Aorta Detection with Fetal Echocardiography Images Using Faster Regional Convolutional Neural Network (R-CNNs)

Ade Iriani Sapitri, Annisa Darmawahyuni

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya
*adeirianisapitri13@gmail.com

ABSTRACT

The fetal heart structure has an important role in analyzing the location of abnormalities in the heart. The aorta is one of the fetal heart structures, which has an essential part in exploring how the fetal heart is structured. To see the fetal heart structure can be seen with the help of an echocardiography tool in the form of ultrasound to see ultrasound images of the fetal heart. In ultrasound image data, detection is challenging because of its low image features, shadows, and contrast levels. So that is the first to do it yourself in one of the points of the culture in the culture in the aorta. The approach in this study uses deep learning in cases using Faster Regional Convolutional Neural Network (R-CNNs) with the R-CNNs mask method. The proposed approach has been applied to 151 ultrasound images of the fetal heart for the aortic region. The evaluation results were tested by evaluating metrics on the detection object with an mAP value of 83.71%.

Keywords: Detection, Aorta, Fetal Echocardiography, CNN, Faster RCNN.

1. INTRODUCTION

Echocardiography is an ultrasound image that is used routinely to visualize the heart chambers and valves. This can be used to diagnose abnormalities in the heart. In addition, fetal Echocardiography was introduced to assess fetal heart function, which has been done in previous studies [1]. The occurrence of cardiac anatomical abnormalities leading to a quantitative assessment of the heart's dimensions, shape, and function has proven helpful in diagnosing and monitoring fetuses with fetal conditions [2][3]. In addition, several cardiac parameters already exist shown to help predict perinatal problems and long-term cardiovascular outcomes. The aorta is a part of the fetal heart that has an essential role for the whole body. The location of the aorta can play a crucial role in further diagnosis in determining the occurrence of abnormalities and the position of the heart chamber in the fetus [4].

Some researchers have very often used deep learning to diagnose structurally [5]. Deep learning algorithms have also been widely used to analyze medical images such as classification, object detection, and segmentation [6][4]. Based on previous research, detection of fetal objects is still a limited case, so further research is needed to diagnose abnormalities in the heart by analyzing the results of the heart structure in the fetal heart [7][8]. Based on previous research, image detection has an important role. In contrast, the one that can support the image detection process from deep learning algorithms is CNN, which is the part of the neural network most commonly used for the image analysis process. Several previous studies have successfully carried out the segmentation and detection process with fairly good

results as has been done by He et al [9]. So this study proposes a detection process in the fetal heart aorta using the Faster R-CNNs method.

2. METHODOLOGY

The purpose of this study was to detect the fetal heart chamber in the aorta using the CNN method with the Faster R-CNNs architecture [10][11][12]. This study will see whether performance can provide good results from training data in the learning process.

2.1. DATASET

Data from abnormal patients detected defects and normal using a 4CH view of the fetal heart. The data is in the form of ultrasound videos as much as four videos. The video description used will be explained in table 1 below. All videos are used, and preprocessing will be carried out before testing the proposed method and model.

TABLE 1.
Raw Data Description

Deskripsi	Type of File	Length	Dimension
Abnormal 1	.mp4 (3.8 mb)	15s	976 x 726
Abnormal 2	.mp4 (636 kb)	7s	480 x 360
Abnormal 3	.mp4 (1.14 mb)	13s	640 x 360
Normal	.mp4 (331 kb)	5s	480 x 360

2.2 PREPROCESSING STEPS

Based on the video data obtained, the next step is to preprocess the data before it is carried out and tested on the proposed method framework. The data preprocessing stages will be explained in the flow chart described in Figure 1 below.

FIGURE 1. Preprocessing Steps

Based on Figure 1, the data stages are carried out in 4 steps. The first stage of the video to frame based on the patient's video will be converted into a frame using the python library using cv2.VideoCapture () so that the results of the framing process will be presented in Table 2 below.

TABLE 2.
Description Data Video to Frame

Description	Length	Dimension	Frame
Abnormal 1	15s	976 x 726	944
Abnormal 2	7s	480 x 360	210
Abnormal 3	13s	640 x 360	388
Normal	5s	480 x 360	149

The second stage performs the data filtering process. For the data filtering stage, the data used for the next step is filtered because each video used has a different length of time, resulting in different frames. This triggers the need for a data filtering process so that the data to be used at the next stage has a balance of data to be tested. The results of filtering the data that will be used are described in Table 3 below.

TABLE 3.
Filtering Data

Description	Dimension	Frame	Filter
Abnormal 1	976 x 726	944	172
Abnormal 2	480 x 360	210	125
Abnormal 3	640 x 360	388	156
Normal	480 x 360	149	149

Based on the results of filtering the data, the third stage is doing the resizing process. This is done because each dimension of the frame results obtained has a different size. So that this stage of the data process is resized on all images with dimensions of 976 x 726, then the resize results will be used to go to the next step of the label annotation stage. The fourth stage is carrying out the label annotation stage. This stage is carrying out the aortic labeling process using the python labelImg library. This label annotation process stores the converted information into data (.json) which contains label information in the image according to the needs used. This data labeling uses data that has been previously processed based on the description in Table 3 above. The following is the label annotation process which can be seen in Figure 2 below. This label will be used at a later stage to be studied by the model to predict the prediction results that will be tested. The final stage is to do data splitting, and this research is done by splitting data by dividing 80% training and 20% testing. The learning process will carry out training data, and data testing will be tested.
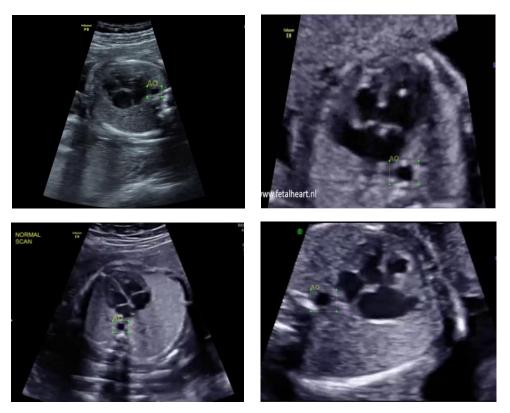
FIGURE 2. Annotation Label

2.3 MODEL ARCHITECTURE

The method framework proposed in this study uses the FR-CNNs method. FR-CNNs performs well even though the processing speed is lower than YOLO and SSD [13][14]. The FR-CNNs is high performance and does not specify the image size of the input in handling large image images. The main idea of our method is based on Faster R-CNNs, in which the four modules of feature extraction, proposal generation, RoI Pooling, classification, and regression are organically combined to form an end-to-end object detection system. After the great success of CNN [15][16] in the object detection also achieves remarkable result by the region-based CNN (RCNNs). The faster R-CNNs comes out, combined with the Regional Proposal Network (RPN) with Fast R-CNNs. As shown in Figure 3, training is the RPN is the Fast R-CNNs. The advantage of Faster R-CNNs is to extract proposal region by RPN, which input an image and output a series of bounding box proposals with object scores. To share convolutional layers between RPN and Fast R-CNNs model, a 4-step training algorithm was developed in step 1, training RPN independently; step 2, training a separate detection network based on Fast R-CNNs using the proposed regions generated in step 1 and step 3, employing the step 2 network to initialize RPN. Only the layers unique to RPN were fine-tuned during training and sharing convolutional layers between the two networks. Step-4, keeping the convolutional layers fixed, only the total connected layers for Fast R-CNNs was fine-tuned.
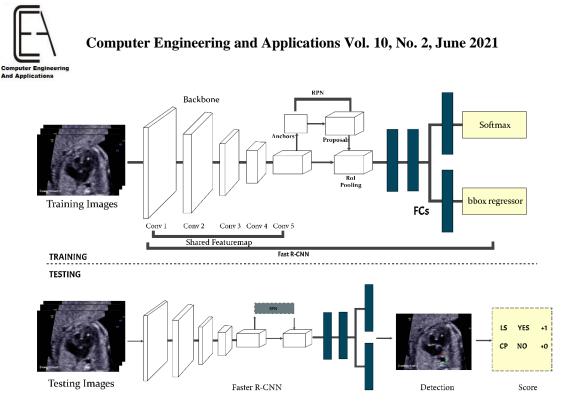
FIGURE 3. Architecture Faster **R-CNNs**

The input images are represented as *Height* x *Width* x *Depth*, passed through a pre-trained CNN until an intermediate layer, ending up with a convolutional feature map. We use this as a feature extractor for the next part. Next, we used a Regional Proposal Network. Using the CNN computed features it is used to find a predefined number of regions (bounding box) as an object. We used Faster R-CNNs implementation provided by Tensorflow Object Detection, a deep learning framework with python language. The pre-trained model had been trained on Common Objects in Context (COCO) [17][18]. We applied a technique called fine-tuning that takes a pre-trained model. Fine-tuning is an essential technique for training the networks. Several studies have reported the effectiveness of the fine-tuning of medical images. In this research, for all layers except the fully connected layer. For each test image, after the Non-Max-Suppression (NMS) processing, 2k regions of positive samples were generated by RPN. Then we selected 300 proposal regions of the highest scores as inputs of Fast R-CNNs, and marks for each of regions were given by the classification branch of Fast R-CNNs. Meanwhile, Regression branch precisely localization the bounding box of object. We perform a NMS operation with the threshold 0.3 on Fast R-CNNs results. We determined the left region whose score greater than 0.6 as the final bounding box of aorta.

2.4 PERFORMANCE METRIC

To determine the best network for aorta extraction with hyperparameters were created with the following characteristics: Input size, number of pooling layers, kernel size, and number of filters of the first convolution. To evaluate the performance of each network, we used mean Average Precision (mAP) to detect the aorta. The mean Average Precision (mAP) and intersection over union (IoU), also

known as the Jaccard index, were employed to evaluate the performance and accuracy of the CNN models in the detection of the aorta region [17]. To assess the performance of our model, we mainly use metrics from standard object detection and segmentation evaluation: AP (average precision) [13]. The definitions are as follows:

$$AP = \int_0^1 P(R)d(R) \tag{1}$$

where p represents precision, which indicates the ratio of the true positives to all predicted positives, R represents recall, which means the percentage of the true positives to all essential elements [14][19]. The mAP metric is used to evaluate both classification and segmentation for all object classes. The mAP is presented in (2) and defined as:

$$mAP = \sum_{i-1}^N \frac{AP_i}{N} \tag{2}$$

where AP_i is the AP in the i class, and n is the total number of courses being evaluated. mAP is a standard metric used for assessing, at the same time, the performance of object detection networks in detection and classification tasks [6]. It is based on the Intersection over Union (IoU) value, which represents the ratio between overlapping area and the area of the union of the predicted box B_p and the ground truth box B_g:

$$IoU = \frac{area(B_p \cap B_g)}{area(B_d \cup B_g)} \tag{3}$$

Based on an IoU-threshold (IoU), it is possible to define true positive, false positive, false negative, and precision and recall.

2.5 TRAINING PROCESS

A transfer learning approach was used to pre-train the model on the MS COCO dataset. Input images for training and testing processes were maintained to their original image without any preprocessing step. All the experiments were performed on TensorFlow 1.14.0 using AMD Ryzen CPU with 6 GB of RAM and NVidia GeForce GTX 1060 6GB.

3. RESULT AND DISCUSSION

The first TensorFlow model used for training was Faster RCNN with VGG16 and ResNet50. Figure 5 shows the loss versus several training steps using TensorBoard, a visualization tool for machine learning. Figure 4 show the sample test image using Faster RCNN with VGG16, and the mode successfully detected aorta
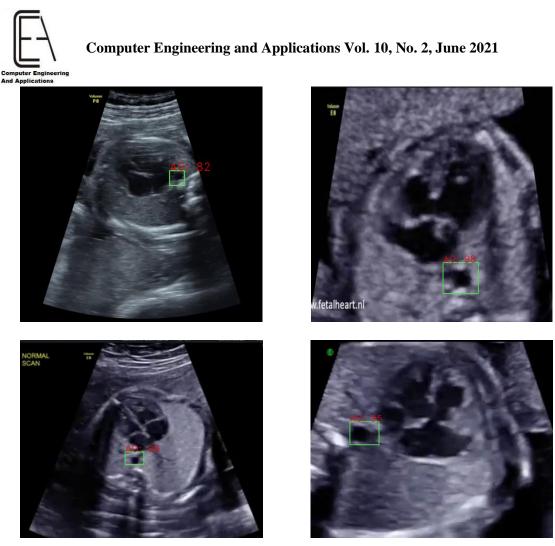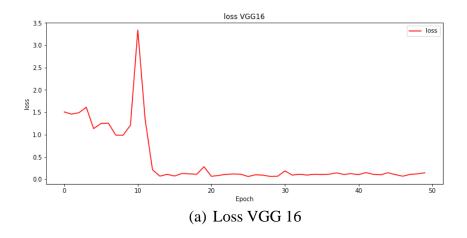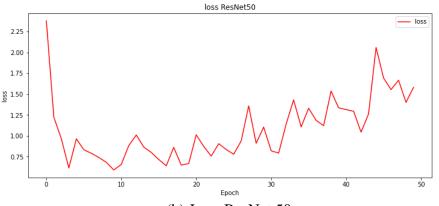
FIGURE 4. Result Predic Aorta

Table 4 shows the loss comparison of two TensorFlow models during the faster RCNN with VGG 16 and ResNet50. Based on the prediction of the Faster RCNN model that has been carried out on the VGG 16 and ResNet 50 backbone, the results of each loss can be seen in table 4 and the loss graph Figure 5 below. The loss value in VGG 16 is lower than the loss value in the ResNet 50 model. The prediction results in detecting the aorta are better in the VGG 16 model, where the aorta is seen significantly.

(a) Loss VGG 16

(b) Loss ResNet 50
FIGURE 5. Grafik Loss model architecture VGG 16 and ResNet 50

TABLE 4.
Loss Comparison FRCNN VGG 16 and ResNet 50

Description	VGG 16	ResNet 50
Loss RPN Classifier	1.31731	0.04606
Loss RPN Regression	0.00054	0.03204
Loss Detector Classifier	0.13646	1.06688
Loss Detector Regression	0.12836	0.29588

Table 5 shows the mean average precision (mAP) of two models using the test set (*images). The mAP was based on detection evaluation metrics used by COCO. The mAP IoU thresholds were used, such as 0.50 (PASCAL VOC metric)[20]. Table 5 shows the model comparisons of the two model predictions mAP. The model was tested in a sample image, from the comparison VGG16 to high from ResNet50.

TABLE 5.
mean Average Precision (mAP) two models

Model	mAP
VGG 16	**83.71**
ResNet 50	71.94

4. CONCLUSION

The detection of the aorta has a significant role in the next stage in analyzing research in the medical field. However, it is complicated to accurately find the aorta location, especially in the fetal heart, because it does not structure the movement of the fetus in the womb. In this study, detecting the aorta in the fetal heart based on the proposed R-CNNs Faster framework to be applied in medical image processing was analyzed to detect the aorta. All experiments were carried out with several methodologies based on the framework, and the results were good enough that the detected aorta gave 83.71% success in carrying out the detection process.

ACKNOWLEDGEMENTS

REFERENCES

[1] E. A. Bhalerao and N. Rajpoot, *Medical Image Understanding and Analysis 2010*, no. July. 2010.

[2] X. Wang, S. Han, Y. Chen, D. Gao, and N. Vasconcelos, "Volumetric Attention for 3D Medical Image Segmentation and Detection," *arXiv*, p. 12, 2020.

[3] G. Mcleod *et al.*, "Echocardiography in Congenital Heart Disease.," *Prog. Cardiovasc. Dis.*, vol. 61, no. 5–6, pp. 468–475, 2018, doi: 10.1016/j.pcad.2018.11.004.

[4] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-Octob, no. October, pp. 2023–2027, 2019, doi: 10.1109/TENCON.2018.8650517.

[5] X. Zeng, L. Wen, B. Liu, and X. Qi, "Deep learning for ultrasound image caption generation based on object detection," *Neurocomputing*, vol. 392, no. xxxx, pp. 132–141, 2020, doi: 10.1016/j.neucom.2018.11.114.

[6] S. Nurmaini *et al.*, "Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation," *IEEE Access*, vol. 8, pp. 196160–196174, 2020, doi: 10.1109/ACCESS.2020.3034367.

[7] F. Shenavarmasouleh and H. Arabnia, *DRDr: Automatic Masking of Exudates and Microaneurysms Caused By Diabetic Retinopathy Using Mask R-CNN and Transfer Learning*. 2020.

[8] S. Gandhi, W. Mosleh, J. Shen, and C.-M. Chow, "Automation, machine learning, and artificial intelligence in echocardiography: A brave new world," *Echocardiography*, vol. 35, Jul. 2018, doi: 10.1111/echo.14086.

[9] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, no. October, pp. 179424–179436, 2020, doi: 10.1109/ACCESS.2020.3026658.

[10] Y. Kawazoe *et al.*, "Faster R-CNN-based glomerular detection in multistained human whole slide images," *J. Imaging*, vol. 4, no. 7, 2018, doi: 10.3390/jimaging4070091.

[11] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASL4GUP 2017, Biometrics Wild, Bwild 2017, Heteroge*, pp. 650–657, 2017, doi: 10.1109/FG.2017.82.

[12] A. Cvpr and P. Id, "View Classification and Object Detection in Cardiac Ultrasound to Localize Valves via Deep Learning," pp. 1–9, 2019.

[13] R. Li *et al.*, "Automatic localization and identification of mitochondria in

cellular electron cryo-tomography using faster-RCNN," *BMC Bioinformatics*, vol. 20, no. Suppl 3, 2019, doi: 10.1186/s12859-019-2650-7.

[14] R. Rosati, L. Romeo, S. Silvestri, F. Marcheggiani, L. Tiano, and E. Frontoni, "Faster R-CNN approach for detection and quantification of DNA damage in comet assay images," *Comput. Biol. Med.*, vol. 123, no. July, 2020, doi: 10.1016/j.compbiomed.2020.103912.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceeding NIPS'12 Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 12, p. 04015009, 2015, doi: 10.1061/(ASCE)GT.1943-5606.0001284.

[16] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2, 2013.

[17] J. Zhang, H. Hu, S. Chen, Y. Huang, and Q. Guan, "Cancer Cells Detection in Phase-Contrast Microscopy Images Based on Faster R-CNN," *Proc. - 2016 9th Int. Symp. Comput. Intell. Des. Isc. 2016*, vol. 1, no. i, pp. 363–367, 2016, doi: 10.1109/ISCID.2016.1090.

[18] E. Avşar and K. Salçin, "Detection and classification of brain tumours from MRI images using faster R-CNN," *Teh. Glas.*, vol. 13, no. 4, pp. 337–342, 2019, doi: 10.31803/tg-20190712095507.

[19] S. A. Yoon, W. H. Hong, and H. J. Cho, "Congenital heart disease diagnosed with echocardiogram in newborns with asymptomatic cardiac murmurs: a systematic review," *BMC Pediatr.*, vol. 20, no. 1, p. 322, 2020, doi: 10.1186/s12887-020-02212-8.

[20] Haerbin gong ye da xue, Aomen da xue, M. IEEE Systems, M. IEEE Systems, B. International Federation of Automatic Control. Technical Committee on Economic, and Institute of Electrical and Electronics Engineers, *Faster R-CNN Based Microscopic Cell Detection*. .