IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 7, 2020, accepted October 23, 2020, date of publication October 28, 2020, date of current version November 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034367

Accurate Detection of Septal Defects With Fetal Ultrasonography Images Using Deep Learning-Based Multiclass Instance Segmentation

SITI NURMAINI, (Member, IEEE), MUHAMMAD NAUFAL RACHMATULLAH, ADE IRIANI SAPITRI, ANNISA DARMAWAHYUNI, ADITHIA JOVANDY, FIRDAUS FIRDAUS, BAMBANG TUTUKO, AND ROSSI PASSARELLA

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang 30139, Indonesia

Corresponding author: Siti Nurmaini (siti_nurmaini@unsri.ac.id)

This work was supported by the Ministry of Research, and Technology (RISTEKBRIN), Indonesia, through the Applied Research, under Grant 096/SP2H/LT/DRPM/2020.

ABSTRACT Accurate screening for septal defects is important for supporting radiologists' interpretative work. Some previous studies have proposed semantic segmentation and object detection approaches to carry out fetal heart detection; unfortunately, the models could not segment different objects of the same class. The semantic segmentation method segregates regions that only contain objects from the same class. In contrast, the fetal heart may contain multiple objects, such as the atria, ventricles, valves, and aorta. Besides, blurry boundaries (shadows) or a lack of consistency in the acquisition ultrasonography can cause wide variations. This study utilizes Mask-RCNN (MRCNN) to handle fetal ultrasonography images and employ it to detect and segment defects in heart walls with multiple objects. To our knowledge, this is the first study involving a medical application for septal defect detection using instance segmentation. The use of MRCNN architecture with ResNet50 as a backbone and a 0.0001 learning rate allows for two times faster training of the model on fetal heart images compared to other object detection methods, such as Faster-RCNN (FRCNN). We demonstrate a strong correlation between the predicted septal defects and ground truth as a mean average precision (mAP). As shown in the results, the proposed MRCNN model achieves good performance in multiclass detection of the heart chamber, with 97.59% for the right atrium, 99.67% for the left atrium, 86.17% for the left ventricle, 98.83% for the right ventricle, and 99.97% for the aorta. We also report competitive results for the defect detection of holes in the atria and ventricles via semantic and instance segmentation. The results show that the mAP for MRCNN is about 99.48% and 82% for FRCNN. We suggest that evaluation and prediction with our proposed model provide reliable detection of septal defects, including defects in the atria, ventricles, or both. These results suggest that the model used has a high potential to help cardiologists complete the initial screening for fetal congenital heart disease.

INDEX TERMS Congenital heart disease, fetal echocardiography, mask-RCNN, septal defects, multiclass instance segmentation.

I. INTRODUCTION

Congenital heart diseases (CHDs) are the most common malformations and occur in 0.8% of the general population [1]. The most common type of CHDs is a septal defect [2]. Defects in septation between the cardiac chambers constitute the largest single group of congenital cardiac malformations [2]. These developmental anomalies may involve the atrial septum, the ventricular septum, or the conotruncus [2]. Septal defects leave holes in the septum. Such a condition can cause the blood to go in the wrong direction or to the wrong place, or it can cause blood to be pumped to the

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

lungs [2]–[4]. Hole positions can occur in three places: in the wall that separates the right and left atria, known as atrial septal defects (ASDs); in the wall that separates the right and left ventricles, known as ventricular septal defects (VSDs); and between the chambers of the right and left sides of heart, known as atrioventricular septal defects (AVSDs) [3]–[5]. To significantly improve the baby's prognosis, the detection of CHDs at early stages is essential since it can involve abnormal fetal heart structures [6]; thus, enabling medical treatment as soon as possible (usually within a week after birth) is necessary.

It is possible to screen CHDs during fetal life with a routine prenatal ultrasound examination. Such a process can identify only 60% of CHDs [1]. An ultrasound device is a non-invasive, low cost, a radiation-free imaging modality that is an indispensable part of modern cardiology techniques for diagnosing CHDs [6], [7]. To produce a complete diagnosis and medical counseling, specialist fetal echocardiography is needed. Fetal ultrasonography (US) diagnosis of a CHD is challenging to accomplish because US imaging is susceptible to blurry boundaries (shadows), which corrupt the image and reduce its quality [6]–[9]. Echocardiography depends not only on the medical professional's skill in image acquisition but also on a highly evolved method of human pattern recognition for image analysis. The human interpretation system's potential limitations for routine US examinations include fatigue or distraction, inter- and intraobserver variability, and the tedious, time-consuming interpretation of large datasets [10]. Furthermore, the heart septum defects are relatively small, with anatomical structures with unclear appearances that are not apparent to the naked eye [11]. Hence, automated detection with a low-quality image of a septal defect requires thorough investigation.

Improving the screening examination process with advanced technology to achieve accurate automatic abnormality detection in fetal hearts using US has become a significant issue [12], [13]. By standardizing the maternal-placental-fetal unit's clinical evaluation using a diagnostic protocol shared by obstetricians, genetics, and fetal echocardiographers, the methods used to screen high-risk groups of pregnancies can be improved. The data collected can then be used as a basis for defining the specific parameters for formulating a clinical, sonographic score and a flowchart, which is performed to decide such a condition, as guides for the diagnosis and therapeutic management of CHDs in pregnancies. CHDs with septal defect conditions particularly require two-dimensional (2D) US imaging. Interpreting images of fetal heart defects is too complicated [8], [14]. Object recognition improves with practice, but echocardiogram perception remains highly subjective today. Artificial intelligence (AI) with a computer-aided diagnosis holds promise for echocardiographic analysis as it can retrieve information that is not readily apparent to the observer [15].

Machine learning (ML) is one approach for computer-aided diagnosis that has been utilized to overcome some problems in medical imaging and has provided excellent results [16], [17]. ML allows diagnostic systems to be faster and more accurate than humans. However, this requires thorough training and practice, as well as a process that is time-consuming and complex [6], [18], [19]. To produce a high prediction accuracy, the ML process involves collecting a sufficient amount of data with over 100,000 observations of both normal and abnormal data for feature learning [19]. Furthermore, CHDs occur very infrequently, making it challenging to collect the necessary amount of abnormal data [20]. Moreover, US imaging is susceptible to blurriness, which causes even abnormal data sometimes to be interpreted as normal. To detect abnormalities with high precision using ML technology, a massive amount of normal data with a variety of blurry patterns is needed. Hence, a technology that can accurately predict CHDs using relatively small and incomplete datasets with low-quality images is desirable.

Deep learning (DL) has already been used for limited echocardiographic observations to diagnose structural heart disease [21]. This has also allowed for the use of cardiac landmarks to evaluate the left endocardial ventricular boundary segments for wall movements, volume assessments [22], [23], chamber size assessments, valve mobility statuses, the presence of pericardial effusion, and several more areas of automated interpretation [24]. All the results achieved satisfactory performance. Unfortunately, the research was not conducted on fetal hearts and only performed binary segmentation. Segmentation and defect detection in the fetal heart septum is hard to accomplish due to the heterogeneity of specific lesion images, the diversity of heart anatomies from one individual to another, and the small objective of detecting a defect of less than 2 mm in the heart wall with low-quality images. These can result in low performance and significant error segmentation and detection [25]. Furthermore, the previous study of fetal object detection is still a limited case [26]. Hence, a deep investigation of septal defect detection as a fetal heart abnormality is desirable.

To our knowledge, we are the first to conduct this comprehensive investigation of the fetal heart with multiclass segmentation and detection of the ''hole-in-heart'' septal defect by using MRCNN. In conclusion, our novel contributions can be summarized as follows:

- We propose a combination of multiclass segmentation of and object detection in the fetal heart for CHDs decisions;
- We use six objects in the fetal heart, including the left atrium (LA), right atrium (RA), left ventricle (LV), right ventricle (RV), aorta, and hole, for multiclass segmentation;
- To evaluate the proposed model, an experiment is conducted with three septal defects (abnormal) conditions, ASDs, VSDs, and AVSDs, and a normal condition;
- To ensure the performance of the proposed model, the architecture of MRCNN is compared to the FRCNN architecture;

- The robustness of the proposed model is evaluated by using the mean intersection over union (mIoU) and Dice score similarity (DSC) for segmentation and mean average precision (mAP) for object detection; and
- We use unseen data to improve the generalization of the proposed model with a normal image.

Apart from novel contributions, the remainder of this article is organized as follows. In Section II, we explain the related works. We then propose instance segmentation for a septal defect in Section III and evaluate our method's performance in Section IV. Finally, we conclude this article in Section V.

II. RELATED WORKS

Computer-based fetal echocardiography diagnostic systems have been developed. In other words, fetal echocardiography interpretations are made with the digital aided of a computer device using AI [22]. With AI-based technology development, an echocardiogram examination for the segmentation and detection of CHDs, especially fetal heart septal defects, previously performed manually by cardiologists, can be performed automatically. Segmentation of the fetal heart, detection of septal defects, and accurate evaluation of their defects' sizes are crucial for tracking CHDs diagnoses. An automatic fetal echocardiogram examination can assist physicians in early detection before referral to a cardiologist for further management.

The segmentation process is the key to exploring fetal heart abnormalities, especially defect conditions [27]. It can aid doctors in making more accurate treatment plans [27]. Nonetheless, manual segmentation can be a very time-consuming process because a radiologist needs to mark target regions in hundreds of frames or images for one patient. Hence, the need for more accurate automated segmentation tools is apparent [10], [12], [13]. Considerable work has been performed in recent years towards the automatic segmentation of the fetal heart to diagnose defect conditions [7], [14], [16], [18]. Previous works have mostly focused on the utilization of conventional learning with supervised and unsupervised training methods, including threshold-based methods, region-based methods, clustering-based methods, edge-detection methods, and deformable modeling methods [4], [6], [8], [13]. Unfortunately, such methods (with threshold-based techniques, for example) yield the best results when the regions of interest in an image exhibit a massive difference in strength from the background of the image, but this results in more similar images with problems, dramatically reducing the efficiency and decreasing the applicability of these methods [6], [27]. Another limitation of conventional learning methods is that they depend on a specific function to conclude, and essential features must be identified by an expert [28].

More recently, the application of DL for medical image segmentation, i.e., convolutional neural networks (CNNs), has gained increased momentum [29]. CNNs can be implemented for segmenting the nuclei of cells [30], brain tumors on MRI scans [31], livers and tumors on CTs [32], the different lobes of the lungs [33], cataract surgery instruments [34], and multiple organs in laparoscopic surgery images [35]. All approaches that offer an end-to-end analysis (from raw images to segmented images) to overcome any previous methods' difficulties suffer. The main challenge of using DL for medical segmentation is that current CNNs do not generalize well to previously unseen object classes that are not present in the training set [36], [37]. However, there has been limited research on fetal heart segmentation and septal defect detection by using DL. In [38], CNNs for segmentation of the left ventricle were presented, and the results showed that dynamic CNNs could achieve good performance and provide robust segmentation. CNNs produce good segmentation in images with leakage, blurry boundaries, and subject-to-subject variations with a Dice score of the ground-truth image of approximately 94%. In [17], a full CNNs architecture was applied to detect the fetal heart and classify each of the ultrasound frames into standard viewing planes. The proposed CNNs relied on 16 layers based on the VGG architecture. The authors' model obtained a classification error of approximately 23.48%. Another DL approach offered a combination of CNNs and Recurrent Neural Network (RNNs) architectures, and the proposed model yielded an error rate of 27.7% [36]. A fully end-to-end, two-stream CNNs has been developed for temporal sequence learning to recognize, characterize, and fuse spatiotemporal fetal heart representations [37]. The CNNs architecture achieved 90% accuracy, 85% precision, and 89% sensitivity.

Automated methods for analyzing fetal echocardiography have been investigated recently [37]. Notably, the research on detecting a defect in the heart septum is very limited, with considerable errors in the segmentation results. Adding bounding boxes to an object can also be the first step before applying other image processing methods such as segmentation [39], [40]. Segmenting medical images can be a challenging problem, especially when lacking enough high- and low-resolution data. This scenario can result in degraded performance. 2D segmentation of fetal heart images using DL provides new opportunities for fetal echocardiography research to contribute precise, reliable, and automated detection for providing interpretations. It can potentially reduce the risk of human errors [7], [41]. Such an approach can accurately measure a wide range of fetal heart features [17], [38]. Several semantic segmentation models have been proposed in fetal organ detection, i.e., the head, heart, kidney, and other organs [42]. However, they do not provide instance-level information and only perform binary segmentation. This means that the model is unable to segment different instances of the same class.

Moreover, limited research on fetal heart segmentation, especially defect detection and low performance, has been performed using DL. The detection of defects can be challenging considering the variety of forms, textures, positions, and contrasts found in US. Therefore, improving the 2D segmentation performance for defect detection of the fetal

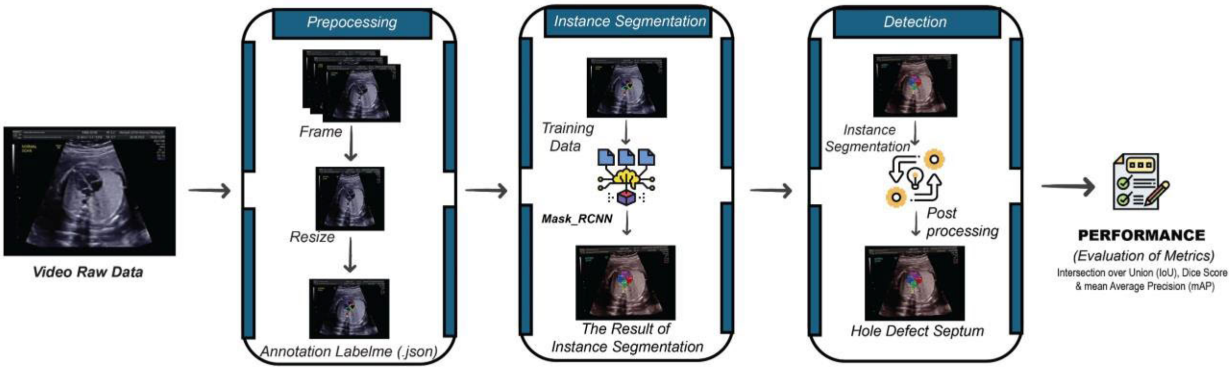


FIGURE 1. The stages of MRCNN architecture for hole detection as a defect. From left to right the i.e. raw data, pre-processing is manual segmentation for preparing the ground truth, wall-heart segmentation is divided in three datasets, i.e. training, validation and testing, and based on segmentation result the hole detection is created. The performance of the proposed algorithm is evaluated in terms of IoU, dice score, and mAP.

heart is essential, and an algorithm performs an in-depth investigation based on limited data must be developed.

III. THE PROPOSED METHOD

The general purpose of object detection is to recognize the desired object, quickly and accurately. However, most of these detection methods are still limited to multiple objects. MRCNN architectures provide a flexible and effective multi-objective framework with a parallel assessment of the regional proposal for object detection and segmentation [43]. MRCNN was initially developed for clear and large natural images [43]. In this study, MRCNN is used to predict septal defect conditions using US images at approximately mid-pregnancy between 18 and 24 weeks of gestation. In that condition, the object is very small, with an unclear image caused by the background shadow.

Furthermore, the US data are acquired during real clinical abnormality screening examinations in a freehand fashion in this work. Freehand scans are acquired without any constraints on the probe's motion, and the operator moves from view to view in no particular order. To our knowledge, the automated segmentation of the fetal standard scan plane for detecting a hole in the septum has never been performed in this challenging scenario. This is the first study in a medical application for septal defect detection using MRCNN.

To ensure that the proposed deep learning-based MRCNN architecture can work properly, four stages are proposed: (i) preconfigured bounding boxes are with various image shapes, and resolutions are established; (ii) the highest boundary boxes are defined to generate regional proposals; (iii) composite region proposals are pruned using non-maximum suppression and used to determine the presence or absence of a hole in a septum; and (iv) segmentation masks are produced for cases in which septal defects are positive, i.e., ASDs, VSDs, and AVSDs. In all stages, as seen in Fig. 1, hole detection is an essential component of this study for making septal defect decisions. To produce the best MRCNN architecture model, the parameter is chosen based on the hole detection's highest value in both the segmentation and object detection processes.

A. DATA PRE-PROCESSING

The input is ultrasound video data, obtained from pregnant women who were 18 – 24 weeks of pregnancy, in 2D images of the four-chamber view. In fetal echocardiography on a standard second-trimester anatomy scan, the four-chamber cardiac view is significant and routinely performed. The US videos are processed into frame form. The process is carried out using an open-source computer vision and machine learning software library, namely as OpenCV. In this study, three septal defect conditions with the ‘‘hole-in-heart'' septum are investigated, including holes in the atria (ASDs) in approximately 154 images, holes in the ventricles (VSDs) in about 178 images, holes in both the atria and ventricles (AVSDs) in approximately 184 images, wall chambers without holes (normal condition) in around 248 images. All data about 764 fetal heart images are splitting into 693 images for training, validation, and testing, and 71 images of normal conditions are employed as unseen data. All data were retrieved for retrospective analysis using a Digital Imaging and Communications in Medicine (DICOM) taken from Mohammad Hoesin Hospital in Indonesia and the Radiopaedia website [44]. The data distribution is described in Table 1, and the sample raw images data are presented in Fig. 2. This study separates the data into 80% for training, 10% for validation, and the remaining 10% for testing. To validate the proposed model's usability, we conduct experiments for the testing phase with unseen data consisting of 71 images under normal conditions.

In general, the 2D ultrasound images do not provide complete information about the structure of the fetal heart. This is because the US images are obtained from a cross-sectional sample of the 3D anatomic volume, which means that the acquired images depend on the probe's placement relative to the body and target structure. To perform a quantitative

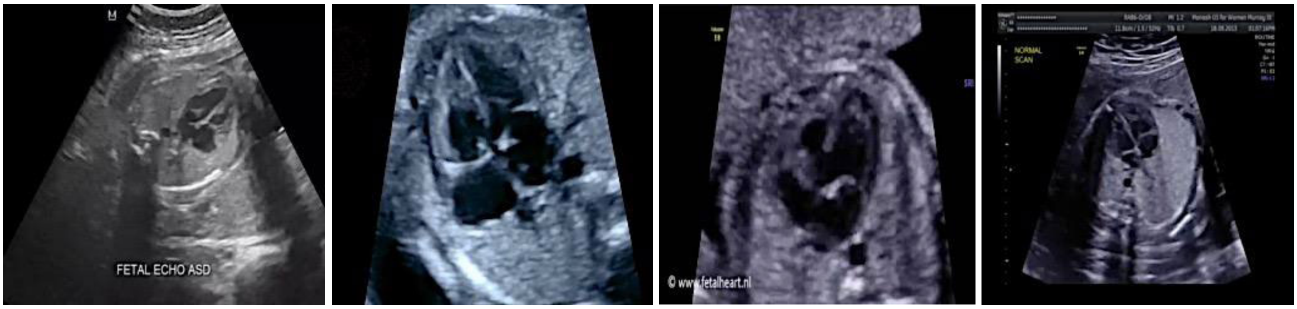


FIGURE 2. Sample of raw data for ASDs, VSDs, AVSDs, and normal conditions using fetal echocardiography (from the left to the right). The illustrated of the fetal heart is taken only four-chamber view. The four-chamber view of the fetal heart shows the majority of structures. It is an important view for diagnosis of the heart defect. It has a higher sensitivity and a very high specificity for the identification of CHDs.

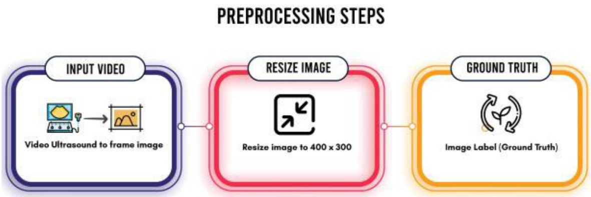

PREPROCESSING STEPS
INPUT VIDEO
Video Ultrasound to frame image
RESIZE IMAGE
Resize image to 400 x 300
GROUND TRUTH
Image Label (Ground Truth)

FIGURE 3. Pre-processing steps. The fetal heart was taken from echocardiogram devices with several size variations from 1.02 MB to 331 kb in DICOM format. All videos are transformed into images with the same resized 400 x 300-pixels resolution, and they are annotated and manually segmented by an expert using Adobe Photoshop CS 6 to mark the ground truths.

TABLE 1. The amount of raw Data for ASDs, VSDs, AVSDs, and normal. All data are splitting into three process, training, validation, and testing. In this study, testing process also use an unseen data for normal and abnormal detection.

Condition	Training	Validation	Testing	Unseen	Raw images
ASDs	120	13	21	-	154
VSDs	142	16	20	-	178
AVSDs	143	16	25	-	184
Normal	139	16	22	71	248
Total					764

analysis of the clinical parameters related to normal and abnormal conditions, segmentation of the fetal heart image is mandatory. Before the fetal heart is segmented, the US's raw data must be preprocessed to obtain a feature map for the processing stage, as depicted in Fig. 3. The preprocessing procedure is divided into three stages, including converting the US video to an image, resizing the image to approximately 400 x 300 pixels, and performing manual segmentation by two cardiologists to produce the ground truths of the wall chamber (atria and ventricles), aorta, and hole object.

B. INSTANCE SEGMENTATION FOR SEPTAL DEFECT DETECTION

MRCNN is a state-of-the-art model for instance segmentation. MRCNN extends the FRCNN architecture by introducing a parallel branch to predict segmentation masks [43]. There are two components of the MRCNN architecture. First, MRCNN creates ideas about the regions where an object might be presented on the input image. Second, MRCNN predicts the object's class, refines the bounding box, and creates a mask at the pixel level based on the first stage proposal. Both stages are related to the structure of the backbone. The backbone consists of a bottom-up pathway, a top-bottom pathway, and lateral connections. Any convolutional network can be the bottom-up pathway that extracts features from raw images. The top-bottom pathway produces a pyramid map function that is similar in size to the bottom-up pathway. On the other hand, lateral connections are convolutional and add operations between two corresponding levels of the two pathways.

The process of the region proposal network (RPN) contains a feature extractor and a region proposal. The feature extractor function serves to extract high-level features from the raw images. Such features require a cardiologist to determine the fetal heart region (RoI) through manual segmentation. The cardiologist draws the precise boundaries surrounding the RoI with correct annotations and draws each of the fetal image's RoI. The significant variations in shapes, sizes, textures, and, in certain cases, RoI colors between patients and those with poor contrasts between regions are used to create a database of ground truths. MRCNN is a 2-stage object-detecting RPN followed by a region-based CNN (RCNN) and a segmentation model as a mask, as seen in Fig. 4. The multitask losses in the RPN and RCNN are minimized for each image by using the following loss function:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*)+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \tag{1}$$

where i is an index of each anchor in the minibatch, p_i is the probability of prediction of the ith anchor to be an object, p_i^* is the ground-truth label that has a value of one if the anchor is positive and 0 if the anchor is negative, and t_i is a vector that represents the four coordinates of the ground-truth box referring to a given anchor. For the classification function, L_{cls} as the log loss over two classes is used. One of the classes is the object, and the other one indicates what is not an object in the form $L_{cls} = R(t_i - t_i^*)$ where R is the same robust, smooth function L_i defined in the Fast-RCNN. It is

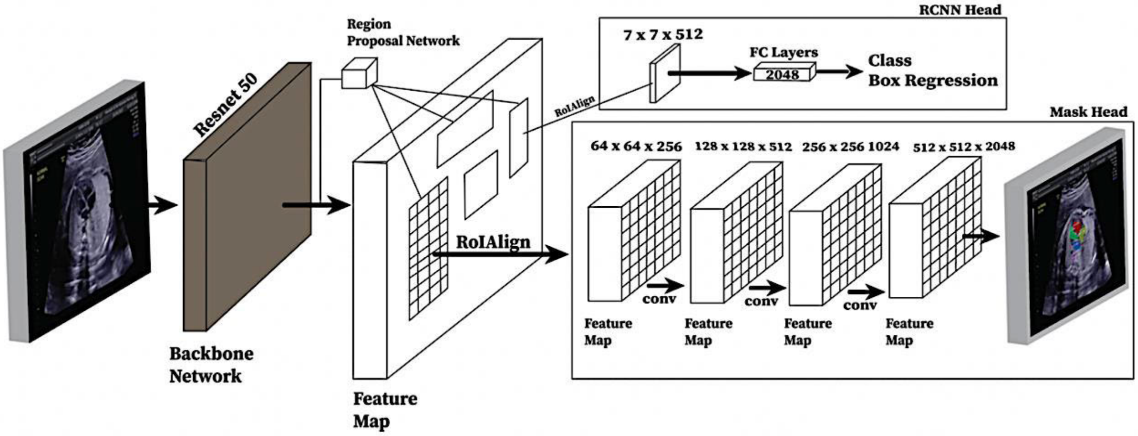


FIGURE 4. **Proposed MRCNN architecture. Such architecture is composed of four parts: backbone, region, RPN, RCNN for class prediction, bounding box regression, and another CNNs for pixel segmentation of objects, which refer to as mask.**

important to note that the term $p_i^* L_{reg}$ activates the regression loss only for positive anchors (that is, for $p_i^* = 1$) and turns it off when $p_i^* = 0$. In addition, the classification and regression bounding boxes are associated with p_i and t_i, respectively, while in MRCNN, the multiple task loss is defined in each RoI as L:

$$L = L_{cls} + L_{reg} + L_{mask} \tag{2}$$

where L_{cls} is the classification loss in the RCNN, L_{reg} is the regression bounding box loss in the RPN, and L_{mask} is the mask loss. Thus, a sigmoid function is applied per pixel that defines the average binary cross-entropy loss L_{mask} for an RoI and a ground-truth class k.

In the proposed architecture (refer to Fig. 4), all input images with the same size, i.e., 400 x 300 pixels, are rescaled to 512 x 512 pixels in the backbone network. Whereas to reduce false positives, negative patches are concatenated. The positive patches contain at least one object in the fetal heart. The concatenated patches are normalized to zero mean and unit variance. In the proposed architecture, ResNet50 architecture is implemented as the backbone in the RPN. The output of the network is a feature map comprised of the outputs of the first and last reduction steps. Each pixel in the feature map is scanned with two anchors with an anchor ratio [0.5, 1, 2] and RPN anchor scales [8, 16, 32, 64, 128]. Bounding boxes for regression and classification are applied to every anchor. RPN proposals with scores higher than 0.1 are passed to the RoI align layer. If no proposals that score higher than 0.1 are found, the ten anchors with the highest scores are passed. The low threshold is chosen to reduce false negatives. Proposals and anchors are considered positive (negative) if their IoU with a ground truth box is higher (lower) than 0.5 (0.1). The RoI align layer crops the proposals from the feature maps and rescales them to a fixed size.

To achieve high accuracy in object detection and small error rates in segmentation, we conduct several hyperparameter tuning trials to find the best model of MRCNN in terms of the learning rates, epoch, IoU baseline, momentum, and RPN backbone. Several learning rates are applied from 0.1 to 0.00001, 100 epoch, intersection over union baselines are set from 0.5 to 0.7, momentum is set from 0.7 to 0.9, and three backbones, the ResNet50, ResNet101, and MobilenetV1 architectures, are implemented with several hidden layers. Each backbone performs several strides [4, 8, 16, 32, 64]. To minimize the objective function, the stochastic gradient descent (SGD) optimizer is used with a batch size of one image. Approximately 256 anchor samples are randomized on all images to compute the loss function of a minibatch. In addition, these anchors have ratios of up to 1:1, 1:2, and 2:1 for sampled positives and negatives. All the layers' weights are randomly initialized from a zero-mean Gaussian distribution with a standard deviation of 0.01, and the ImageNet weights initialize the shared convolutional layers. To implement the proposed MRCNN model that can work properly, we used Python 3 with Anaconda, Keras 2.0.0, Keras Applications 1.0.5, and Keras Preprocessing 1.1.2. Keras was set to work in the TensorFlow backend with the TF-Nightly-GPU (version 2.2.0). The training was run on a 64-bit Windows 10 system with an Intel® core TM i9-9900K CPU @ 3.60 GHz (16 CPUs), a single NVIDIA GeForce RTX 2080 Ti (12 GB) GPU, and 32 GB RAM. The training took approximately 2 hours for eight epoch with 500 steps per epoch. After the prediction process on each training set was performed, the model with the lowest validation loss was selected as the final model.

C. EVALUATION AND VALIDATION

The creation of such models involves a set of structured image collections in three essential sub-datasets, which are widely used in various stages of model development. In the training phase, this first type consists of paired inputs: the image and the corresponding response (which can be a label, image, or mask), generally referred to as the target. In the validation phase, a subset is used to observe the evolution of the learning process while adjusting the model's parameters. Last, in the testing phase, an independent dataset of unlearned images is used to provide an objective evaluation of the fit of the target model on the first training dataset. In this study, the proposed MRCNN contains two loss functions, namely, categorical cross-entropy (CCE) in the RPN for the regression bounding box and image classification, and binary cross-entropy (BCE) for the mask as follows:

$$CCE = -\log q(x) \tag{3}$$

and

$$BCE = -(p(x).\log q(x) + (1-p(x)) \bullet \log(1-q(x)) \tag{4}$$

where $p(x)$ is the probability of class x in a target, and $q(x)$ is the probability of class x in the prediction. The predictions of the target outputs are in the form of probabilities that each image belongs to the foreground, the background for CCE is achieved by SoftMax activation, and the prediction of the target output, achieved by sigmoid activation for BCE, is either 0 or 1.

MRCNN makes predictions in terms of a bounding box, class label, and class segmentation. To measure the three predictions, namely mIoU, mAP, and DSC are used [37], [45]. The proposed model involves an element of confidence that implements a trade-off between precision and sensitivity by adjusting the confidence level needed to make a prediction. The predicted images (P) are compared to the ground truth image (G) with region-based metrics given as percentages to measure the segmentation results. The DSC is calculated, as shown in (5) as follows:

$$DSC(P, G) = 2\frac{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} P_{ij}G_{ij}}{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} P_{ij} + \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} G_{ij}} \tag{5}$$

where i and j represent the pixel indices for the height N and width M, respectively. The range of the DSC is [0, 1], and a higher DSC corresponds to a better match between the predicted image P and the ground truth image G.

The mIoU is also known as the Jaccard index. This is a metric used to calculate the intersection percentage between the labeled mask and the predicted output. The intersection over union is calculated for each class, and the values of all classes are averaged. The mIoU is an extremely effective and very straightforward metric. The mIoU is presented in (6) and defined as:

$$MIoU = \frac{1}{N_{cls}}\sum_{x=1}^{N_{cls}}\frac{N_{xx}}{\sum_{y=1}^{N_{cls}} N_{xy} + \sum_{y=1}^{N_{cls}} N_{yx} - N_{xx}} \tag{6}$$

Finally, the mAP metric is used to evaluate both classification and segmentation for all object classes. Such a metric is used to accurately measure the correctly predicted images after the IoU is obtained. The proposed model must avoid false positives. The confidence threshold is set high to encourage the model to only produce high precision predictions at the expense of lowering its amount of overlapping coverage. The mAP is presented in (7) and defined as:

$$mAP = \frac{1}{n}\sum\nolimits_{i=1}^{N} AP_i \tag{7}$$

where AP_i is the AP in the *ith* class and N is the total number of classes being evaluated.

IV. RESULTS AND DISCUSSION

Segmenting fetal heart images is essential for interpreting US screenings. In Fig. 5, the heart wall's RoI is segmented manually by two cardiologists to obtain a ground truth with the same size to increase the processing speed. There are six classes, including the left and right atria, left and right ventricles, aorta, and hole, that must be annotated as the ground truth. The markers are placed in the wall chamber, aorta, and hole for ASDs, VSDs, AVSDs, and normal conditions. For ASDs and VSDs, there are six objects in the segmented area; for AVSDs, there are seven objects in the segmented area; and for normal conditions, there are five objects in the segmented area.

In this study, holes in the septum must be detected accurately. Thus, all hole's detection performed is highly considered to enable the selection of the best model. The learning rate varies from 0.1 to 0.00001, and the momentum varies from 0.7 to 0.9. However, these two parameters, learning rate and momentum must be comparatively selected. It was observed that momentum of 0.9 produced good results, outperforming the values of 0.8 and 0.7. The convergence time increased by almost two times, but the IoU values were similar for each class. The multiclass segmentation is conducted with a momentum value of 0.9, 100 epoch, and an IoU baseline of 0.5, but in that model, we use a varied learning rate. Fig. 6 shows the IoU result from the predicted image caused by the learning rate's tuning. The MRCNN model produces good performance in terms of IoU in the aorta, hole, LA, LV, RA, and RV at all learning rate values. Fig. 6, shows that both learning rate at 0.001 and 0.0001 produce a relatively same value, 0.5. However, by using learning rate 0.001, it produces a slightly higher by 0.57 IoU in the hole object than the other. Therefore, in the proposed MRCNN model to ensure the robustness a learning rate 0.001 is selected as the value in best MRCNN model.

While selecting the hyperparameters, all the values were changed manually; therefore, the error reached minimum in terms of the cost achieved for a lower convergence time of the system. Initially, ResNet50 was utilized as the RPN backbone. In this instance, segmentation yields three outputs, i.e., the predicted class of RoI, the predicted bounding box from the RoI, and the final segmentation prediction, which provides highly detailed object detection. To select the best model of MRCNN, the RPN backbone use three architectures, Resnet50, ResNet101 and MobilenetV. The architecture chosen is an architecture that can produce the best performance in the segmentation process so that it can

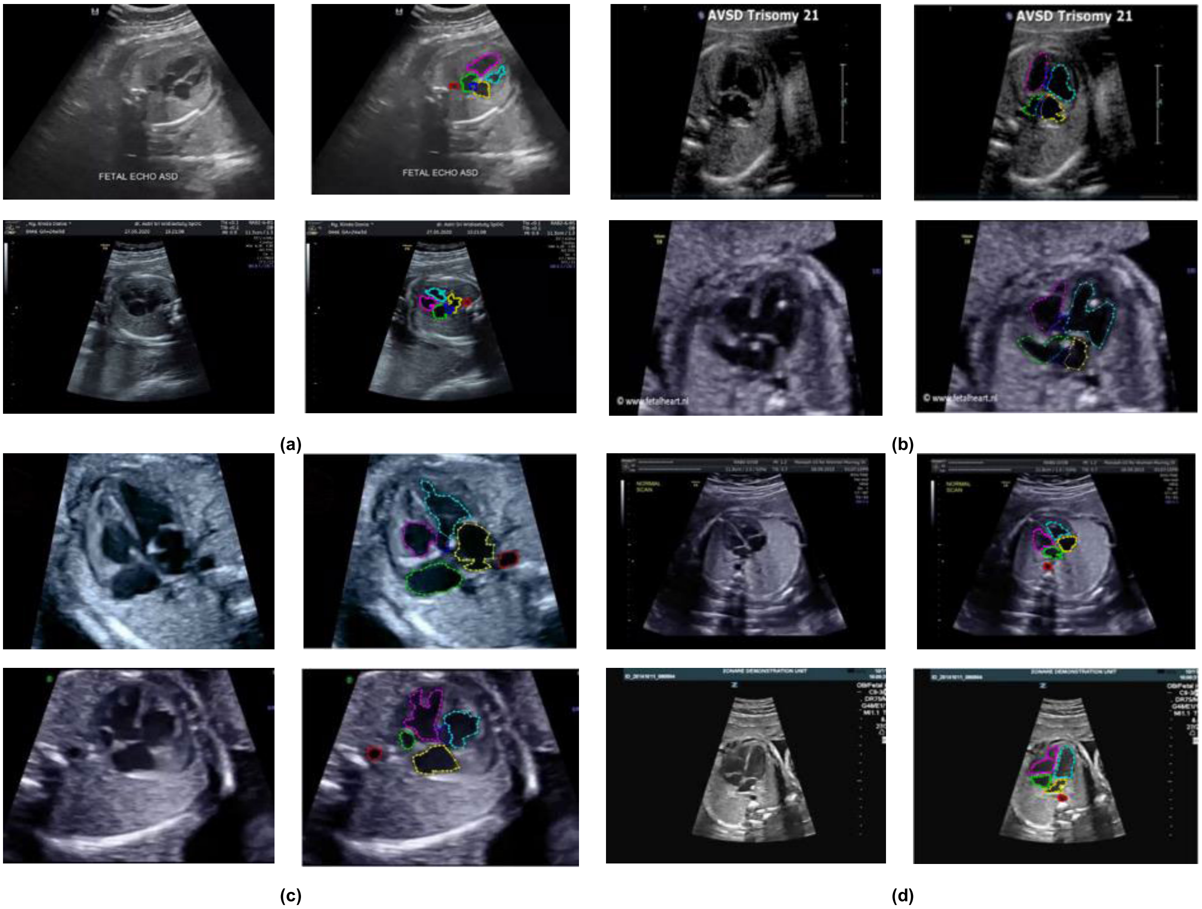
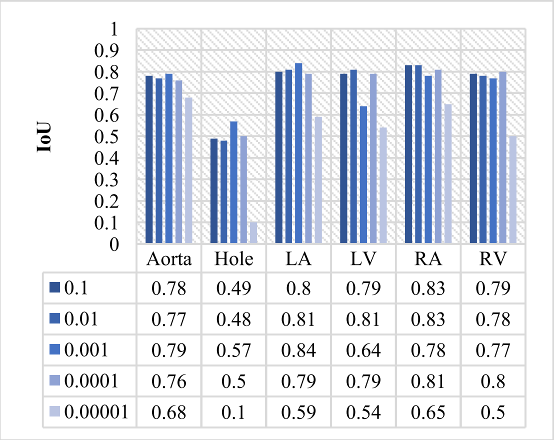


FIGURE 5. Data ground truth in (a) ASDs, (b) VSDs, (c) AVSDs, and (d) Normal. The raw data is annotated with several colors for marking the heart- chamber, aorta, and hole. There are six objects as a class for segmenting, such as right atria (green color), left atria (yellow color), right ventricle (blue color), left ventricle (purple color), aorta (red color), and hole (dark blue color). The aorta is selected as the class to assign the position of atria, caused by the fetus move to several directions. Based on the aorta position, it can be a diagnosis of the defect in atria or ventricle.



	Aorta	Hole	LA	LV	RA	RV
■ 0.1	0.78	0.49	0.8	0.79	0.83	0.79
■ 0.01	0.77	0.48	0.81	0.81	0.83	0.78
■ 0.001	0.79	0.57	0.84	0.64	0.78	0.77
■ 0.0001	0.76	0.5	0.79	0.79	0.81	0.8
■ 0.00001	0.68	0.1	0.59	0.54	0.65	0.5

FIGURE 6. Learning rate tuning to find the best result with 0.5 IoU baseline, 100 epoch, 0.9 momentum, and Resnet50 as RPN back bone model. It uses for six classes, namely Aorta, defect (hole), LA, LV, RA, and RV.

provide a robust structure for the proposed MRCNN model. The performances are measured in mIoU, mAP and DSC, based on the selected baseline IoU. As seen in Table 3, the comparison of the segmentation results, based on the chosen values of IoU for the aorta, hole, LA, RA, LV, and RV using ResNet50, ResNet101, and MobilenetV1 architectures. The difference between them lies in the number of hidden layers.

The IoU is used to measure how much of the predicted image boundary overlaps with the ground truth image. In this study, the hole class is important to accurate detection and is related to the type of septal defect condition. To achieve this goal, the IoU of the hole must be greater than the baseline, or the IoU value must be higher than 0.5. Regarding all the results obtained (shown in Table 2), MRCNN with the ResNet50 architecture outperforms MRCNNs with the ResNet101 and MobilenetV1 backbone architectures in detecting septal defects with IoUs greater than 50% (0.5). The hole segmentation results for ASDs are approximately 0.51, approximately 0.50 for VSDs, and approximately 0.59 for AVSDs. The ResNet101 and MobilenetV1 architectures produce hole segmentations below 50% (0.5) for three conditions. This means that MRCNN cannot detect holes in the heart septum. Based on this result, ResNet50 is selected as the RPN backbone model to ensure that the hole can be detected perfectly for every CHDs condition.

TABLE 2. Segmentation performances for three backbones architecture based on 0.5 IoU to select the best MRCNN model.

IoU performance for each class						
MobilenetV1 backbone has 28 layers						
	Aorta	Hole	LA	LV	RA	RV
ASDs	0.72	0.45	0.80	0.72	0.87	0.75
AVSDs	0.75	0.44	0.22	0.38	0.34	0.61
VSDs	0.75	0.52	0.56	0.43	0.47	0.57
Normal	0.74	no hole	0.77	0.65	0.82	0.72
Resnet50 backbone has 50 layers						
	Aorta	Hole	LA	LV	RA	RV
ASDs	0.73	0.51	0.80	0.75	0.86	0.79
AVSDs	0.76	0.59	0.77	0.77	0.77	0.79
VSDs	0.75	0.50	0.80	0.82	0.86	0.83
Normal	0.74	no hole	0.78	0.70	0.82	0.75
Resnet101 backbone has 101 layers						
	Aorta	Hole	LA	LV	RA	RV
ASDs	0.47	0.23	0.77	0.43	0.85	0.73
AVSDs	0.76	0.29	0.71	0.73	0.78	0.65
VSDs	0.75	0.29	0.44	0.22	0.58	0.56
Normal	0.43	no hole	0.69	0.28	0.56	0.56

TABLE 3. Object detection performances for six classes based on mAP for three backbones architecture to select the best MRCNN model.

	mAP performance (%)		
Class	Resnet 50	Resnet 101	MobilenetV1
Aorta	99.97	84.48	87.89
Hole	87.10	59.66	53.88
LA	100.0	89.29	73.22
LV	99.19	87.55	69.07
RA	99.99	86.92	68.39
RV	99.03	88.94	74.22

TABLE 4. Object detection performances based on mAP for three conditions IoU base line to select the best MRCNN model with Resnet50 architecture.

Class	mAP performance (%)		
	0.5 IoU	0.6 IoU	0.7 IoU
Aorta	99.97	98.68	87.74
Hole	87.10	65.21	48.84
LA	100.0	99.17	89.49
RA	99.99	99.81	99.01
LV	99.19	98.04	92.18
RV	99.03	97.94	93.99

Table 3 shows the object detection results with mAP performances for the three architectures as backbones. The mAP measures the bounding box prediction of the heart-chamber and hole RoI compared to the ground truth. Using the ResNet50 architecture, the detection results for aorta and heart-chamber detection produce an mAP over 99%, while hole detection reaches an mAP of approximately 87.10%. This means that multiclass segmentation by MRCNN achieves a satisfactory result, and the RoI of the hole is predicted perfectly, with an 87.10% overlap with the ground truth. It is concluded that the proposed model possesses the ability to segment the fetal heart chamber and aorta, and it also succeeds in detecting the "hole-in-heart" septum, outperforming the ResNet101 and MobilenetV1 architectures.

As seen in Table 4, the proposed MRCNN model can detect all wall chambers, under both normal and abnormal conditions, with defects in the atria and ventricles. However, the defect or hole detection performance in the wall chamber produces unsatisfactory results when the overlapping requirement between the ground truth and predicted (IoU) is increased to 70% (0.7), which results in an mAP of approximately 48.84%. This means that MRCNN does not detect the hole because it is very small with a large background. In contrast, using an IoU of 50% (0.5) produces 87.10% mAP and using an IoU of 60% (0.6) produces 65.21% mAP. Based on this result, a baseline IoU of 50% (0.5) is utilized in this septal defect problem. In the future, hole detection precision should be increased by adding a postprocessing algorithm. Therefore, the image quality can be further improved, and holes can be detected perfectly with a high IoU baseline.

The Dice score similarity is also used for calculations to ensure the selected object detection method's satisfactory performance. As shown in Table 5, the MRCNN performance produces satisfactory results based on IoU and DSC values for segmenting the wall chambers in ASDs, VSDs, AVSDs, and normal conditions with a high overlap between the ground truth and predicted image. This means that the predicted results can recognize each chamber in the fetal heart. All performances reach IoU and DSC values over 50%. Especially in hole segmentation, MRCNN produces a lesser IoU performance. However, it is still able to detect holes in the wall chamber, which means that the hole can be detected with very minimal overlapping results. Such conditions are undesirable because they can decrease septal defect detection performance.

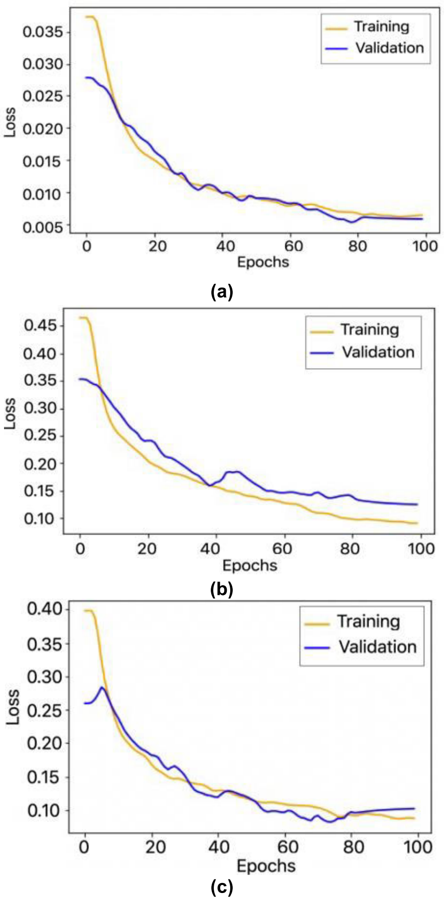
In this study, we utilize 764 US images for training, validation, and testing process. However, to avoid large gap fluctuation between training and validation results (overfitting), we perform a data augmentation to increase our dataset. The fluctuation in the validation loss can occur for several different reasons, such as (i) the learning rate that is too high (making the stochastic gradient descent overshoot when

TABLE 5. Segmentation performances based on DSC with Resnet50 backbones, and 0.5 IoU to measure the overlapping value between ground truth and predicted result.

	DSC performance (%)					
Condition	Aorta	Hole	LA	LV	RA	RV
ASDs	84.17	65.72	88.90	85.73	92.60	87.95
VSDs	83.70	67.15	88.91	90.00	92.28	90.42
AVSDs	86.44	73.42	86.72	85.23	84.87	86.42
Normal	84.94	no hole	87.74	88.90	89.71	85.40

trying to find a minimum); (ii) limited amount of data training and validation; (iii) not all patterns generalizing because the validation data are not a subset of the training data; and (iv) predictions that are sensitive to small changes in the input image caused by wide variations in the source image with a large shadow in the background. The off-line data augmentation technique is used, which indicates the dataset was augmented and annotated before the training process. We performed random image rotations, random zoom, random US image dimension, random brightness with gaussian blur, and luminosity scaling in the range of [0.8, 1.5].

After the augmentation process, all data become 1200 US images for MRCNN segmentation process. The total image augmentation consists of 280 ASDs, 306 VSDs, 316 AVSDs, and 298 Normal conditions. The learning rate value is reduced from 0.001 to 0.0001 to eliminate the SGD overshoot when trying to find a minimum value. To minimized the computation time, 100 epoch, a momentum value 0.9, and 0.5 IoU with ResNet50 as the backbone. Fig. 7 is shown the MRCNN loss graph, including the classification loss, bounding box loss, and segmentation loss using data augmentation. All loss tends to zero in training and validation. The overfitting problem can be overcome, as seen from the training and validation graphs that are very close and tend towards zero. The summary all result in terms of mIoU, mAP, and DSC is presented in Table 6. It shows six classes of the aorta, hole, LA, LV, RA, and RV, respectively; performances are increased significantly. In particular, defect (hole) segmentation performance achieves 76.52% mIoU, 99.84 % mAP, and 87.78% DSC, whereas before augmentation only achieve 53.33% mIoU, 87.10 % mAP, and 69.76% DSC. It can be concluded that data augmentation can improve the performance of the proposed model.



(a)

(b)

(c)

FIGURE 7. MRCNN loss. It describes about loss in object detection part with learning rate 0.0001 and 100 epoch; (a) Classification loss, (b). Bounding-box loss, and (c) Segmentation loss. The orange color is training loss graph, and the blue color is validation loss graph.

TABLE 6. The summary of mIoU, mAP, and DSC from the best model of MRCNN before and after data augmentation.

Class	mIoU	mAP	DSC
Before Data Augmentation			
Aorta	75.40	99.97	84.81
Hole	53.33	87.10	69.76
LA	78.75	100.0	88.07
LV	76.00	99.19	87.47
RA	82.75	99.99	89.86
RV	79.00	99.03	87.55
After Data Augmentation			
Aorta	78.50	99.97	88.19
Hole	76.00	99.48	87.78
LA	84.50	99.67	91.70
LV	64.50	86.17	76.71
RA	77.75	97.59	87.75
RV	76.75	98.83	87.19

To evaluate the method quantitatively, Fig. 8 presents septum defects in the atrial septum and ventricular septum. The proposed segmentation method provides a simple but effective way to detect septal defects automatically. The proposed method was implemented under validation, testing, and evaluation with unseen data to assess the robustness of the proposed model. The results revealed that the heart wall had been accurately detected with small errors around the boundary wall. However, those errors did not affect the final detection (refer to Fig. 8). Each class's confidence image value was over 90% and was used to ensure that the predicted result was comparable to the ground truth. The MRCNN method was shown to be capable of recognizing the multi segmentation of objects that lacked consistency in the acquisition of their source data from the 2D fetal USG. Moreover, a radiologist might take a reasonably decision from such data and need to determine what conclusion would be chosen. In this study, the

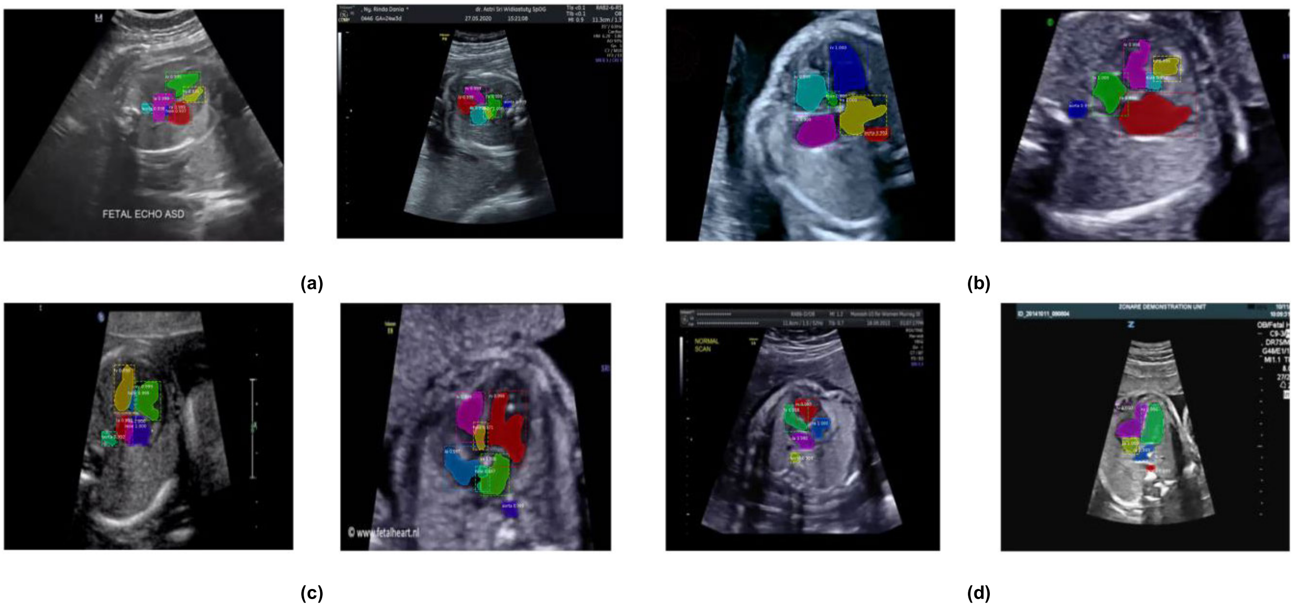


FIGURE 8. The instance segmentation result with data testing. It visualized for (a) ASDs, (b) VSDs, (c) AVSDs, and (d) Normal. The proposed model can segment the heart-chamber, aorta, and hole. The result has validated with a cardiologist to ensure the measurement.

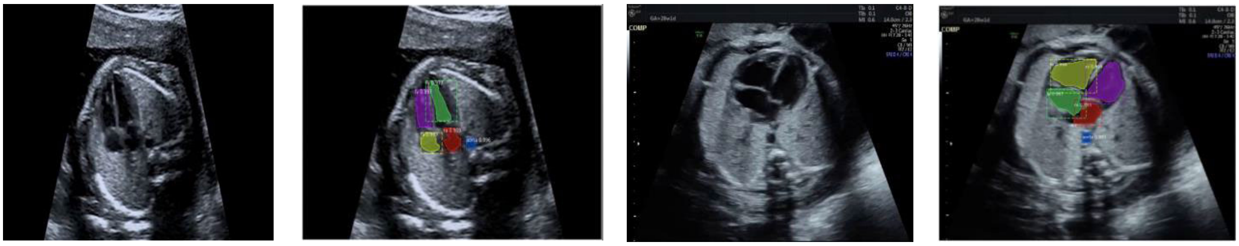


FIGURE 9. Instance segmentation result for data unseen with normal condition to evaluate the hole detection. From left to the right shows raw images and object detection result. The image is used as unseen to evaluated the proposed model. The result has validated with a cardiologist to ensure the measurement.

data of a few abnormal images were used for analysis; thus, the unseen data only used the normal condition. Fig. 9 shows the segmentation result with the unseen data used as the testing data. The proposed model was tested on a real dataset from the Mohammad Hoesin Indonesian Hospital. The evaluation by using the proposed model shows 100% sensitivity for normal conditions. The experimental findings, confirmed by a radiologist, indicate that the proposed model is capable of identifying septal defects and providing the radiologist with visual guidance to decide on a fetal heart condition. There are results with no holes detected based on the proposed model. This means that the MRCNN possesses the capability to recognize the heart wall in normal or abnormal conditions.

The MRCNN architecture is based on the FRCNN [46] that introduced an efficient RPN by using a sliding window approach to make approaches translation-invariant. The system produces precisely the same response regardless of how its input is shifted, and it produces a good recognition process even though the actual pixel values are quite different. MRCNN is a simple but effective addition to the FRCNN architecture. Table 7 shows that MRCNN outperforms FRCNN in defect detection. It increases the detection performance by 5% when based on ResNet50 using 100 epoch for each image before augmentation, and increase 17% after augmentation. Such a method has additional prospects, for instance, accurately predicting the pixel-level instance mask.

TABLE 7. MRCNN versus FRCNN architecture to predict the hole as a defect in wall-chamber with mAP performance.

Object	Mean Average Precision (%)		
	MRCNN with Raw data	MRCNN with Augmentation	FRCNN with Raw data
Hole Detection	87.10	99.48	82.00

The semantic segmentation model places strict spatial restrictions on the boundary boxes to predict the septal defects because each grid cell predicts only two boxes and can only have one class (refer to Fig. 10). This spatial constraint limits the number of nearby objects that our proposed model can

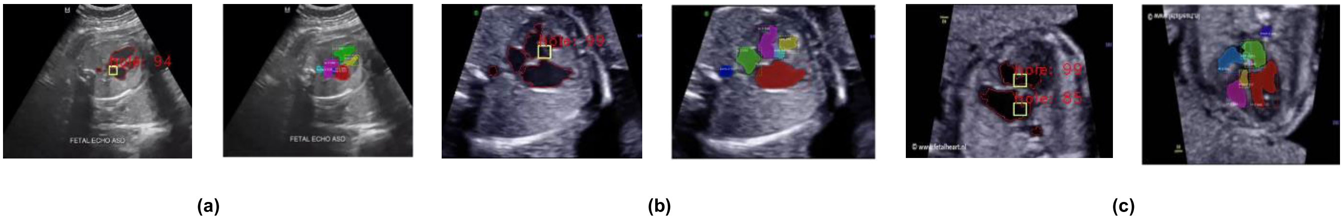


FIGURE 10. Result comparison between semantic segmentation and instance segmentation to septal defect detection in the fetus for (a) ASDs, (b) VSDs, and (c) AVSDs. The red lines are fetal-heart contour by semantic segmentation, yellow boxes are hole detection by FRCNN, and colorful areas are fetal-heart instance segmentation by MRCNN.

TABLE 8. The comparison mAP performances between MRCNN and FRCNN architecture in fetal object with 0.5 IoU.

Method	Object	mAP (%)
FRCNN with a fetal heart for valves detection [26]	Apical 2	49
	Apical 3	55
	Apical 4	89
	Apical 5	52
	PLAX	95
	PLAX- RVIF	73
Proposed MRCNN with a fetal heart for septal defect detection	Aorta	99.97
	Hole	99.48
	LA	99.67
	LV	86.17
	RA	97.59
	RV	98.83

TABLE 9. The comparison object detection with MRCNN architecture for medical imaging.

Medical Object	Performance (%)		
	mIoU	mAP	DSC
Hole segmentation and detection (current work)	76.00	99.48	87.78
Nucleus segmentation and detection [47]	70.54	59.40	-
Lung Nodules segmentation and detection [48]	-	-	70.0
Oral diseases segmentation and detection [49]	-	-	74.0

predict. In the proposed model of instance segmentation, the segmentation process struggles with small objects in groups. It learns to predict bounding boxes from data with a high generalization of objects in new or unusual aspect ratios or configurations. Furthermore, the model also uses relatively coarse features for predicting bounding boxes since our architecture contains multiple down-sampling layers from the input image. The FRCNN-based regional hole detection approach produces an mAP of 82% with a confidence value above 85% for the three conditions, namely, ASDs, VSDs, and AVSDs. It utilizes the best model of the U-Net segmentation architecture with an IoU of 0.7. However, using MRCNN with data augmentation, an mAP of 99.48% is achieved with a confidence value above 90% and the same IoU value. This shows that the proposed model succeeds in detecting hole as a septal defect in the wall chambers, both in atria and ventricles.

Each pixel is assigned to an object category in the semantic model; therefore, the U-Net segmentation produces a binary mask of 1 and 0, where 1 indicates a fetal heart, and 0 suggests the background. The segmentation task only receives the same label as that of the fetal heart. It does not possess different labels to distinguish between different objects as humans would during segmentation. As seen in Fig. 8, the MRCNN model annotates each pixel as an individual object, and it produces a categorical mask between 1 and 0. Therefore, the proposed model is able to segment the wall-chamber, aorta, and hole defect for an input image that contains an object with high probability. In Table 8, the proposed MRCNN model is compared to the FRCNN model for a fetal object's multiclass segmentation. Our model produces high mAPs for atrium, ventricle, aorta, and hole detection. In [26], six classes are segmented. However, the mAP results are unsatisfactory, and only three classes achieve good results. Apical 2, Apical 3, and Apical 5 produce mAPs with only a 50% overlap between the predicted image and ground truth. Large variations in mAP values from 0.49 to 0.95 are based on FRCNN because, in such a model, only bounding box detection is performed without creating RoI in the fetal heart. Our proposed model produces mAPs over 99% for all classes, except LV, only 86.17%. At the same time, MRCNN contains two processes: class segmentation and object detection. MRCNN is fully trainable from end to end. Nonetheless, convergence is faster when training the backbone and RPN and then training only the second-stage heads. Training both the segmentation and detection tasks simultaneously improves the detection rate.

Further research on using the MRCNN model in fetal heart applications is needed to make the model feasible for clinical settings. Table 9 shows the comparison of results of the MRCNN for object detection in other medical applications based on our proposed model. MRCNN can detect holes in the fetal heart with 76% IoU, 99.48% mAP, and 87.78% DSC. All metrics show good results on the segmentation and object detection processes for hole detection. However, the IoU and DSC values obtain the minimum overlap between the ground truth and predicted image. In [47], the authors achieved only 59.40% mAP, which means that many images only partially overlapped, whereas mAP indicated how the model prediction was not highly accurate. In [48] and [49], their proposed MRCNN model produced over 70% DSC, meaning that the prediction data were similar to the ground truth data with a similarity index of approximately 0.7. Such values depend on the RoI. If the RoI is precisely predicted, then the DSC is increased, and vice versa.

Based on all the results, the proposed echocardiographic interpretation instance segmentation is likely to improve not only the IoU and mAP of the reading but also its timeliness. The proposed model can support a physician reading an echocardiogram obtained to evaluate mitral regurgitation before and intervention, which could require a program to recognize all views of mitral regurgitation. This can save the physician time by screening through a study of likely hundreds of images, enabling physicians to easily visualize all relevant details, enabling them to be more effective. The advantage of segmentation in medical imaging, especially in septal defect detection, is detection speed due to the simple and small architecture. It is suitable for real-time detection with a network that understands generalized object representation.

Large backgrounds with shadows or a lack of consistency in data can also cause significant differences in the source image, as is often occurred in the case in real applications. For this reason, several ML approaches have the fundamental problem for a lack of global applicability that limits their utility to a limited number of applications. In this study, MRCNN can overcome this problem and produce several advantages, such as high inference speed, high mean average precision accuracy, an intuitive and easily implementable approach, and extension capability. However, the proposed model has several limitations:

(i) More varied training data are needed to increase the septal defect detection performance, especially under abnormal conditions;

(ii) To ensure the robustness of the proposed model, cross-fold validation can be applied;

(iii) Evaluations with large amounts unseen data are necessary to increase the generalization ability of the proposed model; and

(iv) An extension of our work should be considered to enable the model to work with more cardiac views, such as left and right ventricular outflow tracts (LVOTs and RVOTs), views with three vessels trachea images (3VT).

V. CONCLUSION

Cardiac screening in fetuses remains a problem requiring a team of experts. To improve the diagnosis of a cardiac defect and make it manageable in utero, plan delivery, and identify CHDs that may progress in utero to heart defects, an object detection approach is proposed. This study presents instance multiclass segmentation, which produces automated segmentation of the atria, ventricles, aorta, defects, and object detection, which predicts the hole position. Based on MRCNN, this approach has been shown to successfully segment heart chamber, aorta, and defects with limited image data due to the rare incidence of CHD in fetal cases. By using ResNet50 as the best backbone network, MRCNN can force different scale feature learning by using different layers in the network and using anchors and RoI align instead of treating layers as black boxes. To evaluate our proposed model, we compare it to FRCNN with U-Net segmentation. The results indicate that the proposed model produces a two times faster processing time than FRCNN, with satisfactory multiclass segmentation and hole detection performance. All results have been validated by experts to ensure the appropriate achievement of hole detection. A known problem with the septal defect dataset is the ground truth consists of a very limited and strict dataset. In the future, we plan to train MRCNN on a more extensive dataset with several views (four-chamber, three-vessel, and trachea views, and images of the left and right ventricular outflow tracts) to generalize our model.

REFERENCES

[1] B. J. Bouma and B. J. M. Mulder, "Changing landscape of congenital heart disease," *Circulat. Res.*, vol. 120, no. 6, pp. 908–922, 2017.

[2] W. Dakkak and T. I. Oliver, "Ventricular septal defect," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls Publishing, 2018.

[3] C. Mavroudis, J. A. Dearani, and R. H. Anderson, "Ventricular septal defect," in *Atlas Adult Congenital Heart Surgery*. Springer, 2020, pp. 91–115.

[4] Z. Jalal, "Long-term complications after transcatheter atrial septal defect closure: a review of the medical literature," *Can. J. Cardiol.*, vol. 32, no. 11, p. 1315, 2016.

[5] V. L. Vida, C. Tessari, B. Castaldi, M. A. Padalino, O. Milanesi, D. Gregori, and G. Stellin, "Early correction of common atrioventricular septal defects: A single-center 20-year experience," *Ann. Thoracic Surg.*, vol. 102, no. 6, pp. 2044–2051, Dec. 2016.

[6] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, and B. Bijnens, "Machine Learning in Fetal Cardiology: What to Expect," *Fetal Diagnosis Therapy*, vol. 47, no. 5, pp. 363–372, vol. 2020, doi: 10.1159/000505021.

[7] J. Espinoza, "Fetal MRI and prenatal diagnosis of congenital heart defects," *Lancet*, vol. 393, no. 10181, pp. 1574–1576, 2019, doi: 10.1016/S0140-6736(18)32853-8.

[8] V. Rawat, A. Jain, and V. Shrimali, "Automated techniques for the interpretation of fetal abnormalities: A review," *Appl. Bionics Biomech.*, vol. 2018, pp. 1–11, Jun. 2018, doi: 10.1155/2018/6452050.

[9] L. Allan, J. Dangel, V. Fesslova, J. Marek, M. Mellander, I. Oberhänsli, R. Oberhoffer, G. Sharland, J. Simpson, and S.-E. Sonesson, "Recommendations for the practice of fetal cardiology in europe," *Cardiol. Young*, vol. 14, no. 1, pp. 109–114, Feb. 2004, doi: 10.1017/s1047951104001234.

[10] L. Saba *et al.*, "Intra-and inter-operator reproducibility analysis of automated cloud-based carotid intima media thickness ultrasound measurement," *J. Clin. Diagnostic Res.*, vol. 12, no. 2, pp. 649–664, 2018.

[11] A. Natale, O. M. Wazni, K. Shivkumar, and F. Marchlinski, *Handbook of Cardiac Electrophysiology*. Boca Raton, FL, USA: CRC Press, 2016.

[12] U. Gembruch, "Prenatal diagnosis of congenital heart disease," *Prenatal Diagnosis*, vol. 17, no. 13, pp. 1283–1298, Dec. 1997.

[13] N. J. Bravo-valenzuela, A. B. Peixoto, and E. Araujo Júnior, "Prenatal diagnosis of congenital heart disease: A review of current knowledge," *Indian Heart J.*, vol. 70, no. 1, pp. 150–164, Jan. 2018.

[14] C. P. Bridge, C. Ioannou, and J. A. Noble, "Automated annotation and quantitative description of ultrasound videos of the fetal heart," *Med. Image Anal.*, vol. 36, pp. 147–161, Feb. 2017.

[15] A. Davis, K. Billick, K. Horton, M. Jankowski, P. Knoll, J. E. Marshall, A. Paloma, R. Palma, and D. B. Adams, "Artificial intelligence and echocardiography: A primer for cardiac sonographers," *J. Amer. Soc. Echocardiography*, vol. 33, no. 9, pp. 1061–1066, Sep. 2020.

[16] K. C. Kaluva, C. Shanthi, A. K. Thittai, and G. Krishnamurthi, "CardioNet: Identification of fetal cardiac standard planes from 2D Ultrasound data," North Northlake Way, Seattle, WA, USA, Semant. Scholar, 2018.

[17] V. Sundaresan, C. P. Bridge, C. Ioannou, and J. A. Noble, "Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 671–674, doi: 10.1109/ISBI.2017.7950609.

[18] S. Nurmaini, A. Darmawahyuni, A. N. Sakti Mukti, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification," *Electronics*, vol. 9, no. 1, p. 135, Jan. 2020, doi: 10.3390/electronics9010135.

[19] S. Nurmaini, R. Umi Partan, W. Caesarendra, T. Dewi, M. Naufal Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus, "An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique," *Appl. Sci.*, vol. 9, no. 14, p. 2921, Jul. 2019, doi: 10.3390/app9142921.

[20] C. F. Baumgartner, K. Kamnitsas, J. Matthew, T. P. Fletcher, S. Smith, L. M. Koch, B. Kainz, and D. Rueckert, "SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2204–2215, Nov. 2017, doi: 10.1109/TMI.2017.2712367.

[21] A. Ghorbani, D. Ouyang, A. Abid, B. He, J. H. Chen, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Deep learning interpretation of echocardiograms," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–10, Dec. 2020.

[22] M. Alsharqi, W. J. Woodward, J. A. Mumith, D. C. Markham, R. Upton, and P. Leeson, "Artificial intelligence and echocardiography," *Echo Res. Pract.*, vol. 5, no. 4, pp. R115–R125, Dec. 2018, doi: 10.1530/ERP-18-0056.

[23] F. M. Asch, "Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert," *Circulat., Cardiovascular Imag.*, vol. 12, no. 9, 2019, Art. no. e009303.

[24] N. Poilvert, "Deep learning algorithm for fully-automated left ventricular ejection fraction measurement: P2-45," *J. Amer. Soc. Echocardiography*, vol. 31, no. 6, 2018.

[25] G. Carneiro, *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings* vol. 10008. Athens, Greece: Springer, Oct. 2016.

[26] D. G. Gungor, B. Rao, C. Wolverton, and I. Guracar, "View classification and object detection in cardiac ultrasound to localize valves via deep learning," in *Proc. Mach. Learn. Res.*, London, U.K., 2020. [Online]. Available: https://openreview.net/forum?id=6WmtOMMzn-

[27] J. Torrents-Barrena, G. Piella, N. Masoller, E. Gratacós, E. Eixarch, M. Ceresa, and M. Á. G. Ballester, "Segmentation and classification in MRI and US fetal imaging: Recent trends and future prospects," *Med. Image Anal.*, vol. 51, pp. 61–88, Jan. 2019, doi: 10.1016/j.media.2018.10.003.

[28] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, Inc., 2017.

[29] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018, doi: 10.1109/TMI.2018.2791721.

[30] M. Kowal, M. Żejmo, M. Skobel, J. Korbicz, and R. Monczak, "Cell nuclei segmentation in cytological images using convolutional neural network and seeded watershed algorithm," *J. Digit. Imag.*, vol. 33, no. 1, pp. 231–242, Feb. 2020.

[31] S. Hussain, S. M. Anwar, and M. Majid, "Segmentation of glioma tumors in brain using deep convolutional neural network," *Neurocomputing*, vol. 282, pp. 248–261, Mar. 2018.

[32] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.

[33] H. Tang, C. Zhang, and X. Xie, "Automatic pulmonary lobe segmentation using deep learning," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1225–1228.

[34] H. Al Hajj, M. Lamard, K. Charriere, B. Cochener, and G. Quellec, "Surgical tool detection in cataract surgery videos through multi-image fusion inside a convolutional neural network," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2002–2005.

[35] B. Choi, K. Jo, S. Choi, and J. Choi, "Surgical-tools detection based on Convolutional Neural Network in laparoscopic robot-assisted surgery," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1756–1759.

[36] G. Padmavathi, P. Subashini, and A. Sumi, "Empirical evaluation of suitable segmentation algorithms for IR images," *Int. J. Comput. Sci. Issues*, vol. 7, no. 4, pp. 22–29, 2010. [Online]. Available: https://search.proquest.com/docview/755586660?pq-origsite=gscholar&fromopenview=true

[37] Y. Gao and J. A. Noble, "Detection and characterization of the fetal heartbeat in free-hand ultrasound sweeps with weakly-supervised two-streams convolutional networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 305–313, doi: 10.1007/978-3-319-66185-8_35.

[38] S. Rueda, S. Fathima, and C. L. Knight, "Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 797–813, Apr. 2014, doi: 10.1109/TMI.2013.2276943.

[39] E. Ricci, S. R. Bulò, C. Snoek, O. Lanz, S. Messelodi, and N. Sebe, *Image Analysis and Processing–ICIAP 2019: 20th International Conference, Trento, Italy, September 9–13, 2019, Proceedings*, vol. 11752. Cham, Switzerland: Switzerland: Springer, 2019.

[40] S. Afshari, A. BenTaieb, and G. Hamarneh, "Automatic localization of normal active organs in 3D PET scans," *Computerized Med. Imag. Graph.*, vol. 70, pp. 111–118, Dec. 2018.

[41] I. Young, J. Gerbrands, and L. van Vliet, "Fundamentals of image processing," Delft Univ. Technol., South Holland, The Netherlands, Tech. Rep., 2009.

[42] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019.

[43] P. Ammirato and A. C. Berg, "A mask-RCNN baseline for probabilistic object detection," 2019, *arXiv:1908.03621*. [Online]. Available: http://arxiv.org/abs/1908.03621

[44] Radiopaedia.org. *Radiopaedia*. Accessed: Mar. 7, 2020. [Online]. Available: https://radiopaedia.org/

[45] Z. Lin *et al.*, "Quality assessment of fetal head ultrasound images based on faster R-CNN," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*. Wiesbaden, Germany: Springer, 2018, pp. 38–46.

[46] Y. Ren, C. Zhu, and S. Xiao, "Object detection based on fast/faster RCNN employing fully convolutional architectures," *Math. Problems Eng.*, vol. 2018, Jan. 2018, Art. no. 3598316.

[47] J. W. Johnson, "Adapting mask-RCNN for automatic nucleus segmentation," 2018, *arXiv:1805.00500*. [Online]. Available: http://arxiv.org/abs/1805.00500

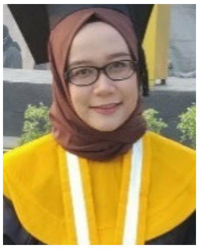
[48] E. Kopelowitz and G. Engelhard, "Lung nodules detection and segmentation using 3D mask-RCNN," 2019, *arXiv:1907.07676*. [Online]. Available: http://arxiv.org/abs/1907.07676

[49] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing Mask R-CNN for detection and segmentation of oral diseases," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 2197–2204.

SITI NURMAINI (Member, IEEE) received the master's degree in control system from the Institut Teknologi Bandung (ITB), Indonesia, in 1998, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotics.

MUHAMMAD NAUFAL RACHMATULLAH is currently a Research Assistant with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include medical imaging, biomedical signal and engineering, deep learning, and machine learning.

ADE IRIANI SAPITRI is currently pursuing the master's degree with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include medical imaging, deep learning, and machine learning.



ANNISA DARMAWAHYUNI is currently a Research Assistant with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include biomedical signal and engineering, deep learning, and machine learning.

ADITHIA JOVANDY is currently pursuing the bachelor's degree with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. He is also a member with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya. His research interests include medical imaging, deep learning, and machine learning.

FIRDAUS FIRDAUS is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include text processing, deep learning, and machine learning.

BAMBANG TUTUKO is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include robotics, deep learning, and machine learning.

ROSSI PASSARELLA is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include image forensic, deep learning, and machine learning.