# Multidimensional Echocardiography Image Segmentation using Deep Learning Convolutional Neural Network

Hasan Imaduddin[1], Riyanto Sigit[1] and Anhar Risnumawan[2]
[1]Computer Engineering Division, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia
[2]Mechatronics Engineering Division, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia

Keywords: Echocardiography, Deep Learning, Segmentation, Convolutional Neural Network, Multidimensional.

Abstract: One of the most dangerous diseases that threaten human life is heart disease. One way to analyze heart disease is by doing echocardiography. Echocardiographic test results can indicate whether the patient's heart is normal or not by identifying the area of the heart cavity. Therefore, many studies have emerged to analyze the heart. Therefore I am motivated to develop a system by inputting four points of view of the heart, namely 2 parasternal views (long axis and short axis) and 2 apical views (two chambers and four chambers) with the aim of this study being able to segment the heart cavity area. This research is part of a large project that aims to analyze the condition of the heart with 4 input points of view of the heart and the project is divided into several sections. For this research, it focuses on the process of echocardiographic image segmentation to obtain images of the heart cavity with 4 input points of view of the heart using the Deep Learning method by using the VGG-16 and RESNET-18 architecture. The training process is done using 30 epochs with 50 iterations per epoch and 1 batch size so that the total iteration is 7500 iterations. It can be seen that during the training process, the percentage accuracy is already high, reaching 95% -99%. On the VGG-16 architecture, it has an average accuracy in each viewpoint of around 83% -93%. The architecture of RESNET-18 has an average accuracy in every point of view which is around 76% -92%.



## 1 INTRODUCTION

According to the World Health Organization (WHO), until 2018 heart disease is still one of the diseases that causes the most deaths in the world. WHO data in 2015 states that more than 17 million people have died from heart and blood vessel disease (about 30% of all deaths in the world), of which the majority or about 8.7 million are caused by coronary heart disease. More than 80% of deaths from heart and blood vessel disease occur in low to moderate income developing countries. In Indonesia, the results of the 2018 Basic Health Research show that 1.5% or 15 out of 1,000 Indonesians suffer from coronary heart disease. Meanwhile, when viewed from the highest cause of death in Indonesia, the 2014 Sample Registration System Survey showed 12.9% of deaths were due to coronary heart disease. By 2030, it is estimated that the death rate from heart and blood vessel disease will increase to 23.6 people (WHO, 2019). Therefore, various efforts to anticipate and treat heart disease have been developed.

Therefore, echocardiography emerged as one of the non-invasive and painless technologies for depicting images of the human heart. Echocardiographers diagnose heart conditions based on several symptoms that appear on the image. One example of symptoms is heart wall movement. Movement of the heart wall can give an indication of whether the heart is healthy or not. However, echocardiography has several limitations including image quality, operator dependency, and interpreter dependency. These limitations have an impact on the accuracy of the doctor's diagnosis. The accuracy of a doctor's diagnosis depends on the doctor's knowledge and experience.

Echocardiographic videos are used by doctors to analyze the patient's heart performance. Echocardiography uses techniques by emitting ultrasonic waves to determine the size, shape, function of the heart, and the condition of the heart valves which are visualized into video images. The results of the examination on the echocardiography test can only be read by the doctor so that the analysis of the condition of the heart cavity depends on the



Imaduddin, H., Sigit, R. and Risnumawan, A.
Multidimensional Echocardiography Image Segmentation using Deep Learning Convolutional Neural Network.
DOI: 10.5220/0010963200003260
In *Proceedings of the 4th International Conference on Applied Science and Technology on Engineering Science (iCAST-ES 2021)*, pages 1250-1254
ISBN: 978-989-758-615-6; ISSN: 2975-8246

doctor's accuracy and experience so that it will cause a different analysis from each doctor in determining the true condition of the heart..

## 2 RELATED WORK

Automatic cardiac segmentation using triangle and optical flow has been done (Sigit et al., 2019). The proposed method presents a solution for segmentation of echocardiography image for heart disease. This research has produced a Median High Boost Filter which is able to reduce noise and preserve image information. Segmentation using the triangle equation method has the smallest error value. Performance segmentation for the assessment of cardiac cavity errors obtained an average of 8.18% triangles, 19.94% snakes, and 15.97% DAS. The experimental results show that the extended method is able to detect and improve the image segmentation of the heart cavity accurately and more quickly.

Improved segmentation of cardiac image using active shape model has been done (Sigit and Saleh, 2017). The purpose of making this system is to segment the heart cavity and then calculate its area to determine the performance of the heart. Pre-processing to enhance and enhance image quality using Median Filtering, erosion, and dilation. Segmentation using Active Shape Model. Then to calculate the area using Partial Monte Carlo. This system has an error of 4.309%.

Deep learning, (LeCun et al., 2015), recently has shown excellent results in image classification and recognition by, this method started (Krizhevsky et al., 2012), (Simonyan and Zisserman, 2014) since the win in ImageNet challenge. Image segmentation (Farabet et al., 2013), (Graves et al., 2008), (Mohamed et al., 2011) and object detection (Girshick, 2015), (Felzenszwalb et al., 2010). Deep learning such as convolutional neural networks can learn features that exist in visual input, these features are automatically trained from a lot of image data called datasets, and do not need to involve design features created manually. In this work, using in-depth techniques, we show that it is possible to solve these problems to build a multidimensional echocardiographic image segmentation system that is strong in images.

## 3 DEEP LEARNING BASED MULTIDIMENSIONAL ECHOCARDIOGRAPHY IMAGE SEGMENTATION

The video generated from the echocardiography device is converted in the form of images and as input to the neural network. The network then output the segmenting label which shows the area of cardiac cavity.

### 3.1 Multidimensional Echocardiography Image

The heart has four chambers. The two receiving chambers in the superior section are the atrium, while the two pumping chambers in the inferior part are the ventricles. In analyzing the heart's performance, there are several views needed, including a long axis view, a short axis view, a view of 4 apical spaces and 2 apical spaces.


Figure 1: 4 Viewspoints of Cardiac Cavity.

In this final project will use multi-dimensional echocardiographic image input from 4 viewpoints of the heart, namely 2 parasternal views (parasternal long axis and parasternal short axis) and 2 apical views (apical two chambers and apical four chambers) shown in Fig. 1.

### 3.2 Convolutional Neural Network

Convolutional Neural Networks (CNN) as one of the methods in deep learning have shown performance as a prominent approach, which is applied in a variety of computer vision applications, to learn the effective features automatically from training data and train them end-to-end (Krizhevsky et al., 2012).

Basically, CNN consists of several layers that are staged together. Layers usually consist of convolutional, pooling, and fully connected layers that have different roles. During the forward and backward training is carried out. For input patches, the advance stage is carried out at each layer. During training, after the advanced stage, the output is compared to the basic truth and the loss is used to do the backward phase by updating the weight and bias parameters using a general gradient decrease. After several iterations, the process can be stopped when the desired accuracy is achieved. All layer parameters are updated simultaneously based on training data.

### 3.3 Semantic Segmentation

Semantic segmentation draws on the task of classifying each label in an image by connecting each pixel in the image to the label class (Felzenszwalb et al., 2010). These labels can be tailored to the needs of programmers to collect more labels installed in the label class. Semantic segmentation is very useful to know the number and location of objects detected, such as where's the cardiac cavity in muldimetional echocardiography image to help doctors determine which cardiac cavity when diagnosing heart disease patients.

### 3.4 Training

We did the training step for each one of the four perspectives from one viewpoint of the heart because each point of view has different form characteristics, for that we split the dataset to be divided into 4 different sections (Imaduddin et al., 2016).

We train this network for approximately 2 hours with maximum 8.000 iterations and validation accuracy achieve 95 percent on test validations images, then convert the model to perform segmentation.

## 4 EXPERIMENTAL SETUP

System testing is done by running a program on a computer with specifications processor Intel Core i5-9400F CPU 2.90GHz, RAM 8GB, GPU NVIDIA GeForce GTX 1060TI, Operation System Windows 10 Pro 64-bit, run in MATLAB 2019b application using Convolutional Neural Network deep learning.

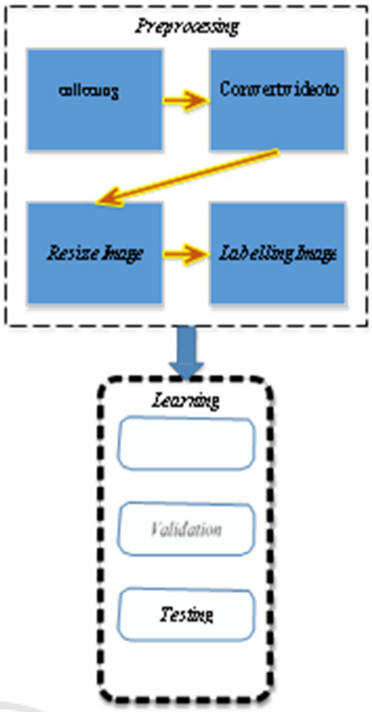


Figure 2: System Diagram.

We use a dataset that we collect from hospital. The dataset we collected in the forms of 4 viewpoints echocardiography video with various conditions, sizes, colors, and brightness levels. From various conditions, we do manual pre-processing by convert video to several pixels according to the duration of the video, changing the size of 432x636 pixels (most original image sizes) to match the input image that has been prepared. The dataset used was 416-468 RGB images which transferred 60% of training and 40% for testing. We wrote a SegNet Implementation that is compatible with Matlab GPU using publicly available optimization library functions. In this work, we divide it into several steps to make it easier to understand the workflow of the system that we have built before segmenting cardiac cavity and classify them based on training data.

The first steps to doing the pre-processing process are to prepare the dataset by doing pre-processing using traditional image processing shown in Fig. 2. then, determine the number of classes and labels to be labeled using the Image Labeler in Matlab. And then, semantic segmentation to take pixel label class decisions using SegNet Layers and Deep Lab V3 Plus Layers.

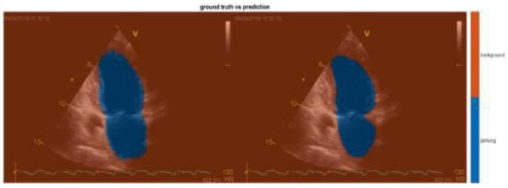

Figure 3: Ground truth vs prediction VGG-16.

It can be seen in Fig. 3. that the comparison between ground truth and prediction for image testing from the point of view 2 chambers randomly selected by the system can detect the location of the heart cavity by segmenting the area of the heart cavity using blue with an accuracy of 0.8805 or 88.05%. This shows that the neural network has been able to study the heart cavity according to the data provided.

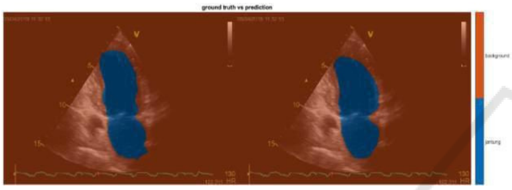

Figure 4: Ground truth vs prediction RESNET-18.

It can be seen in Fig. 4. that the comparison between ground truth and prediction for image testing from the point of view 2 chambers randomly selected by the system can detect the location of the heart cavity by segmenting the area of the heart cavity using blue color with an accuracy of 0.7871 or 78.71%. This shows that the neural network has been able to study the heart cavity according to the data provided.

Table 1: Mean Accuracy Cnn Architecture Heart Viewpoint.

| CNN Architecture | Heart Viewpoint | Mean Accuracy |
|---|---|---|
| VGG- 16 | 2 Chamber | 0.9298 |
| | 4 Chamber | 0.9195 |
| | Long Axis | 0.9388 |
| | Short Axis | 0.8814 |
| RESN ET-18 | 2 Chamber | 0.8899 |
| | 4 Chamber | 0.9285 |
| | Long Axis | 0.7616 |
| | Short Axis | 0.8363 |

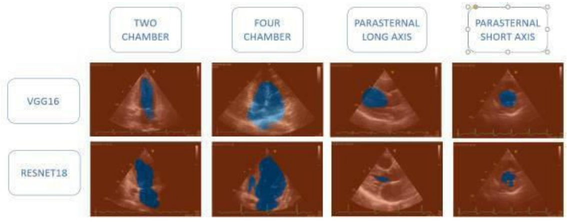


Figure 5: Result of Our Training for 4 Viewpoint using VGG-16 and RESNET-18.

The average accuracy of each viewpoint of the heart can be seen in Table I. The long axis viewpoint has the highest average accuracy rate from the other four viewpoints, which is 0.9388 and the short axis viewpoint has the lowest average accuracy level, which is 0.8814. This is because the VGG-16 architecture is better able to determine the segmented area at the long axis viewpoint is simpler than the short axis viewpoint. And also the segmented area at the long axis viewpoint has a more consistent pattern than the short axis viewpoint which tends to be more abstract in the pattern. So that neural networks that have studied heart cavity patterns following the data provided more easily recognize objects with simpler areas because they are not too large and small are segmented areas shown in Fig. 5.

The average accuracy of each viewpoint of the heart can be seen in Table I. The 4 chamber view angle has the highest average accuracy rate from the other four viewpoints, which is 0.9285 and the long axis viewpoint has the lowest average accuracy level, which is 0.7616. This is because the RESNET-18 architecture is better able to determine the segmented area at the 4 chamber viewpoint is greater than the long axis viewpoint and also the segmented area at the 4 chamber viewpoint has a more consistent pattern than at the long axis viewpoint which tends to be more vary the pattern. So that neural networks that have studied heart cavity patterns following the data provided can more easily recognize objects with larger areas for segmentation shown in Fig. 5.

## 5 CONCLUSIONS

The results of the segmentation of the heart cavity in the VGG-16 architecture have a pattern of segmentation areas that tend to be the same shape, while the results of the segmentation of the heart cavity on the RESNET-18 architecture have a pattern of segmentation areas that tend to vary in shape. The results of cardiac cavity segmentation on the VGG-16 architecture have a segmentation area pattern that tends to exceed the labeled area, while the results of cardiac cavity segmentation on the RESNET-18 architecture have a segmentation area pattern that tends to reduce the area that has been labeled. The VGG-16 has a higher average accuracy at each point of view, which is about 83% -93% than the RESNET-18 architecture with an average accuracy of around 76% -92%. The heart point of view on the VGG-16 architecture which has the highest average accuracy is the long axis and the lowest in the short axis, while on the ResNet architecture the highest is 4 chambers and the lowest in the long axis.

We have presented a multidimensional echocardiography image segmentation using deep learning such as a convolutional neural network with only information from echocardiography image which is converted from the video are able to solve the existing problem such as, helps doctors determine the cardiac cavity when examining cardiac patients from echocardiography used in Indonesia. Those existing works that use features that are manually designed as input to a system to perform segmentation are seemed rather difficult to implement in multidimensional echocardiography image segmentation due to its difference and uncertain quality video. The parameters were studied from the training data but the video produced was based on each echocardiography device in the hospital and the video taking when the heart examination was also carried out by people who sometimes differed from different patients every day made the video results from echocardiography also differed in quality. So it seems a little difficult to implement at this time.

## REFERENCES

Sigit, R., Roji, C. A., Harsono, T., & Kuswadi, S. (2019). Improved echocardiography segmentation using active shape model and optical flow. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(2), 809-818.

Sigit, R., & Saleh, S. N. (2017, October). Improved segmentation of cardiac image using active shape model. In *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 209-214). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521 (7553), 436-444. *Google Scholar Google Scholar Cross Ref Cross Ref*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1915-1929.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2008). A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, *31*(5), 855-868.

Mohamed, A. R., Dahl, G. E., & Hinton, G. (2011). Acoustic modeling using deep belief networks. *IEEE transactions on audio, speech, and language processing*, *20*(1), 14-22.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *38*(1), 142-158.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), 1627-1645.

Imaduddin, H., Anwar, M. K., Perdana, M. I., Sulistijono, I. A., & Risnumawan, A. (2018, October). Indonesian vehicle license plate number detection using deep convolutional neural network. In *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)* (pp. 158-163). IEEE.

World Health Organization (WHO). (2019). *The state of food security and nutrition in the world 2019: safeguarding against economic slowdowns and downturns* (Vol. 2019). Food & Agriculture Org.