Neural Computing and Applications
https://doi.org/10.1007/s00521-022-07217-1

ORIGINAL ARTICLE

An improved semantic segmentation with region proposal network for cardiac defect interpretation

Siti Nurmaini[1] · Bayu Adhi Tama[2] · Muhammad Naufal Rachmatullah[1] · Annisa Darmawahyuni[1] · Ade Iriani Sapitri[1] · Firdaus Firdaus[1] · Bambang Tutuko[1]

Received: 9 August 2021 / Accepted: 29 March 2022

Abstract

Detecting cardiac abnormalities between 14 and 28 weeks of gestation with an apical four-chamber view is a difficult undertaking. Several unfavorable factors can prevent such detection, such as the fetal heart's relatively small size, unclear appearances in anatomical structures (e.g., shadows), and incomplete tissue boundaries. Cardiac defects without segmentation are not always straightforward to detect, so using only segmentation cannot produce defect interpretation. This paper proposes an improved semantic segmentation approach that uses a region proposal network for septal defect detection and combines two processes: contour segmentation with U-Net architecture and defect detection with Faster-RCNN architecture. The model is trained using 764 ultrasound images that include three abnormal conditions (i.e., atrial septal defect, ventricular septal defect, and atrioventricular septal defect) and normal conditions from an apical four-chamber view. The proposed model produces a satisfactory mean intersection over union, mean average precision, and dice similarity component metrics of about 75%, 87.80%, and 96.37%, respectively. Furthermore, the proposed model has also been validated on 71 unseen images in normal conditions and produces 100% sensitivity, which means that all normal conditions without septal defects can be detected effectively. The developed model has the potential to identify the fetal heart in normal and pathological settings accurately. The developed deep learning model's practical use in identifying congenital heart disorders has substantial future promise.

Keywords Cardiac septal defect · Fetal echocardiography · U-Net architecture · Faster-RCNN · Deep Learning

✉ Siti Nurmaini
siti_nurmaini@unsri.ac.id; sitinurmaini@gmail.com

Bayu Adhi Tama
bayuat@ibs.re.kr

Muhammad Naufal Rachmatullah
naufalrachmatullah@gmail.com

Annisa Darmawahyuni
riset.annisadarmawahyuni@gmail.com

Ade Iriani Sapitri
adeirianisapitri13@gmail.com

Firdaus Firdaus
virdauz@gmail.com

Bambang Tutuko
bambangtutuko60@gmail.com

[1] Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang, South Sumatera 30139, Indonesia

[2] Data Science Group, Institute for Basic Science (IBS), Daejeon 34126, Republic of Korea

Published online: 18 April 2022

1 Introduction

Since it is vital to understand congenital heart diseases (CHDs), an accurate examination of fetal heart anatomy and structure is essential [1]. As the most prevalent congenital disability, CHDs account for nearly 48% of deaths during infancy in the USA from 1999 to 2006 [2, 3]. It has been known that early diagnosis and proper treatment would significantly improve the survival of infants with CHDs [4]. Due to their inexpensive cost and radiation-free characteristics, ultrasounds (US) are often utilized during regular obstetrics to identify fetal cardiac abnormalities [1]. Prenatal ultrasound is used to diagnose CHDs between 14 and 28 gestational weeks in clinical practice [1, 5]. A cardiac defect is a clinically significant subtype of CHDs with abnormalities in anatomy and heart structure. This defect is caused by an incomplete septum connection between the lower and upper chambers of the heart either in the atrium, ventricle, or both, which vary in size. The most common types of cardiac defects are atrial septal defect (ASD), ventricular septal defect (VSD), and atrioventricular septal defect (AVSD) [1]. Exploring cardiac defects in a fetus is critical for investigating these complex associations.

Clinicians use four views for fetal heart anatomy and structure examination: four-chambers (4C), left and right ventricular outflow tract (LVOT/RVOT), and three-vessel trachea (3VT). The 4C view is a straightforward normal fetal cardiac scan, while the LVOT, RVOT, and 3VT views are more sophisticated. However, the apical four-chamber (A4C) is only a basic standard view. This indispensable view reveals the embryonic heart's architecture and structure, allowing clinicians to assess the key cardiac structures including the epicardium, levocardia, left ventricle, left atrium, right ventricle, right atrium, and descending aorta [5, 6]. Regrettably, clinicians must possess an extensive theoretical background and clinical competence to comprehend such structures [7–9]. Analyzing fetal hearts manually interpreting US images is difficult due to the following factors: (i) signal dropout, artifacts, missing boundaries, attenuation, shadows, and speckle noise may all influence the image formation features inherent to US images.; (ii) the fetal heart is small-sized with a complex structure, fast movements, and a low signal-to-noise ratio; and (iii) full fetal heart structure detection is complicated, even for experienced physicians. As a result, fetal heart examination using US would not be optimal [5, 6, 10, 11], with prenatal detection rates of only 30–50% reported in developed countries [12]. Many undetected cardiac septal defects occur because the structural fetal heart view was not captured properly and precisely, or because the defect was exhibited but not identified by the operator due to the low-quality picture.

Accurate identification of cardiac defects on A4C may give pathological information and save physicians significant time examining and quantifying defect severity. Several technique-based computer-aided diagnosis solutions have been proposed to solve medical imaging problems [13–18]. However, such methods use mathematical models to process medical imaging, which have high computational complexity and require an image pre-processing stage before reaching a medical diagnosis. Methods with multiple levels of medical image representation, which process raw data automatically without the image pre-processing stage, are desirable for performing accurate segmentation, classification, or detection tasks. Deep learning (DL) approaches have been proposed to overcome the challenges above, and they have shown promising results [15, 19]. For example, the fully convolutional networks (FCNs) model developed for real-world clinical US data, achieves a classification error rate of 23.48% [8]. To cope with unconstrained psoriasis pictures for computer-aided diagnosis, deep convolutional neural networks (CNNs) are developed, followed by an MV base point tracking technique to fetal left ventricle segmentation in echocardiographic sequences [16]. The performance indices of the proposed scheme were higher than 93%, and its mean average precision (mAP) rates were at least 85.9% [10]. A novel DL-based approach has been proposed to identify fetal cardiac standard planes from 2D US videos [6]. For the task of identifying the four standard planes, the proposed network achieved an accuracy of 92% and an average F1-score of 0.919.

However, the fetal cardiac defects are infrequently detected during routine obstetric visits, significantly impacting the construction of a DL model for defect detection. Bear in mind that DL is highly dependent on the availability of huge labeled datasets since the limited amount of data may decrease its performance accuracy. CNNs are one of the most common and efficient methods that address medical imaging problems [9, 20, 21]. Moreover, CNNs are suited for recognizing the patterns in US images [6], even for relatively small datasets that produce excellent performance [19, 22]. Most existing researches have been undertaken to explore the use of CNNs for classification [23], object detection [24, 25], and segmentation [26, 27].

Image segmentation for fetal echocardiography using CNNs provides a new research opportunity to produce a precise and reliable method. CNNs constructed from the segmentation results are used to track the fetal heart's organ growth, especially its anatomy measurements, and can potentially reduce human errors in monitoring heart abnormalities in the fetus [5, 15, 19, 28]. Several

segmentation models have been developed to overcome inherent artifacts and low image quality [11, 21, 29–31]. Medical images with leakage, blurred borders, and subject-to-subject variations can reach a 94% detection rate [19, 21]. Some attempts have been made to address these issues, but more focus has been on adult heart problems. Pixel-based image segmentation has been successfully developed for fetal heart segmentation, and it is even effective with a small number of images [9, 20, 22].

However, the segmentation result cannot completely interpret an abnormality, as it only recognizes the heart structure's contour and heart-chamber (i.e., atrium, ventricle, aorta, valve, and levocardia). The incomplete wall-chamber boundary should be diagnosed in a defective fetal heart in terms of the defect position and size. Furthermore, the lack of septal defect in the fetal heart is another notable limitation of previous studies [8]. Therefore, there is a need to consider US fetal heart anatomy segmentation for further investigation. The main objective of cardiac septal defect detection is to detect the position of a defect in the septum. Such a position would determine the abnormality type. Therefore, the improved object detection model with a segmentation approach can facilitate an accurate decision in septal defects interpretation. In this study, two CNNs architectural models are combined. The proposed method produces accurate defect detection via a pre-trained and fine-tuned end-to-end network. In short, this article makes the following contributions:

- An improved semantic segmentation with region proposal network architecture for fetal heart defect interpretation from an A4C view is proposed.
- A pre-trained semantic segmentation model and fine-tuned object detection model are taken into account to simplify the algorithm.
- Three different conditions of cardiac defect (i.e., ASD, VSD, and ASD) and normal conditions are incorporated.
- The robustness of the proposed approach is demonstrated by carrying out inter-patient images in normal conditions.

We break down the remainder of this work as follows. Section 2 outlines pertinent works in the existing body of knowledge. Section 3 details the technique used in this research, while Sect. 4 explains the findings and discussion. Finally, Sect. 5 summarizes the findings and makes recommendations for further research.

2 Apical four-chamber view segmentation and detection approaches

This section summarizes various approaches reported in the literature. Zheng et al. [32] formulated a heart chamber segmentation as a twofold learning problem for anatomical structure localization and boundary delineation. Numerous authors have addressed ultrasound image segmentation from the A4C view. Nakphu et al. [33] proposed a Marker-controlled Watershed segmentation method. Pre-processing and filtering parameters have affected segmentation, where the morphological technique was found to be the most effective pre-processing approach. Additionally, dummy images were used for testing the segmentation algorithm. Cao et al. [34] used a new convexity pursuit segmentation algorithm. Syeda-Mahmood et al. [35] incorporated a modified parametric distorted elliptic shape model to capture aberrations from the heart's typical appearance to discriminate between normal and pathological left ventricular shapes. Kang et al. [36] proposed a novel framework (e.g., gradient-assisted localized active contour model) to segment the four chambers of the heart automatically. To verify the proposed four-chamber segmentation approach, the authors asked a cardiologist with nine years of clinical experience to manually segment the four chambers in each dataset. Zheng and Comaniciu [37] developed an automatic object identification and segmentation system based on marginal space learning. The locations can be consistently identified and used to guide the automatic model fitting process. During mesh modification, the authors enforced mesh point correspondence. As a result, creating a statistical form to guide the automated segmentation process is straightforward.

A large body of work has been completed using deep learning-based methods, which provide a different insight into US image segmentation. In [38], the authors suggested a strategy for segmenting the left ventricle (LV) in the A4C view that combines deep, fully convolutional networks with optical flow estimates. Subsequently, Jafari et al. [39] adopted a semi-supervised approach for LV segmentation. This approach is based on the development of an inverse mapping between segmentation masks and their corresponding echo frames using a generative model. This generator is subsequently leveraged to enhance a U-Net-generated LV segmentation mask. Work in Alsharqi et al. [40] employed a mask R-CNN model for recognizing the geometrical features of the LV. Fully convolutional network (FCN) and adversarial training were employed in Arafati et al. [41], while Painchaud et al. [42] introduced a restricted variational autoencoder for the purpose of learning a representation of viable cardiac structures. Leclerc et al. [43] proposed an encoder-decoder deep CNN

method to segment cardiac structures from 2D echocardiographic images. Yang et al. [44] examined the segmentation performance of DeeplabV3 and U-Net in the A4C view of fetal echocardiography. Dong et al. [45] presented the aggregated residual visual block network (ARVBNet) as a new real-time detection model for anatomical features in A4C planes.

A method of lesion detection (e.g., where the lesion is localized within the image) can be done by manual or automated feature engineering. Some prior works had proposed feature extraction techniques before applying machine learning algorithms. For instance, Zhou et al. [46] considered the LogitBoost algorithm to automatically classify cardiac view as a multiclass object detection. Similarly, Park et al. [47] presented an automatic system for cardiac view classification by characterizing local and global evidence, specific knowledge, and applying LogitBoost. A machine learning approach called shape regression machine was adopted for efficient segmentation of the LV endocardial [47]. Two feature extraction techniques, i.e., statistical features and histogram features, were considered in Balaji et al. [48] to automate cardiac view classification. Lili et al. [49] adopted a modified imbalanced Adaboost algorithm called ImAdaboost for four-chamber plane recognition in cardiac ultrasound images where SIFT features were used as features. Work in Khamis et al. [50] combined spatiotemporal feature selection and a dictionary learning-based classifier to enhance the classification accuracy of echocardiograms. Lastly, Yang et al., 2020 utilized RetinaNet, comprised of the residual network (ResNet), the feature pyramid network, and two fully convolutional networks for classifying A4C images [44].

3 Material and method

Detecting cardiac defects from the A4C view between 14 and 28 weeks' gestation is a very difficult task. The segmentation process views abnormal fetal heart anatomy using contour boundaries. Still, using only object segmentation cannot produce an interpretation of the defect's position and severity. This section outlines the specific material and methods used within this study. The whole process workflow, shown in Fig. 1, is comprised of five phases: data acquisition, data pre-processing, segmentation, defect detection, and model evaluation. All stages are described in the following section.

3.1 Data acquisition

The datasets were gathered in an Indonesian hospital. The raw videos are extracted from video echocardiography equipment and range in size from 1.02 to 331 MB in the digital imaging and communications in medicine (DICOM) format. All videos are converted to two-dimensional US images of the same quality and size (e.g., 400 × 300). The dataset employed in this study comprises 764 fetal heart images only in an A4C view with four conditions: ASD, VSD, AVSD, and normal. The distribution of the fetal US images for ASD, VSD, AVSD, and normal is outlined in Table 1.

3.2 Data pre-processing

To produce reliable results, segmentation and object recognition algorithms should first comprehend the entire image's global spatial features. However, to detect the position of the defects in the septum, two different labels are needed as the ground truth for segmentation (GT-S) and the ground truth for object detection (GT-O). In the segmentation process, each GT-S label is made based on the fetal heart contour (see Fig. 2b, f, j, m), while for the object detection process, each GT-O label is made based on the defect region of interest (RoI) (see Fig. 2d, h, l, o). All GTs were labeled according to clinical characteristics by qualified clinicians from the hospital's echocardiography department.

3.3 Proposed stacking U-Net and faster-RCNN architecture

This section discusses the overall technique and procedures used to develop a more accurate DL model for fetal heart segmentation and defect diagnosis using the A4C view. Fetal heart segmentation utilizes U-Net architecture [19], while defect detection uses Faster-RCNN architecture [51]. The U-Net model only has convolution and max-pooling layers without a fully connected layer [19]. Likewise, Faster-RCNN architecture consists of region proposal networks (RPNs) and the detector network [51]. The detector also includes a CNNs backbone, a region of interests (RoIs) pooling layer, fully convolutional networks (FCNs), two sibling branches for classification, and bounding box regression. The proposed DL model is depicted in Fig. 3, which can be described as follows:

- U-Net architecture is leveraged to generate the cardiac segmentation model. The architecture is made up of an encoder acting as a contraction path and a decoder acting as an expansion path, which are linked by a sequence of layered dense convolutional blocks. The structure is comprised of an encoder and a decoder subnetwork with the objective of bridging the semantic disparity between the encoder and decoder feature maps preceding to fusion.

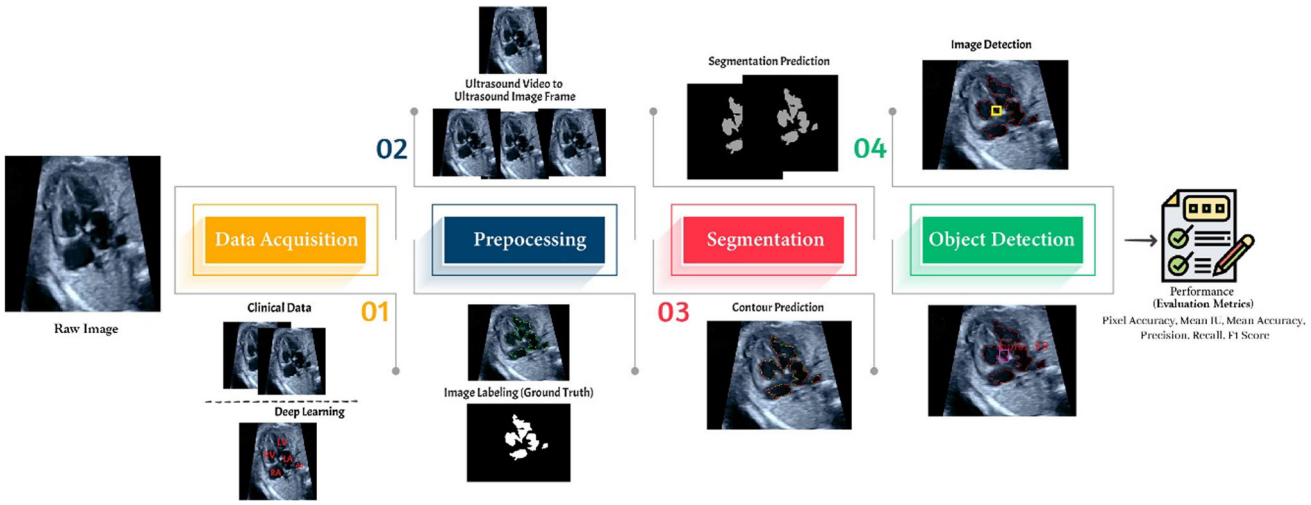

Fig. 1 Main automated fetal cardiac defect detection processes based on a deep learning technique

Table 1 The distribution of fetal heart images utilized throughout the training, validation, and testing stages

Condition	Training samples	Validation samples	Testing samples	Total
ASD	133	21	–	154
VSD	158	20	–	178
AVSD	159	25	–	184
Normal	155	22	71	248
Total	605	88	71	764

- A ReLU function activates each convolution process in U-Net, and the output layer uses the sigmoid function. The U-Net hyperparameters include 1000 epochs, a batch size of 64, a learning rate of 0.00001, a smoothing loss of 1–5, and a threshold of 0.5 with Adam optimizer. The U-Net model is designed based on two models of down filter for encoder and decoder structure. The filter structure is (16, 32, 64, 128, 256, 128, 64, 32, 16). Three types of loss function were used: binary cross-entropy (BCE), dice coefficient (DC), and Soft-Max with ranges between 0 and 1. It selected the best model that produces satisfactory performances.
- The output of the U-Net model is an image in which each pixel is assigned to either 0 or 255, indicating the foreground or background object. The semantic segmentation result of fetal heart contour from U-Net becomes an input to the Faster-RCNN model for detecting the defect position. Faster-RCNN is composed of two components. The first module is RPNs that propose RoIs, and the second module is the Fast R-CNN detector, which utilizes the RPNs model's proposed regions. The RPNs and detection network exchange full-image convolutional features, allowing practically cost-free region proposals, while the FCNs concurrently predict object boundaries and abjectness indices at each position.
- The Faster-RCNN model is developed using four different backbone architectures: VGG16, VGG19, Resnet50, and MobilenetV1. All architectures are compared to select the best detection, regression, and classification performance. Several hyperparameters in the whole structure are selected, such as a single batch size with a 0.001 learning rate and a 0.9 momentum. The pre-trained model uses two hundred feature learnings to achieve the best defect detection sensitivity. Moreover, the model was trained with 100 epochs equipped with the Stochastic gradient descent (SGD) and Adam as optimizer. The anchor box scales were fine-tuned using 128, 256, and 512 pixels, while the aspect ratio was 1:1, 1:2, and 2:1, respectively. Finally, all bounding boxes are predicted with a non-max suppression overlapping threshold of 0.7, and the overlapping classifier is 0.5.

The proposed model was developed on a computer equipped with an NVIDIA GeForce RTX 2080 GPU and a 3.60 GHz Intel(R) Core^{TM} version 9 CPU. Each image is processed in 9 s, while testing takes 147 ms.

3.4 Evaluation metrics

This study validates the proposed model using precision, sensitivity, pixel accuracy, mean intersection over union,

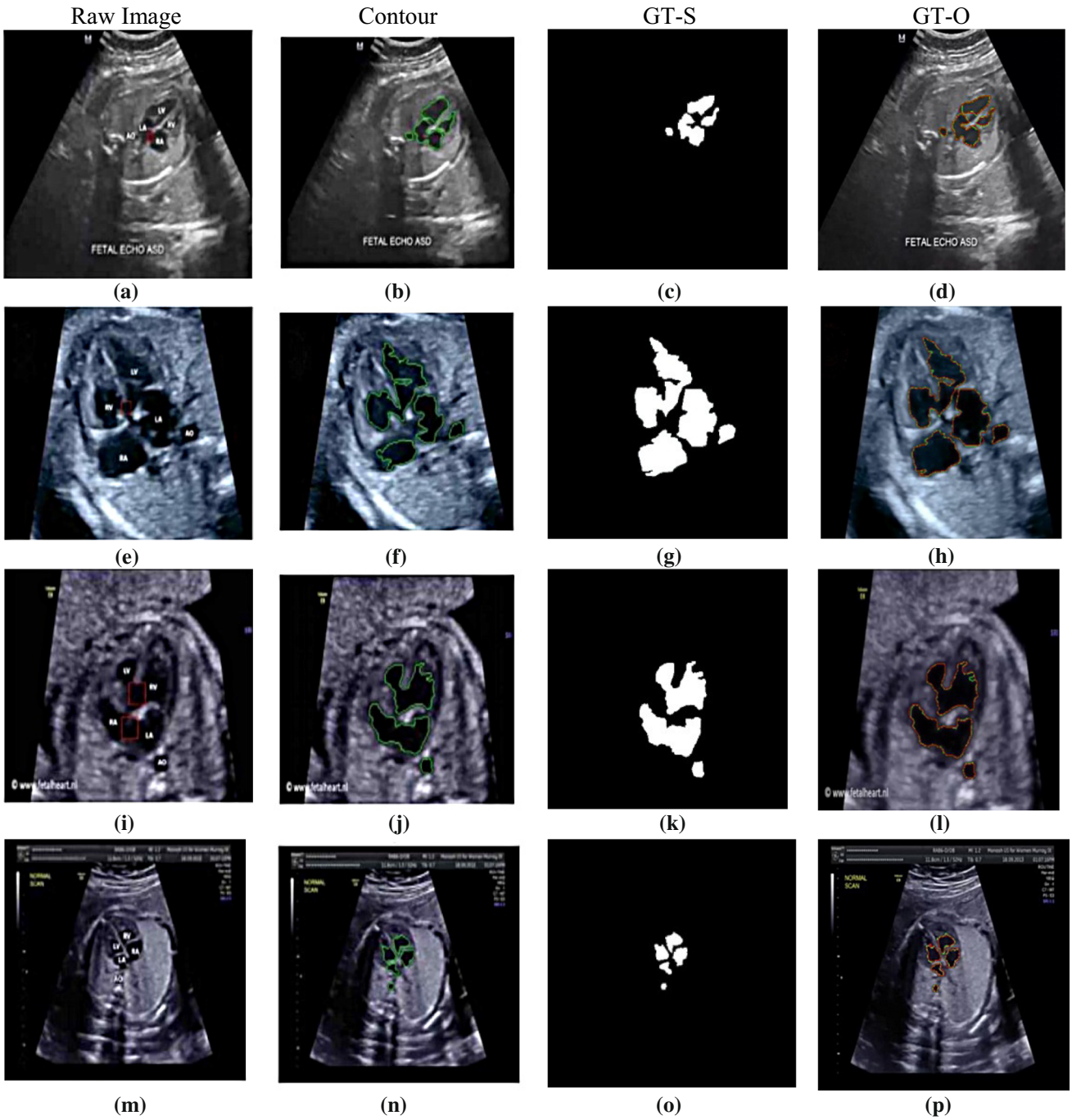

Fig. 2 Steps involved in data pre-processing (from left to right): raw image, fetal heart contour, ground truth for segmentation (GT-S), and ground truth for object detection (GT-O). Three septal defects, i.e., ASD (**a–d**), VSD (**e–h**), and AVSD (**i–l**), and normal conditions (**m–p**), are shown. Green lines are wall-chamber contour, red boxes are the hole, white color objects are the septum, and overlap between green and red lines are the segmented result by U-Net (Color figure online)

mean accuracy, and dice similarity coefficient [15]. The performance of a classifier is commonly represented in a confusion (e.g., contingency) matrix, which has four elements: true positive, false positive, true negative, and false negative. Let $C_{ii}, C_{ij}, C_{jj,}$ and C_{ji} be true positive, false positive, true negative, and false negative, respectively. Precision and sensitivity scores are obtained as follows.

$$\text{Precision} = \frac{\sum_{i=1}^{n} C_{ii}}{\sum_{i=1}^{n} C_{ii} + \sum_{y=1}^{n} C_{ij}} \quad (1)$$

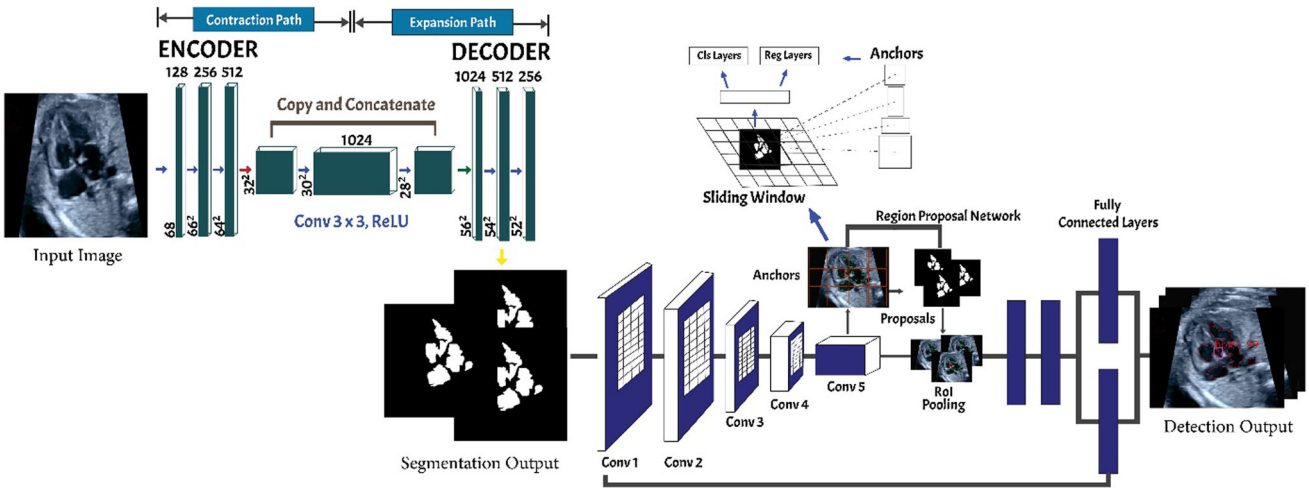

Fig. 3 Proposed architecture. Input images from US are segmented, while the output becomes an input of RPNs to detect the heart defect

$$\text{Sensitivity} = \frac{\sum_{i=1}^{n} C_{ii}}{\sum_{i=1}^{n} C_{ii} + \sum_{y=1}^{n} C_{ji}} \tag{2}$$

The pixel accuracy (PA) is specified as the ratio between correctly classified pixels and the total number of pixels. It is formally defined as:

$$\text{PA} = \frac{\sum_{i=1}^{n} C_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}} \tag{3}$$

where n is the number of classes and C_{ij} is the number of pixels in class i that are predicted as class j.

In addition, mean intersection over union (mIoU) depicts the intersection percentage between the labeled mask and the predicted output. More formally, it is acquired by taking the average of each class's IoU value over all classes as follows.

$$\text{mIoU} = \frac{1}{n} \sum_{x=1}^{n} \frac{C_{ii}}{\sum_{j=1}^{n} C_{ij} + \sum_{y=1}^{n} C_{ji} - C_{ii}} \tag{4}$$

Mean accuracy (MA) specifies the accuracy ratio for each class and the overall average (e.g., n) It can be expressed as:

$$\text{MA} = \frac{1}{n} = \sum_{x=1}^{n} \frac{C_{ii}}{\sum_{j=1}^{n} C_{ij}} \tag{5}$$

Finally, the dice similarity coefficient (DSC) metric that has been widely known in medical imaging applications is also employed. The DSC between a predicted image D and a ground truth image R, both of size k × l is specified as:

$$\text{DSC}(D, R) = 2 \frac{\sum_{a=0}^{k-1} \sum_{b=0}^{l-1} D_{ab} R_{ab}}{\sum_{a=0}^{k-1} \sum_{b=0}^{l-1} D_{ab} + \sum_{a=0}^{k-1} \sum_{b=0}^{l-1} R_{ab}} \tag{6}$$

where a and b denote the pixel, indicating the height k and width l, respectively. The range of the DSC is $[0, 1]$, and a greater DSC depicts to a better fit between D and R.

4 Result and discussion

The primary finding of this work is outlined and discussed in this section. The A4C segmentation is conducted by using the CNNs-based U-Net architecture. The model is validated by other two-loss functions, DC and BCE. Before the defect detection process is carried out, the wall chamber and aorta must be segmented. The segmentation approach is constructed using three different down-filtering (DF) schemes and an up-filtering (UF) scheme for the encoder and decoder structure, respectively. The process is performed to choose the optimal architecture of the encoder-decoder structure. Model 1 is developed with DF (8, 16, 32, 64, 128, 64, 32, 16, 8); model 2 is developed with DF (16, 32, 64, 128, 256, 128, 64, 32, 16); and model 3 is developed with DF (32, 64, 128, 256, 512, 256, 128, 64, 32); while model 4 is developed with UF (128, 256, 512, 1024, 2064, 1024, 512, 256, 128). All performance results in terms of precision, sensitivity, PA, mIoU, MA, and DSC are shown in Table 2. Model 3 with the DF scheme performs best compared to other models. Moreover, the loss function is tuned to select the best U-Net model so that a small number of false segmented pixels can be obtained during the testing process. Three different loss functions, i.e., dice coefficient, BCE, and SoftMax, are tested in the experiment. Figures 4 and 5 show the performance results of U-Net models with various loss functions in terms of precision, sensitivity, PA, mIoU, MA, and DSC. Based on the results, BCE outperforms dice and SoftMax's overall performance metrics.

Table 2 Segmentation results for four conditions using the best U-Net model

Condition	Performance (%)					
	Precision	Sensitivity	PA	mIoU	MA	DSC
ASD	95.73	96.14	99.50	95.82	97.92	95.89
VSD	94.58	97.74	99.51	96.00	98.68	96.08
AVSD	96.64	98.58	99.82	97.56	99.22	97.60
Normal	92.94	96.67	99.84	94.96	98.28	94.76

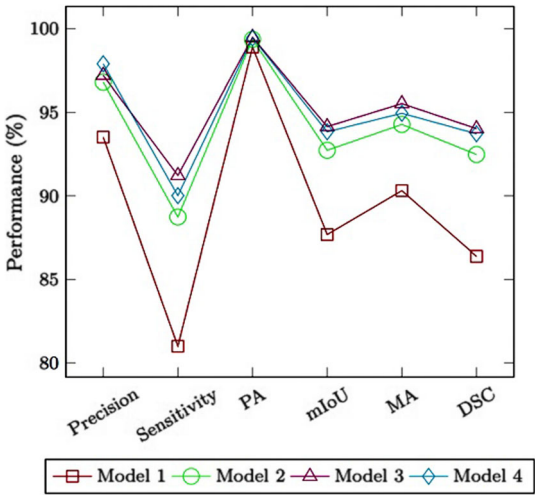

Fig. 4 Filter tuning to search the optimal U-Net model

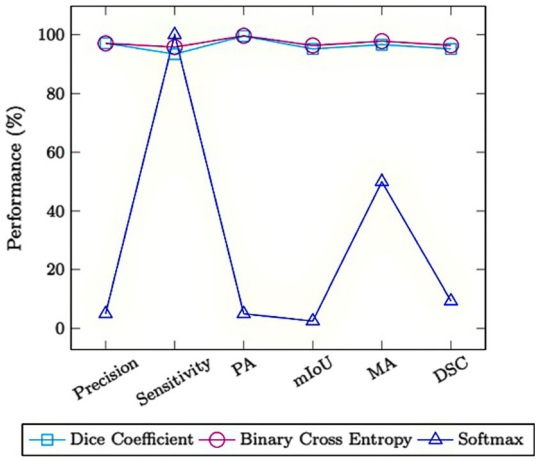

Fig. 5 Loss function tuning to search the optimal U-Net model

Figure 6 visualizes the contour segmentation results in four types of images: raw image, GT image with green lines in contour, the predicted image with red contour lines, and an overlap contour area between GT and the predicted image. The proposed segmentation model achieves above 90% in all performance metrics (see Table 2), implying that the proposed U-Net architecture demonstrates a promising result in automatic segmentation of the contour region. Furthermore, segmentation performance is good for each class, with false-positive rates of 0.32%, 0.10%, 0.08%, and 0.09% for ASD, VSD, AVSD, and normal condition, respectively. These rates indicate that the number of falsely segmented pixels is quite small since the proposed model can accurately predict all heart-chamber contour conditions. The precision and sensitivity metrics are significant in measuring the segmentation results because both are sensitive to over-segmentation and under-segmentation. If the prediction is over-segmented, the precision score will be low, while an under-segmented prediction will result in a low sensitivity value. Therefore, high precision and sensitivity values mean the boundaries in both segmentations are in the correct positions. Sensitivity and precision are evaluation metrics that measure how pixel boundaries match the predicted segmentation and ground truth.

The PA and mIoU are the most common metrics used for evaluating how well the segmentation model performs. The PA metric is used to assess the percentage of pixels in each correctly classified image. In this study, the proposed U-Net architecture predicts each pixel's image class and reports each class's prediction separately and globally across all categories. Figure 7 shows each class's pixel accuracy for ASD, VSD, AVSD, and normal conditions. While using the proposed segmentation model, each class's pixel accuracy tends to be near-perfect accurate (e.g., achieving 100% accuracy).

To validate the proposed model, the results are compared with other segmentation architectures such as U-Net++ [52] and V-Net [53]. These models are chosen since they have shown remarkable performance in the medical field. Figure 8 compares the prediction results of three architectures. The experimental results show that the U-Net architecture provides major efficiency gains over U-Net++ and V-Net, yielding average improvements concerning the precision, sensitivity, PA, mIoU, MA, and DSC metrics. It is worth mentioning that U-Net produces 97.06% precision and 95.76% sensitivity, which are better than those produced by V-Net. It exhibits an average improvement of 23% in sensitivity and precision metrics over V-Net. Moreover, the U-Net produces a mIoU 16% over that of V-Net. It is shown that U-Net is likely better adequate for semantic segmentation of 2D images than V-Net. The pixel accuracy of the U-Net architecture is higher (e.g., around 0.36 and 4.27%) than U-Net++ and V-Net, respectively (see Fig. 8).

This study represents fetal heart defect detection based on the Faster-RCNN approach to effectively screen for abnormalities in a fetal heart. Faster-RCNN is trained with four different backbones in the RPNs such as VGG16,

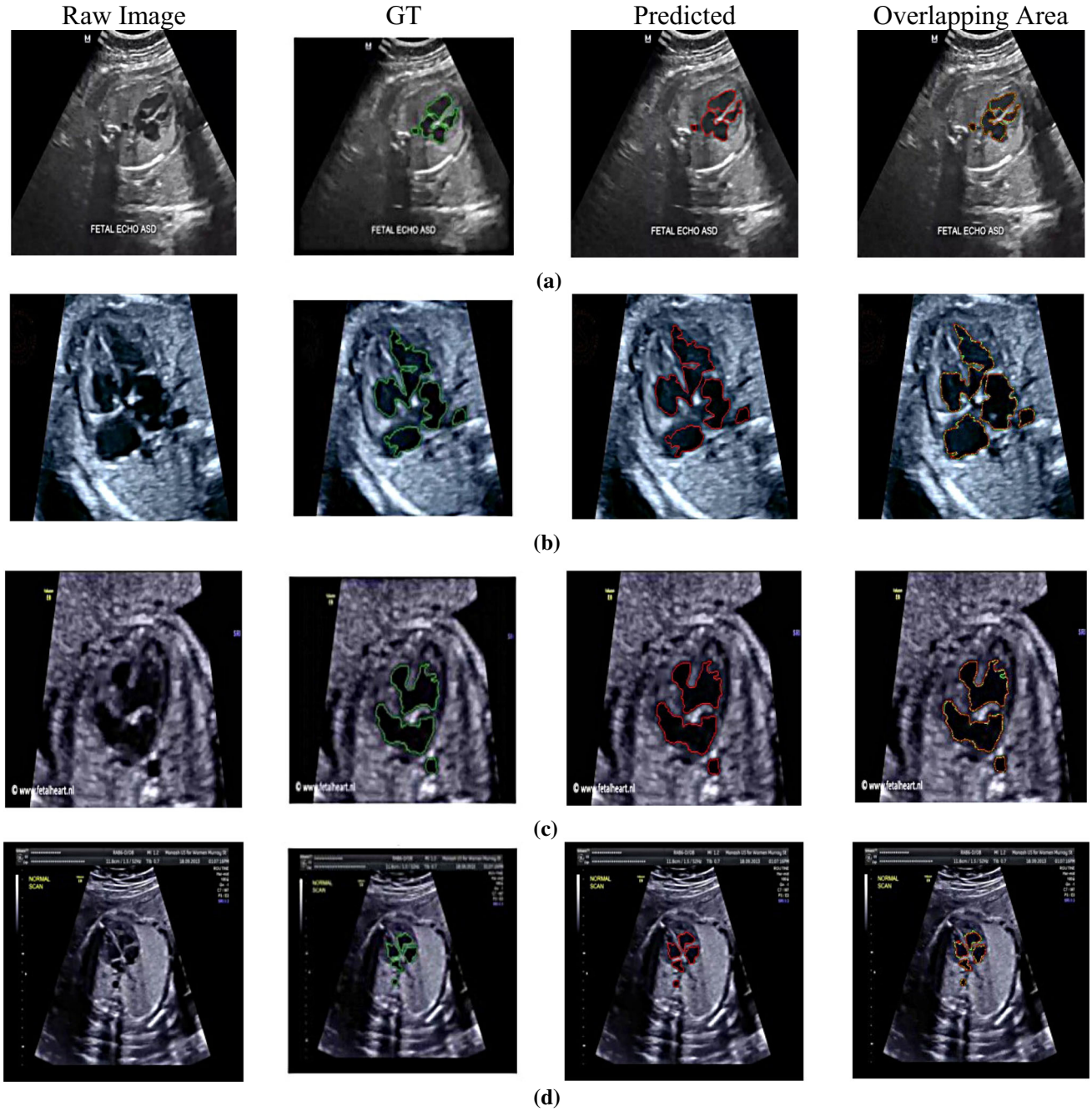

Fig. 6 Segmentation results of U-Net with respect to the following conditions: **a** ASD, **b** VSD, **c** AVSD, and **d** Normal. The green lines depict the expert's GT characterization, the red lines indicate the U-Net segmentation's wall-chamber contour image, and the green–red lines represent the overlap between the two (Color figure online)

VGG19, ResNet50, and MobilenetV1. RPNs are an important part of Faster-RCNN. Using a translational invariant, the model has a bounding box around the identified objects (e.g., defect). RPNs are trained until a small loss is achieved. The score is based on the projected box that overlaps the most with a ground truth box in terms of IoU. As seen in Table 3, Faster-RCNN with VGG16 architecture as the backbone can detect a heart defect in the septum with large mAP (87.80%). Therefore, VGG16 is the best backbone model in the proposed DL model.

In summary, Table 3 and Fig. 8 demonstrate that the proposed DL model (e.g., a combination of Faster-RCNN and U-Net) produces 96.33% mIoU, 96.37% DSC, and 87.80% mAP for three septal defect conditions, i.e., ASD, VSD, AVSD, and normal conditions, respectively. Figure 9 shows the results of the detection of a septal defect.

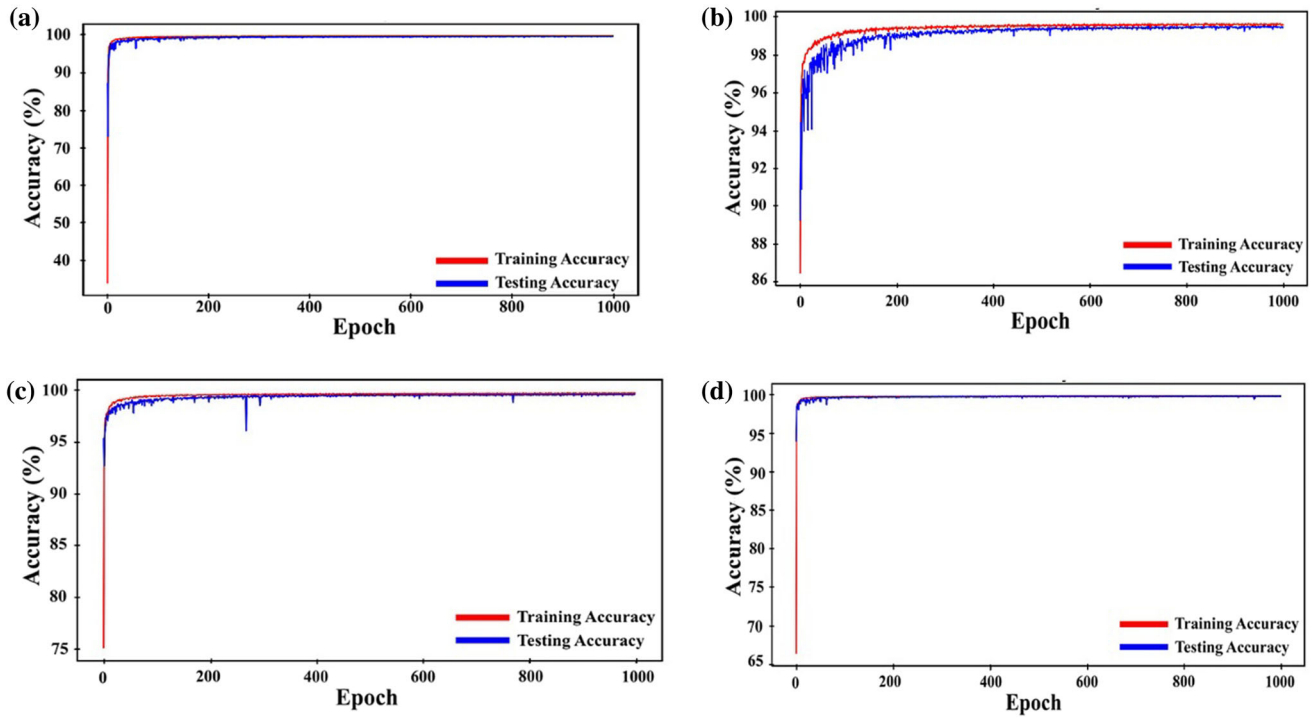

Fig. 7 Pixel accuracy performance of the proposed model for **a** ASD, **b** VSD, **c** AVSD, and **d** normal conditions

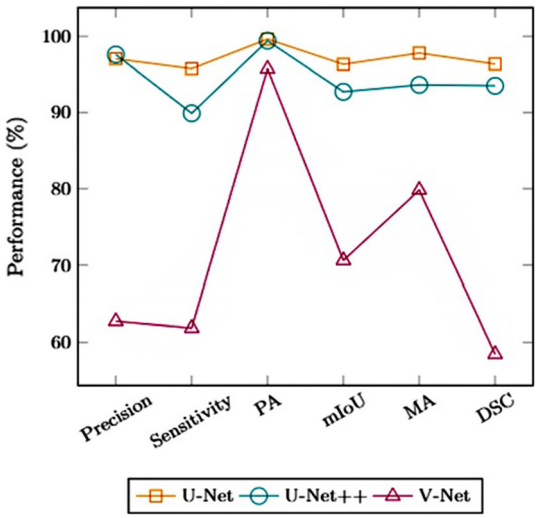

Fig. 8 Performance comparison of three different architectures for segmentation task

Table 3 Performance results of Faster-RCNN with different backbones

Backbone	mAP defect (%)
VGG16	87.80
VGG19	73.17
ResNet50	61.82
MobilenetV1	70.18

All examples are from the U-Net model test set. The proposed model can predict defective RoI with more than a 90% confidence value (see Fig. 9a–d). In addition, using 71 images of the invisible data, the proposed model successfully distinguishes a normal from an abnormal fetal heart. In Fig. 9, it is shown that the absence of a bounding box is detected in normal data. In contrast, the presence of a bounding box is detected based on the location of the defect where the abnormality occurs for abnormal data. Based on the performance obtained after analysis with this approach, the performance is not good enough for unseen data due to the multidimensional structure of the fetus.

As for benchmark, we compare the proposed model with three cases: (i) individual Faster-RCNN, (ii) other segmentation models, and (iii) other object detection model:

- Individual Faster-RCNN cannot detect the defect in the atrium with mAP 20.10% and both atrium and ventricle with mAP 30.45% (see Table 4). A septal defect is likely to be detected during the second and third trimester with the grayscale US. The defect size is not more than 3–4 mm, with many shadows [5]. The fetal heart images are natural with a wide range of semantic patterns, color, and intensities; they also contribute to detection difficulty [6, 10]. Without contour segmentation, the Faster-RCNN detected the small object like a defect. However, using an improved model provides more efficient performance results with the ability to detect all defect positions in the atrium, ventricle, and both.

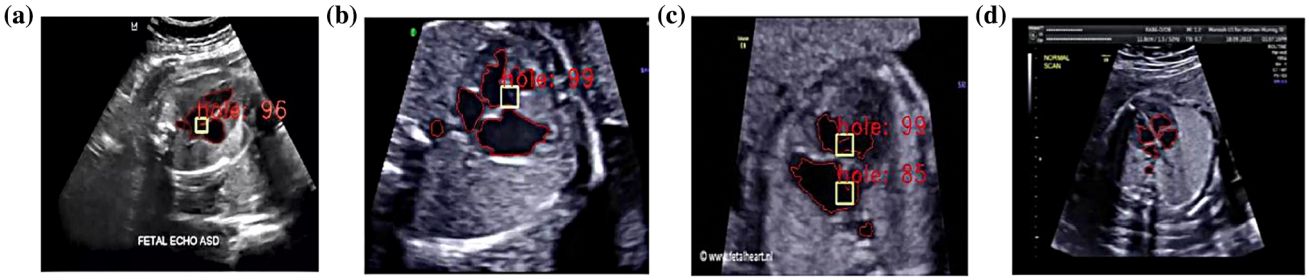

Fig. 9 Cardiac defect detection using the proposed model. **a** ASD, **b** VSD, **c** AVSD, and **d** normal. Red color denotes the segmentation result by U-Net, and yellow color refers to object detection by Faster-RCNN (Color figure online)

Table 4 Comparison results between the improved model and individual Faster-RCNN

Defect Position	mAP (%)	
	Proposed model	Faster-RCNN
Atrium	91.07	20.10
Ventricle	82.14	67.66
Atrium and ventricle	65.76	30.45

- The proposed approach is validated against existing object detectors in clinical applications. As shown in Table 5, three previous works (e.g., [10, 28, 54, 55] segmenting the left ventricle, head, thorax, and abdomen of the fetus produce 94.5%, 96.84%, 93.30%, and 90% w.r.t DSC metric, respectively, while one work (e.g., [54]) for fetal cardiac views reaches 95%. Our proposed model for fetal heart segmentation achieves a good performance of 96.37% w.r.t DSC metric. Further research is needed to measure US image quality to better understand data with varying image quality.
- The contour segmentation process is combined with Faster-RCNN to detect the defect in the septum. This process provides more effective performance results, and the detection model is relatively faster than other object detectors (see Table 6). Compared with some previous works in [56–59], their performance results are lower than 80%, while our proposed model produces 87.80% and 96.37% in terms of mAP and DSC, respectively. This means that the predicted cardiac defect in the septum has a large overlap with the ground truth. However, it is necessary to conduct an extensive experiment using unseen image samples regarding septal defect conditions.

Table 5 Performance comparison with other segmentation techniques

Method	Fetal object segmentation	DSC (%)
Dynamic CNNs [10]	Fetal left ventricle	94.50
CNNs-based Link-Net [28]	Fetal head	96.84
DW-Net [54]	Fetal thorax	93.30
CNNs + HOG [55]	Fetal abdomen	90
Proposed model	Fetal Heart	96.37

Table 6 Performance comparison with other detection tasks

Object detection	Performance (%)		
	mIoU	mAP	DSC
Breast lesions [56]	–	–	91.42
Lung nodules [57]	–	–	70.00
Oral diseases [58]	–	–	74.00
Nucleus [59]	70.54	59.40	–
Septal defect with Faster-RCNN	–	37.57	–
Proposed model	75	87.80	96.37

The CNNs approach with stacking U-Net and Faster-RCNN architecture is applied layer by layer in this research, with no features provided beforehand. Consequently, the proposed approach has the benefit of providing a complete picture of the fetal heart's anatomical structure, indicating that the developed model accurately segments the fetal heart in A4CH while also identifying a septal heart septum defect. As far as we are concerned, neither study on segmenting the fetal heart image nor identifying cardiac defects utilizing this stacking architecture has been attempted. Although the findings show promise in terms of segmentation and diagnosis of fetal cardiac abnormalities, there are major drawbacks to this study: (i) to identify CHDs, solely the fetal heart on a 4CH view is employed, (ii) the extent of the defect is not properly considered while assessing the degree of CHD, and (iii) to improve performance, the number of fetal heart imaging data collections with normal and defective fetal heart structures should be increased.

5 Conclusions

This paper proposes an improved Faster-RCNN model to automate heart-chamber segmentation and automatically detect the septum defect using 764 US images. The proposed deep learning was constructed via the combination of architecture U-Net and Faster-RCNN. Based on the experiment, the proposed model achieved good performance by 75.00%, 96.37%, and 87.80%, with respect to mIoU, DSC, and mAP metrics, respectively, surpassing other similar segmentation techniques, e.g., CNNs. The proposed model was also successful in a testing scenario using 71 unseen normal class images, accomplishing 100% sensitivity, showing that the proposed model could recognize the normal class effectively without any misclassification error. However, there are several limitations to this work. According to the defect detection result, performance accuracy must be improved. The lower accuracy happens mainly due to the data set's limitation, inconsistent image quality, variances in specific anatomic structures, and variability of the image size. Moreover, the proposed model was trained on images from US equipment, e.g., GE Volluson P8, without considering the echocardiogram variability. Further experiments with larger datasets from different US equipment are necessary to improve the robustness of the model. The proposed deep learning model's generalizability and adaptability should be considered for future work. All experiments were carried out on the same dataset, where the training set was randomly split and would have an identical distribution with the testing set. Although the results look promising, the view of the fetal heart discussed in this work was limited to the A4C view. Therefore, future research should incorporate other views, such as right RVOT, LVOT, and 3VT.

Acknowledgements We thank the Intelligent System Research Group (ISysRG), Faculty of Computer Science, Universitas Sriwijaya, Indonesia.

Author contribution Conceptualization was contributed by Siti Nurmaini, Ade Iriani Sapitri, and Bayu Adhi Tama; Methodology was contributed by Ade Iriani Sapitri and Muhammad Naufal Rachmatullah; Formal analysis and investigation were contributed by Siti Nurmaini; Writing—original draft preparation was contributed by Siti Nurmaini; Writing—review and editing was contributed by Bayu Adhi Tama and Annisa Darmawahyuni; Funding acquisition was contributed by Siti Nurmaini; Resources were contributed by Firdaus Firdaus and Bambang Tutuko.

Funding This work was supported by the Ministry of Research and Technology, Indonesia, through Applied Research, under Grant 096/SP2H/LT/DRPM/2021 and Professional Grant 2022 from Universitas Sriwijaya, Indonesia. This work is also funded by Institute for Basic Science (IBS) under grant No. IBS-R029-C2-001.

Availability of data and material All data considered for this study are available at https://github.com/ISySRGg/U-FRcnns/Image Data.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Consent to participate Not applicable.

Ethical approval Not applicable.

References

1. Garcia-Canadilla P, Sanchez-Martinez S, Crispi F, Bijnens B (2020) Machine learning in fetal cardiology: what to expect. Fetal Diagn Ther 47(5):363–372. https://doi.org/10.1159/000505021
2. Lopez KN, Morris SA, Sexson Tejtel SK, Espaillat A, Salemi JL (2020) US mortality attributable to congenital heart disease across the lifespan from 1999 through 2017 exposes persistent racial/ethnic disparities. Circulation 142(12):1132–1147
3. Pace ND, Oster ME, Forestieri NE, Enright D, Knight J, Meyer RE (2018) Sociodemographic factors and survival of infants with congenital heart defects. Pediatrics 142(3)
4. Han B, Tang Y, Qu X, Deng C, Wang X, Li J (2021) Comparison of the 1-year survival rate in infants with congenital heart disease diagnosed by prenatal and postnatal ultrasound: a retrospective study. Medicine 100(4)
5. Espinoza J (2019) Fetal MRI and prenatal diagnosis of congenital heart defects. Lancet (London, England) 393(10181):1574–1576. https://doi.org/10.1016/S0140-6736(18)32853-8
6. Kaluva KC, Shanthi C, Thittai AK, Krishnamurthi G (2018) CardioNet: identification of fetal cardiac standard planes from 2D Ultrasound data
7. Rawat V, Jain A, Shrimali V (2018) Automated techniques for the interpretation of fetal abnormalities: a review. Appl Bionics Biomech 2018
8. Sundaresan V, Bridge CP, Ioannou C, Noble JA (2017) Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), pp 671–674. https://doi.org/10.1109/ISBI.2017.7950609
9. Harangi B (2018) Skin lesion classification with ensembles of deep convolutional neural networks. J Biomed Inform 86:25–32. https://doi.org/10.1016/j.jbi.2018.08.006
10. Yu L, Guo Y, Wang Y, Yu J, Chen P (2016) Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. IEEE Trans Biomed Eng 64(8):1886–1895. https://doi.org/10.1109/TBME.2016.2628401
11. Ha VK, Ren J-C, Xu X-Y, Zhao S, Xie G, Masero V, Hussain A (2019) Deep learning based single image super-resolution: a survey. Int J Autom Comput 16(4):413–426. https://doi.org/10.1007/s11633-019-1183-x
12. Patel N, Narasimhan E, Kennedy A (2017) Fetal cardiac US: techniques and normal anatomy correlated with adult CT and MR imaging. Radiographics 37(4):1290–1303
13. Xia Z, Wang X, Wang C, Zhang C (2018) Subpixel-based accurate and fast dynamic tumor image recognition. J Med Imaging Health Inform 8(5):925–931. https://doi.org/10.1166/jmihi.2018.2390
14. Xia Z, Wang X, Zhou W, Li R, Wang C, Zhang C (2019) Color medical image lossless watermarking using chaotic system and

accurate quaternion polar harmonic transforms. Signal Process 157:108–118. https://doi.org/10.1016/j.sigpro.2018.11.011
15. Nurmaini S, Rachmatullah MN, Sapitri AI, Darmawahyuni A, Tutuko B, Firdaus F, Partan RU, Bernolian N (2021) Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection. Sensors 21(23):8007
16. Yang R, Yu Y (2021) Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. Front Oncol 11:573
17. Rachmatullah MN, Nurmaini S, Sapitri AI, Darmawahyuni A, Tutuko B, Firdaus F (2021) Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view. Bull Electr Eng Inform 10(4):1987–1996
18. Ma P, Li Q, Li J (2022) Application of artificial intelligence in cardiovascular imaging. J Healthc Eng
19. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 234–241.https://doi.org/10.1007/978-3-319-24574-4_28
20. Hunter LE (2018) Screening views of the fetal heart. In: Fetal cardiology. Springer, Berlin, pp 9–20. https://doi.org/10.1007/978-3-319-77461-9_2
21. Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, Foi A, Maggioni M, Pepe A, Tohka J, Stebbing RV, McManigle JE, Ciurte A, Bresson X, Cuadra MB, Sun C, Ponomarev GV, Gelfand MS, Kazanov MD, Noble JA (2014) Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge. IEEE Trans Med Imaging 33(4):797–813. https://doi.org/10.1109/TMI.2013.2276943
22. Gao Y, Noble JA (2017) Detection and characterization of the fetal heartbeat in free-hand ultrasound sweeps with weakly-supervised two-streams convolutional networks. In: International conference on medical image computing and computer-assisted intervention, pp 305–313. https://doi.org/10.1007/978-3-319-66185-8_35
23. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312. https://doi.org/10.1109/TMI.2016.2535302
24. Xie HN, Wang N, He M, Zhang LH, Cai HM, Xian JB, Lin MF, Zheng J, Yang YZ (2020) Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol 56(4):579–587. https://doi.org/10.1002/uog.21967
25. Komatsu M, Sakai A, Komatsu R, Matsuoka R, Yasutomi S, Shozu K, Dozen A, Machino H, Hidaka H, Arakaki T et al (2021) Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Appl Sci 11(1):371. https://doi.org/10.3390/app11010371
26. Al-Bander B, Alzahrani T, Alzahrani S, Williams BM, Zheng Y (2020) Improving fetal head contour detection by object localisation with deep learning. In: Communications in computer and information science, 1065 CCIS, pp 142–150. https://doi.org/10.1007/978-3-030-39343-4_12
27. Liu K, Ye Z, Guo H, Cao D, Chen L, Wang FY (2021) FISS GAN: A generative adversarial network for foggy image semantic segmentation. IEEE/CAA J Automatica Sinica 8(8):1428–1439
28. Sobhaninia Z, Rafiei S, Emami A, Karimi N, Najarian K, Samavi S, Soroushmehr SMR (2019) Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: 2019 41st Annual international conference of the ieee engineering in medicine and biology society (EMBC), pp 6545–6548. arXiv:1909.00273
29. Senouf O, Vedula S, Zurakhov G, Bronstein A, Zibulevsky M, Michailovich O, Adam D, Blondheim D (2018) High frame-rate cardiac ultrasound imaging with deep learning. In: International conference on medical image computing and computer-assisted intervention, pp 126–134. https://doi.org/10.1007/978-3-030-00928-1_15
30. Wang C, Pedrycz W, Li Z, Zhou M (2020) Residual-driven fuzzy C-means clustering for image segmentation. IEEE/CAA J Automatica Sinica 8(4):876–889. https://doi.org/10.1109/JAS.2020.1003420
31. Wang C, Pedrycz W, Yang J, Zhou M, Li Z (2020) Wavelet frame-based fuzzy C-means clustering for segmenting images on graphs. IEEE Trans Cybern 50(9):3938–3949. https://doi.org/10.1109/TCYB.2019.2921779
32. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D (2008) Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 27(11):1668–1681. https://doi.org/10.1109/TMI.2008.2004421
33. Nakphu N, Dewi DEO, Rizqie MQ, Supriyanto E, Kho DCC, Kadiman S, Rittipravat P et al (2014) Apical four-chamber echocardiography segmentation using Marker-controlled Watershed segmentation. In: 2014 IEEE Conference on biomedical engineering and sciences (IECBES), pp 644–647. https://doi.org/10.1109/IECBES.2014.7047583
34. Cao Y, McNeillie P, Syeda-Mahmood T (2014) Segmentation of anatomical structures in four-chamber view echocardiogram images. In: 2014 22nd International conference on pattern recognition, pp 568–573. https://doi.org/10.1109/ICPR.2014.108.
35. Syeda-Mahmood T, Wang Q, McNeillie P, Beymer D, Compas C (2014) Discriminating normal and abnormal left ventricular shapes in four-chamber view 2D echocardiography. In: 2014 IEEE 11th International symposium on biomedical imaging (ISBI), pp 401–404. https://doi.org/10.1109/ISBI.2014.6867893
36. Kang HC, Kim B, Lee J, Shin J, Shin Y-G (2015) Accurate four-chamber segmentation using gradient-assisted localized active contour model. J Med Imaging Health Inform 5(1):126–137
37. Zheng Y, Comaniciu D (2014) Marginal space learning for medical image analysis: efficient detection and segmentation of anatomical structures. In: Marginal space learning for medical image analysis: efficient detection and segmentation of anatomical structures, vol 9781493906, Issue Mdl. https://doi.org/10.1007/978-1-4939-0600-0
38. Jafari MH, Girgis H, Liao Z, Behnami D, Abdi A, Vaseli H et al (2018) A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, pp. 29–37. https://doi.org/10.1007/978-3-030-00889-5
39. Jafari MH, Girgis H, Abdi AH, Liao Z, Pesteie M, Rohling R, Gin K, Tsang T, Abolmaesumi P (2019) Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as Prior University of British Columbia, Vancouver, Canada. Vancouver General Hospital, Vancouver, Canada. In: 2019 IEEE 16th International symposium on biomedical imaging (ISBI 2019), Isbi, pp 649–652. https://doi.org/10.1109/ISBI.2019.8759292
40. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P (2018) Artificial intelligence and echocardiography. Echo Res Pract 5(4):R115–R125. https://doi.org/10.1530/ERP-18-0056
41. Arafati A, Morisawa D, Avendi MR, Amini MR, Assadi RA, Jafarkhani H, Kheradvar A (2020) Generalizable fully automated

multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. J R Soc Interface 17(169):20200267. https://doi.org/10.1098/rsif.2020.0267
42. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin P-M (2020) Cardiac segmentation with strong anatomical guarantees. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2020.3003240
43. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg E, Jodoin PM, Grenier T, Lartizien C, Dhooge J, Lovstakken L, Bernard O (2019) Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 38(9):2198–2210. https://doi.org/10.1109/TMI.2019.2900516
44. Yang M, Xiao X, Liu Z, Sun L, Guo W, Cui L, Sun D, Zhang P, Yang G (2020) Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification. Sci Program
45. Dong J, Liu S, Wang T (2019) 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. Springer, Berlin, pp 130–137. https://doi.org/10.1007/978-3-030-33327-0_16
46. Zhou SK, Park JH, Georgescu B, Comaniciu D, Simopoulos C, Otsuki J (2006) Image-based multiclass boosting and echocardiographic view classification. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), vol 2, pp 1559–1565. https://doi.org/10.1109/CVPR.2006.146
47. Park JH, Zhou SK, Simopoulos C, Otsuki J, Comaniciu D (2007) Automatic cardiac view classification of echocardiogram. In: 2007 IEEE 11th international conference on computer vision, pp 1–8. https://doi.org/10.1109/ICCV.2007.4408867
48. Balaji GN, Subashini TS, Chidambaram N (2015) Automatic classification of cardiac views in echocardiogram using histogram and statistical features. Procedia Comput Sci 46:1569–1576. https://doi.org/10.1016/j.procs.2015.02.084
49. Lili W, Zhongliang F, Pan T (2016) Four-chamber plane detection in cardiac ultrasound images based on improved imbalanced AdaBoost algorithm. In: 2016 IEEE International conference on cloud computing and big data analysis (ICCCBDA), pp 299–303. https://doi.org/10.1109/ICCCBDA.2016.7529574
50. Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D (2017) Automatic apical view classification of echocardiograms using a discriminative learning dictionary. Med Image Anal 36:15–21. https://doi.org/10.1016/j.media.2016.10.007
51. Lin Z, Le MH, Ni D, Chen S, Li S, Wang T, Lei B (2018) Quality assessment of fetal head ultrasound images based on faster R-CNN. In: Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation. Springer, Berlin, pp 38–46. https://doi.org/10.1007/978-3-030-01045-4_5
52. Zhou Z, Siddiquee M, Tajbakhsh N, Liang J (2018) 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, S..., vol 11045, pp 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
53. Milletari F, Navab N, Ahmadi SA (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings—2016 4th international conference on 3D vision, 3DV 2016, pp 565–571. https://doi.org/10.1109/3DV.2016.79
54. Xu L, Liu M, Shen Z, Wang H, Liu X, Wang X, Wang S, Li T, Yu S, Hou M, Guo J, Zhang J, He Y (2020) DW-Net: a cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. Comput Med Imaging Graph 80:101690. https://doi.org/10.1016/j.compmedimag.2019.101690
55. Ravishankar H, Prabhu SM, Vaidya V, Singhal N (2016) Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI), pp 779–782. https://doi.org/10.1109/ISBI.2016.7493382
56. Yap MH, Goyal M, Osman FM, Martí R, Denton E, Juette A, Zwiggelaar R (2019) Breast ultrasound lesions recognition: end-to-end deep learning approaches. J Med Imaging (Bellingham, Wash.), 6(1):011007. https://doi.org/10.1117/1.JMI.6.1.011007
57. Kopelowitz E, Engelhard G (2019) Lung nodules detection and segmentation using 3D mask-RCNN. ArXiv Preprint arXiv:1907.07676
58. Anantharaman R, Velazquez M, Lee Y (2018) Utilizing mask R-CNN for detection and segmentation of oral diseases. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM), pp 2197–2204. https://doi.org/10.1109/BIBM.2018.8621112
59. Kowal M, Żejmo M, Skobel M, Korbicz J, Monczak R (2020) Cell Nuclei Segmentation in cytological images using convolutional neural network and seeded watershed algorithm. J Digit Imaging 33(1):231–242. https://doi.org/10.1007/s10278-019-00200-8