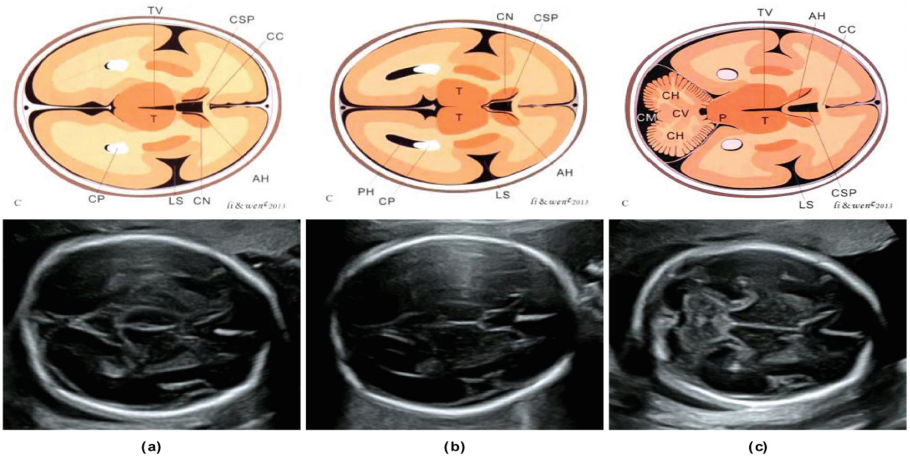

Fig. 1. The ultrasound map and the specific pattern of three fetal head plane. (a) transthalamic plane; (b) transventricular plane; (c) transcerebellar plane.

denote the detected regions and the quality of the detected regions, respectively. The output results are used to determine whether the ultrasound image is a standard plane. To the best of our knowledge, our proposed method is the first fully automatic deep learning based method for quality assessment of the fetal transthalamic plane in ultrasound images.

Overall, our contributions can be mainly highlighted as follows: (1) This is the first Faster R-CNN based method for the quality assessment of transthalamic plane of fetal; (2) The proposed framework could effectively assist doctors and reduce the workloads in the quality assessment of the transthalamic plane in ultrasound images; (3) Experimental results suggest that Fast R-CNN can be feasibly applied in many applications of ultrasound images. The proposed technique is generalized and can be easily extended to other medical image localization tasks.

2 Methodology

Figure 2 illustrates the framework of the proposed method for quality assessment of the fetal transthalamic plane. Faster R-CNN contains Fast R-CNN and RPN module. Images are cropped with a fixed-size of 224 × 224. The shared feature map, Fast R-CNN and RPN module of Faster R-CNN are explained in detail in this section.

2.1 Shared Feature Map

To achieve a fast detection while ensuring the accuracy of positioning results, the RPN module and Fast R-CNN [8] module share the first 5 convolutional layers of the convolutional neural network. However, the final effect and outputs of RPN and Fast R-CNN are different since the convolutional layers are modified in different ways.

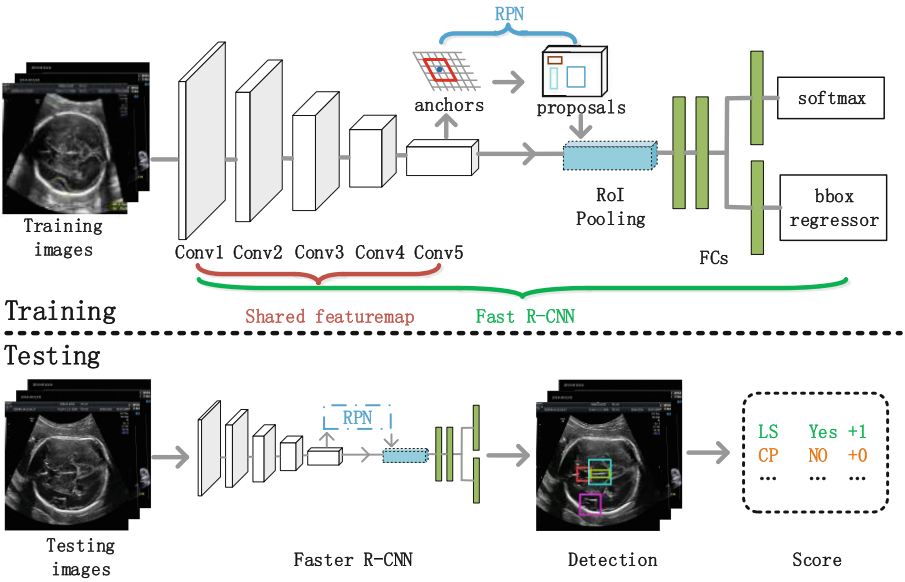

Fig. 2. The framework of our method based on Faster R-CNN.

At the same time, the feature map of the shared convolutional layer extraction must include the features required by both modules. This requirement cannot be easily obtained by just only using back propagation, which is in combination with the loss function optimization of the two modules. Fast R-CNN may not converge when the RPN could not provide fixed sizes of predicted bounding boxes.

To tackle the mentioned difficulties, Faster R-CNN learns the shared features through joint training and alternative optimization. Specifically, the pre-trained model of VGG16 is initialized and fine-tuned for training the RPN module. The generated bounding boxes are used as inputs to Fast R-CNN module. A separate detection network is then trained by Fast R-CNN. The pre-trained model of Fast R-CNN is the same as the pre-trained model of RPN module. However, these two networks are trained separately and do not share parameters. Next, the detection network is used to initialize the RPN training, but we fix the shared convolutional layer and only fine tune the RPN-specific layers. Then, we still keep the shared convolutional layer fixed and the RPN result is used to fine-tune the full connection layer of the Fast R-CNN module again. As a result, the two networks keep sharing the same convolutional layer until the end of the network training. Also, the detection and identification sets form a unified network.

2.2 Fast R-CNN Module

The structure of Fast R-CNN is designed based on R-CNN. In R-CNN, the processing steps (e.g., region proposal extraction, CNN features extraction, support vector machine (SVM) classification and box regression) are separated from each other that

causes the training process hardly to optimize the network performance. By contrast, the training process of Fast R-CNN is executed in an end-to-end manner (except for the region proposal step). Fast R-CNN directly adds an region of interest (ROI) pooling layer, which is essentially equivalent to the simplification of spatial pyramid pooling (SPP). With ROI layer, Fast R-CNN convolutes an ultrasound image only once. Then, it extracts feature from the original image and locates its region proposal boxes, which greatly improves the speed of the network. Fast R-CNN eventually outputs the localization scores and the detected bounding-boxes simultaneously.

Base Network: Fast R-CNN is trained on VGG16 and the network is modified to be able to receive both input images and the annotated bounding boxes. Fast R-CNN preserves 13 convolutional layers and 4 max pooling layers of the VGG-16 architecture. In addition, the last fully connected layer and softmax of VGG16 are replaced by two sibling layers.

ROI Pooling Layer: The last max pooling layer of VGG16 is replaced by an ROI pooling layer to extract the fixed-length of feature vectors from the generated feature maps. Fast R-CNN is able to convolute an image only once. It extracts feature from the original image and locates its region proposal boxes, which boosts the speed of the network. Since the size of the ROI pooling input is varying, each pooling grid size needs to be designed, which ensures that the subsequent classification in each region can be normally preceded. For instance, the input size of a ROI is $h \times w$, the output size of the pooling is $H \times W$, and the size of each grid is designed as $h/H \times w/W$.

Loss Function: Two output layers of Fast R-CNN include the classification probability score prediction for each ROI region *p,* and the offset for each ROI region's coordinate $t^u = \left(t_x^u, t_y^u, t_w^u, t_h^u\right), 0 \leq u \leq U$, where U is the number of object classes. The loss function of Fast R-CNN is defined as follows:

$$L = \begin{cases} L_{cls}(p,u) + \lambda L_{loc}(t^u, v), & \textit{if } u \textit{ is a structure}, \\ L_{cls}(p,u), & \textit{if } u \textit{ is a background}, \end{cases} \tag{1}$$

where L_{cls} is the loss function of the classification, and L_{loc} is the loss function for the localization. It is worthy mentioned that we do not consider the loss function of the bounding boxes location if the classification result is misclassified as the background. The loss function of L_{cls} is defined as follows:

$$L_{cls}(p,u) = \log p_u, \tag{2}$$

where L_{loc} is also described as the difference between the predicted parameter t^u corresponding to the real classification and the true translation scaling parameter υ. L_{loc} is defined as follows:

$$L_{loc}(t^u, v) = \sum\nolimits_{i=1}^{4} g\left(t_i^u - v_i\right), \tag{3}$$

where g is the smooth deviation, which is more sensitive to the outlier. g is defined as

$$g(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & otherwise. \end{cases} \tag{4}$$

2.3 RPN Module

The role of RPN module is to output the coordinates of a group of rectangular predicted bounding boxes. The implementation of RPN module did not slow down the training and detection process of the entire network because of the shared feature map. By taking the shared feature map as input of the RPN network, repetitive feature extraction is avoided and the calculation of regional attention is nearly cost-free. The RPN module performs convolution with a 3×3 sliding window on the incoming convolutional feature map and generate a 512-dimension feature matrix.

Then, RPN module also takes advantage of the principle of parallel output and accesses both branches after generating a 512-dimensional feature. The first branch is used to predict the upper left coordinates x, y, width w, and height h of the predicted bounding boxes corresponding to the central anchor points of the bounding boxes. For the diversity of predicted bounding boxes, the multi-scale method commonly is used in the RPN module. In order to obtain the more accurate predicted bounding boxes, the parameterizations of bounding box's coordinates are introduced. The second branch classifies the predicted bounding regions by the softmax classifier, which obtains a foreground bounding boxes and a background predicted bounding boxes (detection target is a foreground predicted bounding boxes). The last two branches converge at the FC layer, which is responsible for synthesizing the foreground predicted bounding box scores and the bounding box regression offsets, while removing the candidate boxes that are too small and out of bounds. In fact, the RPN module can get about 20,000 predicted bounding boxes, but there are many overlapping boxes. Here, a non-maximum suppression method is introduced to set the Intersection over Union (IOU) to a threshold of 0.7, i.e., preserving only predicted bounding boxes with local maximum score not exceeding 0.7. Finally, RPN module only passes 300 bounding boxes with higher score to the Fast R-CNN module. The RPN module not only simplifies the network input and improves the detection performance, but also enables the end-to-end training of the entire network, which is important for performance optimization.

3 Experiments

3.1 Dataset

The ultrasound images, which contain one single fetus, are collected from a local hospital. The gestation age of the fetus varies from 14 to 28 weeks. The most clearly visible images are selected in the second trimester. As a result, a total of 513 images which clearly visualize the 5 anatomical structures of LS, CP, T, CSP and TV are selected.

Due to the diversity of image sizes in the original dataset, the images are resized to 720 × 960 for further processing. Since the training for Faster R-CNN requires a large number of images, we increase the numbers and varieties of images by adopting a commonly used data augmentation method (e.g., random cropping, rotating and mirroring). As a result, a total of 4800 images are finally selected for training and the remaining 1153 images are used for testing. All the training and testing images are annotated and confirmed by an 8 years clinical experienced ultrasound doctor. All experiments are performed on a computer with CPU Inter Xeon E5-2680 @ 2.70 GHz, GPU NVIDIA Quadro K4000, and 128G of RAM.

3.2 Results

The setting of the training process is kept the same whenever possible for fair comparison. Recall (Rec), Precision (Prec) and Average Precision (AP) are used as performance evaluation metrics. We adopt 2 popular object detection methods including Fast R-CNN and Yolov2 [10] for performance comparisons. Table 1 summarizes the experimental results of each network. We observe that the detection results for single anatomical structure of the LS and CP are the best. This is because LS and CP have distinct contour, moderate size with high contrast and less surrounding interference. Another reason is that LS and CP classes contain more training samples than other classes, making the detection biased to detect these classes and misdetect other classes. The results of TV are quite low due to its blurry anatomical structure, small size, and structure similarity of other tissues.

Table 1. Comparison of the proposed method with other methods (%).

Method	Value	LS	CP	T	CSP	TV
Fast R-CNN	Rec	87.6	63.7	62.6	44.2	–
	Prec	84.7	57.0	60.8	29.3	–
	AP	70.6	36.3	39.5	19.8	–
YOLOv2	Rec	90.4	83.7	34.7	48.6	4.2
	Prec	99.6	97.2	79.9	94.1	85.2
	AP	90.3	82.9	30.3	46.9	3.6
Faster R-CNN (VGG16)	Rec	96.8	96.0	89.6	89.3	56.5
	Prec	96.6	96.7	77.1	94.6	72.8
	AP	**94.9**	**93.8**	**81.0**	**87.1**	**44.1**

Generally, the detection performance of Faster R-CNN is better than Fast R-CNN and Yolov2. In particular, Faster R-CNN has significantly improved the detection performance of TV. The running time per image from Fast R-CNN, YOLOv2, and Faster R-CNN is 2.7 s, 0.0006 s, and 0.27 s, respectively. Although the running time of Faster R-CNN is not the fastest, its speed still satisfies the clinical requirements.

Figure 3 shows the structure localization results using the proposed technique compared with other methods. The green, red, yellow, blue and green bounding boxes

indicate the LS, CP, T, CSP and TV, respectively. As shown in Fig. 3, our method can simultaneously locate multiple anatomical structures in an ultrasound image and achieve the most superior localization results.

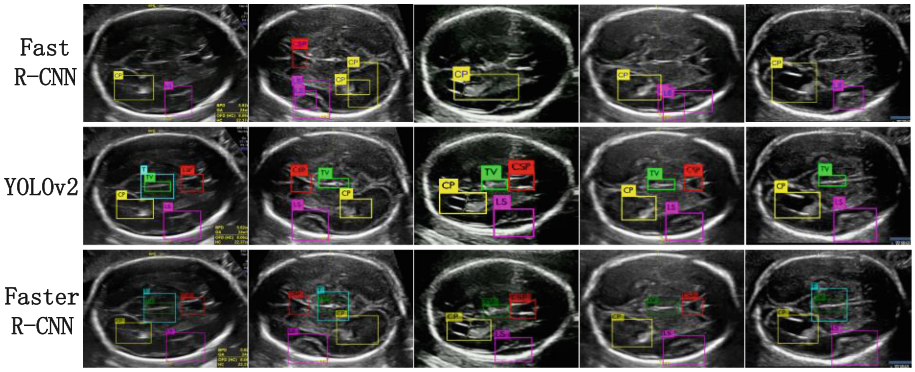

Fig. 3. The detection results of Fast R-CNN, YOLOv2, and Faster R-CNN (VGG16), respectively. The purple, yellow, cyan, red, and green boxes locate the lateral fissure, choroid plexus, thalamus, transparent compartment, and third ventricle, respectively. (Color figure online)

4 Conclusion

In this paper, we propose an automatic detection technique for quality assessment of fetal head in ultrasound images. We utilize Faster R-CNN to automatically locate five specific anatomical structures of the fetal transthalamic plane. Accordingly, the quality of the ultrasound image is scored and the standard plane is determined based on the number of detected regions. Experimental results demonstrate that it is feasible to employ deep learning for the quality assessment of fetal head ultrasound images. This technique can be also extended to many ultrasound images tasks. Our future work will tackle the existing problem of inhomogeneity of image contrast in ultrasound images, which will apply intensity enhancement method to enhance the contrast between the anatomical structures and the background. The clinical prior-knowledge will be utilized to achieve better detection and localization.

Acknowledgement. This work was supported partly by National Key Research and Develop Program (No. 2016YFC0104703).

References

1. Namburete, A., Xie, W., Yaqub, M., Zisserman, A., Noble, A.: Fully-automated alignment of 3D Fetal brain ultrasound to a canonical reference space using multi-task learning. Med. Image Anal. **46**, 1 (2018)
2. Li, J., et al.: Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. IEEE J. Biomed. Health Inf. **17**, 1–12 (2017)
3. Sundaresan, V., Bridge, C.P., Ioannou, C., Noble, J.A.: Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: ISBI, pp. 671–674 (2017)
4. Ni, D., et al.: Standard plane localization in ultrasound by radial component model and selective search. Ultrasound Med. Biol. **40**, 2728–2742 (2014)
5. Chen, H., et al.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J. Biomed. Health Inf. **19**, 1627–1636 (2015)
6. Li, S., et al.: Quality control of training prenatal ultrasound doctors in advanced training. Med. Ultrasound Chin. J. **6**, 14–17 (2009)
7. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**, 1137–1149 (2015)
8. Girshick, R.: Fast R-CNN. In: CVPR, pp. 1440–1448 (2015)