Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 2684~2696
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9244

Residual pixel-wise semantic segmentation for assessing enlarged fetal heart: a preliminary study

Muhammad Taufik Roseno[1,4], Siti Nurmaini[2], Dian Palupi Rini[3], Tommy Saputra[5], Putri Mirani[6,7], Muhammad Naufal Rachmatullah[2], Annisa Darmawahyuni[2], Ade Iriani Sapitri[2], Hadi Syaputra[4]
[1]Doctoral Program in Engineering Science, Universitas Sriwijaya, Indonesia
[2]Intelligent System Research Group, Universitas Sriwijaya, Indonesia
[3]Department of Informatic Engineering, Universitas Sriwijaya, Indonesia
[4]Department of Computer Science, Universitas Sumatera Selatan, Indonesia
[5]Department of Information System, Institut Teknologi dan Sains Nahdlatul Ulama Sriwijaya, Indonesia
[6]Department of Obstetrics and Gynecology, Dr. Mohammad Hoesin General Hospital, Palembang, Indonesia
[7]Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

Article Info

Article history:

Received Sep 2, 2024
Revised Apr 5, 2025
Accepted May 27, 2025

Keywords:

Cardiomegaly
Fetal heart
Medical imaging
Residual network
Semantic segmentation

ABSTRACT

The four-chamber view is a crucial scan plane routinely employed in both second-trimester perinatal screening and fetal echocardiographic examinations. Sonographers typically measure biometrics in this plane, such as the cardiothoracic ratio (CTR) and heart axis, to diagnose fetal heart anomalies. However, due to the echocardiographic artifacts, the assessment not only suffers from low efficiency but also inconsistent results depending on the operators' skills. This study proposes a residual pixel-wise semantic segmentation, which segmented the fetal heart and thoracic contours in a 4-chamber view for assessing an enlarged fetal heart condition. The accuracy of intersection-over-union (IoU) and dice coefficient similarity (DCS) is used for model validation to further regulate the evaluation procedure. We use 1174 US images, comprising about 560 enlarged heart images, and about 614 normal heart images. Out of these data, 248 images are used for unseen data, and the remaining for training/validation processes. The performance of the proposed model, when tested on unseen data, achieved satisfactory results with 97.71% accuracy, 90.36% IoU, and 94.93% DCS. These metrics collectively demonstrate the satisfactory performance of the proposed model compared to existing segmentation models. The outcomes underscore that the proposed model establishes a state-of-the-art standard for enlarged fetal heart detection.

Corresponding Author:

Siti Nurmaini
Intelligent System Research Group, Universitas Sriwijaya
Palembang, South Sumatera, Indonesia
Email: siti_nurmaini@unsri.ac.id

1. INTRODUCTION

The routine anatomy scan, typically conducted between 18 and 22 weeks of pregnancy using ultrasound (US), is a standard component of prenatal care [1], [2]. This scan plays a crucial role in identifying fetal heart structural abnormalities. Enlarged heart or cardiomegaly is a sign of a cardiovascular disease condition in which the anatomical size of the heart is larger than the normal heart size [3], [4]. The cardiothoracic ratio (CTR) in fetuses is an important metric used to assess heart size relative to the chest cavity during prenatal US. The normal CTR in a fetus typically ranges between 0.45 and 0.50, meaning that the transverse diameter of the heart should occupy 45% to 50% of the transverse diameter of the thorax [5].

This ratio is relatively consistent throughout gestation, although minor variations may occur depending on the gestational age [5]. A CTR greater than 0.50 is generally considered abnormal and indicative of fetal cardiomegaly. This condition may result from a variety of causes, including structural heart defects, genetic syndromes, fetal anemia, or maternal conditions such as diabetes or infections [3], [5]. The severity and implications of fetal cardiomegaly depend on its underlying cause, which can range from benign to life-threatening. Early detection through detailed US images and fetal echocardiography is crucial for diagnosing the specific cause and guiding management [5]–[11].

Clinicians can use US medical pictures to make critical judgments about diagnosis and treatment [12]–[15]. The US photos are manually interpreted as part of the decision-making process. The field of medical image processing is committed to comprehending and improving the clinical interpretation process in order to further enhance decision-making and hence improve patient care [16]. Deep learning (DL) methods for medical image analysis have been developed by researchers in the last few years as artificial intelligence (AI) has advanced [15]–[17]. Using US medical pictures, categorization of the entire image, region-based object recognition, and semantic segmentation have been proposed to support the clinical decision, specifically on cardiomegaly condition [4], [18]–[21]. Cardiomegaly illness classification using a transfer learning method that observes many ConvNet designs [19]. Another study used gradient-weighted class activation mapping (Grad-CAM) for model interpretability and the routing-by-agreement method to categorize thoracic illnesses, including cardiomegaly [18]. Using the DenseNet model as a baseline and the U-Net model for picture segmentation, a method for detecting cardiomegaly is provided. As a diagnostic measure, they compute the CTR using U-Net [20]. Based on an edge detection technique and a gray-level histogram for feature analysis, a computerized system has been suggested to calculate CTR [5]. For the categorization and localization of cardiomegaly on chest X-ray images, a convolutional attention mapping deep neural network has been proposed [4].

All of the earlier DL-based techniques, however, were created expressly to identify cardiomegaly in adult cardiac structures. Cardiomegaly in fetuses with US pictures has not been automatically detected using a DL model. Fetal cardiothoracic (C/T) size assessment is used to identify enlarged hearts [21]. This procedure calculates the circumference ratio, which is a useful metric for evaluating the fetal heart in the thoracic/chest wall and 4-chamber view [16]. Only US samples of normal hearts were used in the prior study, which effectively investigated the fetal heart using fetal C/T employing DL techniques [7]. The largest obstacle is the scarcity of US heart pictures with cardiomegaly situations, so we need to select a robust segmentation model independent of data volume [22], [23]. The small size of the fetal heart, its dynamic position, and the noise produced by US instruments during the test are further obstacles to creating a reliable model [15], [24]. Raising awareness is crucial for identifying cardiomegaly in fetuses so that the newborns can be treated quickly and well. With the ultimate goal of enabling the prompt detection and efficient treatment of this illness, we are innovative in this study by suggesting the residual pixel-wise semantic segmentation for fetal cardiomegaly assessment. In order to ensure that newborns with enlarged hearts receive timely treatment and comprehensive care for a healthy start in life, we work to improve their prospects for the future.

2. MATERIALS AND METHOD

This section presents the model training setup and evaluation techniques, as well as a description of the US dataset used in this work. The data pre-processing procedures are also covered in this section, and the suggested approaches are finally covered.

2.1. Data collection

In this study, 12 fetal heart US videos with cardiomegaly and normal were used. These videos were obtained from Bunda Hospital, Palembang, Indonesia. The data available is not extensive due to the hard for detecting such cardiomegaly abnormalities in fetus. The US video recording is then saved in the digital imaging and communications in medicine (DICOM) format, with a duration ranging from 15 to 40 seconds, and a size ranging from 2 to 50 kilobytes. All examinations were performed with USG GE Voluson V8 for measuring fetal heart size dividing heart area (HA) by chest area (CA). The sample of US videos as visualized in Figure 1(a) (enlarged heart) and Figure 1(b) (normal).

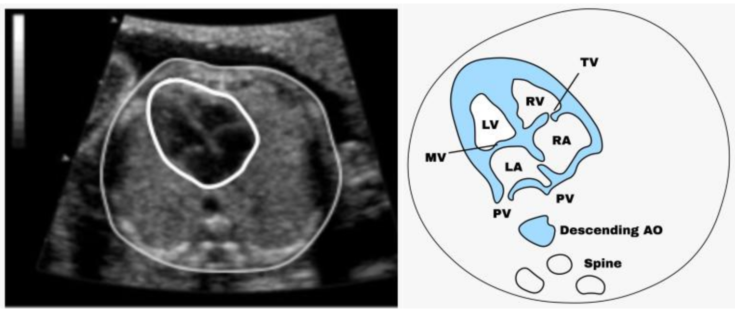
2.2. Pre-processing and ground truth

The US video is converted frame by frame using a Python software to produce two-dimensional image. The library feature is used to capture multiple frames with the cap.get(propld) method. All images produced amount to 1174, with cardiomegaly comprising about 560 images, and normal images about 614. Out of these data, 248 images are used for unseen data, and the remaining for training/validation processes. Stratified random sampling was used for training and testing data with a size of 970×672 pixels for all frames. The next stage is to performs the process of labelling the fetal heart region against the thoracic cavity to determine the occurrence of cardiomegaly and normal (Figures 2(a)-(c)).

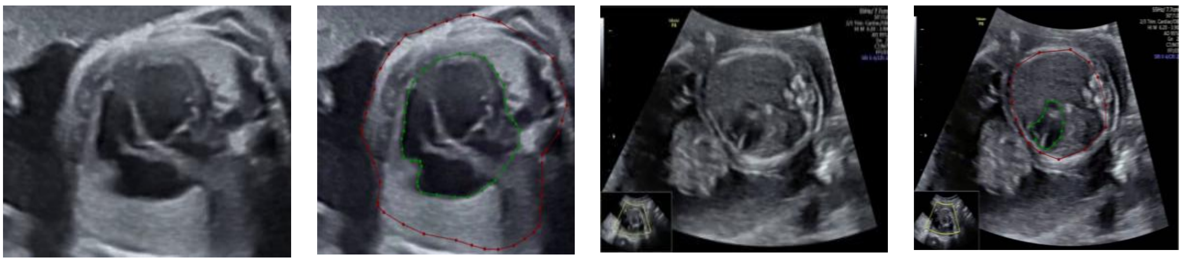
(a)

(b)

Figure 1. The sample of fetal heart US images; (a) enlarged heart and (b) normal



(a)

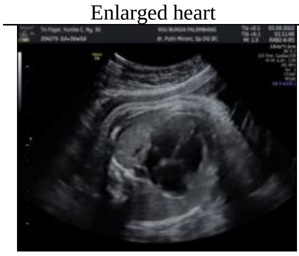
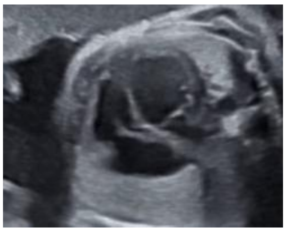
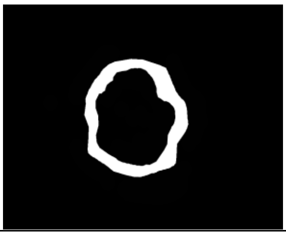
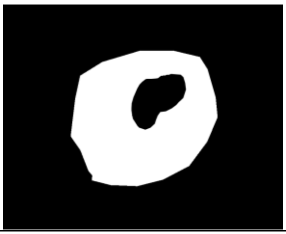


(b) (c)

Figure 2. Sample of fetal heart annotation; (a) references fetal (C/T) to assess fetal heart size on 4-chamber view, (b) enlarged heart, and (c) normal

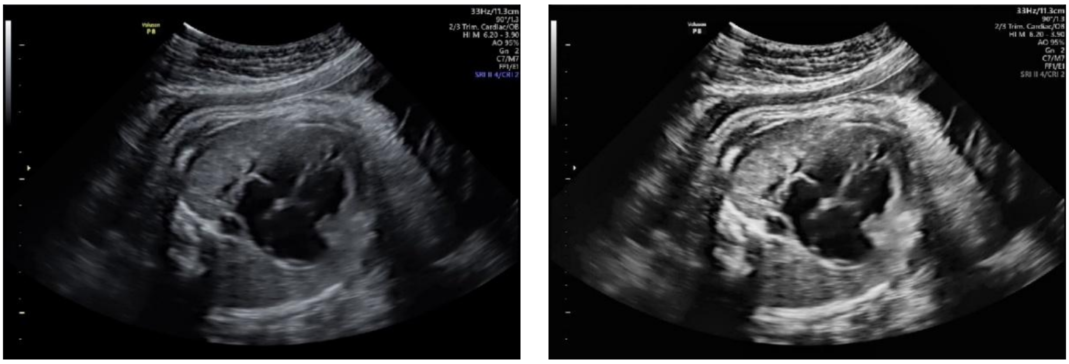
The label consisted of a fetal heart and a fetal thoracic cavity. The annotation label production process generates ground truth data. This involves utilizing the Python LabelMe library to transform images into semantic segmentation. The output of labels was saved in image thresholding, and it converted into data (.json). The resulting ground truth data is stored in the form of (.png) format. The annotation process was assisted and validated by a fetomaternal from Bunda Hospital, Indonesia with above 15 years experience. Actually, fetal cardiothoracic (C/T) circumference ratio is used to assess fetal heart size by fetometernal. In this study, we estimate the size of the fetal heart in relation to the thoracic wall. If the fetal heart experiences cardiomegaly, the heart size will increase and closer to the thoracic wall. In the Table 1 visualized the annotation for the ground truth data, white color in the back ground denotes the thoracic wall, and the black color in front ground is a fetal heart on 4-chamber view.

Table 1. Raw image and ground truth

Enlarged heart	Ground truth	Normal	Ground truth
			
			
			
			
			
			

2.3. Image enhancement

An enhancement process is needed to prevent low quality image, due to noise in fetal heart US images. It conducted using contrast-limited adaptive histogram equalization (CLAHE) [20]. Such technique has contrast amplification around certain pixel values based on the slope of the transformation axis. Such method to limit the amplification by photographing the histogram at a predetermined value before calculating the cumulative probability. The sample US image after enhancement as depicted in Figures 3(a) before and Figure 3(b) after enhancement.



(a) (b)

Figure 3. Sample of enhancement result; (a) before and (b) after

2.4. Residual pixel-wise semantic segmentation

In this paper, the pixel-wise segmentation is proposed using residual-U-Net (Res-U-Net) architecture for image segmentation tasks, due to it robust encoder-decoder network, faster processing, and high performance [25]. It is particularly popular in the field of medical image segmentation task. The architecture consists of a contracting path, a bottleneck, and an expansive path. This design allows U-Net to effectively segment fetal cardiomegaly objects in images. To address the vanishing gradient and over fitting, the residual network strategy is created connecting the input layer to its output layers and facilitating the training of very deep networks (Figure 4). It is hoped, such modification can enhance the performance of the original U-Net, particularly when dealing with deeper architectures where training might become challenging due to small size of fetal heart challenging object.

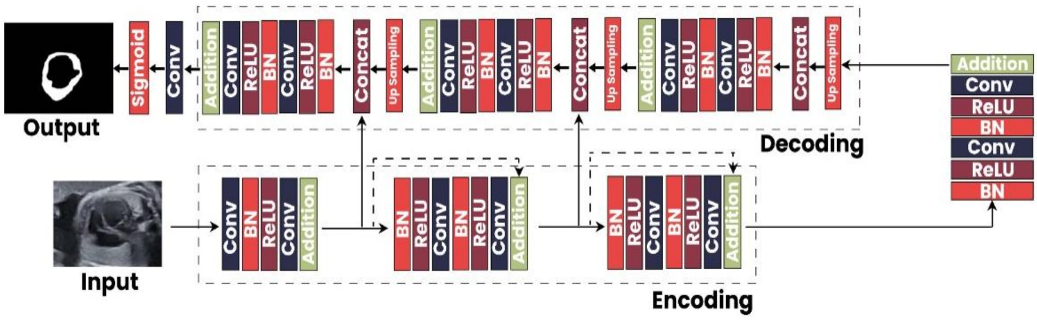


Figure 4. Residual-pixel-wise semantic segmentation

The initial phase involves inputting the original image into the encoder path. Subsequently, the image undergoes processing within the residual block across the first to fourth. Each residual block comprises a set of layers, including batch normalization, rectified linear units (ReLU) activation, and a weight (W) layer, all stacked together [17]. The residual block consists of a series of stacked residual units, and each block is represented as (1) and (2):

$$y_l = h(x_l) + \mathcal{F}(x_l, W_l), \quad (1)$$

$$x_{l+1} = f(y_l), \quad (2)$$

where the input is x_l and the output is x_{l+1} of the l-th residual unit, $\mathcal{F}(.)$ denotes the residual function, $f(y_l)$ is the activation function, and $h(x_l)$ is an identity mapping function.

To minimize the loss between the predicted segmentation and the ground truth segmentation for a set of training images $\{I_i, s_i\}$, the parameter W should be calculated. The mean squared error (MSE) is chosen as the loss function.

$$\mathcal{L}(W) = \frac{1}{n}\sum_{i=1}^{N}\left\|Net(I_{i;}W) - s_i\right\|^2 \quad (3)$$

Training on large datasets with batch processing can be slow. To address this limitation, stochastic gradient descent (SGD) optimization is employed [25]. In contrast to batch gradient descent, which computes the gradient from the entire batch of N training samples, SGD processes individual samples randomly. Additionally, pixel-wise cross entropy is chosen as the loss function for optimizing the model. The processed data then advances through the decoder path, involving a sigmoid operation. In the final convolution layer, data from both the encoder and decoder stages of each block are processed [26]. The proposed network structure as depicted in Table 2.

Table 2. The network structure of proposed Res-U-Net

	Unit level	Conv layer	Filter	Stride	Output size
Input					224×224×3
Encoding	Level 1	Conv 1	3×3/64	1	224×224×64
		Conv 2	3×3/64	1	224×224×64
	Level 2	Conv 3	3×3/128	2	112×112×128
		Conv 4	3×3/128	1	112×112×128
Bridge	Level 3	Conv 5	3×3/512	2	28×28×512
		Conv 6	3×3/512	1	28×28×512
Decoding	Level 4	Conv 7	3×3/128	1	112×112×128
		Conv 8	3×3/128	1	112×112×128
	Level 5	Conv 9	3×3/64	1	224×224×64
		Conv 10	3×3/64	1	224×224×64
Output		Conv 11	1×1	1	224×224×1

A learning rate of 0.001 and 150 epochs were used in the experiment. The suggested model was put into practice with the PyTorch 1.7.1 library and trained on a computer that had the following specs: an Ubuntu 18.04.5 LTS operating system, an NVIDIA Corporation GV102 (rev a1) GeForce 2080 RTX Ti, an Intel Core i9-9920X CPU processor operating at 3.50 GHz, and 490191 MB of RAM.

2.5. Model evaluation

Pixel-wise semantic segmentation algorithms are evaluated using two points: pixel accuracy and intersection over union (IoU) [15], [27]. In essence, IoU—also referred to as the Jaccard index—is a technique for calculating the percentage overlap between the target mask and the anticipated result. Pixels found in both the ground truth mask and the predicted mask make up the intersection (target ∩ prediction), whereas all pixels found in either the target or the predicted mask make up the intersection (target ∩ prediction). In (4) to (6) provides the mean IoU, dice coefficient similarity (DCS), and pixel accuracy:

$$\text{Pixel accuracy}=\frac{\sum_i n_i i}{\sum_i t_i} \tag{4}$$

$$\text{Mean accuracy}=\frac{1}{n_c i}\sum\frac{n_c i}{t_i} \tag{5}$$

$$\text{Mean IoU}=\frac{1}{n_c i}\sum\frac{n_c i}{t_i+\sum j n_{ji}-n_{ii}} \tag{6}$$

where n_{ij} be the number of pixels of class I predicted to belong to class j, where there are $n_c i$ different classes, and $t_i=\sum_j nij$ be the total number of pixels of class i. For the DCS is used as (7):

$$D=\frac{2\sum_i^N pi\, gi}{\sum_i^N pi^2+\sum_i^N gi^2} \tag{7}$$

3. RESULTS AND DISCUSSION

In this study, we have designed a DL model to accurately segment the fetal heart and thoracic cavity. To conduct a comprehensive comparative analysis, our proposed model is built on the U-Net architecture. In order to enhance IoU performance, we have incorporated a residual network for deep network processing, addressing the challenge of accurately delineating the smaller structures within the fetal heart. Prediction result of our proposed model is benchmarking with three architectures of pixel wise segmentation including, original (Ori)-U-Net, Res-U-Net, and attention (Att)-U-Net (Figure 5). It can be observed that Res-U-Net architecture outperform other architectures, with an accuracy of 97.71% and IoU of 84.11% for predicting the cardiomegaly abnormality. Meanwhile, for the normal condition, it achieves an accuracy of 95.01% and an IoU of 84.12%.

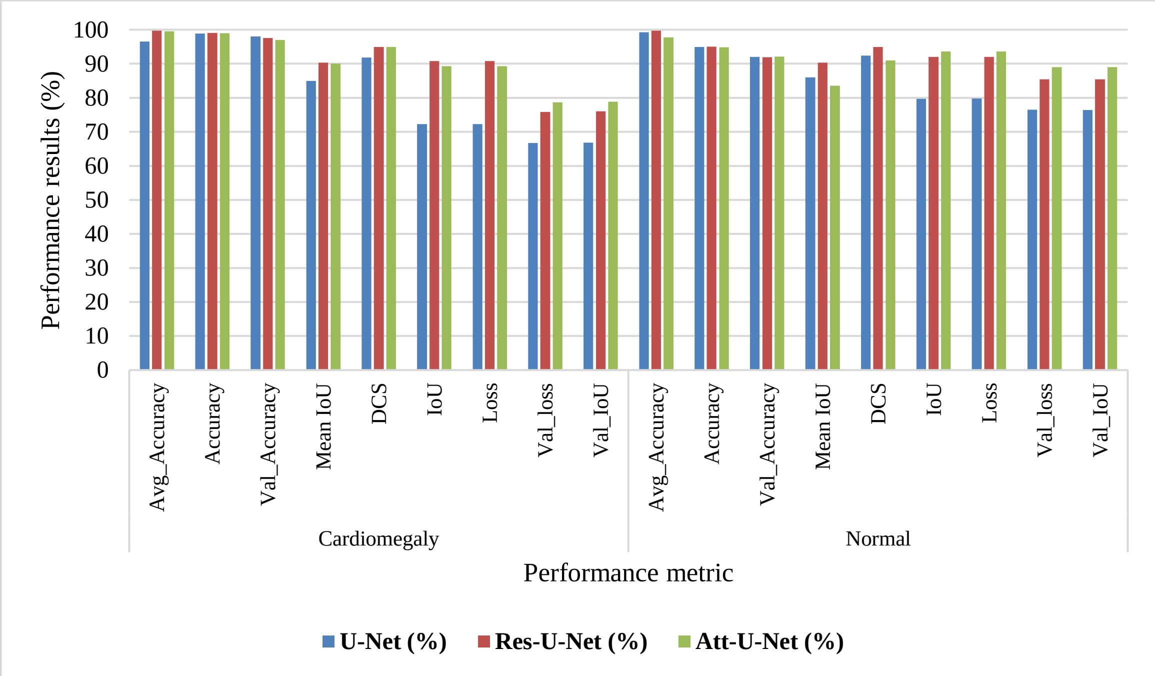


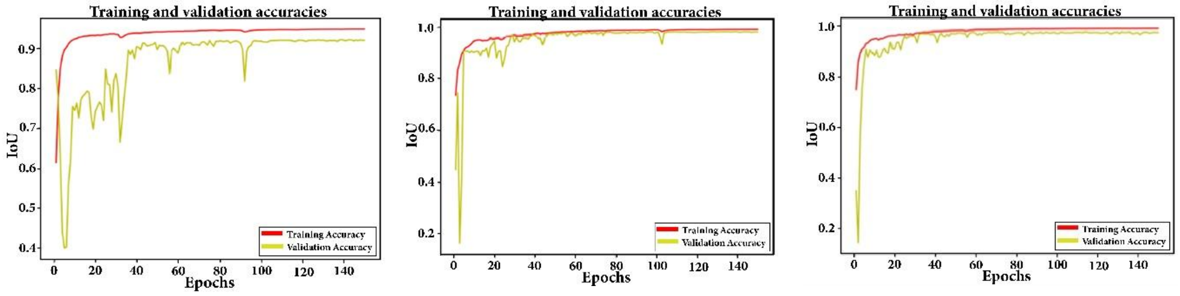
Figure 5. The proposed model performance for training/validation data

In this study, the fetal heart image segmentation task is what U-Net was originally developed for small number of training US images. Having a complex model with limited data can lead to overfitting, where the model performs well on the training data but fails to generalize to new, unseen data [23]. However, it can be observed from the results that the two architectures, Res-U-Net and Att-U-Net, demonstrate nearly identical performance levels, as both models effectively train the network (Figure 5). Previous research indicates that the residual technique addresses degradation performance issues by introducing shortcut connections [25], while the attention technique eliminates irrelevant activations, aiding the network in achieving generalization [26]. Despite these similarities, the selection of the superior model necessitates testing on previously unseen data (Table 2). Upon validation testing, Res-U-Net attains an accuracy of 99.09%, outperforming Att-U-Net, which achieves 98.95%. Similarly, during testing on unseen data, Res-U-Net achieves an accuracy of 97.71%, surpassing Att-U-Net with its accuracy of 96.87%. This translates to a decrease of 1.38% for Res-U-Net and 2.08% for Att-U-Net. Additionally, in terms of IoU performance, Res-U-Net exhibits a marginal decrease of 0.1%, while Att-U-Net experiences a more substantial decrease of 5.44%. These findings suggest that, for a more robust model in recognizing patterns of cardiomegaly conditions in fetal heart, Res-U-Net is the preferred choice.

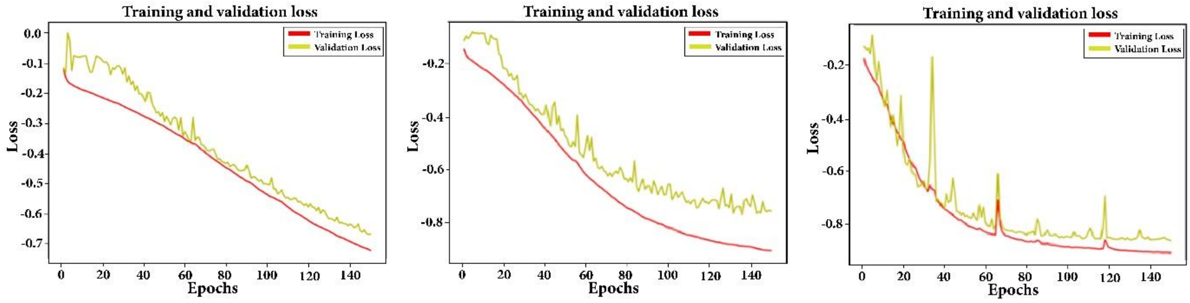
Table 2. The proposed model performance for unseen data

Fetal heart	Architecture	Accuracy (%)	IoU (%)	DCS (%)
Enlarged heart	U-Net	92.04	76.26	86.55
	Res-U-Net	97.71	90.36	94.93
	Att-U-Net	96.87	84.11	89.48
Normal	U-Net	93.58	73.34	87.51
	Res-U-Net	96.88	84.12	90.91
	Att-U-Net	92.05	76.27	89.78

The accuracy of IoU and loss curve in the segmentation tasks is typically a measure of dissimilarity between the predicted and ground truth segmentations. As training progresses, it observes that decreasing trend in the loss and increasing trend in the accuracy, indicating that the model is learning to generate more accurate segmentations. We compared two curves of accuracy and loss, when the model learns the US image. The accuracy serves as the most straightforward metric for assessing the efficacy of an image segmentation model [27]-[29]. As illustrated in Figures 6(a) accuracy and Figure 6(b) loss, the utilization of Res-U-Net demonstrates a smooth learning process, gradually converging to stability. In contrast, the two alternative models exhibit a transient response before attaining stability. This underscores the significance of monitoring the model's performance over time, treating it as a dynamic and ongoing process.



(a)



(b)

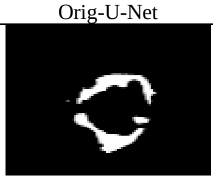
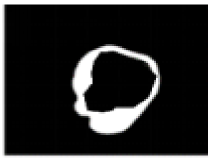
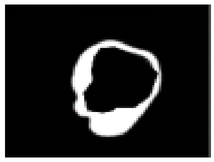
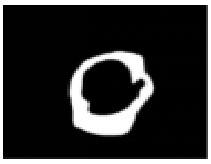
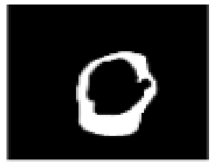
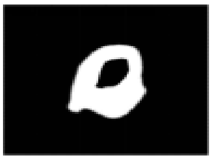
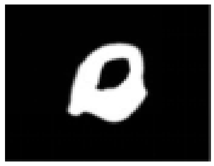
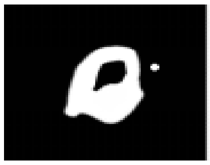
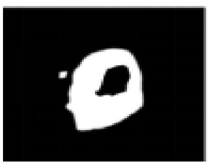
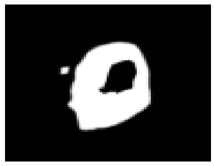
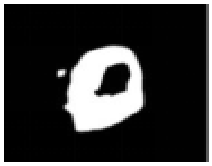
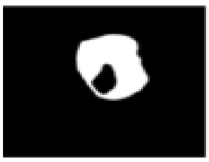
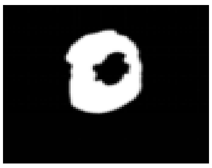
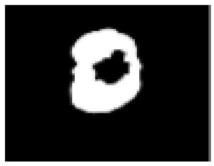
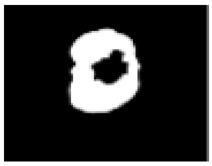
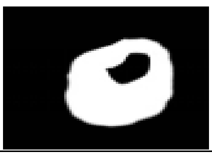
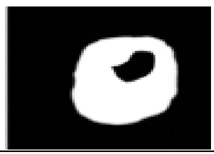
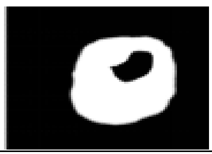
Figure 6. The accuracy curve of three architectures, Ori-U-Net, Att-U-Net, and Res-U-Net: (a) accuracy and (b) loss

One measure for evaluating fetal heart in 4-chamber view and thoracic/chest wall abnormalities is the fetal C/T circumference ratio [21]. Traditionally, throughout pregnancy, the C/T circumference ratio should be less than 0.5. On fetal US/echocardiography, it is the ratio of the fetal heart circumference to the thoracic circumference. An increased fetal heart is indicated when the C/T rises above 0.5. The Res-U-Net shows competence in segmenting the thoracic/chest wall and the heart under two particular situations. The fetal heart object's segmentation is precise and closely matches the ground truth (Table 3). In instances of cardiomegaly, where the C/T circumference ratio exceeds 0.5, and in normal conditions, where the C/T ratio is <0.5, the proposed model consistently and accurately segments the desired regions. Notably, the outcomes of this segmentation are consistently evaluated based on IoU and DCS performance metrics, with Res-U-Net showcasing superior performance compared to Orig-U-Net and Att-U-Net.

As this is the first study in this domain, lacking any precedent, we aimed to ensure the fairness of our results through a comparative analysis. To achieve this, we benchmarking our findings with a study that employed DL to measure fetal C/T, although solely utilizing normal US images [7], [30]. In their approach, they employed a deep layer aggregation network, supplementing it with a sigmoid activation function at the end of the ellipse axis heads for prediction. Their reported results yielded a DCS of 93.36%. In contrast, our proposed model, utilizing Res-U-Net, achieved a higher DCS of 94.94%. It's noteworthy that our research not only provides comparable results but also includes testing with unseen data, demonstrating a DCS result of approximately 90.91%.

To the best our knowledge, this is the first experiment with fetal cardiomegaly using residual pixel-wise semantic segmentation, and although there are indications of promising results, we acknowledge several limitations. These include the limited dataset, the need for further refinement of methods, and the improvement of performance during unseen tests. Nevertheless, despite these challenges, this study can be considered a pioneering effort that may serve as a foundation for supporting clinical practices in detecting abnormalities in fetal heart structures.

Table 3. The sample of fetal heart size prediction results with three architectures

Fetal heart	Origin	Orig-U-Net	Res-U-Net	Att-U-Net
Enlarged heart				
				
				
				
				
Normal				
				
				
				
				

4. CONCLUSION

The Res-U-Net architecture operation, which allows the classification of cardiomegaly from the fetal heart US images addressing both image-level classification and pixel-wise localization using only the image-level labels, was used in this study to investigate pixel-wise classification. We take advantage of some of the well-performing state-of-the-art Origin-U-Net and Att-U-Net to assess the empirical evidence of the suggested Res-U-Net with CLAHE model. Our experiment demonstrated that the suggested model outperforms the other models by a significant margin, looks to be the well-generalized model, and delivered impressive results when handling cardiomegaly class information and localization simultaneously. In actual practice, where comprehensive annotations are scarce, the suggested model can be a wonderful AI tool for radiologists to employ in medical diagnosis.

ACKNOWLEDGEMENTS

This work was supported by Universitas Sumatera Selatan and the Intelligent Systems Research Group (ISysRG) at Universitas Sriwijaya.

FUNDING INFORMATION

This research was funded by a Professional Grant from Universitas Sriwijaya Number 0016/UN9/SK.LP2M.PT/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Taufik Roseno	✓	✓			✓				✓					
Siti Nurmaini		✓							✓			✓		
Dian Palupi Rini					✓					✓		✓		
Tommy Saputra		✓	✓							✓				
Putri Mirani				✓	✓					✓				
Muhammad Naufal Rachmatullah		✓	✓							✓				
Annisa Darmawahyuni						✓				✓				
Ade Iriani Sapitri		✓	✓					✓		✓				
Hadi Syaputra		✓	✓							✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

There is no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

All methods were conducted according to relevant guidelines and regulations. All experimental protocols were approved by Bunda Hospital, Indonesia. Informed consent was obtained from all subjects and/or their legal guardians.

DATA AVAILABILITY

- The dataset used in this study is available at https://github.com/ISySRGg/Fetal_Enlarge/tree/main/dataset.
- The deep learning model and code for Computer-aided Assessment for Enlarged Fetal Heart are available at https://github.com/ISySRGg/ Fetal_Enlarge/tree/main/dataset or upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

REFERENCES

[1] O. Dudnikov, A. E. Quinton, and J. Alphonse, "The detection rate of first trimester ultrasound in the diagnosis of congenital heart defects: A narrative review," *Sonography*, vol. 8, no. 2, pp. 36–42, Jun. 2021, doi: 10.1002/sono.12253.

[2] E. Hildebrand, "Sonographic evaluation of fetal cardiac anatomy in the late first trimester," *Journal of Diagnostic Medical Sonography*, vol. 36, no. 2, pp. 140–150, 2020, doi: 10.1177/8756479319891127.

[3] S. S. Alghamdi, I. Abdelaziz, M. Albadri, S. Alyanbaawi, R. Aljondi, and A. Tajaldeen, "Study of cardiomegaly using chest x-ray," *Journal of Radiation Research and Applied Sciences*, vol. 13, no. 1, pp. 460–467, 2020, doi: 10.1080/16878507.2020.1756187.

[4] M. Innat, M. F. Hossain, K. Mader, and A. Z. Kouzani, "A convolutional attention mapping deep neural network for classification and localization of cardiomegaly on chest X-rays," *Scientific Reports*, vol. 13, no. 1, p. 6247, 2023, doi: 10.1038/s41598-023-32611-7.

[5] T. Ishida, S. Katsuragawa, K. Chida, H. MacMahon, and K. Doi, "Computer-aided diagnosis for detection of cardiomegaly in digital chest radiographs," in *Proceedings of the SPIE*, 2005, pp. 914–920, doi: 10.1117/12.595755.

[6] Y. Gong *et al.*, "Fetal Congenital Heart Disease Echocardiogram Screening Based on DGACNN: Adversarial One-Class Classification Combined with Video Transfer Learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1206–1222, 2020, doi: 10.1109/TMI.2019.2946059.

[7] J. Chen, Y. Zhang, J. Wang, X. Zhou, Y. He, and T. Zhang, "EllipseNet: anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 (MICCAI 2021)*, 2021, pp. 218–227, doi: 10.1007/978-3-030-87234-2_21.

[8] S. Qiao *et al.*, "RLDS: An explainable residual learning diagnosis system for fetal congenital heart disease," *Future Generation Computer Systems*, vol. 128, pp. 205–218, 2022, doi: 10.1016/j.future.2021.10.001.

[9] L. Xu *et al.*, "DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography," *Computerized Medical Imaging and Graphics*, vol. 80, 2020, doi: 10.1016/j.compmedimag.2019.101690.

[10] T. G. Day, B. Kainz, J. Hajnal, R. Razavi, and J. M. Simpson, "Artificial intelligence, fetal echocardiography, and congenital heart disease," *Prenatal Diagnosis*, vol. 41, no. 6, pp. 733–742, 2021, doi: 10.1002/pd.5892.

[11] T. Yang *et al.*, "Segmentation of Five Components in Four Chamber View of Fetal Echocardiography," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI),* Iowa City, IA, USA, 2020, pp. 1962-1965, doi: 10.1109/ISBI45749.2020.9098726.

[12] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[13] 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 Pediatrics*, vol. 20, no. 1, p. 322, 2020, doi: 10.1186/s12887-020-02212-8.

[14] M. T. Roseno and H. Syaputra, "Classification of Fetal Heart Based on Images Augmentation Using Convolutional Neural Network Method," *Journal of Information Systems and Informatics*, vol. 4, no. 2, pp. 252–265, 2022, doi: 10.51519/journalisi.v4i2.247.

[15] 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.

[16] M. J. Schuuring, I. Išgum, B. Cosyns, S. A. J. Chamuleau, and B. J. Bouma, "Routine echocardiography and artificial intelligence solutions," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 648877, 2021, doi: 10.3389/fcvm.2021.648877.

[17] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.

[18] Y. Shen and M. Gao, "Dynamic routing on deep neural network for thoracic disease classification and sensitive area localization," in *Machine Learning in Medical Imaging (MLMI 2018)*, 2018, pp. 389–397, doi: 10.1007/978-3-030-00919-9_45.

[19] S. Candemir, S. Rajaraman, G. Thoma, and S. Antani, "Deep learning for grading cardiomegaly severity in chest x-rays: an investigation," in *2018 IEEE Life Sciences Conference (LSC), Montreal, QC, Canada,* 2018, pp. 109-113, doi: 10.1109/LSC.2018.8572113.

[20] Q. Que *et al.*, "CardioXNet: automated detection for cardiomegaly based on deep learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA, 2018, pp. 612–615, doi: 10.1109/EMBC.2018.8512374.

[21] D. Paladini, S. K. Chita, and L. D. Allan, "Prenatal measurement of cardiothoracic ratio in evaluation of heart disease.," *Archives of Disease in Childhood*, vol. 65, no. 1, pp. 20–23, 1990, doi: 10.1136/adc.65.1_Spec_No.20.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[24] S. Nurmaini *et al.*, "An improved semantic segmentation with region proposal network for cardiac defect interpretation," *Neural Computing and Applications*, vol. 34, pp. 13937–13950, 2022, doi: 10.1007/s00521-022-07217-1.

[25] L. Xiang, Y. Li, W. Lin, Q. Wang, and D. Shen, "Unpaired deep cross-modality synthesis with fast training," in *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support*, 2018, pp. 155–164, doi: 10.1007/978-3-030-00889-5_18.

[26] S. Katakis, N. Barotsis, A. Kakotaritis, G. Economou, E. Panagiotopoulos, and G. Panayiotakis, "Automatic Extraction of Muscle Parameters with Attention UNet in Ultrasonography," *Sensors*, vol. 22, no. 14, p. 5230, 2022, doi: 10.3390/s22145230.

[27] S. Nurmaini *et al.*, "Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21238007.

[28] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[29] N. J. Bravo-valenzuela, A. B. Peixoto, and E. A. Júnior, "Prenatal diagnosis of congenital heart disease: A review of current knowledge," *Indian Heart Journal*, vol. 70, no. 1, pp. 150–164, 2018, doi: 10.1016/j.ihj.2017.12.005.

[30] J. Xu, E. A. Turk, P. E. Grant, P. Golland, and E. Adalsteinsson, "STRESS: Super-Resolution for Dynamic Fetal MRI Using Self-supervised Learning," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 (MICCAI 2021),* vol. 12907, pp. 197–206, 2021, doi: 10.1007/978-3-030-87234-2_19.

BIOGRAPHIES OF AUTHORS

Muhammad Taufik Roseno obtained a bachelor's degree in industrial engineering from Telkom Institute of Technology in 2001. Obtained a master's degree in Informatics Engineering from Bina Darma University, Indonesia in 2012. Currently a researcher of USS Artificial Intelligence Research Group at the Universitas Sumatera Selatan, South Sumatra, Palembang, Indonesia. His research interests include deep learning, machine learning, image processing, and software engineering. He can be contacted at email: mtroseno@gmail.com.

Siti Nurmaini 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 the Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic. She can be contacted at email: siti_nurmaini@unsri.ac.id.

Dian Palupi Rini is currently a lecturer at Department of Informatic Engineering, Faculty of Computer Science, Universitas Sriwijaya. Her research interest include machine learning and artificial intelligence. She can be contacted at email: dprini@unsri.ac.id.

Tommy Saputra received a bachelor's degree in Information Systems from Universitas Sriwijaya in 2018 and received a master's degree in Computer Science at Universitas Sriwijaya, Indonesia in 2024. Currently a lecturer at Information System Department, Institut Teknologi dan Sains Nahdlatul Ulama Sriwijaya, Indonesia. His research interests include deep learning, machine learning, image processing, and software engineering. He can be contacted at email: tommysaputra@uss.ac.id.

Putri Mirani received her medical doctor degree from Faculty of Medicine Universitas Sriwijaya in 2002, medical specialty training in Obstetrics and Gynecology from Universitas Sriwijaya in 2008 and obtained fellowship of Maternal-Fetal Medicine from Universitas Indonesia in 2014 and Doctoral degree in sciences and biomedics from Universitas Sriwijaya in 2024. She is currently a staff and lecturer at Department of Obstetrics and Gynecology, Dr. Mohammad Hoesin General Hospital/Faculty of Medicine Sriwijaya University, Palembang, South Sumatera, Indonesia. She has vast experience and knowledge in the area of high-risk pregnancies. Her research interest includes placenta accreta and prenatal ultrasound. She has conducted several research studies which have resulted in multiple recent publications. She can be contacted at email: putrimirani@fk.unsri.ac.id.

Muhammad Naufal Rachmatullah is currently a lecturer and researcher in the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes biomedica image processing, deep learning, and machine learning. He can be contacted at email: naufalrachmatullah@gmail.com.

Annisa Darmawahyuni is currently a lecturer and researcher in the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes biomedical signal/image processing, deep learning, and machine learning. She can be contacted at email: riset.annisadarmawahyuni@gmail.com or annisadarmawahyuni@unsri.ac.id.

Ade Iriani Sapitri is currently a lecturer and researcher in the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes biomedical image processing, deep learning, and machine learning. She can be contacted at email: adeirianisapitri13@gmail.com.

Hadi Syaputra received a bachelor's degree in computer engineering from Universitas Bina Darma in 2006. Obtained a master's degree in Informatics Engineering at Universitas Bina Darma, Indonesia in 2011. Currently a researcher of USS Artificial Intelligence Research Group at the Universitas Sumatera Selatan, South Sumatra, Palembang, Indonesia. His research interests include deep learning, machine learning, image processing, and software engineering. He can be contacted at email: hadisyaputra@uss.ac.id.