BioMedical Engineering OnLine

RESEARCH

Open Access

A deep learning framework for identifying and segmenting three vessels in fetal heart ultrasound images

Laifa Yan[1,2], Shan Ling[2], Rongsong Mao[1,2], Haoran Xi[2] and Fei Wang[3*]

*Correspondence: wangfeidyh@163.com

[1] College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, China
[2] Hangzhou Institute of Medicine, Chinese Academy of Sciences, Hangzhou, Zhejiang, China
[3] The Center of Four-Dimensional Ultrasound, Affiliated Xiaoshan Hospital, Hangzhou Normal University, Hangzhou, Zhejiang, China

Abstract

Background: Congenital heart disease (CHD) is one of the most common birth defects in the world. It is the leading cause of infant mortality, necessitating an early diagnosis for timely intervention. Prenatal screening using ultrasound is the primary method for CHD detection. However, its effectiveness is heavily reliant on the expertise of physicians, leading to subjective interpretations and potential underdiagnosis. Therefore, a method for automatic analysis of fetal cardiac ultrasound images is highly desired to assist an objective and effective CHD diagnosis.

Method: In this study, we propose a deep learning-based framework for the identification and segmentation of the three vessels—the pulmonary artery, aorta, and superior vena cava—in the ultrasound three vessel view (3VV) of the fetal heart. In the first stage of the framework, the object detection model Yolov5 is employed to identify the three vessels and localize the Region of Interest (ROI) within the original full-sized ultrasound images. Subsequently, a modified Deeplabv3 equipped with our novel AMFF (Attentional Multi-scale Feature Fusion) module is applied in the second stage to segment the three vessels within the cropped ROI images.

Results: We evaluated our method with a dataset consisting of 511 fetal heart 3VV images. Compared to existing models, our framework exhibits superior performance in the segmentation of all the three vessels, demonstrating the Dice coefficients of 85.55%, 89.12%, and 77.54% for PA, Ao and SVC respectively.

Conclusions: Our experimental results show that our proposed framework can automatically and accurately detect and segment the three vessels in fetal heart 3VV images. This method has the potential to assist sonographers in enhancing the precision of vessel assessment during fetal heart examinations.

Keywords: Medical image segmentation, Congenital heart disease, Ultrasound image analysis, Deep learning, Prenatal heart examination

Introduction

Congenital heart disease (CHD) is the most prevalent birth defects, accounting for approximately 28% of all congenital abnormalities [1–3]. Research has reported that the incidence of CHD is around 1% in the global birth population [1–3]. CHD encompasses a spectrum of anatomical anomalies that result from disruptions or developmental irregularities in the formation of the fetal heart and major blood vessels during embryonic development [4]. Notably, it is the leading cause of neonatal mortality [4, 5]. Disturbances in fetal heart and major blood vessels can occur during the early stages of pregnancy, typically within the first 2–3 months, and have the potential to impair the normal growth and function of the infant's heart. Therefore, the early diagnosis of CHD is imperative and holds paramount significance in providing essential medical intervention and mitigating health risks for infants affected by this condition.

Ultrasound has emerged as the primary imaging modality for fetal heart examination thanks to its cost-effectiveness, lack of radiation exposure, and minimal side effects [6, 7]. During an ultrasound examination of the fetal heart, multiple cardiac planes should be carefully examined to thoroughly assess the cardiac four-chamber and vessels [8, 9]. The three-vessel view (3VV) is a critical cardiac plane that reveals the structure and function of the three major vessels of the fetal heart—the pulmonary artery, aorta, and superior vena cava [10]. Some typical cardiac anomalies that may appear normal in the four-chamber view are frequently identified in the 3VV, such as complete transposition of the great arteries, Tetralogy of Fallot, and pulmonary atresia with ventricular septal defect [11]. Therefore, a precise evaluation of the 3VV can improve the detection rate of significant cardiac malformations. However, the effectiveness of ultrasound diagnoses heavily relies on the experience and expertise of physicians, often leading to subjective ultrasound interpretations [6]. In cases where physicians lack sufficient experience, there is a risk of underdiagnosis or misdiagnosis. Experienced sonographers may also face challenges in making accurate diagnoses when dealing with complex examination procedures and a large volume of patients. Therefore, the development of an automated and reliable diagnostic tool capable of assessing cardiac vascular structures during fetal heart examinations is highly desired. Such a tool could significantly alleviate the workload of physicians and assist clinical physicians in performing more precise and efficient early CHD screening.

In recent years, deep learning has made remarkable progress in the field of medical image segmentation because of its powerful ability to autonomously learn image features and perform pixel-level classification [12–14]. One notable architecture is the Fully Convolutional Network (FCN), which consists of multiple convolutional layers and fully connected layers [15]. FCN leverages the deconvolution technique to restore the final feature map to the dimension of the input image, enabling pixel-level predictions and effectively addressing the challenge of semantic image segmentation. U-Net is another network that has been widely adopted for various segmentation tasks [16]. It is named from its U-shaped architecture characterized by an end-to-end encoder–decoder structure. The encoder gradually reduces the spatial dimension of the input image while extracting features. The decoder is responsible for upsampling the feature maps and progressively increasing the spatial dimension with the help of transposed convolutional layers. Recently, researchers have introduced several innovative techniques to enhance feature extraction and decoding capabilities, including the integration of multiple model architectures, the incorporation of residual pathways, and the utilization of attention mechanisms, etc. [17–19]. For example, Zhou et al. [20] proposed U-Net++, which integrates features of varying scales through a cascade of densely interconnected skip connections. This method minimizes semantic loss between feature maps and labels, enhancing the network's ability to capture salient information effectively. Additionally, Oktay et al. [21] proposed Attention U-Net, which introduces attention gates to suppress irrelevant regions and enhance valuable salient features crucial to the target. To address the challenge of multi-scale image segmentation Chen et al. [22] proposed Deeplabv3, which introduces dilated convolutions and ASPP (Atrous Spatial Pyramid Pooling) techniques to maintain the feature map size while effectively controlling the receptive field.

Deep learning has been applied in the field of fetal echocardiography for various tasks, including standard plane identification from a sequence of fetal heart images [23–25], detection of abnormal structures [26–29], and segmentation of cardiac structures [30–33]. Most studies designed for fetal cardiac structure segmentation have focused on the four-chamber view of the fetal heart [30, 31, 33]. In this study, we aim to develop a deep-learning based framework for the accurate segmentation of the three vessels within the three-vessel plane of the fetal heart, namely, the aorta, the pulmonary artery and the superior vena cava. We began by carefully selecting the most promising baseline model for the segmentation of three-vessel cross-sectional images. Subsequently, we validated the effectiveness of region-of-interest (ROI) detection before segmentation, as many studies have demonstrated that ROI detection followed by segmentation can significantly enhance the segmentation of small objects [34, 35]. Lastly, we devised an attention-based multi-scale feature extraction module to address the challenge posed by the large variation in vessel sizes. In comparison to several existing deep learning methods, our proposed framework demonstrates best performance in segmenting the three-vessel plane of the fetal heart. Our method holds the potential to assist sonographers in enhancing the effectiveness and accuracy of vessel assessment during fetal heart examinations.

Results

Comparison of baseline models for full-size 3VV image segmentation

We first conducted a comparative experiment to assess the performance of several baseline segmentation models on full-size fetal 3VV ultrasound images, including FCN [15], U-Net [16], U-Net++ [20], Attention U-Net [21] and Deeplabv3 [22]. Table 1 presents the results of these baseline segmentation models on our collected dataset of fetal heart 3VV images. As shown in Table 1, Deeplabv3 surpasses the other models in segmenting

Table 1 Comparative analysis of baseline models for three-vessel segmentation in full-size 3VV images

Baseline	Dice			Mean		
	PA	Ao	SVC	IoU	HD	Dice
FCN [15]	78.81	75.84	54.55	58.61	3.95	69.73
U-Net [16]	80.75	85.10	68.40	67.56	3.65	78.08
U-Net++ [20]	<u>82.38</u>	<u>86.60</u>	69.45	<u>69.78</u>	<u>3.56</u>	79.48
Attention U-Net [21]	81.73	84.91	<u>72.80</u>	69.29	3.63	<u>79.82</u>
Deeplabv3 [22]	**83.49**	**86.61**	**73.36**	**70.74**	**3.50**	**81.15**

PA: pulmonary artery; Ao: aorta; SVC: superior vena cava; Mean: the average value of the three vessels

The optimal value is highlighted in bold, while the second-best value is underscored (in column) all three vessels, achieving the highest Dice and IoU scores and the lowest HD scores [36]. While Deeplabv3 has exhibited strong performance in PA and Ao segmentation, its performance in segmenting the SVC remains suboptimal, primarily due to the SVC's small size.

Evaluation of the two-stage framework: ROI detection followed by segmentation

In this section, we evaluated the effectiveness of ROI detection in relation to its subsequent segmentation. We compared three ROI localization strategies for our task, and the results are presented in Table 2. "Deeplabv3 + Deeplabv3" represents a two-stage framework comprising two Deeplabv3 models, with the first performing a binary segmentation for ROI localization and the second performing a multi-class segmentation for a fine extraction of the three vessels. "Faster-RCNN + Deeplabv3" is a framework where ROI detection is carried out by Faster-RCNN, followed by subsequent segmentation with Deeplabv3. "Yolov5 + Deeplabv3" is a framework in which ROI detection is performed using Yolov5, followed by subsequent segmentation with Deeplabv3. As shown in Table 2, the two-stage framework "Yolov5 + Deeplabv3" demonstrates superior performance compared to other frameworks for our task. In comparison to the baseline one-stage method, which involves using Deeplabv3 for full-size image segmentation, "Yolov5 + Deeplabv3" enhances the Dice scores of PA, Ao, and SVC by 1.95%, 0.74%, and 3.18%, respectively. Figure 1 provides exemplary results of various ROI detection strategies.

Comparison of our method with state-of-the-art segmentation models

Our experiments have demonstrated that the two-stage framework, where in Yolov5 detection followed by Deeplabv3 segmentation, exhibits superior performance in our task. To further enhance the segmentation performance, we replace the Atrous Spatial Pyramid Pooling (ASPP) module in Deeplabv3 with a novel module called "Attentional Multi-scale Feature Fusion (AMFF)". As shown in Table 3, our proposed method significantly outperforms existing CNN-based segmentation models in the segmentation of all three vessels in fetal heart ultrasound images across all evaluation metrics. Compared to the original Deeplabv3, our model increases the Dice score for Ao by 1.77% and for SVC by 1.02%. Figure 2 provides a visual comparison of the segmentation performance of different methods in our task.

Table 2 Comparative analysis of two-stage frameworks for vessel segmentation in fetal 3VV ultrasound images using varied ROI localization strategies

Method	Dice			Mean		
	PA	Ao	SVC	IoU	HD	Dice
Deeplabv3 (full-size)	83.49	86.61	73.36	70.74	3.50	81.15
Deeplabv3 + Deeplabv3	82.71	83.79	66.15	67.92	3.58	77.55
Faster-RCNN + Deeplabv3	72.03	67.36	61.42	57.59	4.14	66.93
Yolov5 + Deeplabv3	**85.44**	**87.35**	**76.52**	**73.25**	**3.30**	**83.11**

The optimal value is highlighted in bold (in column)

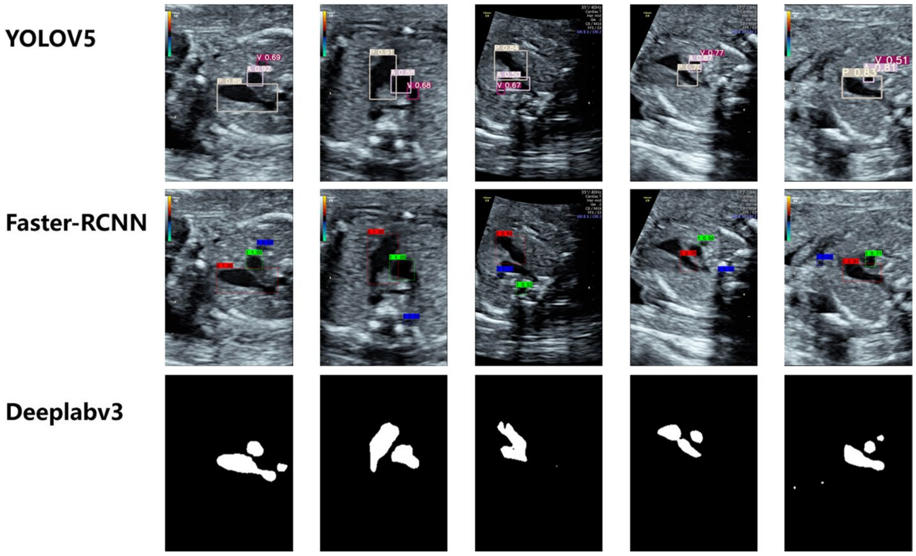

Fig. 1 Visual comparison of different ROI localization strategies

Table 3 Comparison of different segmentation models for vessel segmentation in YOLOv5-generated ROIs

Method	Dice			Mean		
	PA	Ao	SVC	IoU	HD	Dice
FCN [15]	82.90	81.64	69.03	67.40	3.54	77.86
U-Net [16]	83.80	83.28	72.94	70.18	3.48	80.01
U-Net++ [20]	83.31	84.25	72.41	69.92	3.47	79.99
Attention U-Net [21]	82.93	81.41	66.72	67.13	3.61	77.02
Deeplabv3 [22]	85.44	87.35	76.52	73.25	3.30	83.11
T-S-deeplab [32]	81.36	86.89	74.24	71.38	3.46	81.35
Ours	**85.55**	**89.12**	**77.54**	**74.51**	**3.25**	**84.07**

The optimal value is highlighted in bold, while the suboptimal value is underlined (in column)

Comparing with the underlined values, this further illustrates the superiority of our proposed method

Discussion

In this study, we propose a deep-learning based framework for the automatic identification and segmentation of the PA, Ao, and SVC in fetal heart 3VV ultrasound images. We conducted a performance comparison of several baseline segmentation models in our specific task and our method has the potential to assist physicians in diagnosing congenital heart defects more effectively and objectively in clinical practice. Specifically, the automatic segmentation of vessels in the 3VV of the fetal heart could assist in-experienced physicians to efficiently localize the three vessels. In addition, it was a prerequisite for developing a technique for the automatic measurement of the size ratio between the pulmonary artery and the aorta, a biometric essential for screening CHD.

One challenge in our task is the large variation in the size of vessels. Compared with the other two vessels, the SVC is much smaller, thus often being overlooked by a multi-object segmentation model. We conducted a performance comparison of several baseline segmentation models, including FCN, U-Net, U-Net++, Attention U-Net, and

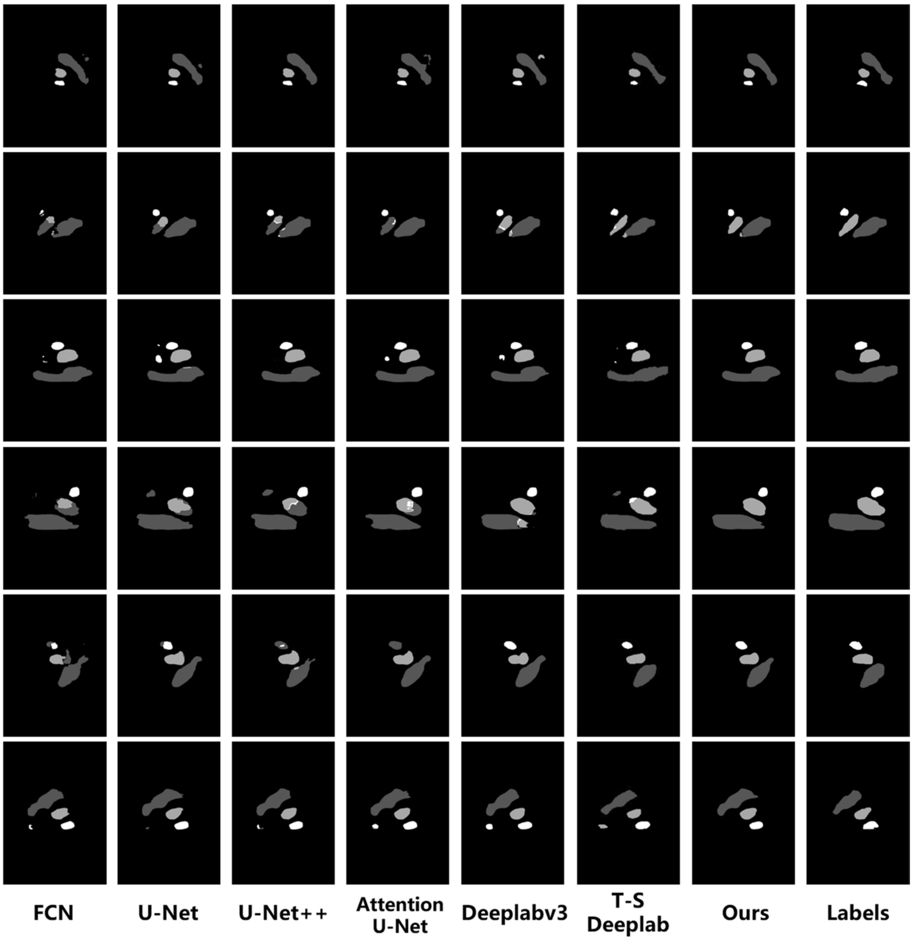

Fig. 2 Visual comparison of the performance of different segmentation models in our task

Deeplabv3, using our collection of full-sized ultrasound images. As shown in Table 1, Deeplabv3 outperforms the other models thanks to its multiscale-feature extraction mechanism that incorporates varying receptive field sizes, rendering it advantageous for capturing features of vessels with diverse scales.

Another challenge in segmenting the 3VV images is the interference caused by the surrounding background [37]. This is particularly problematic for small-sized SVC vessels, which often occupy a limited region in the image. To mitigate the adverse effects of the irrelevant background information and allow the segmentation model to concentrate on vessel details, we employed a two-stage framework, with the first stage for detecting the ROI, and the second stage for a fine segmentation of the three vessels in the cropped ROI images. We experimented with different ROI extraction strategies in combination with the Deeplabv3 segmentation model. Our findings, shown in Table 2, indicate that the combination of Yolov5 and Deeplabv3 produced the most optimal segmentation performance, while the combination of Faster R-CNN and Deeplabv3 yielded the poorest results. A visual analysis of Fig. 1 revealed that the discrepancy can be attributed to Faster R-CNN's higher tendency for false positives during the detection phase, particularly in the inaccurate recognition of the smallest SVC. This inaccuracy led to a significant error in ROI extraction, subsequently hindering satisfactory segmentation results in the second stage. Notably, the performance of Deeplabv3 combined with Deeplabv3 was even worse than directly performing a multi-class segmentation task on full-sized images. Similarly, based on visual analysis of the results in Fig. 1, Deeplabv3 exhibited a higher degree of difficulty in identifying the SVC during initial ROI extraction for foreground–background separation. The loss of SVC in the ROI extraction stage makes it challenging for Deeplabv3 to achieve effective segmentation in the second stage.

To further improve the segmentation performance, we replaced the ASPP module in the Deeplabv3 it with our designed AMFF (Attentional Multi-scale Feature Fusion) module. Specifically, we devised multi-scale feature extraction branches with varied dilation rates, and flow small-scale features through different branches using hierarchical connections. We also introduced a spatial attention mechanism at the end of each branch to further enhance feature representations. These modifications allowed the model to effectively capture multi-scale features of all blood vessels in our task. As demonstrated in Table 3, within the two-stage framework, the integration of our AMFF module into Deeplabv3 resulted in significant improvements in the segmentation of both the Ao and SVC.

While our framework has exhibited promising outcomes, this study has two limitations. First, the data size is restricted, and all data are obtained from a single hospital, lacking external validation data from other medical facilities. As a result, one of our future objectives involves validating our model on a larger and more diverse dataset. Second, our 3VV segmentation framework is not trained end-to-end. It consists of an ROI extraction model and a segmentation model, each trained independently. In the future, we aim to improve our method by transitioning to a unified framework, thereby enhancing efficiency in both training and application.

Conclusions

In this study, we propose a two-stage deep learning framework for vessel segmentation in fetal 3VV ultrasound images, involving a Yolov5 for ROI localization and a Deeplabv3 model equipped with our novel AMFF module for segmentation within the ROI regions. Our proposed framework has exhibited remarkable performance in segmenting all three vessels with average HD value of 3.25 and Dice value of 84.07% and IoU value of 74.51%, surpassing other state-of-the-art segmentation models. Our future work includes validating our method on a larger and more diverse dataset collected from multiple hospitals to enhance the generalizability of our approach.

Methods

In this paper, we developed a two-stage deep-learning framework for the identification and segmentation of the vessels in fetal heart 3VV images to assist radiologists in diagnosing vascular structural abnormalities. The overall workflow of our method is illustrated in Fig. 3.

The code and detailed instructions for users have been made available to the public at the following link: https://github.com/ylfas/3VV_demo/blob/master/README.md.

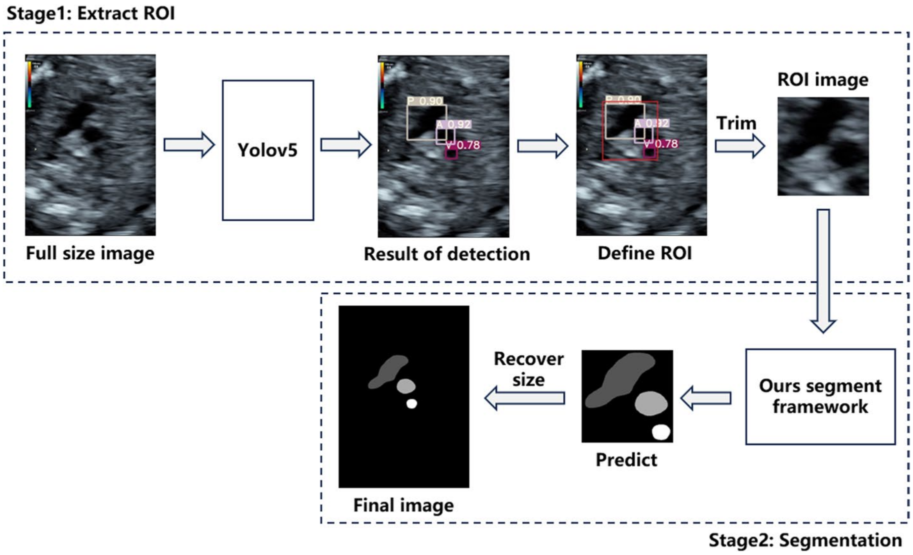

Fig. 3 Workflow of our proposed strategy

Table 4 Types of CHD reported in our dataset

Data type	Number
Normal images	413
Abnormal images	98
Abnormal vessel diameter ratio	67
Cardiac chamber abnormality	6
Arterial vascular abnormality	15
Outflow tract abnormality	3
Tetralogy of Fallot	7

Clinical dataset

The dataset used in this study was obtained from Hangzhou Normal University Affiliated Xiaoshan Hospital, China, with ethical approval granted by the hospital's Ethics Committee. The dataset comprises images acquired from 607 pregnant women in mid-term pregnancy, with gestational ages ranging from 20 to 40 weeks. The images were captured using a GE Voluson E8 ultrasound machine equipped with a 2–5 MHz linear ultrasound transducer. During the examination process, physicians conducted a comprehensive assessment of the fetal cardiac structure and function. Standard cardiac planes, including the 3VV, were captured and stored. All patient information in the images was de-identified. After excluding non-standard or challenging-to-interpret sectional images, a refined set of 511 images was obtained, including 413 normal cases and 98 abnormal cases. Table 4 presents the distribution of various types of CHD data reported within our final dataset. Subsequently, these images were labeled by two experienced physicians with 15 and 20 years of expertise, respectively. The two physicians independently annotated the boundaries of the vessels within the 3VV images. If there were notable disparities between their annotations, a consensus was reached through a comprehensive discussion. The final mask label for each image was derived by averaging the annotations of the two physicians.

Data preprocessing

The dataset was divided into training, validation, and test sets in a ratio of 7:1:2. Before being fed into a network, all images were resized to 256 × 256 and subjected standard normalization. Data augmentation techniques such as random horizontal flipping, random angle rotations, and random scale adjustments are applied to mitigate over-fitting [38].

ROI localization

Many studies have demonstrated that a two-stage deep learning framework that involves ROI detection followed by segmentation, can significantly enhance the final segmentation performance, particularly for small objects [34, 35]. Specifically, the first stage of this framework aims to localize the target objects in the original images. This localization can be accomplished using an object detection model, such as Faster-RCNN [39], and YoLo-series models [40–42]. It can also be achieved through a segmentation model applied to full-size images to obtain coarse object masks [43, 44]. The areas containing the coarse masks are treated as ROIs. In the second stage, the ROI regions are cropped from the original images and fed into a second model to achieve fine object segmentation. This two-stage approach can mitigate the adverse effects of irrelevant background and enable the model to concentrate on object details, thus improving the segmentation performance.

We compared three different ROI localization strategies on our dataset. The first strategy involves using Deeplabv3 to roughly segment the full-size image. The ROI region was then defined by expanding the segmented vessel masks. The ground truth labels in this method were binary vessel masks, where all three vessels were labeled with 1 and the background was labeled with 0. The second strategy utilized YOLOv5 to identify the three vessels within the full-size image. From the predicted candidate boxes for each class, the one with the highest confidence was selected as the final output. This process yielded three candidate boxes, each containing one of the three vessels. The minimum bounding rectangle that enclosed all three predicted boxes was extended by 5 pixels to obtain the final ROI of the image (as shown in Fig. 3). The third strategy was similar to the second strategy, but instead employed Faster RCNN [39] as the detection model. In the latter two strategies, ground truth labels were bounding boxes of the three vessels.

Attention-based multiscale feature fusion framework for vessel segmentation

The second stage of our framework is a modified Deeplabv3 equipped with our novel AMFF module for instance segmentation of the three vessels. The ROI regions are cropped from the original images and fed into the second model to achieve fine object segmentation. For the second model training, the label format comprises boundary masks for each blood vessel within the cropped region of each ultrasound image from the initial stage, with individual differentiation of each blood vessel as a distinct category. The AMMF's architecture is illustrated in Fig. 4. A cascade of ResNet34 [17] blocks are used to encode image features. To be concrete, the initial

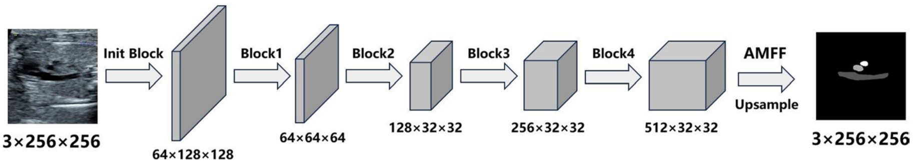

Fig. 4 The framework of modified network based on deeplabv3. AMFF: Attentional Multi-scale Feature Fusion module

phase involves an initialization block, which consists of a 7 × 7 convolution with a stride of 2, a padding of 3, and a Batch Normalization (BN) layer. Following the initialization block, multiple copies of the last ResNet34 block that referred to as blocks 1 to 4 in Fig. 4 are employed and organized in a cascading manner. These blocks contain four 3 × 3 convolutions, with the first convolution having a stride of 2, except for block 3 and block 4. The resulting deep features are subsequently input into our specially designed AMFF module to enhance feature representations across various scales.

Figure 5 displays the structure of our AMFF module. It consists of multiple feature extraction branches with convolutions of different dilation rates to obtain features with diverse receptive fields. To ensure that each branch preserves small object features, we encourage interaction among branches by integrating features through hierarchical connections. Furthermore, we introduce spatial attention operations to selectively enhance the most effective features of each branch, thereby improving feature representations at multiple scales. Subsequently, the features from all branches are concatenated to create fused features that retain information related to multi-scale targets. The fused feature (2048 × 32 × 32) is then dimensionally reduced to three channels through two convolutional layers, with each channel predicting one type of vessel. Finally, the prediction is upsampled eight times through bilinear interpolation to restore it to the original image resolution.

The loss function for training the model is a combination of cross-entropy loss and dice loss, which is defined as:

$$L = 0.5 * L_{\text{dice}} + 0.5 * L_{ce} \tag{1}$$

$$L_{ce} = -\frac{1}{MN}\sum_{k}^{M}\sum_{i}^{N} W_k \cdot g(k,i) \cdot \log(p(k,i)), W_k = \frac{N_{\text{total}}}{N_k} \tag{2}$$

$$L_{\text{dice}} = \frac{1}{MN}\sum_{k}^{M}\left(1 - 2\frac{\sum_{i}^{N} g(k,i) \cdot p(k,i)}{\sum_{i}^{N} g(k,i) + \sum_{i}^{N} p(k,i)}\right) \tag{3}$$

In above equations, the symbol k represents the index of the k-th channel, and the symbol i denotes the position of the i-th pixel within each channel. Therefore, the symbol $p(k, i)$ is used to denote the prediction of the i-th pixel in the k-th channel of the matrix, while the symbol $g(k, i)$ is employed to represent the i-th pixel in the k-th channel of the actual segmentation mask.

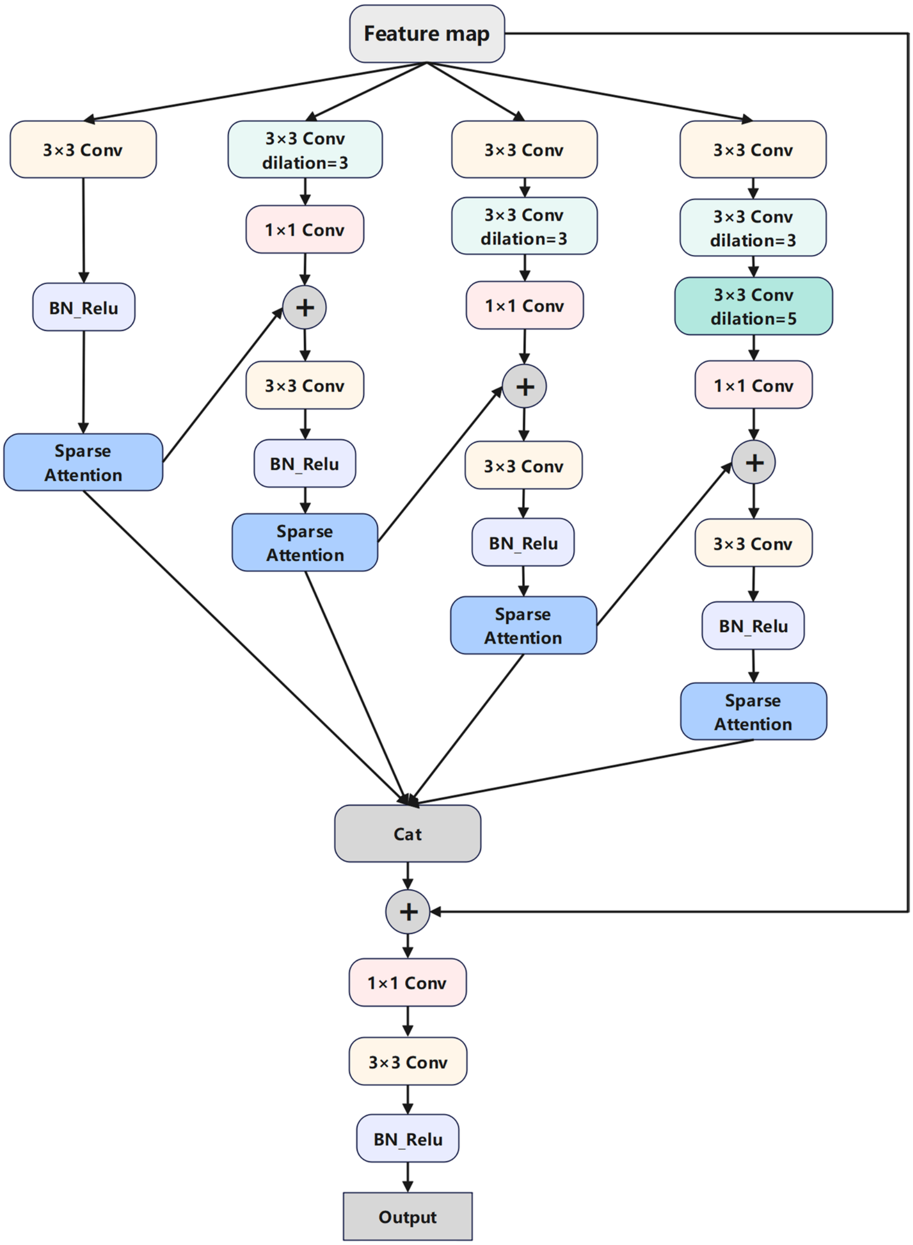

Fig. 5 Structure of AMFF (Attentional Multi-scale Feature Fusion) module

Model comparison and evaluation metrics

To select the best-performing baseline model as the foundation for this study, we first conducted a comparative experiment to assess the performance of several baseline segmentation models on full-size fetal 3VV ultrasound images, including FCN [15], U-Net [16], U-Net++ [20], Attention U-Net [21] and Deeplabv3 [22]. All models underwent training and evaluation using the identical data-splitting strategy (70% training, 10% validation, 20% test) and hyperparameter settings. To be specific, all models were trained for a total of 35 epochs using the Adam optimizer [45]. The learning rate was decreased from the initial value of 0.001to 0.0001 in the final 10 epochs, with a decay rate of 1e−8.

The loss function is a combination of cross-entropy loss and Dice loss, with both losses weighted equally at 0.5 [46].

In this study, we employed several metrics to evaluate the segmentation performance of different methods, including IoU, HD, and Dice coefficient. IoU is defined as the ratio of the overlap area between the predicted and ground truth masks to the area of their union. It quantifies the spatial overlap between the two masks, providing insight into segmentation accuracy. HD represents the maximum distance between the predicted and ground truth boundaries. It offers a measure of the maximum segmentation error, helping us understand the extent of boundary discrepancies. The Dice coefficient is calculated as twice the intersection of the predicted and true masks divided by the sum of their areas. It serves as a metric for assessing the agreement between the predicted and true masks and is commonly used in medical segmentation tasks. These metrics collectively offer a comprehensive evaluation of the segmentation performance, aiding in the assessment of the accuracy and effectiveness of our method across different vessels [36]. The three metrics are defined as follows:

$$\mathrm{IOU} = \frac{\text{Intersection Area}}{\text{Union Area}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{4}$$

$$\mathrm{Dice} = \frac{2 * \mathrm{TP}}{2 * \mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{5}$$

In the Eqs. (4) and (5), TP (True Positives) represents the number of observations correctly predicted as the positive class. TN (True Negatives) represents the number of observations correctly predicted as the negative class. FP (False Positives) indicates the number of observations that were incorrectly predicted as the positive class. FN (False Negatives) represents the number of observations that were incorrectly predicted as the negative class.

$$\mathrm{HD}_A = \max_{a \in A}(\min_{b \in B} d(a, b)) \tag{6}$$

$$\mathrm{HD}_B = \max_{b \in B}(\min_{a \in A} d(b, a)) \tag{7}$$

$$\mathrm{HD(A,B)} = \max(\mathrm{HD}_A, \mathrm{HD}_B) \tag{8}$$

In the Eq. (8), A represents the set of points in our predicted matrix, while B represents the set of points in the actual mask label matrix. HD_A calculates the maximum value of the shortest distances between all points in set A to the points in set B, whereas HD_B computes the maximum value of the shortest distances between all points in set B to the points in set A. The final Hausdorff distance value, denoted as HD, is determined by selecting the larger of the two calculated values.

Abbreviations

CHD	Congenital heart disease
PA	Pulmonary artery
Ao	Aorta
SVC	Superior vena cava
3VV	Three-vessel view
ROI	Region of Interest

AMFF	Attentional Multi-scale Feature Fusion
FCN	Fully Convolutional Network
ASPP	Atrous Spatial Pyramid Pooling
IOU	Intersection over Union
HD	Hausdorff Distance
BN	Batch Normalization

Acknowledgements
Not applicable.

Author contributions
YLF made contributions to research design, computer program algorithms and modifications, experiment, and writing. LS contributed to research design, manuscript revision, and conceptualization. MRS contributed to experiments, data processing and analysis. XHR contributed to data processing and analysis, and program algorithms. WF contributed to data collection, annotation of data, and provided critical guidance to the article. All authors read and reviewed the final manuscript.

Funding
Not applicable.

Availability of data and materials
The data sets during the current study are not publicly available due to hospital information protection mechanism, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the hospital's Ethics Committee of the Hangzhou Normal University Affiliated Xiaoshan Hospital, China.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 7 November 2023 Accepted: 18 March 2024
Published online: 02 April 2024

References

1. Hoffman JIE, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol. 2002;39(12):1890–900.
2. Williams K, Carson J, Lo C. Genetics of congenital heart disease. Biomolecules. 2019;9(12):879.
3. Qu Y, Liu X, Zhuang J, et al. Incidence of congenital heart disease: the 9-year experience of the Guangdong registry of congenital heart disease, China. PLoS ONE. 2016;11(7): e0159257.
4. Becker R, Wegner RD. Detailed screening for fetal anomalies and cardiac defects at the 11–13-week scan. Ultrasound Obstet Gynecol. 2006;27(6):613–8.
5. Gilboa SM, Devine OJ, Kucik JE, et al. Congenital heart defects in the United States: estimating the magnitude of the affected population in 2010. Circulation. 2016;134(2):101–9.
6. Qiu X, Weng Z, Liu M, et al. Prenatal diagnosis and pregnancy outcomes of 1492 fetuses with congenital heart disease: role of multidisciplinary-joint consultation in prenatal diagnosis. Sci Rep. 2020;10(1):7564.
7. Menahem S, Sehgal A, Meagher S. Early detection of significant congenital heart disease: the contribution of fetal cardiac ultrasound and newborn pulse oximetry screening. J Paediatr Child Health. 2021;57(3):323–7.
8. Ogge G, Gaglioti P, Maccanti S, et al. Prenatal screening for congenital heart disease with four-chamber and outflow-tract views: a multicenter study. Ultrasound Obstet Gynecol. 2006;28(6):779–84.
9. Van Nisselrooij AEL, Teunissen AKK, Clur SA, et al. Why are congenital heart defects being missed? Ultrasound Obstet Gynecol. 2020;55(6):747–57.
10. Yoo SJ, Lee YH, Kim ES, et al. Three-vessel view of the fetal upper mediastinum: an easy means of detecting abnormalities of the ventricular outflow tracts and great arteries during obstetric screening. Ultrasound Obstet Gynecol. 1997;9(3):173–82.
11. Tanaka T, Inamura N, Kawazu Y, et al. A rapid and easy objective evaluation of the three vessel view to enhance diagnostic confidence in fetal echocardiography. J Fetal Med. 2022;9(01):1–5.
12. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48.
13. O'Shea K, Nash R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458, 2015.
14. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1–9.
15. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431–3440.

16. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234–241.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–778.
18. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017:30.
19. Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). 2018: 3–19
20. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. Unet++: a nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing, 2018: 3–11.
21. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.
22. Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.
23. Rachmatullah MN, Nurmaini S, Sapitri AI, et al. Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view. Bull Electr Eng Inform. 2021;10(4):1987–96.
24. Nurmaini S, Rachmatullah MN, Sapitri AI, et al. Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection. Sensors. 2021;21(23):8007.
25. Li F, Li P, Wu X, et al. FHUSP-NET: a multi-task model for fetal heart ultrasound standard plane recognition and key anatomical structures detection. Comput Biol Med. 2023;168: 107741.
26. Torrents-Barrena J, Piella G, Masoller N, et al. Segmentation and classification in MRI and US fetal imaging: recent trends and future prospects. Med Image Anal. 2019;51:61–88.
27. Arnaout R, Curran L, Zhao Y, et al. Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning. medRxiv, 2020: 2020.06. 22.20137786.
28. Nurmaini S, Partan RU, Bernolian N, et al. Deep learning for improving the effectiveness of routine prenatal screening for major congenital heart diseases. J Clin Med. 2022;11(21):6454.
29. Zhang Y, Zhu H, Cheng J, et al. Improving the quality of fetal heart ultrasound imaging with multihead enhanced self-attention and contrastive learning. IEEE J Biomed Health Inf. 2023;27:5518–29.
30. An S, Zhu H, Wang Y, et al. A category attention instance segmentation network for four cardiac chambers segmentation in fetal echocardiography. Comput Med Imaging Graph. 2021;93: 101983.
31. Dong J, Liu S, Wang T. ARVBNet: real-time detection of anatomical structures in fetal ultrasound cardiac four-chamber planes. In: Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting: First International Workshop, MLMECH 2019, and 8th Joint International Workshop, CVII-STENT 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 1. Springer International Publishing, 2019, 130–137.
32. Cai Q, Chen R, Li L, et al. The application of knowledge distillation toward fine-grained segmentation for three-vessel view of fetal heart ultrasound images. Comput Intell Neurosci. 2022;2022:1–7.
33. Xu L, Liu M, Shen Z, et al. DW-Net: a cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. Comput Med Imaging Graph. 2020;80: 101690.
34. Pollicelli D, Coscarella M, Delrieux C. RoI detection and segmentation algorithms for marine mammals photo-identification. Eco Inform. 2020;56: 101038.
35. Vu K, Hua KA, Tavanapong W. Image retrieval based on regions of interest. IEEE Trans Knowl Data Eng. 2003;15(4):1045–9.
36. Polak M, Zhang H, Pi M. An evaluation metric for image segmentation of multiple objects. Image Vis Comput. 2009;27(8):1223–7.
37. Kremkau FW, Taylor KJ. Artifacts in ultrasound imaging. J Ultrasound Med. 1986;5(4):227–37.
38. Ying X. An overview of overfitting and its solutions. Journal of physics: Conference series. IOP Publishing, 2019, 1168: 022022.
39. Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015:28.
40. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779–788.
41. Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
42. Bochkovskiy A, Wang C Y, Liao HYM. Yolov4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
43. He K, Gkioxari G, Dollár P, et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017: 2961–2969.
44. Yan K, Wang X, Lu L, et al. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J Med Imaging. 2018;5(3):036501–036501.
45. Kingma D P, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
46. Jadon S. A survey of loss functions for semantic segmentation. In:2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB). IEEE, 2020: 1–7.

Publisher's Note