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

Simultaneous Segmentation of Four Cardiac Chambers in Fetal Echocardiography

Shan An[1,*], Xiaoxue Zhou[2,*], Haogang Zhu[1,†], Fangru Zhou[1], Yuduo Wu[2], Tingyang Yang[1], Xiangyu Liu[1], Yingying Zhang[1], Zhicheng Jiao[3] and Yihua He[2]

Abstract— **Accurate segmentation of cardiac chambers is helpful for the diagnosis of Congenital Heart Disease (CHD) in fetal echocardiography. Previous studies mainly focused on single cardiac chamber segmentation, which cannot provide sufficient information for the cardiologists. In this paper, we present an instance segmentation approach capable of segmenting four cardiac chambers accurately and simultaneously. A novel object proposal recovery strategy is further deployed to retrieve possible missing objects. To alleviate the shortage of medical data and further improve the segmentation performance, we utilize a rotation and distortion method for data augmentation. Experiments on a fetal echocardiography dataset of 319 fetuses demonstrate that the proposed approach can achieve superior performance according to common-used evaluation metrics.**

Clinical relevance—**This can be used to help the cardiologists to better analyze the structure and function of the fetal heart.**

I. INTRODUCTION

Congenital heart disease (CHD) is one or more abnormalities in the heart's structure and function present at birth. The analysis of the fetal echocardiography images will be useful for cardiologists to measure the fetal heart's size and function and make a heart disease diagnosis [1]. The segmentation of four cardiac chambers can be used as auxiliary information to diagnose several kinds of CHD.

In the past few years, there have been many studies on cardiac chambers segmentation in echocardiography [2], [3], [4]. However, several studies have focused on single ventricular segmentation, such as the left ventricle (LV) [2], the left atrium (LA) [3], and the right ventricle (RV) [4] segmentation. There are few studies on the simultaneous segmentation of the four cardiac chambers. In DW-Net [5], the segmentation of four cardiac chambers is achieved through multi-scale contextual information aggregation. Furthermore, most of the work focused on adult patients [2], [3], [4], [6], [7], but less on fetal echocardiography [8].

This paper proposes a generic and novel system framework for cardiac chamber segmentation in fetal echocardiography. Unlike previous research, which uses semantic segmentation method [6], [7], we utilize the instance segmentation to simultaneously segment four cardiac chambers and a novel object proposals recovery strategy to retrieve missing parts. This method will help cardiologists better to analyze the structure and function of the fetal heart.

The paper is organized as follows: In Section 2, we describe related work. Section 3 discusses the technical details of the framework. Experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

II. RELATED WORK

The echocardiographic segmentation methods can be divided into two main categories: non-deep learning techniques and deep learning techniques. Early methods use the formal, while the more recent methods use the latter. Early echocardiographic segmentation methods used non-deep learning techniques. A level set deformable model was proposed to segment four cardiac chambers using region-based information simultaneously [9]. Similar method proposed in [10] can automatically segment the small fetal cardiac chambers. This paper will focus on the use of deep learning techniques for echocardiographic segmentation. In 2012, a two-stage deep learning method for the LV segmentation in four-chamber view was proposed [11]. The results showed the robustness to imaging conditions and shape variations compared to traditional methods, such as level-sets and deformable templates. A multi-texture active appearance model (AAM) based on the Hermite transform (HT) was proposed to achieve an efficient segmentation of the LV in fetal echocardiography [12]. ACNN [7] utilized anatomical prior knowledge for cardiac image segmentation and enhancement. In [6], six different methods were evaluated using an open large-scale dataset (CAMUS: Cardiac Acquisitions for Multi-structure Ultrasound Segmentation), which concluded that encoder-decoder based neural network outperforms the non-deep learning methods.

Recently, a method [8] focuses on fetal cardiac structure localization problem, which aggregates global spatial context and detects anatomical structures on spatial region proposals. However, this method can only localize anatomical structures without pixel-level segmentation. A semi-supervised learning algorithm was proposed in [13], which used conditional deep generative models as prior to improve the performance of LV segmentation. A Bilateral Segmentation Network (BiSeNet) [14] was used to automatic segment pediatric echocardiography images. DW-Net [5] performs apical four-chamber view segmentation on healthy hearts, which is easier

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

[2]Xiaoxue Zhou, Yuduo Wu and Yihua He are with Beijing Anzhen Hospital affiliated to Capital Medical University, Beijing, China, 100029.

[3]Zhicheng Jiao is with Perelman School of Medicine at University of Pennsylvania, PA, USA.

*Equal contribution

†Corresponding Author

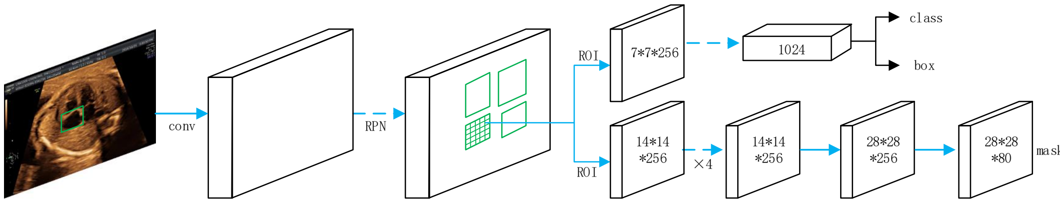

Fig. 1. The segmentation network architecture

than the segmentation of abnormal hearts in this study. This paper studies the cardiac chambers segmentation in fetal echocardiography, which is more complicated than the adult and pediatric echocardiography.

III. METHODOLOGY

Our work formalizes the cardiac chambers segmentation as an instance segmentation problem, which utilizes Mask R-CNN [15] to segment four cardiac chambers simultaneously. We propose using a rotation and distortion strategy for data augmentation and an object proposals recovery strategy to obtain better results.

A. Data Augmentation

Due to the complexity of manual annotation by the cardiologists, the echocardiology datasets commonly have less than one thousand annotated images, insufficient for deep neural network training. Therefore, data augmentation strategies should be performed to increase the diversity of the data.

There are recurrent rhythmical movements in the periodic expansion and contraction of the fetal heart. We propose to utilize a rotation and a distortion operation on images to mimic the fetal heart's movements. The rotation operation can be realized by applying an affine transformation to an image. We adopt randomized and elastic distortions on images.

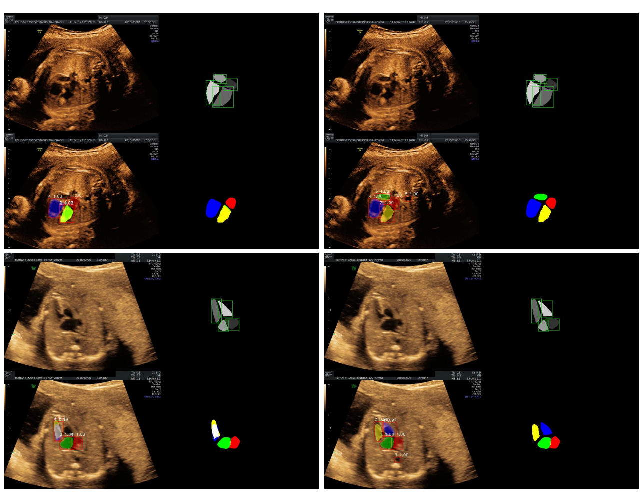

Fig. 2. Two segmentation results with (right) and without (left) our object proposals recovery strategy. In the bottom, the echocardiagraphy shows a complex congenital heart disease: dextrocardia with inverted ventricular.

B. Segmentation Network

Instance segmentation is the task of assigning a class-aware and instance-aware label to each pixel of the image. To segment four cardiac chambers, we utilize Mask R-CNN [15] to produce accurate segmentation masks. Mask R-CNN is based on Faster R-CNN [16], which has two outputs, a class label, and a bounding box offset. A mask branch is added for predicting segmentation masks on each Region of Interest (ROI). A RoIAlign layer is used to replace the commonly used RoIPool layer to extract an object's finer spatial layout. The network architecture is shown in Fig. 1.

C. Object Proposals Recovery

One heart chamber class should appear once in an echocardiographic frame because we only reserve the object proposal with the highest score in each class. Some heart chambers may be lost in the final segmentation results of Mask R-CNN. This is because an object with a higher score may be incorrectly classified as a specific class and edge out the object with a lower score but belongs to this class. We propose a novel strategy to tackle this problem. Firstly, the center of each ROI and L2 distances between centers are computed. Secondly, if two centers' distance is below a threshold of ε, we treat these ROIs as a cluster. ε is set to an empirical value since the class boundary is clearly distinguishable. Finally, if a cluster's ROIs belong to the same class, they are labeled to this class. Otherwise, we calculate the overall scores of these ROIs and decide the class according to the largest score. If a class exists in the previous procedure, it will not be considered in the score computation step. Using this strategy, we can retrieve the missing part and obtain better results, as shown in Fig. 2.

We assume a ROIs cluster contains N ROIs, and the score of each ROI in the four categories is $\{s_i^1, s_i^2, s_i^3, s_i^4\}, i = 1 \ldots N$, then the score of ROIs for category j is:

$$S^j = \frac{1}{N}\sum_i^N s_i^j \tag{1}$$

Then we compute each ROIs cluster's four category scores and rank all category scores of all ROIs clusters. Finally, Algorithm 1 generates the categories of the clusters, which are regarded as the categories of all ROIs in the clusters. In this way, the predictions of the network are updated by Algorithm 1.

TABLE I

THE DETAILS OF THE FETAL ECHOCARDIOGRAPHY DATASET

Types of fetal heart disease	Train set(#frames)	Test set(#frames)
Healthy	142	16
Ebstein's Anomaly (EA)	28	4
Cardiac Rhabdomyomas (CR)	87	12
Atrioventricular Septal Defect (AVSD)	114	14
Hypoplastic Left Heart Syndrome (HLHS)	73	6
Pulmonary Atresia with Intact Ventricular Veptum (PA/IVS)	62	8
Total Anomalous Pulmonary Venous Connection (TAPVC)	66	6
Total	572	66

Algorithm 1

Input: ROIs cluster's number M; *score*, all ROIs clusters all categories score indices, whose size is $M \times 4$; $ROIs\_id$, indices of ROIs after sorting by scores, whose size is $M \times 4$; *Class*, indices of categories after sorting by scores, whose size is $M \times 4$.

Output: $ROIs\_class$: category of every ROIs cluster.

initialize every ROIs cluster's category to -1, $ROIs\_class \leftarrow -ones(M)$.
for $i = 1$ to $M * 4$ **do**
    if $ROIs\_class[ROIs\_id[i]] == -1$ **then**
        $ROIs\_class[ROIs\_id[i]] = Class[i]$
    end if
    if $ROIs\_class.min() > -1$ **then**
        break
    end if
end for

D. Implementation Details

The proposed method is implemented using PyTorch and trained on an NVIDIA P40 GPU. We use the pre-trained weights of the ImageNet dataset and COCO dataset. The models are then fine-tuned on our fetal echocardiography 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 0.01, and the training epoch is 100. Other loss functions are the same as that in Mask R-CNN [15]. The threshold ε for object proposals recovery is empirically set as 10 pixels.

IV. EXPERIMENTS AND RESULTS

We will first introduce the clinical datasets and give the evaluation metrics used in our experiments. Then we evaluate our method using different settings and demonstrate the superior performance of our method.

A. Datasets

A dataset of fetal echocardiography is collected, which contains echocardiographic sequences with four-chamber views of 319 fetuses. The dataset was acquired using two types of equipment, the General Electric Voluson E8 ultrasound system with transabdominal 2-to 4-MHz curvilinear transducers (GE Healthcare Ultrasound, Milwaukee, WI, USA) or the Aloka SSD ultrasound system (Aloka, Tokyo, Japan) using transabdominal 3-to 6-MHz curvilinear transducers. All ultrasound examinations were performed by experienced operators in the same medical center. 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). The cases in this dataset include most types of congenital heart disease. More than ten cardiologists participate in the manual annotation of the contour of the cardiac chambers. Each fetal echocardiography sequence is annotated by one cardiologist and is examined by another cardiologist. There are two frames of annotation in each echocardiographic sequence, one is in the end-diastolic, and the other is in the end-systolic. This dataset is randomly divided into train set and test set, i.e., 286 sequences are used for training and 33 for testing. The overall number of labeled frames is 638, with the details of the dataset given in Table I.

In the following experiments, we use the data augmentation methods described below. The train set images are rotated 360°, with the interval of 30°, which generates 12x more images. The distortion operations are performed multiple times, with different times for different kinds of disease, to balance the overall images in each category.

B. Evalutation metrics

To evaluate the accuracy of our method's segmentation output, we use four metrics in the experiments, which are frequently used in medical image segmentation. Specifically, the dice similarity coefficient (DSC), sensitivity, specificity, and Hausdorff distance (HD) are evaluated.

The dice similarity coefficient (DSC) is defined as:

$$D = 2 \times \frac{|S_{seg} \cap S_{gt}|}{|S_{seg} + S_{gt}|} \tag{2}$$

S_{seg} is the area of the segmentation result of a method and S_{gt} is the area of the ground truth. The DSC is a measure of overlap between the ground truth and the segmented result, with 1.0 represents perfect overlap, and 0.0 represents no overlap.

TABLE II

THE SETTINGS OF OUR METHOD

Different Settings	M-1	M-2	M-3	M-4
ImageNet pre-trained	✓	✓	✓	✓
COCO pre-trained			✓	✓
with data augmentation		✓	✓	✓
with recovery strategy				✓

TABLE III

RESULTS OF FOUR CARDIAC CHAMBERS SEGMENTATION.

Parts	Metrics	M-1	M-2	M-3	M-4
LA	Dice coefficient ↑	0.6886	0.6780	0.7064	**0.7192**
	Sensitivity ↑	0.6835	0.6836	0.7113	**0.7305**
	Specificity ↑	0.9972	**0.9982**	0.9978	0.9978
	Hausdorff distance ↓	18.9924	22.2033	17.0644	**16.7062**
LV	Dice coefficient ↑	0.6927	**0.7005**	0.6653	0.6759
	Sensitivity ↑	0.6965	0.6911	0.6553	**0.6669**
	Specificity ↑	0.9961	0.9971	0.9980	**0.9980**
	Hausdorff distance ↓	**24.8708**	25.7863	29.2935	29.6767
RA	Dice coefficient ↑	0.7630	0.7711	0.7798	**0.7941**
	Sensitivity ↑	0.7794	0.7975	0.8127	**0.8205**
	Specificity ↑	**0.9976**	0.9969	0.9968	0.9971
	Hausdorff distance ↓	19.3185	17.0840	15.8978	**15.7556**
RV	Dice coefficient ↑	0.6578	0.6863	**0.7095**	0.7076
	Sensitivity ↑	0.6450	0.6874	**0.7088**	0.7078
	Specificity ↑	0.9963	**0.9969**	0.9968	0.9967
	Hausdorff distance ↓	26.3090	**24.0388**	24.7103	24.8770

The sensitivity and specificity of the segmentation methods are computed using the manually segmented ground truth mask, as shown in Eqn. 3 and Eqn. 4.

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

TP is the true positive and FP is the false positive, while TN and FN stand for true negative and false negative.

HD is defined as Eqn. 5, with the distance computed using the Euclidean distance.

$$HD(X,Y) = \max_{x\in X} \min_{y\in Y} \|x-y\|_2 \tag{5}$$

C. Method Evaluation

We conduct several experiments using different settings to validate the effectiveness of the method. To pre-train the model, we use only the ImageNet dataset or use both ImageNet and COCO datasets. The data augmentation method and the object proposals recovery strategy are also evaluated. We denote the different methods as Method 1 to Method 4 (M-1 to M-4 for short), as shown in Table. II.

Firstly, we evaluate the importance of the data augmentation strategy. The model is pre-trained using ImageNet with and without the proposed data augmentation (M-2 vs. M-1). As shown in Table III, the Dice coefficients of M-2 is higher than M-1 for LV, RA, and RV segmentation. The HD of M-2 is lower than M-1 for RA and RV segmentation, in which lower distance means better accuracy. For the evaluation of sensitivity and specificity, M-2 also has generally better performance than M-1, which indicates the effectiveness of the data augmentation strategy.

Secondly, we compare the method pre-trained with ImageNet and pre-trained with both ImageNet and COCO datasets (M-2 vs. M-3). We can see, M-3 has higher Dice coefficients for LA, RA, and RV. For HD, M-3 gets a lower distance for LA and RA. The sensitivity and specificity of M-2 and M-3 are similar to each other.

Finally, the object proposals recovery (M-3 vs. M-4) will improve the Dice coefficient. As shown in Table III, M-4 has higher Dice coefficients for LA, LV, and RA. Furthermore, for LA and RA segmentation, M-4 has lower HD values (16.7062 and 15.7556) than M-3 (17.0644 and 15.8978). The sensitivity and specificity of M-4 are also better than M-3 for LA, LV, and RA segmentation. The effect of using the object proposals recovery to retrieve missing parts can be seen in Fig. 2. We show several segmentation results in Fig. 3, which demonstrate our method's good segmentation performance.

We also compare our method with U-Net [17], which is the most frequently used segmentation method in medical image analysis. As shown in Table IV, our method outperform U-Net in almost all metrics. Our method's Hausdorff distance is much lower than U-NET, which demonstrates the superior performance of the proposed method.

TABLE IV

COMPARISON OF OUR METHOD WITH U-NET [17].

Parts	Metrics	U-NET	Our Method
LA	Dice coefficient ↑	0.6318	**0.7192**
	Sensitivity ↑	0.6312	**0.7305**
	Specificity ↑	**0.9980**	0.9978
	Hausdorff distance ↓	61.9734	**16.7062**
LV	Dice coefficient ↑	0.6174	**0.6759**
	Sensitivity ↑	0.6155	**0.6669**
	Specificity ↑	0.9970	**0.9980**
	Hausdorff distance ↓	62.4760	**29.6767**
RA	Dice coefficient ↑	0.7048	**0.7941**
	Sensitivity ↑	0.7456	**0.8205**
	Specificity ↑	0.9971	**0.9971**
	Hausdorff distance ↓	41.7324	**15.7556**
RV	Dice coefficient ↑	0.6515	**0.7076**
	Sensitivity ↑	0.6725	**0.7078**
	Specificity ↑	0.9957	**0.9967**
	Hausdorff distance ↓	62.5010	**24.8770**

V. CONCLUSION

We present an improved instance segmentation approach capable of segmenting four cardiac chambers accurately at the same time and thus provide more information for cardiologists. A rotation and distortion operation is used to perform data augmentation, which better mimics the movements of the fetal heart and improves the segmentation results. Our method is based on the well-known Mask R-CNN, with a novel object proposal recovery method to retrieve the missing parts. Experiments conducted on a fetal echocardiography dataset fully demonstrate the superior performance of our proposed framework. In the future, we will support the segmentation of more components of the fetal heart.

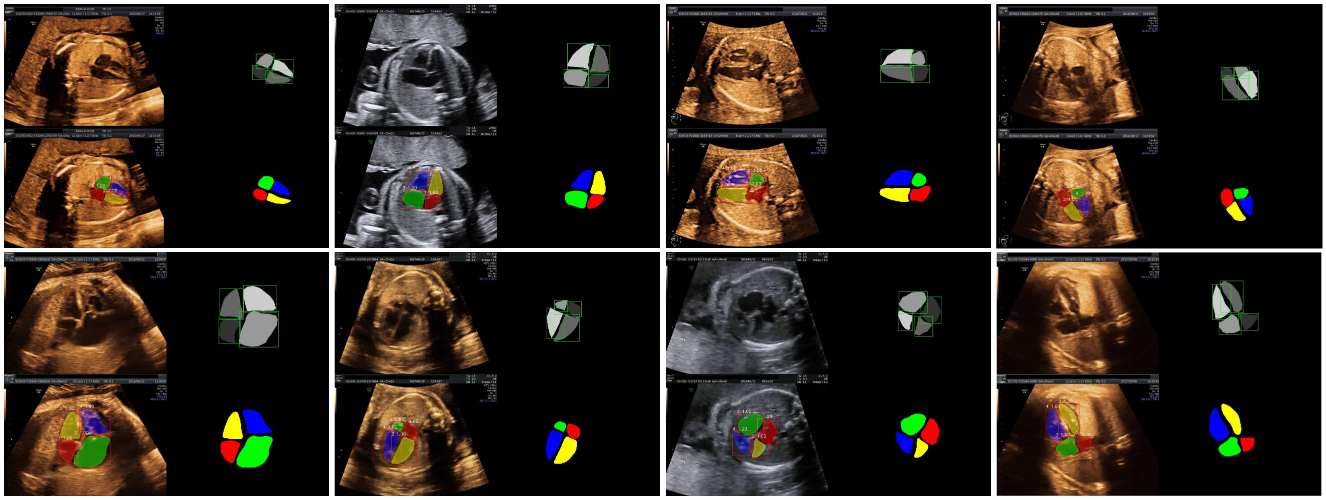

Fig. 3. These are eight segmentation results. For each result, four figures are shown. (Top left) The original fetal echocardiographic image. (Top right) The ground-truth segmentation. (Botton left) The original image covered by the segmentation result. (Bottom right) The segmentation results. The cardiac chambers are indicated using different colors(red for LA, yellow for LV, green for RA, blue for RV).

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] J. Pedrosa, S. Queiros, and et al., "Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based b-spline explicit active surfaces," *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2287–2296, 2017.

[3] G. Allan, S. Nouranian, and et al., "Simultaneous analysis of 2d echo views for left atrial segmentation and disease detection," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 40–50, 2017.

[4] J. Bersvendsen, F. Orderud, and et al., "Automated segmentation of the right ventricle in 3d echocardiography: A kalman filter state estimation approach," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 42–51, 2015.

[5] L Xu, M Liu, Z Shen, and et al., "DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography," *Computerized Medical Imaging and Graphics*, 2020.

[6] S. Leclerc, E. Smistad, and et al., "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE Transactions on Medical Imaging*, 2019.

[7] O. Oktay, E. Ferrante, and et al., "Anatomically constrained neural networks (ACNN): Application to cardiac image enhancement and segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384, 2018.

[8] A. Patra and J. A. Noble, "Sequential anatomy localization in fetal echocardiography videos," *CoRR*, vol. abs/1810.11868, 2018.

[9] Irving Dindoyal, Tryphon Lambrou, Jing Deng, Cliff F Ruff, Alf D Linney, CH Rodeck, and Andrew Todd-Pokropek, "Level set segmentation of the fetal heart," in *International Workshop on Functional Imaging and Modeling of the Heart*. Springer, 2005, pp. 123–132.

[10] Irving Dindoyal, Tryphon Lambrou, Jing Deng, and Andrew Todd-Pokropek, "2d/3d fetal cardiac dataset segmentation using a deformable model," *Medical physics*, vol. 38, no. 7, pp. 4338–4349, 2011.

[11] C. Gustavo, J. C. Nascimento, and F. António, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 968–982, 2012.

[12] L. Vargas-Quintero, B. Escalante-Ramírez, and et al., "Left ventricle segmentation in fetal echocardiography using a multi-texture active appearance model based on the steered hermite transform," *Computer Methods & Programs in Biomedicine*, vol. 137, pp. 231–245, 2016.

[13] M. H. Jafari, H. Girgis, A. H Abdi, and et al., "Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as prior," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 649–652.

[14] 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.

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[16] S. Ren, K. He, R Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks.," *IEEE TPAMI*, vol. 39, no. 6, pp. 1137–1149, 2017.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.