2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Fetal Heart and Descending Aorta Detection in Four-Chamber View of Fetal Echocardiography

Shan An[1], Jing Lv[2], Haogang Zhu[1,†], Jingyi Wang[2], Xiaoxue Zhou[2], Qining Liu[1], Yier Shu[1], Zhengyu Liu[1], Yingying Zhang[1], Xiangyu Liu[1] and Yihua He[2]

***Abstract*— Automatic analysis of fetal heart and related components in fetal echocardiography can help cardiologists to reach a diagnosis for Congenital Heart Disease (CHD). Previous studies mainly focused on cardiac chamber segmentation, while few researches deal with the cardiac component detection. In this paper, we tackle the task of simultaneous detection of the fetal heart and descending aorta in four-chamber view of fetal echocardiography, which is useful to analyze some kinds of CHD, such as left/right atrial isomerism, dextroversion of heart, etc. Several CNN-based object detection methods with different backbones are thoroughly evaluated, and finally, the Hybrid Task Cascade method with HRNet is selected as the detection method. Experiments on a fetal echocardiography dataset show that the method can achieve superior performance according to common-used evaluation metrics.**

***Clinical relevance*—This can be used to help the cardiologists to estimate the position of the fetal heart and the descending aorta, which is also useful to estimate the direction of the cardiac axis and apex and analyze some kinds of CHD, such as left/right atrial isomerism, dextroversion of heart, etc.**

I. INTRODUCTION

Congenital heart disease (CHD) is the most important factor in fetal death [1]. Fetal echocardiography can help doctors diagnose congenital heart disease. Segmentation of the atria and the ventricles in a four-chamber view enhances the doctors' ability make a correct diagnosis. The descending aorta (DA) is an important component of the four-chambers view of a fetal heart, located next to the left atrium(LA), as shown in Fig. 1. There are two benefits to detect DA: one is to help the cardiologists to estimate the position of LA and then position of the four cardiac chambers; the other is to estimate the direction of the cardiac axis and apex. Furthermore, automatic and simultaneous localization of the fetal heart and DA could provide additional information to help the cardiologists to analyze some kinds of CHD, such as left/right atrial isomerism, dextroversion of heart, etc.

In the past few years, with the application of deep learning techniques, there have been some works to segment the cardiac chambers of the fetal cardiogram. However, there are few studies on fetal heart and DA detection. We are interested in the simultaneous localization of the fetal heart and DA, because the location of the fetal heart and DA could help the cardiologists to a certain extent.

We tackle the problem to accurately and simultaneously detect and locate fetal heart and DA. Three state-of-the-art methods with three different backbones for general object detection are evaluated. The experimental results show the superior performance of Hybrid Task Cascade(HTC) [2] with HRNet [3] as the backbone, which could be a starting point for future improvements.

The paper is organized as follows: In Section 2, we give an overview of the related work. Section 3 discusses the dataset and the technical details of the framework. The experimental results are presented in Section 4. Finally, we conclude the paper in Section 5.

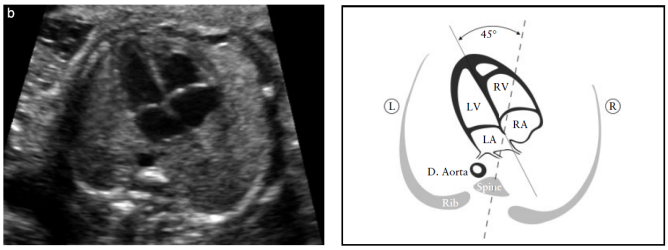

Fig. 1. An illustration of cardiac position and axis in four-chamber view of fetal echocardiography [1]. The descending aorta (DA) is near the left atrium (LA).

II. RELATED WORK

Generic Object Detection: In object detection, based on R-CNN [4], which combines a proposal detector and a region-wise classifier, the Faster R-CNN [5] introduced a region proposal network (RPN) then has accelerated the CNN computations significantly. The FPN [6], who addressed to the scale-invariant problem, detected high-recall proposals at multiple output layers by using a top-down architecture with lateral connections to achieve high-level semantic feature maps at all scales. Cascade R-CNN [7] proposed a multi-stage architecture composed of a sequence of detectors trained with increasing IoU thresholds for high-quality detection. Hybrid Task Cascade [2] introduced cascade to instance segmentation by performing jointly cascaded refinement on segmentation and detection tasks. One-stage architecture is proposed to realize real-time detection, but with reduced

[1]Shan An, Haogang Zhu, Qining Liu, Yier Shu, Zhengyu Liu, Yingying Zhang and Xiangyu Liu are with School of Computer Science and Engineering, Beihang University, Beijing, China, 100191, (e-mail: haogangzhu@buaa.edu.cn).

[2]Jing Lv, Jingyi Wang, Xiaoxue Zhou and Yihua He are with Beijing Anzhen Hospital affiliated to Capital Medical University, Beijing, China, 100029.

[†]Corresponding Author

accuracy. So we will eventually compare the two-stage and multi-stage methods.

Anatomy localization and detection in echocardiography: In echocardiography image processing domain, many studies are focused on the segmentation problem, while few methods are applied to detection and localization. For segmentation tasks in echocardiographic images, Y. Hu et al. [8] present a convolutional network model based on BiSeNet to segment left ventricle and left atrium without manual intervention. As for echocardiography detection and localization, a Naïve-Bayes classifier model [9] was used for mitral valve annulus localization. It demands a prior manual aortic reference for initialization use. D. Bibicu et al. [10] propose an artificial Neural Network method for detection of the left ventricle without using any prior knowledge of the shape. There are only one research using CNN method for anatomy detection in echocardiography, which uses Single Shot Multibox Detector (SSD) [11] and Faster R-CNN [5] to detect Aortic Valve [12].

III. METHODS

A. Dataset

A dataset of fetal echocardiography, which contains echocardiographic sequences with a four-chamber view of 319 fetuses, is collected following the requirements according to the ethic committee. This dataset contains fetal echocardiography of healthy fetuses and unhealthy fetuses with six types of fetal heart diseases, which are Ebstein's Anomaly (EA), Cardiac Rhabdomyomas (CR), Atrioventricular Septal Defect (AVSD), Hypoplastic Left Heart Syndrome (HLHS), Pulmonary Atresia with Intact Ventricular Septum (PA/IVS) and Total Anomalous Pulmonary Venous Connection (TAPVC). Twelve cardiologists participate in the manual annotation of the contour of the cardiac chambers and the descending aorta. Each echocardiographic sequence has two frames of annotation, one in the end-diastolic and the other in the end-systolic. Therefore, the overall number of labeled frames is 638.

B. Model Architecture

We use Hybrid Task Cascade (HTC) [2] as the detection model. HTC has proposed a multi-stage and multi-task hybrid cascade structure, and a branch of semantic segmentation is integrated to enhance the spatial context. At first, HTC combines Cascade RCNN [7] and Mask R-CNN [13], which strongly improves the box AP of mask R-CNN. Then by adding the interleaved execution module which obtains the mask using box information, and mask information flow, HTC has shown great progress in segmentation task. Finally, HTC introduces a semantic segmentation subnet and semantic feature fusion to get the better spatial context; this method further boosts the performance for detection.

Previous works in the object detection domain have shown that good segmentation and high-resolution representation can improve the performance of detection. Using parallel high-to-low resolution subnetworks and multi-scale fusion,

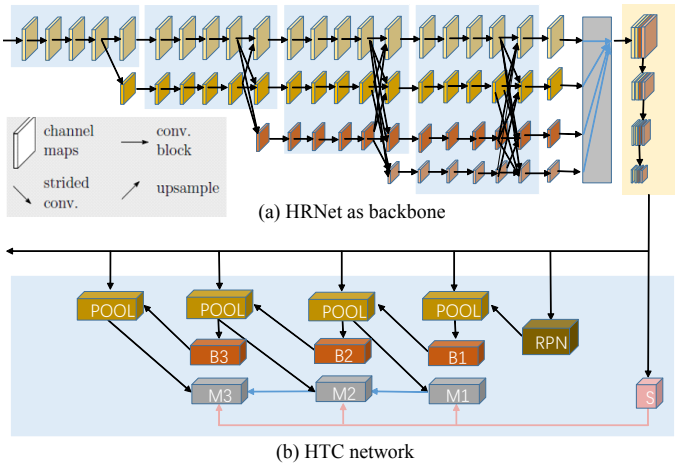

Fig. 2. The model architecture of HTC with HRNet backbone.

HRNet [3] has successfully learned the high-resolution representation for the input image and demonstrated its superiority on various detection datasets. Therefore, our approach uses HRNet as the backbone. This design enables the model to better localize the object, especially the small object (DA in our dataset) by high-resolution feature and fine-grained box. The network structure is shown in Fig. 2.

C. Data Augmentation

The echocardiology datasets are commonly limited, because of the complexity of manual annotation by the cardiologists. However, for the training of the deep neural network, a large number of annotated medical images are demanded. Therefore, we perform the data augmentation strategies to increase the diversity of the data. We adopt the online data augmentation method. During training, half of data are randomly selected to perform a horizontal flip, then scaled with a ratio randomly selected in range 0.9 to 1.1. Finally, with 50% probability, the data are rotated -10 to 10 degrees. We believe that rotation and scale operation on an image can mimic the movements of the fetal heart. All the methods above are demonstrated to result in positive effects in optimizing evaluation metrics.

D. Implementation Details

The proposed method is implemented using pytorch. All the models are trained and tested on a Tesla P40 GPU. For backbones, we use the pre-trained weights of ImageNet [16] dataset and COCO [17] dataset. The optimizer is stochastic gradient descent(SGD) with a momentum of 0.9 and a weight decay of 0.0001. The base learning rate is 0.0025, and the training epoch is 24.

IV. EXPERIMENTS AND RESULTS

To evaluate the detection accuracy of different methods, we use three commonly used metrics in the experiments. The ratio between the intersection and the union of the predicted boxes and the ground truth boxes is defined as Intersection over Union (IoU). The metric AP_{50} means the detection is

TABLE I

RESULTS OF FETAL HEART AND DA DETECTION WITH AND WITHOUT DATA AUGMENTATION.

Method	with scale	with rotation	box AP(DA)	AP_{50}(DA)	AP_{75} (DA)	box AP(HEART)	AP_{50}(HEART)	AP_{75} (HEART)
HTC+HRNet	✗	✗	0.327	0.738	0.191	0.690	0.979	0.866
	✓	✗	0.355	0.807	0.217	**0.696**	0.980	0.826
	✗	✓	0.340	0.736	0.242	0.686	0.987	**0.871**
	✓	✓	**0.364**	**0.816**	**0.244**	0.677	**0.997**	0.861

TABLE II

RESULTS OF FETAL HEART AND DA DETECTION USING DIFFERENT SETTINGS

Method	Center of Rotation	With Expanding	box AP(DA)	AP_{50}(DA)	AP_{75} (DA)	box AP(HEART)	AP_{50} (HEART)	AP_{75}(HEART)
HTC+HRNet	(240,0)	✗	0.356	0.806	0.225	0.685	0.989	0.834
	(240,180)	✗	**0.364**	**0.816**	0.244	0.677	**0.997**	**0.861**
	(240,180)	✓	0.363	0.788	**0.292**	**0.692**	0.986	0.842

TABLE III

RESULTS OF FETAL HEART AND DA DETECTION USING DIFFERENT METHODS

Method	Backbone	box AP(DA)	AP_{50}(DA)	AP_{75} (DA)	box AP(HEART)	AP_{50} (HEART)	AP_{75}(HEART)	#frame/s
Faster R-CNN [5]	ResNet [14]	0.305	0.744	0.205	0.612	0.996	0.666	**21.2**
	ResNeXt [15]	0.290	0.778	0.189	0.632	0.980	0.758	12.8
	HRNet [3]	0.334	0.798	0.179	0.619	0.980	0.688	7.8
Cascade R-CNN [7]	ResNet [14]	0.296	0.733	0.167	0.623	0.970	0.767	13.8
	ResNeXt [15]	0.314	0.783	0.172	0.652	0.970	0.808	9.8
	HRNet [3]	0.341	0.758	0.213	0.679	0.980	0.818	6.8
Hyrbrid Task Cascade [2]	ResNet [14]	0.343	0.765	0.228	0.640	0.996	0.778	6.7
	ResNeXt [15]	0.352	0.768	0.227	**0.684**	0.980	0.782	6.6
	HRNet [3]	**0.364**	**0.816**	**0.244**	0.677	**0.997**	**0.861**	5.2

considered as True Positive if the IoU is above 0.5, and AP_{75} means the IoU is above 0.75. The metric box AP computes the average precision value for recall value over 0 to 1. AP_{50} is important for the DA detection, because the size of DA is relatively small, we can easily identify the DA when the detected box is 50% overlapped with the GT box. Besides, we think AP_{75} is important for fetal heart detection, for the reason that the fetal heart is large in the images, which should be accurately localized. The fetal echocardiography dataset is randomly divided into the train set and the test set according to subjects with a ratio of 0.8 and 0.2.

We conduct several experiments using three object detection methods, which are Faster R-CNN [5], Cascade R-CNN [7], and Hybrid Task Cascade [2]. Three different CNN networks are incorporated as backbones, which are ResNet [14], ResNeXt [15], and HRNet [3].

Firstly, we compare different data augmentation methods using the model described in Section III-B. In Table I, we can see that using only scaling or only rotation for data augmentation, the detection accuracy of DA is improved compared with no augmentation. Furthermore, we can see that with data augmentation, box AP, AP_{50} and AP_{75} are higher than the method without augmentation, which demonstrates the effectiveness of our data augmentation method. For fetal heart detection, using only rotation results in a higher AP_{75} than using both scaling and rotation. Because the detection accuracy of DA is significantly improved, we use both scaling and rotation for data augmentation in the following experiments.

Secondly, we try to use different points in the image as the center of rotation. The size of the images of our fetal echocardiography is $480 * 360$. One way is to use the center of the image, which is $(240, 180)$, as the center of rotation. Another way is to use $(240, 0)$ as the center of rotation. In Table II, it can be seen that the former approach performs better than the latter.

Thirdly, the ground truth box is expanded to $1.1\times$ larger than the original. We hope to include more information in the larger box. As shown in Table II, AP_{50} for DA and AP_{75} for heart are both better when no expanding deployed.

The complete results of three methods with different backbones are shown in Table III. For DA detection, AP_{50} reaches 0.816 using HTC with HRNet, which means the algorithm can accurately detect the DA. For fetal heart detection, AP_{75} is 0.861 when using HTC with HRNet, which also means the algorithm is good for fetal heart detection. We show several detection results in Fig. 3, which demonstrate the good segmentation performance of our method.

We also evaluate the detection time of different methods. It can be seen in Table III, Faster R-CNN with ResNet as the backbone is the fastest method among these methods, which can process about 21.2 frames per second. HTC with HRNet as the backbone is the slowest, which cost nearly 200 milliseconds to detect cardiac components in an image.

According to the extensive experiments, we can see among these generic object detection methods, HTC with HRNet gets the best detection accuracy in this specific task, which could be used as a starting point for the future improvements.

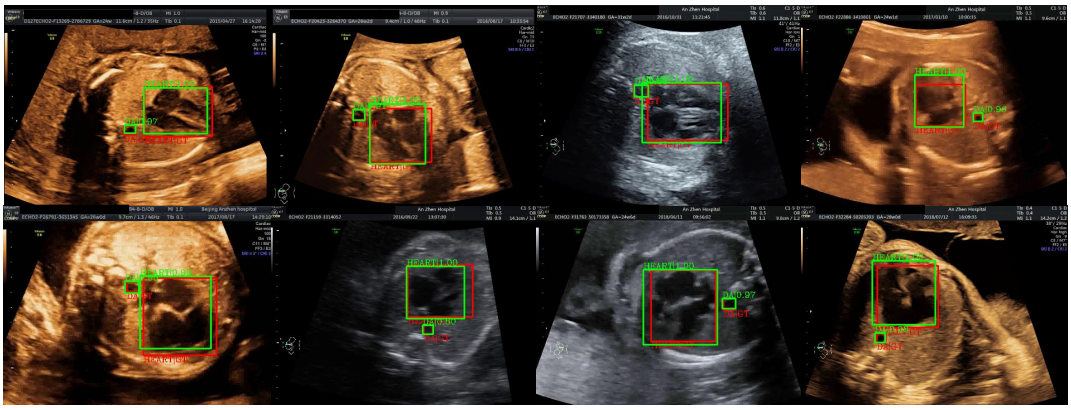

Fig. 3. These are eight detection results. For each result, the red box indicates the ground truth (GT) bounding box, and the green box indicates the detected bounding box. The larger object is the fetal heart and the smaller one is the DA.

V. CONCLUSIONS

Fetal heart and DA detection will provide more information for cardiologists to diagnose some types of CHD. We evaluated the mainstream generic object detection methods on a fetal echocardiography dataset of 319 fetuses. The experimental results show that Hybrid Task Cascade with HRNet backbone has the ability to detect fetal heart and DA accurately and simultaneously. In the future, we plan to employ Generative Adversarial Networks (GAN) to generate new images with bounding boxes for the small-data object detection in fetal echocardiography.

VI. ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China under Grant 2020YFC2006200, the Major Project of Science and Technology of Yunnan Province under Grant No. 2019ZE005, the Beijing Municipal Science & Technology Commission under Grant Z181100001918008, and the Beijing Lab for Cardiovascular Precision Medicine under Grant PXM2018014226000013 (*Corresponding author: Haogang Zhu*).

REFERENCES

[1] J.S. Carvalho, L.D. Allan, and et al., "ISUOG practice guidelines (updated): Sonographic screening examination of the fetal heart," *Ultrasound in Obstetrics and Gynecology*, vol. 41, no. 3, pp. 348–359, 2013.

[2] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al., "Hybrid task cascade for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4974–4983.

[3] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al., "Deep high-resolution representation learning for visual recognition," *arXiv preprint arXiv:1908.07919*, 2019.

[4] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[6] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[7] Zhaowei Cai and Nuno Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6154–6162.

[8] Yujin Hu, Libao Guo, Baiying Lei, Muyi Mao, Zelong Jin, Ahmed Elazab, Bei Xia, and Tianfu Wang, "Fully automatic pediatric echocardiography segmentation using deep convolutional networks based on bisenet," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6561–6564.

[9] Abhishek Tiwari and Kedar A Patwardhan, "Mitral valve annulus localization in 3d echocardiography," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1087–1090.

[10] Dorin Bibicu and Luminita Moraru, "Cardiac cycle phase estimation in 2-d echocardiographic images using an artificial neural network," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1273–1279, 2012.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[12] Muhammad Hanif bin Ahmad Nizar, Chow Khuen Chan, Ahmad Khairuddin Mohamed Yusof, Azira Khalil, and Khin Wee Lai, "Detection of aortic valve from echocardiography in real-time using convolutional neural network," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2018, pp. 91–95.

[13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[15] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, 2009, pp. 248–255.

[17] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft COCO: Common objects in context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755.