Segmentation atrioventricular septal defect by using convolutional neural networks based on U-NET architecture

Ade Iriani Sapitri¹, Siti Nurmaini², Sukemi³, M. Naufal Rachmatullah⁴, Annisa Darmawahyuni⁵
¹Department of Informatic, Faculty of Computer Science, Universitas Sriwijaya, Indonesia.
², ⁴Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia.
³Faculty of Computer Science, Universitas Sriwijaya, Indonesia

ABSTRACT

Congenital heart disease often occurs, especially in infants and fetuses. Fetal image is one of the issues that can be related to the segmentation process. The fetal heart is an important indicator in the process of structural segmentation and functional assessment of congenital heart disease. This study is very challenging due to the fetal heart has a relatively unclear structural anatomical appearance, especially in the artifacts in ultrasound images. There are several types of congenital heart disease that often occurs namely in septal defects it consists of the atrial septal defect, ventricular septal defect, and atrioventricular septal defect. The process of identifying the standard of the heart, especially the fetus, can be identified with a 2D ultrasound video in the initial steps to diagnose congenital heart disease. The process of diagnosis of fetal heart standards can be seen from a variety of spaces, i.e., 4 chamber views. In this study, the standard semantic segmentation process of the fetal heart is abnormal and normal in terms of the perspective of 4 chamber views. The validation evaluation results obtained in this study amounted to 99.79% pixel accuracy, mean iou 96.10%, mean accuracy 97.82%, precision 96.41% recall 95.72% and F1 score 96.02%.

Keywords:
Congenital heart disease
Fetal ultrasound images
Semantic segmentation
Convolutional neural networks
U-NET

1. INTRODUCTION

Congenital 0.1 of 1.000 births in nearly all countries [1]. CHD has been estimated to be 32-40 thousand per year, in which the birth rate is 4 million per year in Indonesia. The venous pole defects can cause some abnormalities, such as atrioventricular defect (AVSD), atrial defect (ASD), and ventricular defect (VSD) [2]. Still, AVSD is a relatively common congenital heart defect, caused by a hole that allows communication between all four chambers [3]. However, the hole detection of AVSD is still a challenging task due to the small size of the heart or unintentional fetal movements [4]. The false detection carries a moderate to high risk for both the mother and her fetus.

CHD can be examined by using ultrasound imaging views. Ultrasound is a type of imaging modality process which often applied because of its non-invasive nature compared to other image modalities [5]. The two-dimensional (2D) fetal ultrasound imaging assists the medician for monitoring the gestational age, size, and weight of the fetus, specifically, heart structure and function [6]. However, the sonographers that have different degrees have a lack of expertise to assess the fetal heart. Experts, generally manually represent objects to define the ground truth to be used in the segmentation process [6]. The heart hole detection can be
addressed by the precise semantic segmentation. To improve the automated segmentation, new technologies standardize measurements for optimal assessment of the fetal heart [4].

As computer science, machine learning can be implemented for segmentation tasks. Nguyen et al. [7] discuss the surface extraction using support vector machine (SVM) based texture classification for fetal ultrasound imaging. Gupta et al. [8] using the conditional random field method for the segmentation process in 2D images of ultrasound fetal with exploiting context information. Rahmatullah et al. [9] conducted a process of automated image analysis method based on a machine learning algorithm for detecting important anatomical landmarks employed in manual scoring of ultrasound images of the fetal abdomen. From the results of the previous research analysis using machine learning can do the segmentation process, especially on the fetus. However, the results obtained do not provide maximum results in the process of segmenting the fetus. Besides, the lack of feature representation of ML can affect the result of segmentation.

To date, deep learning (DL), a widespread algorithm and parts of ML that generalize the layered structure of artificial neural networks (ANN) can be implemented for medical applications [10], [11]. Deep learning has developed into sophisticated algorithms for sharing aspects of image and video processing for segmentation, object detection, and classification [12]. Yu et al. [13] The fetal left ventricle is segmented in an echocardiographic sequence based on a dynamic convolutional neural network. Then, Kaluva et al. [14] identified the standard fetal cardiac planes from 2D ultrasound data. Sundaresan et al. [15] conducted an automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks (CNNs). Baumgartner et al. [16] carried out the process of real-time detection and localization of fetal standards using CNNs. From the aforementioned literature, CNNs are the latest generation of models to overcome in segmentation problems, especially for automated feature extraction.

This research proposes an approach that will be carried out using a deep learning algorithm with a CNNs architecture to conduct the semantic process of ultrasound image segmentation in congenital heart disease affected by septal defects. CNNs have found to be more suitable especially in ultrasound images compared to the traditional contextual model-based approach of all images in the training data process [16]. Furthermore, this study proposes a U-NET-based CNNs method for conducting semantic segmentation processes on ultrasound image data. Overall, this research has a contribution to the semantic segmentation process using U-NET architecture.

2. RESEARCH METHOD

Convolutional neural networks (CNNs) are a method that will be used in this study using U-NET architecture. Figure 1 shows the steps of the research carried out.

![Figure 1. The workflow of fetal echocardiography](image)

2.1. Data preparation

This study using data are normal and abnormal fetal heart ultrasound videos. The video data used in this study was viewed with the view of 4 chamber views. The ultrasound video format used uses the mp4 format with a video range of 0-20 seconds. The video data used are normal and abnormal namely atrioventricular septal defect. Figure 2 is screenshots of the video proposed in this study. Figure 2(a) shows the information from atrioventricular septal defect while Figure 2(b) is the normal one. The video used consists of two videos of normal and abnormal patients. The raw video data details that will be used in this study can be seen in Table 1. This table contains information about the ultrasound video that will be used.
2.2. Preprocessing

The video data obtained will be processed to the next stage. The stages to be carried out consist of four steps which are summarized in Table 2 and will be discussed in detail. The preparatory steps taken are image resize, ground truth and data augmentation.

Table 1. Video data information

| Video   | Size   | Length |
|---------|--------|--------|
| AVSD    | Patient 1 | 1.02 mb | 11 d/s |
|         | Patient 2 | 2.10 mb | 20 d/s |
| Normal  | Patient 1 | 617 kb  | 2 d/s  |
|         | Patient 2 | 331 kb  | 5 d/s  |

Table 2. Preprocessing summary

| Preprocessing steps | Target data                  |
|---------------------|------------------------------|
| Input Video         | Frame (Image)                |
| Image Resize        | Image (Resize)               |
| Ground Truth        | Image Label                  |
| Data Augmentation   | Flipping Image (Horizontal & Both) |

a. Input video. At this stage, the ultrasound video data obtained will be processed into several images. The framing process carried out in this process is carried out using the help of the OpenCV tool in the python programming language. The resulting frame is obtained from the video inputted. Furthermore, the frame will be processed back to the next stage.

b. Image resize. The next stage is to resize the image on the frame results obtained previously by resizing the image to 400x300 pixels. This is done to be able to balance all picture frames that will be used at a later stage. The process of the resize stage is the same as the above frame using the OpenCV tool in python programming.

c. Ground truth. This stage organizes image data in the manual segmentation of images that will be seen in the heart. At this stage, manual segmentation is performed using Adobe Photoshop CS 6 Software Table 3 is the detailed results of manual segmentation.

d. Data augmentation. This stage is a technique used to increase the amount of data in order to produce the best model without losing the essence of the data. This can be done by performing random transformations for the data. The data augmentation used was augmentation flipping technique. The image is randomly fixed at the predetermined initial size of 400x300 and does not change the previous process.
Table 3. The ground truth stage

| Atrioventricular Septal Defect | Normal |
|-------------------------------|--------|
| Original Image                | Ground Truth | Original Image | Ground Truth |

2.3. Convolutional neural networks (CNN)

The CNN architecture is part of a deep learning technique that has variations from the multilayer perceptron (MLP). The CNN has many successful applications like handwritten character recognition, object detection, and classification, where CNN has significantly outperformed traditional methods using handcrafted features or other learning based approaches [17]. The most important part of the CNN consists of nodes that are connected to each other based on the weight of the neural network. CNN's important points are the convolution layer, pooling layers, and fully connected layer. Data which consists of three 2D arrays is CNN input which has height, width, and depth. CNNs introduced by Cun et al. [18] are a class of biologically inspired neural networks of convolutional filters and simple non-linearities [19]. An illustration of the CNNs architecture, in general, is shown in Figure 3. CNNs have a hierarchical architecture.

![Architecture of a convolutional neural networks](image)

Figure 3. Architecture of a convolutional neural networks

The essential components of each convolution consist of three components of the convolution layer, the max-pooling layer, and the activation function, as shown in Figure 3. For the convolutional layer, each channel of its output is computed as:

\[ y(i) = \sum_j (b(i,j) + k(i,j) \ast x(j)) \]  \hspace{1cm} (1)

The pooling operation involves sliding a two-dimensional filter over each channel of the feature map and summarizing the features lying within the region covered by the filter. For the pooling layer, its output is computed as:

\[ y'_i = max (y_i, 0) \]  \hspace{1cm} (2)

CNNs is usually composed of convolution layer, pooling layers and fully connected (FC) layer. The last FC layer, as usually configured in classification problem, we use sigmoid function (3).
\sigma(z) = \frac{1}{1+e^{-z}} \quad (3)

From output of each layer, there is often a nonlinear active function, such as sigmoid and ReLU. In this research, we adopt the basic CNNs architecture U-NET and is shown in Figure 4. CNNs are convolution models that have the main categories for image recognition, namely segmentation, classification, identification of object detection. Segmentation is an area where CNNs are widely used [20]. U-NET is a semantic segmentation architecture on CNNs which is often used to perform segmentation processes using medical data. U-NET has two important points in the convolution process, namely contraction path and expansion path, to improve the accuracy of both architectures using the dice coefficient.

![Figure 4. Architecture for segmentation using U-NET](image)

The network architecture is illustrated in Figure 4 which consists of the contraction path and expansion path. The contraction path follows a typical convolutional network architecture [21]. This consists of a 3x3 repetitive convolution process that is not added. This process is done using ReLU and 2x2 max-pooling to do the downsampling process. Every downsampling process is carried out, this process is duplicated in feature channels. In the expansion path, always consistent with the process of upsampling the feature map with the convolution process 2x2 (up-convolution) which divides the two features with the feature map cut from the contraction path and the 3x3 convolution process using ReLU. At the final layer convolution, 1x1 is used to map each vector feature of 64 components to the desired number of classes.

3. RESULTS AND DISCUSSION

The results of the segmentation research after training the U-net model on deep learning on ultrasound images. U-NET is able to use hyperparameter with activation function (ReLU), sigmoid activation function, and loss using dice coefficient function [22]. In this section, the fetal heart segmentation process uses ultrasound images detected by septal defects and normal. The technique used in this study uses a convolutional neural network method based on U-NET architecture. The U-NET architecture information used is summarized in Table 4.

The training data process is done by using a dataset that has been done preprocessing is 372 image data consisting of original image data and ground truth. Used data contains septal and normal defect images. The data information used can be seen in the Table 5.

a. Training parameters: We did the training process of the U-NET based convolutional neural network model with epoch 1000 and batch size 64. Using epoch 1000 was due to get the right results for...
segmenting the process especially in the analysis of medical data to require a very long process in processing image computing performed training process. We use learning rate 1e-5 with Adam optimizer and smooth loss 1^-5, threshold 0.5. In the training data, the data augmentation process is carried out on the data generator model that is applied in this research, namely U-NET.

b. Post-processing: What is obtained from the prediction data from a model sometimes has several regions with different labels, unlike the ground truth segmentation that has been done before. This study uses postprocessing to help get maximum results, while postprocessing is used using the thresholding algorithm. Thresholding is the process of dividing an image into two or more classes of pixels, as in this case, it is "foreground" and "background". The thresholding algorithm can assist in image processing in terms of eliminating noise and allowing it to increase high accuracy [23]. The predicted results obtained at the next stage of U-NET architecture process the results of the prediction image using the thresholding algorithm used in this study using the fixed thresholding. The next stage is the process of validation and evaluation to see how well the CNNs method works with the U-NET architecture on ultrasound image data affected by septal defects.

c. Metrics evaluation: Segmentation is an important step for pre-processing in image analysis application [24]. From the result of segmentation, it will then be possible to identify the area of interest and the object of the occurrence of an event which is very useful in subsequent image analysis [25]. We conduct test on the semantic segmentation process using metrics evaluation pixel accuracy, mean accuracy, mean iou, precision, recall and F1 score. The results of the segmentation methods that have been carried out above can be seen in Table 6. We conducted training on the U-NET model with random data of 594 picture frames with two patient objects and testing data of 75 picture frames with two objects. Table 6 shows the results of a segmentation sample from U-NET and V-NET.

The results of the U-NET and V-NET training test results based on the plot graph of the accuracy results as a feature extractor with 1000 epochs are shown in Figures 5 and 6. Based on the prediction results of the U-NET model that has been carried out the next step is to test the model with metrics evaluation based on the segmentation case in the image. The performance results for segmentation can be seen in Table 7. Then, in this study summarizes the results of previous studies and makes a comparison of the results with the segmentation case. The results of comparisons in previous studies can be seen in Table 8.

### Table 4. Summaries of architecture layers of U-NET

| Layer                | Kernel Size, Feature Map | Stride | Activation | Output Shape |
|----------------------|--------------------------|--------|------------|--------------|
| Input Layer          | -                        | -      | -          | 256x256x1    |
| Convolution Layer 1  | 64x64x3                  | 1      | ReLU       | 128x128x3    |
| Max Pooling 1        | 2x2                      | 2      | -          | 128x128x3    |
| Convolution Layer 2  | 128x128x3                | 1      | ReLU       | 256x256x3    |
| Max Pooling 2        | 2x2                      | 2      | -          | 256x256x3    |
| Convolution Layer 3  | 256x256x3                | 1      | ReLU       | 512x512x3    |
| Max Pooling 3        | 2x2                      | 2      | -          | 512x512x3    |
| Convolution Layer 4  | 512x512x3                | 1      | ReLU       | 1024x1024x3  |
| Dropout              | p=0.5                    | -      | -          | 1024         |
| Convolution 5        | 2x2                      | 2      | ReLU       | 1024x1024x3  |
| Dropout              | p=0.5                    | -      | -          | 1024         |
| Convolution Layer 6  | 512x512x2                | 3 (axis) | ReLU       | 512x512x3    |
| Convolution Layer 7  | 256x256x3                | 1      | ReLU       | 256x256x3    |
| Up                   | 128x128x2                | 3 (axis) | ReLU       | 128x128x3    |
| Convolution Layer 8  | 128x128x3                | 1      | ReLU       | 64x64x2      |
| Up                   | 64x64x2                  | 3 (axis) | ReLU       | 64x64x3      |
| Convolution Layer 9  | 64x64x3                  | 1      | ReLU       | 2x2x3        |
| Output Layer         | -                        | -      | Sigmoid    | 1            |

### Table 5. Split data training and testing

| Data     | Data Train | Data Test | Total |
|----------|------------|-----------|-------|
|          | Original Image | Ground Truth | Original Image | Ground Truth |
| AVSD     | 151        | 151       | 35    | -       | 337   |
| Normal   | 146        | 146       | 40    | -       | 332   |

Int J Artif Intell, Vol. 10, No. 3, September 2021: 553 - 562
### Table 6. Result predict U-NET & V-NET

| Data  | Original Image | Ground Truth | U-NET | V-NET |
|-------|----------------|--------------|-------|-------|
| AVSD  | ![Original Image](image1) | ![Ground Truth](image2) | ![U-NET](image3) | ![V-NET](image4) |
| Normal| ![Original Image](image5) | ![Ground Truth](image6) | ![U-NET](image7) | ![V-NET](image8) |

**Figure 5.** The result of (a) accuracy and (b) loss of the pretrained U-NET model

*Segmentation atrioventricular septal defect by using convolutional neural... (Ade Iriani Sapitri)*
Figure 6. The result of (a) accuracy and (b) loss of the pretrained V-NET model

Table 7. Comparison performance measures for the segmentation results

| Metrics Evaluation | U-NET  | V-NET   |
|--------------------|--------|---------|
| Pixel Accuracy     | 99.79% | 97.96%  |
| Mean IU            | 96.10% | 74.06%  |
| Mean Accuracy      | 97.82% | 81.57%  |
| Precision          | 96.41% | 65.88%  |
| Recall             | 95.72% | 64.15%  |
| F1 Score           | 96.02% | 62.60%  |

Table 8. The comparison CNNs-based U-NET architecture performance in dice coefficient

| Method                        | Task                      | Dice coefficient |
|-------------------------------|---------------------------|------------------|
| Dynamic CNN                  | Segmentation of fetal left ventricle | 94.50 %          |
| Proposed CNNs-based U-NET    | Segmentation AVSD         | 96.02 %          |

4. CONCLUSION
The research results show the results of segmentation predictions using performance in segmentation cases using the U-NET model as feature extraction. The result after training has a good segmentation with pixel accuracy is 97.79%, mean accuracy is 97.82%, mean IoU is 96.10%, precision is 96.41%, recall 95.72% and F1 score is 96.02%.

ACKNOWLEDGEMENTS
This research is supported by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia under the Applied Research Grants, Number. 096/SP2H/LT/DRPM/2019 and Universitas Sriwijaya, Indonesia under Professional Grants 2020.
BIOGRAPHIES OF AUTHORS

Ade Iriani Sapitri is currently a postgraduate student of Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes Medical Imaging, Deep Learning and Machine Learning.

Siti Nurmaini is currently a professor in the Faculty of Computer Science, Universitas Sriwijaya. She was received her Master's degree in Control system, Institut Teknologi Bandung – Indonesia (ITB), in 1998, and the PhD degree in Computer Science, Universiti Teknologi Malaysia (UTM), in 2011. Her research interest including Biomedical Engineering, Deep Learning, Machine Learning, Image Processing, Control systems, and Robotic.

Sukemi is currently a lecture and researcher in Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Computer Architecture and Machine Learning.

Muhammad Naufal Rachmatullah is currently a Research Assistant with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes medical imaging, biomedical signal and engineering, 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 interest includes biomedical signal and engineering, deep learning, and machine learning.