MobileUNet-FPN: A Semantic Segmentation Model for Fetal Ultrasound Four-Chamber Segmentation in Edge Computing Environments

Bin Pu, *Graduate Student Member, IEEE*, Yuhuan Lu, Jianguo Chen, *Member, IEEE*, Shengli Li, Ningbo Zhu, Wei Wei, *Senior Member, IEEE*, and Kenli Li, *Senior Member, IEEE*

***Abstract*—The apical four-chamber (A4C) view in fetal echocardiography is a prenatal examination widely used for the early diagnosis of congenital heart disease (CHD). Accurate segmentation of A4C key anatomical structures is the basis for automatic measurement of growth parameters and necessary disease diagnosis. However, due to the ultrasound imaging arising from artefacts and scattered noise, the variability of anatomical structures in different gestational weeks, and the discontinuity of anatomical structure boundaries, accurately segmenting the fetal heart organ in the A4C view is a very challenging task. To this end, we propose to combine an explicit Feature Pyramid Network (FPN), MobileNet and UNet, i.e., MobileUNet-FPN, for the segmentation of 13 key heart structures. To our knowledge, this is the first AI-based method that can segment so many anatomical structures in fetal A4C view. We split the MobileNet backbone network into four stages and use the features of these four phases as the encoder and the upsampling operation as the decoder. We build an explicit FPN network to enhance multi-scale semantic information and ultimately generate segmentation masks of key anatomical structures. In addition, we design a multi-level edge computing system and deploy the distributed edge nodes in different hospitals and city servers, respectively. Then, we train the MobileUNet-FPN model in parallel at each edge node to effectively reduce the network communication overhead. Extensive experiments are conducted and the results show the superior performance of the proposed model on the fetal A4C and femoral-length images.**

***Index Terms*—Apical four-chamber view, anatomical structure segmentation, congenital heart disease, fetal echocardiography, Internet of medical things.**

Manuscript received 1 October 2021; revised 18 March 2022 and 30 April 2022; accepted 3 June 2022. Date of publication 14 June 2022; date of current version 7 November 2022. This work was supported in part by the National Key R&D Program of China under Grants 2018YFB0203800 and 2019YFB2103005, in part by the National Natural Science Foundation of China under Grants 62072168, 61860206011, 62002110, and 6217071835, and in part by the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant QL20210079. *(Corresponding author: Kenli Li.)*

Bin Pu, Yuhuan Lu, Jianguo Chen, Ningbo Zhu, and Kenli Li are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: pubin@hnu.edu.cn; hnuluyuhuan@hnu.edu.cn; jianguochen@hnu.edu.cn; quietwave@hnu.edu.cn; lkl@hnu.edu.cn).

Shengli Li is with the Department of Ultrasound, Shenzhen Maternal and Child Healthcare Hospital, Southern Medical University, Shenzhen 518028, China (e-mail: lishengli63@vip.126.com).

Wei Wei is with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: weiwei@xaut.edu.cn).

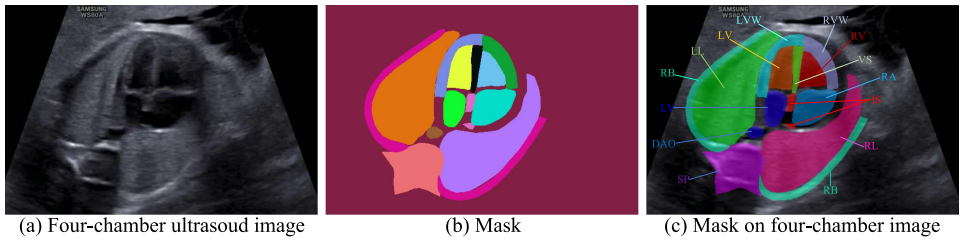
Digital Object Identifier 10.1109/JBHI.2022.3182722

I. Introduction

CONGENITAL Heart Disease (CHD) is a type of heart condition that can be detected as early as the first trimester of life [1]. The defect is characterized by abnormalities in the heart structure, which can be mild or severe. CHD is the leading cause of infant mortality, estimated to cause approximately 4–13 deaths per 1000 live births [2]–[4]. Screening and diagnosing CHD during pregnancy is important so that life-saving treatment can be carried out in time. Newborns with severe heart disease who are not examined and diagnosed before birth may face worsening disease symptoms in the delivery room or in the first month of life. Echocardiography is the primary examination for the diagnosis of CHD. It is non-invasive, low-cost, convenient, and suitable for real-time medical imaging. The basic fetal heart screening is interpreted by the apical four-chamber (A4C) view of fetal echocardiography. In fact, although some fetal diseases can be directly visualized in the A4C view, some functional CHD cases cannot be observed, and require organ segmentation and parametric measurements to diagnose. The A4C view of a fetal ultrasound image and the masks of the key anatomical structures are shown in Fig. 1.

A4C view segmentation has been widely used in the measurement of growth parameters and auxiliary diagnosis of CHD. For example, clinical parameters (e.g., left ventricular length diameter (LVD), mitral valve inner diameter (MVD), left ventricular area (LVA), and left atrial area (LAA)) are the most important indicators for evaluating left heart function [5]. However, the segmentation of the anatomical structure of the A4C view needs to be conducted manually, which is subjective and time-consuming, and requires manual delineation of the anatomical structures. Exploring automatic segmentation algorithms can help reduce the subjective inconsistency of manual segmentation, greatly improve the work efficiency of sonographers, and reduce the workload of obstetricians.

However, the ultrasound four-chamber view segmentation comes with multiple challenges, such as contour information



(a) Four-chamber ultrasoud image (b) Mask (c) Mask on four-chamber image

Fig. 1. The original A4C image and the anatomical structures. (a) is an original fetal four-chamber ultrasound image, (b) is the mask of 13 heart organs, and (c) is the mask superimposed on the ultrasound image.

(a) 17th gestational week (b) 24th gestational week (c) 29th gestational week (d) 32nd gestational week (e) 34th gestational week

Fig. 2. Examples of A4C views scanned from different probe angles and various gestational ages. (a)–(e) Show the fetal four-chamber view at different gestational weeks and the different scanning angles of the sonographer. The red arrow area in (e) is an ultrasonic echo noise and the boundaries of the anatomical structures surrounding this red arrow area are blurred.

loss and noise. In addition, the segmentation of the fetal four-chamber view presents more challenges than the adult view. The fetal four-chamber view is affected by gestational age and fetal movement; the anatomical structures of the result are ambiguous, and the borders may be blurred, as illustrated in Fig. 2. Moreover, the parts of the fetal four-chamber view tend to be relatively small and in a period of growth. In contrast, the parts of the adult heart tend to be fixed.

With the rapid development of computer vision, deep learning (DL)-based methods offer new ideas for coping with the above challenges. Most researchers believe that the A4C segmentation task can be expressed as a semantic segmentation problem, and the DL-based approach is the preferred choice [6]–[9]. For example, UNet [10] and its variants [9], [11], [12] have shown excellent performance in medical image segmentation and have been extended to many other applications. However, there are still some unresolved challenges. The segmentation task of anatomical structures in the four-chamber view requires the extraction of high-level and abstract features. Some anatomical structures, such as SP, LL and RL, are also critical for disease diagnosis, but there has been little research on the segmentation of these structures. In other words, involving the segmentation task of multiple anatomical structures in the A4C view has been ignored. In addition, the multi-organ segmentation task requires the consideration of multi-scale features to segment organs of different magnitudes. This paper is the first to combine the advantages of feature pyramid networks (FPN) [13] and MobileNet [14], and proposes a MobileNet-FPN model for fetal ultrasound four-chamber image segmentation. The proposed MobileNet-FPN model is built based on an encoder–decoder structure [15] and can automatically segment 13 anatomical structures. Moreover, the proposed method is lightweight and has few parameters, making it ideal for deployment in mobile edge environments. To the best of our knowledge, the proposed model involves segmentation of the largest number of anatomical structures in the A4C view of the fetus. The role of the key anatomical structure segmentation is shown in Table III. The main contributions of this work can be summarized as follows:

- We divide MobileNet into four stages and construct different multi-scale features in the four stages as encoders to enhance multi-scale semantic information.
- We combine the advantages of FPN into semantic segmentation tasks for organ segmentation in A4C views and femoral-length images.
- We employ the MobileNet-FPN model to segment 13 anatomical structures of the A4C view. To the best of our knowledge, it is the first time that a DL-based method has performed segmentation in many heart anatomical structures.
- We conduct extensive experiments on actual fetal A4C views and femoral-length images to evaluate the performance of the proposed model. Experimental results show that the proposed model has superior competitive performance compared with state-of-the-art methods.

The rest of the paper is organized as follows. Section II reviews related work in terms of ultrasound image segmentation and fetal four-chamber segmentation. Section III describes the details of the proposed MobileNet-FPN model. Experimental results and comparisons are presented in Section IV with respect to the segmentation results on the A4C view and femoral-length view. Finally, Section V concludes the paper, highlighting future work and research directions.

II. Related Work

A. Ultrasound Image Segmentation

Before DL-based methods emerged, previous approaches were designed for various segmentation methods, including

TABLE I
ABBREVIATIONS AND DESCRIPTIONS OF THE FETAL FOUR-CHAMBER SEGMENTATION METHODS

Abbr	Description	Segmentation structure
DW-Net [6]	A dilated Convolutional Chain and a W-shape net	RA, LA, LV, RV, thorax, epicardium
DP-FEM [7]	A high- and low-level feature fusion	LA, LV
MFP-UNet [11]	UNet merges semantic information	LV
AIDAN [28]	An attention mechanism, spatial and context paths	LA, LV
CNN method [29]	UNet architecture	Atrial septal defect, ventricular septal defect
CU-net [30]	Structural similarity index measure	LV, LA, DAO, RA, RV, thorax, epicardium
MS-Net [31]	A bidirectional feature fusion and a W-UNet	LA, LV
FCN [33], [35]	An encoder–decoder framework	LV

active contouring, Haar-Like features, compound methods, boundary fragment models, and temporal constraints [16]–[18]. These methods are semi-automatic and have limited performance. DL-based methods have been highlighted as a better option [19]. In [20], based on original UNet, Yang *et al.* proposed a residual UNet and an atrous spatial pyramid pooling UNet for the segmentation of fetal abdominal circumference (AC), fetal femur length (FL), and fetal crown-rump length, avoiding gradient disappearance and gradient explosion. Sobhaninia *et al.* [21] presented a multi-task deep convolutional neural network with a minimized composite cost function to segment the head circumference of fetal head images. Mishra *et al.* [22] designed a fully convolutional neural network (FCN) combined with an attention deep supervision mechanism for anatomical structure segmentation. The FCN model can infer high-level features using low-level information. Because breast cancer is common, many researchers have focused on the automatic segmentation of ultrasound breast images to assist in the auxiliary diagnosis of breast cancer [23]–[27]. For example, in [23], Xu *et al.* developed a CNN-based method to solve four major challenges, skin, fibroglandular tissue, masses, and fatty tissue, and it can recognize tissues in breast ultrasound images. Zhou *et al.* [26] also proposed a multi-task learning algorithm for tumor segmentation and classification in breast ultrasound images. The algorithm consists of an encoder–decoder segmentation network and a multi-scale classification network. In summary, previous studies have mostly focused on breast and thyroid nodule segmentation.

B. Fetal Four-Chamber Segmentation

Various studies have been carried out on the segmentation of anatomical structures in the ultrasound A4C view and cardiac MRI images. Hu *et al.* [28] proposed an attention-guided dual-path network for cardiac image segmentation, which addresses the noise ratio and internal variability in the ultrasound heart view. Guo *et al.* [7] presented a dual attention enhancement feature fusion network for ultrasound pediatric echocardiography, which extracts rich features (e.g., high-level features and low-level features) via a channel attention mechanism. Rachmatullah *et al.* [29] used a CNN-based method to automatically segment the small structures of fetal cardiac standard planes. Xu *et al.* [30] proposed a cascaded UNet with structural similarity index measurement loss to solve the problem of low imaging resolution and tissue boundaries. Zhao *et al.* [31] designed a multi-scale wavelet network for echocardiographic segmentation, which can extract features of different scales and use a discrete wavelet transform to process information loss. Andreassen *et al.* [32] developed a 2D CNN-based method for mitral annulus segmentation in 3D transesophageal echocardiography, thereby reducing the need for manual interaction. Moradia *et al.* [11] proposed an improved UNet for left ventricle segmentation, abbreviated as MFP-UNet, which addresses the semantic strength of the origin UNet segmentation process. In addition to segmentation on ultrasound four-chamber images, several studies have also focused on the segmentation of cardiac magnetic resonance imaging (MRI) [11], [33], [34]. For example, Avendi *et al.* [11] designed a combined method based on DL and a deformable model to segment the left ventricle. Given the limited availability of fetal ultrasound images, the manual workload of image annotating is huge, especially the annotation of multiple anatomical structures. Thus, there have been few studies in this area. More methods also focus on the segmentation of anatomical structures, such as the segmentation of the left ventricle [11], [33], right ventricle [34], and small tissues [6]. Moreover, most existing approaches do not explore the advantages of the FPN structure and MobileNet in semantic segmentation tasks. Some popular DL-based methods for four-chamber view segmentation are summarized in Table I.

III. METHODOLOGY

In this section, we first design a distributed computing system based on edge computing and treat it as the computing platform of the proposed lightweight model in the future. Then, we propose a MobileUNet-FPN model for the segmentation of 13 key heart structures in the A4C view. We describe the core components of this model, including the MobileNet backbone, FPN segmentation module, and encoder–decoder module.

A. Edge-Computing System Architecture

Ultrasound image analysis systems face increasing demands for large-scale images, accurate data analysis, and low-latency response. Edge-of-things is a new computing architecture that has been widely adopted in various customer applications, especially in smart healthcare [36]. We design a multi-level edge computing system for the lightweight MobileUNet-FPN model, providing flexible and scalable computing capabilities and effectively reducing network communication overhead. The architecture of the designed multi-level edge-computing system is shown in Fig. 3.

As illustrated in Fig. 3, the edge-computing system consists of a large number of ultrasound imaging equipment, hospital-level

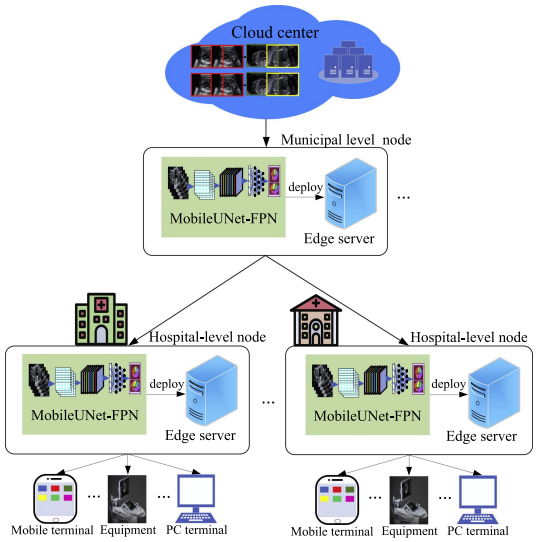


Fig. 3. Architecture of the designed multi-level edge computing system. The proposed system contains various edge nodes deployed in different hospitals and municipal servers, PC terminals, and mobile terminals. Our MobileUNet-FPN model is deployed at each edge node and trained in parallel on local datasets. Ultrasound equipment transmits image data to the edge server. Radiologists and doctors in different hospitals can use PC terminals to access the nearby edge node to obtain multi-anatomical segmentation results. Patient users can access the imaging results from their mobile terminals.

edge nodes, municipal edge nodes, the cloud center, PC terminal, and mobile terminal. Ultrasound imaging equipment in different hospitals produces large-scale ultrasound images. Multi-level edge nodes are deployed at different management levels, such as hospitals, districts, and counties of a city. The hospital-level nodes are responsible for connecting the imaging equipment from different departments in the current hospital and training the model. Each middle-level node is connected to all low-level nodes within its jurisdiction. The proposed MobileUNet-FPN model is deployed at each hospital-level and municipal edge node and trained in parallel by using local datasets. In this way, data transmission and model training tasks are migrated from the traditional network center to the network edge, which greatly reduces the cost of network data communication and delay.

B. MobileUNet-FPN Model

Fig. 4 illustrates the overview of the proposed model. First, we divide the backbone of MobileNet into four stages, which represent the features of different stages. They can be regarded as low-level features, two medium-level features, and high-level features. These four-stage features also have different scales. Second, we use multiple upsampling operations to restore the feature map of the fourth stage to the same size as the first stage. Then, based on the above steps, a 1×1 convolution layer is used to link the features of the same dimension in the encoding and decoding modules. Finally, compared with the prediction on each scale feature map of the original FPN, this paper directly upsamples different scales to the same size as the first stage and merges them. Table II provides some professional terms and corresponding abbreviations.

TABLE II
PROFESSIONAL TERMS AND CORRESPONDING ABBREVIATIONS

Abbr	Description	Abbr	Description
A4C	apical four-chamber	DAO	descending aorta
LA	left atrium	RA	right atrium
LV	left ventricle	RV	right ventricle
VS	inter ventricular-septum	IS	interatrial septum
SP	spine	RB	ribs
LVW	left ventricular wall	RVW	right ventricular wall
LL	left lung	RL	right lung
FL	femoral-length	Me	metaphysis
MVD	mitral valve inner diameter	LVD	left ventricular length diameter
LVA	left ventricular area	LAA	left ventricular area

C. MobileNet Backbone

1) Depthwise Separable Convolution: Depthwise Separable Convolution (DSC) is the basic component of MobileNet, which can be used to reduce the computation workload. It was first introduced in [37] and used in Inception models [38] and Xception [39] to reduce the scale of the model parameters. Recently, DSC has been aimed at reducing the size and computational cost of the CNN [40], [41]. The definition of DSC is as follows:

$$Conv(W,f)_{(i,j)} = \sum_{k,l,m}^{K,L,M} W_{(k,l,m)} \cdot f_{(i+k,j+l,m)}, \quad (1)$$

$$Pointwise\text{-}Conv(W,f)_{(i,j)} = \sum_{m}^{M} W_m \cdot f_{(i,j,m)}, \quad (2)$$

$$Depthwise\text{-}Conv(W,f)_{(i,j)} = \sum_{k,l}^{K,L} W_{(k,l)} \odot f_{(i+k,j+l)}, \quad (3)$$

$$DSC\text{-}Conv\left(W_p, W_d, f\right)_{(i,j)} = Pointwise\text{-}Conv_{(i,j)} \left(W_p, Depthwise\text{-}Conv_{(i,j)}\left(W_d, f\right)\right), \quad (4)$$

where $Conv$ represents the original standard convolution action, $Pointwise\text{-}Conv$ denotes the pointwise convolution operation, $Depthwise\text{-}Conv$ is the depthwise convolution mechanism, and $DSC\text{-}Conv$ refers to the depthwise separable convolution operation. $Depthwise\text{-}Conv$ adopts different convolution kernels for each input channel, that is, one convolution kernel for each input channel, which is a depth-level operation. $Pointwise\text{-}Conv$ adopts a 1×1 window convolution operation, and it is still essentially a standard convolution operation. DSC first uses $Depthwise\text{-}Conv$ to perform convolution operations on different input channels and then uses $Pointwise\text{-}Conv$ to combine the above outputs. DSC reduces the computation workload and the number of model parameters without the loss of model accuracy.

2) Four-Stage Features: Before describing the four-stage features, we first introduce the conv_block and depthwise_conv_block. conv_block is composed of a zero-padding feature, standard convolution operation, batch-normalization, and the Relu function, as shown

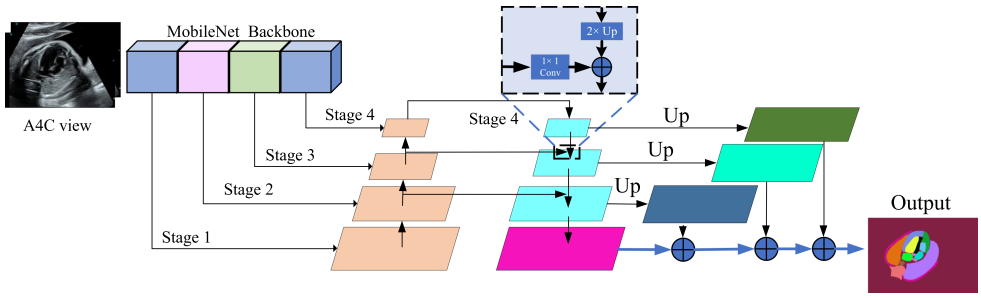


Fig. 4. Overview of the proposed MobileUNet-FPN model.

TABLE III
KEY ANATOMICAL STRUCTURES OF THE FETAL HEART IN THE A4C VIEW

Number	Key structure	Role
1	LA	Computing of left atrial internal diameter, area, volume index, and emptying fraction.
2	LV	Calculation of clinical parameters such as ventricular volume, ejection fraction and left ventricular mass.
3	RA	Evaluation of right atrial internal diameter and area.
4	RV	Estimation of right ventricular volume and assessment of ventricular morphological changes, etc.
5	DAO	Calculation of descending aortic internal diameter.
6	VS	Estimation of septal thickness, amplitude of motion, and assistance in the diagnosis of single ventricle and myocardial infarction.
7	IS	Observe the presence and morphology of the atrial septum. It assists in the diagnosis of single atrium and atrial septal defect.
8	SP	Observation of spinal morphology, vertebral body, and cone arch position.
9	LVW	Calculation of left ventricular wall thickness and left ventricular wall motion amplitude.
10	RVW	Computing of right ventricular wall thickness and right ventricular wall motion amplitude
11	LL	Assessment of left lung area.
12	RL	Assessment of right lung area.
13	FL	Estimation of the length and weight.
14	Me	Assistance in the identification of the humerus and femur.

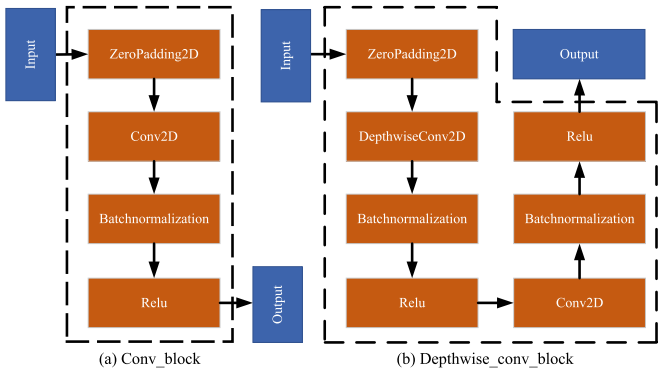


Fig. 5. Embedded conv_block and depthwise_conv_block.

in Fig. 5(a). Moreover, Fig. 5(b) shows the structure of depthwise_conv_block, where $Pointwise\text{-}Conv$ is essentially the same as the original convolution operation (i.e., Conv2D), except $Depthwise\text{-}Conv$ employs a 1×1 convolution, and both the stride and kernel are 1×1. All four stages are serial. The first stage uses a conv_block and a depthwise_conv_block. The second stage uses two depthwise_conv_blocks, and the third stage is the same as the second stage. The fourth stage requires six depthwise_conv_blocks. Through the four stages, we can acquire features of different scales and levels, such as low-level features, intermediate-level features, and high-level features.

D. FPN Segmentation Module

Inspired by PFN, we introduce FPN in the semantic segmentation task, where the purpose is to enhance multi-scale information. As the segmentation task has a small target (e.g., DAO), there is also a large-area object (e.g., LL and RL) and many other anatomical structures in between. Therefore, the use of multi-scale semantic information can better segment various key anatomical structures. As shown in Fig. 4, we build an explicit FPN model with multi-scale features from the four stages of MobileNet and the upsampling operations in different stages. The original FPN directly performs prediction on the different multi-scale feature maps. In the semantic segmentation task, there are two operations for feature maps of different scales. One is to concatenate the next upsampling operation, and the other is to directly upsample to the same size as the input image. Then, the feature maps of the same size are connected to make the final prediction.

E. Encoder-Decoder Module

The proposed semantic segmentation method consists of an encoder–decoder framework. We adopt the feature maps from the first stage to the fourth stage as a continuous encoder process. In this process, the feature map becomes smaller, and the semantic information becomes more abstract. In the decoder procedure, as the feature map becomes larger, the upsampling method is used to gradually restore the original pixel space. The nearest-neighbor interpolation is used as an upsampling function, as defined in

$$O_x = I_x \times (O_{width}/I_{width}), \tag{5}$$

$$O_y = I_y \times (O_{height}/I_{height}), \tag{6}$$

where (I_x, I_y) and (O_x, O_y) denote the coordinates of the target image and the original image, respectively, and O_{height}

TABLE IV
COMPARISON OF THE PROPOSED MODEL AND THE COMPETITIVE METHODS ON FETAL FOUR-CHAMBER IMAGES

Model	LV	RV	LA	RA	LVW	RVW	VS	Average
UNet [10]	0.5487±0.0373	0.4733±0.0409	0.5926±0.0151	0.6641±0.0217	0.4734±0.0228	0.3840±0.0158	0.5610±0.0379	
	IS	LL	RL	SP	RB	DAO		0.5456±0.0277
	0.3672±0.0144	0.6001±0.0439	0.6365±0.0266	0.6910± 0.0155	0.4541±0.0324	0.6229±0.0359		
SegNet [42]	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.4216±0.0169	0.2701±0.0594	0.4213±0.0335	0.4165±0.1136	0.3338±0.0279	0.2105±0.0411	0.4123±0.0398	
	IS	LL	RL	SP	RB	DAO		0.3847±0.0148
	0.2272±0.0263	0.3787±0.0601	0.5394±0.0310	0.5970±0.0958	0.3640±0.0131	0.4094±0.0464		
Resnet50-UNet	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.5845±0.0672	0.5347±0.0855	0.5976±0.0669	0.7080±0.0465	0.5276±0.0225	0.4437±0.0339	0.5892±0.0484	
	IS	LL	RL	SP	RB	DAO		0.5875±0.0329
	0.3481±0.0261	0.7021±0.0189	0.7532±0.0047	0.7327±0.0189	0.5109±0.0167	0.6054±0.1091		
PSNet [43]	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.6212±0.0178	0.5793±0.0184	0.6309±0.0064	0.7046±0.0108	0.4818±0.0341	0.4740±0.0125	0.4841±0.0058	
	IS	LL	RL	SP	RB	DAO		0.5566±0.0109
	0.2389±0.0291	0.7611±0.0116	0.7814±0.0117	0.7256±0.0094	0.4022±0.0256	0.3510±0.0220		
FCN8 [35]	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.5790±0.0456	0.4384±0.0730	0.5785±0.0554	0.6856±0.0313	0.4602±0.0571	0.3873±0.0485	0.4845±0.0695	
	IS	LL	RL	SP	RB	DAO		0.5369±0.0232
	0.2623±0.0724	0.7661±0.0204	0.7919±0.0112	0.6473±0.0477	0.3889±0.0210	0.5093±0.0600		
FCN32 [35]	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.6843±0.0084	0.6341±0.0116	0.6883±0.0244	0.7364±0.0137	0.5707±0.0056	0.5341±0.1150	0.5519±0.0059	
	IS	LL	RL	SP	RB	DAO		0.6163±0.0099
	0.2448±0.0256	0.8071±0.0056	0.8253±0.0043	0.7323±0.0258	0.5213±0.0048	0.4815±0.0445		
VGG-UNet	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.5997±0.0234	0.5226±0.0442	0.5661±0.0495	0.6495±0.0474	0.4773±0.0145	0.4097±0.0170	0.5798±0.0182	
	IS	LL	RL	SP	RB	DAO		0.5313±0.0193
	0.2642±0.0367	0.5840±0.0218	0.6614±0.0405	0.6297±0.0302	0.3739±0.0207	0.5889±0.0382		
Linknet [44]	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.6900±0.0219	0.6400±0.0103	0.7227±0.0215	0.7899±0.0185	0.5559±0.0213	0.4942±0.0434	0.5880±0.0118	
	IS	LL	RL	SP	RB	DAO		0.6374±0.0256
	0.3845±0.0256	0.7763±0.0562	0.7883±0.0414	0.7583±0.0203	0.4702±0.0494	0.6283±0.0314		
Mobile-UNet	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.7358±0.0122	**0.6964±0.0089**	0.7520±0.0144	0.8141±0.0074	**0.6136±0.0058**	**0.5653±0.0078**	0.6685±0.0061	
	IS	LL	RL	SP	RB	DAO		0.6871±0.0041
	0.4111±0.0390	**0.8081±0.0101**	**0.8228±0.0089**	0.7822±0.0155	0.5434±0.0138	0.7192±0.0173		
Ours	LV	RV	LA	RA	LVW	RVW	VS	Average
	0.7345±0.0081	0.6904±0.0123	**0.7589±0.0082**	**0.8187±0.0076**	0.6105±0.0080	0.5607±0.0086	**0.6771±0.0133**	
	IS	LL	RL	SP	RB	DAO		**0.6910 ±0.0034**
	0.4430±0.0177	0.8014±0.0047	0.8194±0.0024	**0.7893±0.0069**	**0.5487±0.0137**	**0.7297±0.0073**		

and O_{width} are the length and the width of the target image, respectively. In the same way, I_{height} and I_{width} are the length and the width of the original image, respectively. Based on the original value of the input fetal image, the nearest-neighbor interpolation is used to insert new values between different regions. The decoder operation can be written as

$$Dec^{i}_{image} = UP_{sampling}(H_{image}), \tag{7}$$

where H_{image} represents the encoded feature map, Dec_{image} denotes the decoded feature map, $UP_{sampling}$ indicates an upsampling operation, and i is the number of encoding. We apply four upsampling operations to the feature map of the fourth stage to restore the pixel space of the feature map to the same as that of the first stage. The training process of the proposed method is listed in Algorithm 1. The training images and annotations are constructed (*lines* 1–4) from the training dataset, and then MobileUNet-FPN is trained employing back-propagation and Adam (*lines* 6–9); the learned MobileUNet-FPN model is returned (*line* 10).

Algorithm 1: Training Process of the MobileUNet-FPN Model.

Input:
$\mathcal{D}_f = [f_1, f_2, f_3 \ldots, f_n]$: Fetal four-chamber images;
Y_s=$[y_1, y_2, y_3 \ldots, y_m]$: Mask of $\mathcal{D}_f$.
Output:
$\mathcal{M}$: The trained MobileUNet-FPN model;
1: $\mathcal{D}_f \leftarrow \varnothing$;
2: **for all** available images $t(1 \leqslant t \leqslant n)$ **do**
3: Put $f_t \rightarrow \mathcal{D}_f$;
4: **end for**
5: **Initialize:** Parameters of $\mathcal{M}$ are θ;
6: **repeat**
7: Randomly select a batch of samples $\mathcal{D}_i$ from $\mathcal{D}_f$;
8: Find θ by minimizing objective function with $\mathcal{D}_f$;
9: **until** Reaching the stopping criteria (usually the pre-defined epoch is reached);
10: **return** $\mathcal{M}$;

IV. EXPERIMENTS

A. Evaluation Metrics

To evaluate the results of all comparison segmentation methods more comprehensively, the Intersection Over Union (IoU) is used as an evaluation index, as defined as:

$$IoU = \frac{Area_{P_c} \cap Area_{G_c}}{Area_{P_c} \cup Area_{G_c}}, \tag{8}$$

where $Area_{G_c}$ and $Area_{P_c}$ represent the area of ground truth and the predicted segmentation results respectively.

TABLE V
PARAMETERS SETTINGS OF THE EXPERIMENTS

Parameter	Configuration
Weight decay	0.0001
Initial learning rate	0.01
Image pixel	416 × 608
Batch size	2
Epoch	30

B. Experimental Setting

This work is approved by the committee of the Shenzhen Maternal and Child Health Hospital with the approval number of SFYLS[2020]019. The approval date is June 17, 2020. Experiments are carried out on fetal ultrasound A4C and fetal femur views are obtained at the Shenzhen Maternal and Child Health Hospital, with ethics approval from the committee of our institution. A total of 677 ultrasound images of fetuses from 18 to 32 weeks of gestation are collected. The 13 key anatomical structures are manually annotated by a sonographer with at least 3 years of experience. Further, 85% of the data (575 annotated images) are used for training, and 15% (102 annotated images) are used for validation. In addition, a total of 1303 fetal femur images are used to further verify the performance of the proposed model. Among them, 1096 images are used for model training, and 207 images are used for testing. There are two key anatomical structures in the fetal femur image: femoral length and metaphysis. All experiments are conducted on a Linux server with the following configuration: Eight Intel(R) Xeon(R) CPUs, E5-2680 v4 @ 2.40 GHZ, 256 GB RAM, and 8 NVIDIA P100 GPUs. The parameter settings of this work are listed in Table V. The proposed MobileUNet-FPN model is implemented in the Python programming language by using the popular Keras deep learning framework.

C. Baselines

Some popular baselines of medical image segmentation are used as comparison methods in our experiments, as summarized as follows:

- *UNet* [10]: A popular semantic segmentation network based on the encoder–decoder framework.
- *SegNet* [42]: It combines the backbone VGG and the idea of FCN, and addresses image semantic segmentation tasks in autonomous driving or intelligent robot scenarios.
- *Resnet50-UNet*: It combines the advantages of Resnet50 [45] and UNet.
- *PSNet* [43]: A pyramid scene resolution network based on an FCN-based pixel prediction framework with deeply supervised loss, which embeds difficult scene contextual features.
- *FCN* [35]: A semantic segmentation model based on fully convolutional networks, which consists of only convolutional networks without fully connected layers. We use two variants of FCN: FCN8 and FCN32.
- *VGG-UNet*: It incorporates VGGNet [46] and UNet.
- *LinkNet* [44]: A lightweight network that explores encoder representations for efficient semantic segmentation tasks.
- *Mobile-UNet*: It incorporates the MobileNet and UNet. All settings are the same for MobileNet and our proposed model, except for those used to build the FPN network.

D. Experimental Results on A4C View

Table IV illustrates the segmentation IoUs of the MobileUNet-FPN model and all baselines on the A4C test dataset. It is clear that MobileUNet-FPN achieves high performance on the 13 key anatomical structures, with an IoU of 0.6910. The IoU of MobileUNet-FPN is 14.54% higher than the popular UNet model.

For small anatomical structures (e.g., DAO), the average IoU reaches 0.7297, which is higher than SegNet's IoU by 37.87%. In larger key anatomical structures (e.g., LL and RL), the segmentation results of MobileUNet-FPN are close to the best performance of Mobile-UNet, but obviously, the segmentation error is much smaller than those of the other baselines. In addition, MobileUNet-FPN achieves the best performance on LA, RA, VS, IS, SP, and RB anatomical structures. Compared to other baselines, Mobile-UNet still achieves good performance, with an average IoU of 0.6871. Its good performance may be due to the fact that Mobile-UNet extracts features of different scales in four stages. Compared with Mobile-UNet, the proposed MobileUNet-FPN model increases the IoU by 0.39%, which shows that the FPN module is effective and can enhance multi-scale semantic information to improve segmentation performance. Moreover, MobileUNet-FPN is more stable than Mobile-UNet and obtains a smaller variance.

Fig. 6 shows the segmentation results of the proposed model and other baselines on the test dataset. It can be clearly seen that the segmentation results of our model are close to the ground truth. The segmentation results of UNet and SegNet are not well fitted, and the contour information of 13 anatomical structures is not good. This may be due to the multiple challenges caused by excessive segmentation of anatomical structures in ultrasound images. Further, the baseline methods are missing some key anatomical structures. It can be observed that PSNet does not fit the contours of anatomical structures well in some segmentation results, and most structures' boundaries show jaggedness, such as LA, RA, RB and SP. The performance of FCN32 outperforms that of FCN8, which may be because the deeper network extracts more abstract and high-level features. However, the FCN32 produces discontinuous results when segmenting the RB anatomical structure. In addition, FCN32 segments the IS structure and loses a part in the second and third A4C views. The segmentation results of Mobile-UNet fit well, but the DAO and IS are missing in the first A4C view. Linknet shows a lack of DAO in the first segmentation mask. Resnet50-UNet also misses a lot of structures, such as LL, SP, and LA. All methods perform poorly when segmenting IS and RB structures, but our method achieves the best segmentation compared to the baseline methods, achieving an IoU of 0.4430 and 0.5487,

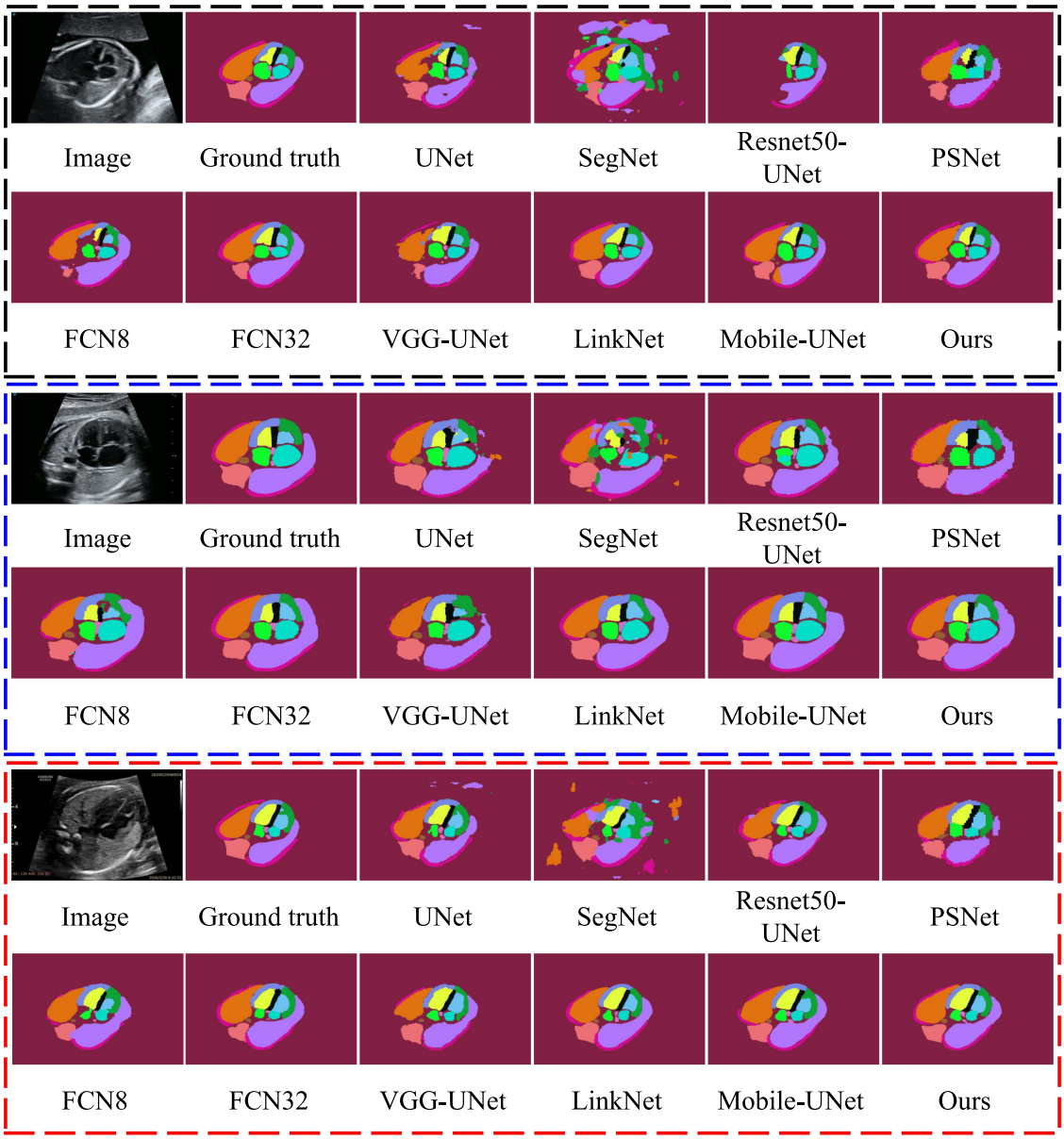


Fig. 6. Segmentation results on the A4C test dataset of the proposed model and all baselines.

respectively. However, the segmentation performance of IS and RB structures is low compared to other anatomical key structures (e.g., LV, RV, LA, and RA). This is mainly because the IS and RB structures are blurred and small, with an unclear boundary and two discontinuous parts (as depicted in Fig. 1).

E. Extension to Fetal Femoral-Length Image Segmentation

To further validate the performance of the proposed MobileUNet-FPN model, we extend our model to segment the femoral-length views. Fig. 8 shows an example of a fetal femoral-length image and the ground-truth mask.

As shown in Table VI, the proposed MobileUNet-FPN model achieves the best performance on the fetal femoral-length view. The best performance is reached on both FL and ME anatomy structures, achieving an average performance IoU of 0.6959. Compared with UNet, the IoU of MobileUNet-FPN is 7.8% higher than that of UNet. Compared with Mobile-UNet, our MobileUNet-FPN model improves the IoU by 2.57%. In addition, the performance of MobileUNet-FPN is more stable and has less fluctuation than other methods. Although the performance of Resnet50-UNet is not as good as that of the proposed model, it produces good performance, which can be attributed to the advantages of the residual network. As illustrated in Fig. 7, MobileUNet-FPN produces good segmentation results on the

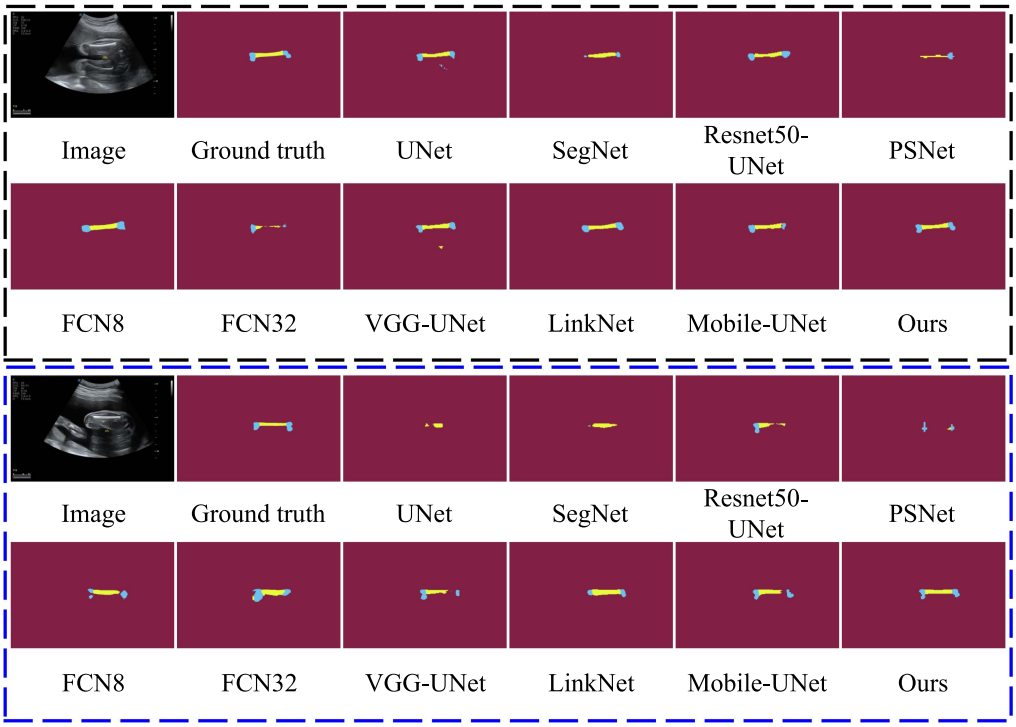


Fig. 7. Segmentation results on the fetal femur test dataset of the proposed model and all baselines.

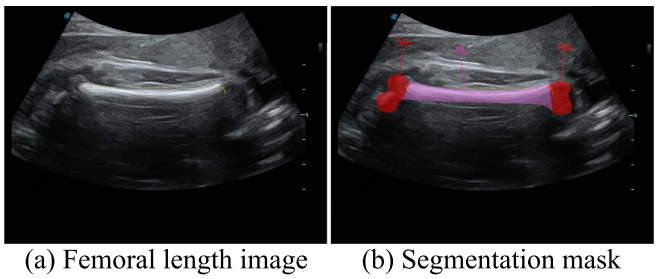


Fig. 8. Example of a femoral-length image and the related mask.

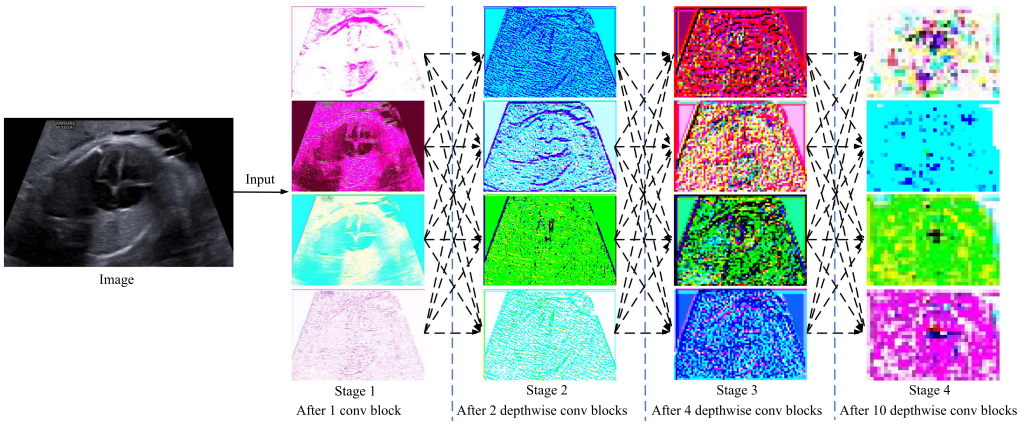


Fig. 9. Feature maps of four stages of the proposed model on fetal A4C views.

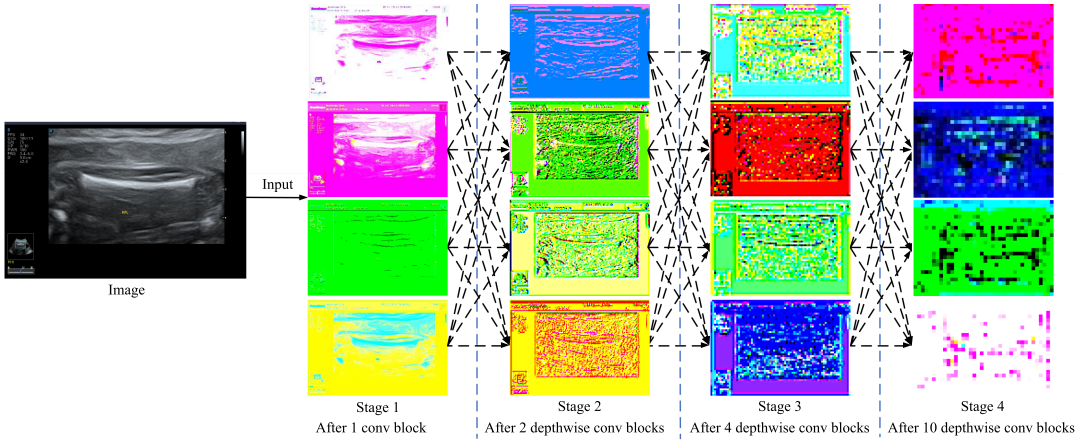


Fig. 10. Feature maps of four stages of the proposed model on fetal femur length views.

TABLE VI
COMPARISON OF THE PROPOSED MODEL AND THE COMPETITIVE BASELINES ON FETAL FEMUR IMAGES

Model	FL	Me	Average
Uet [10]	0.6678±0.0187	0.5680±0.0189	0.6179±0.0100
SegNet [42]	0.6392±0.0171	0.4708±0.0442	0.5550±0.0224
Resnet50-UNet	0.7109±0.0141	0.6539±0.0180	0.6824±0.0103
PSnet [43]	0.5904±0.0124	0.5293±0.0135	0.5598±0.0121
FCN8 [35]	0.6513±0.0183	0.6011±0.0128	0.6262±0.0096
FCN32 [35]	0.6109±0.0338	0.5723±0.0163	0.5916±0.0214
VGG-UNet	0.6794±0.0230	0.6040±0.0187	0.6417±0.0043
LinkNet [44]	0.6866±0.0249	0.6577±0.0062	0.6722±0.0130
Mobile-UNet	0.6967±0.0153	0.6436±0.0260	0.6702±0.0140
Ours	**0.7235±0.0102**	**0.6682±0.0127**	**0.6959±0.0109**

femoral-length image, and the segmentation results of FL and ME anatomical structures fit well with the ground truth. In contrast, other methods, such as PSNet, SegNet, and FCN32, can only divide a portion of the FL and ME structures.

The feature maps of four stages on the A4C and FL views are illustrated in Figs. 9 and 10. It can be clearly observed in the first stage, where the low-level features are visually distinguishable as an A4C view. In the second and third stages, some of these intermediate-level features are identified as the contours of an A4C heart. In the fourth stage, the visual high-level abstract features are no longer visible in the A4C view. From Figs. 9 and 10, the features become increasingly abstract and high-level from the first to the fourth stage. Compared to UNet, it can take full advantage of MobileNet's backbone, extracting to a deeper level features that well segment multiple anatomical structures in the fetal ultrasound view.

V. CONCLUSION

In this paper, we proposed a novel semantic segmentation model, MobileUNet-FPN, based on FPN structure, MobileNet backbone, and UNet module. Specifically, the MobileNet backbone was divided into four stages: low-level features, first intermediate features, second intermediate features, and high-level features. Each stage represented the encoding information of different scales. Then, we first constructed an explicit FPN model for semantic segmentation tasks to enhance semantic multi-scale information and well segment multiple categories of key anatomical structures. To the best of our knowledge, the proposed MobileUNet-FPN model could segment 13 anatomical key structures of a fetal A4C view, the largest number of organ segmentations achieved so far. The performance of the proposed method was evaluated on the fetal four-chamber view (i.e. multi-class segmentation task) and the fetal femoral-length view (i.e. few-class segmentation task). It was clear that MobileUNet-FPN achieved remarkable performance by comparing the related baselines. However, our method had some limitations, with poor performance in segmenting particularly small anatomical structures, such as IS, RB, and RVW. In the future, we will further optimize the MobileUNet-FPN model by incorporating structural information between organs and extend the model to other fetal ultrasound images.

REFERENCES

[1] H. Jicinska *et al.*, "Does first-trimester screening modify the natural history of congenital heart disease? Analysis of outcome of regional cardiac screening at 2 different time periods," *Circulation*, vol. 135, no. 11, pp. 1045–1055, 2017.
[2] J. S. Carvalho *et al.*, "ISUOG practice guidelines (updated): Sonographic screening examination of the fetal heart," *Ultrasound Obstet. Gynecol.*, vol. 41, no. 3, pp. 348–359, 2013.
[3] B. Pu, N. Zhu, K. Li, and S. Li, "Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework," *Future Gener. Comput. Syst.*, vol. 115, pp. 825–836, 2021.
[4] B. Pu, K. Li, S. Li, and N. Zhu, "Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT," *IEEE Trans. Ind. Inform*, vol. 17, no. 11, pp. 7771–7780, Nov. 2021.
[5] R. M. Lang *et al.*, "Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging," *Eur. Heart J.-Cardiovasc. Imag.*, vol. 16, no. 3, pp. 233–271, 2015.
[6] L. Xu *et al.*, "DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography," *Comput. Med. Imag. Graph.*, vol. 80, 2020, Art. no. 101690.
[7] L. Guo *et al.*, "Dual attention enhancement feature fusion network for segmentation and quantitative analysis of paediatric echocardiography," *Med. Image Anal.*, vol. 71, 2021, Art. no. 102042.

[8] X. Wu *et al.*, "CacheTrack-YOLO: Real-time detection and tracking for thyroid nodules and surrounding tissues in ultrasound videos," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 10, pp. 3812–3823, Oct. 2021.

[9] Y. Zhang *et al.*, "Multi-needle detection in 3D ultrasound images using unsupervised order-graph regularized sparse dictionary learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2302–2315, Jul. 2020.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] S. Moradi *et al.*, "MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Phys. Medica*, vol. 67, pp. 58–69, 2019.

[12] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet : A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[13] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[14] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[15] B. Pu, Y. Liu, N. Zhu, K. Li, and K. Li, "ED-ACNN: Novel attention convolutional neural network based on encoder–decoder framework for human traffic prediction," *Appl. Soft. Comput.*, vol. 97, 2020, Art. no. 106688.

[16] J. Perez-Gonzalez, J. B. Muñoz, M. R. Porras, F. Arámbula-Cosío, and V. Medina-Bańuelos, "Automatic fetal head measurements from ultrasound images using optimal ellipse detection and texture maps," in *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 Oct. 2014*. Cham, Switzerland: Springer, 2015, pp. 329–332.

[17] W. Jatmiko, I. Habibie, M. A. Ma'sum, R. Rahmatullah, and I. P. Satwika, "Automated telehealth system for fetal growth detection and approximation of ultrasound images," *Int. J. Smart Sens. Intell. Syst*, vol. 8, no. 1, pp. 697–719, 2015.

[18] X. Wu *et al.*, "Deep parametric active contour model for neurofibromatosis segmentation," *Future Gener. Comput. Syst.*, vol. 112, pp. 58–66, 2020.

[19] K. Lekadir *et al.*, "A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 48–55, Jan. 2017.

[20] Y. Yang, P. Yang, and B. Zhang, "Automatic segmentation in fetal ultrasound images based on improved U-Net," in *Proc. J. Phys.: Conf. Ser.*, vol. 1693, no. 1, IOP Publishing, 2020, Art. no. 012183.

[21] Z. Sobhaninia *et al.*, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2019, pp. 6545–6548.

[22] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin, "Ultrasound image segmentation: A deeply supervised network with attention to boundaries," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1637–1648, Jun. 2019.

[23] Y. Xu, Y. Wang, J. Yuan, Q. Cheng, X. Wang, and P. L. Carson, "Medical breast ultrasound image segmentation by machine learning," *Ultrasonics*, vol. 91, pp. 1–9, 2019.

[24] L. Panigrahi, K. Verma, and B. K. Singh, "Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution," *Expert Syst. Appl.*, vol. 115, pp. 486–498, 2019.

[25] Q. Huang, Y. Huang, Y. Luo, F. Yuan, and X. Li, "Segmentation of breast ultrasound image with semantic classification of superpixels," *Med. Image Anal.*, vol. 61, 2020, Art. no. 101657.

[26] Y. Zhou *et al.*, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101918.

[27] B. Lei *et al.*, "Self-co-attention neural network for anatomy segmentation in whole breast ultrasound," *Med. Image Anal.*, vol. 64, 2020, Art. no. 101753.

[28] Y. Hu *et al.*, "AIDAN: An attention-guided dual-path network for pediatric echocardiography segmentation," *IEEE Access*, vol. 8, pp. 29176–29187, 2020.

[29] M. Rachmatullah, S. Nurmaini, A. Sapitri, A. Darmawahyuni, B. Tutuko, and F. Firdaus, "Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view," *Bull. Elect. Eng. Informat.*, vol. 10, no. 4, pp. 1987–1996, 2021.

[30] L. Xu, M. Liu, J. Zhang, and Y. He, "Convolutional-neural-network-based approach for segmentation of apical four-chamber view from fetal echocardiography," *IEEE Access*, vol. 8, pp. 80437–80446, 2020.

[31] C. Zhao *et al.*, "Multi-scale wavelet network algorithm for pediatric echocardiographic segmentation via feature fusion," in *Proc. IEEE 18th Int. Symp. Biomed. Imag.*, 2021, pp. 1402–1405.

[32] B. S. Andreassen, F. Veronesi, O. Gerard, A. H. S. Solberg, and E. Samset, "Mitral annulus segmentation using deep learning in 3-D transesophageal echocardiography," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 4, pp. 994–1003, Apr. 2020.

[33] M. Nasr-Esfahani *et al.*, "Left ventricle segmentation in cardiac MR images using fully convolutional network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 1275–1278.

[34] J. Chen, H. Zhang, W. Zhang, X. Du, Y. Zhang, and S. Li, "Correlated regression feature learning for automated right ventricle segmentation," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–10, 2018, Art. no. 1800610.

[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[36] B. Prabadevi *et al.*, "Toward blockchain for edge-of-things: A new paradigm, opportunities, and future directions," *IEEE Internet Things Mag.*, vol. 4, no. 2, pp. 102–108, Jun. 2021.

[37] L. Sifre and S. Mallat, "Rigid-motion scattering for image classification," Ph.D. dissertation, 2014.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.

[40] J. Jin, A. Dundar, and E. Culurciello, "Flattened convolutional neural networks for feedforward acceleration," 2014, *arXiv:1412.5474*.

[41] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," 2017, *arXiv:1706.03059*.

[42] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[43] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[44] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.