Automatic Measurement Application of Heart Area from Chest X-Ray Images Using the U-Net Deep Learning Method

Andhika Putra Setianto*, Cahya Damarjati , and Asroni

Universitas Muhammadiyah Yogyakarta, Jln.Brawijaya, Tamantirto, Kasihan, Bantul, Yogyakarta 55183, Indonesia
Corresponding author: andhika.putra.2016@ft.umy.ac.id

Abstract

Heart health is a basic human right and a crucial component of global health justice. In an ever-more-advanced age, every task becomes simple due to science, technology, and information development. However, certain tasks are still performed manually. Therefore, innovation in computerized system design is required. The human heart area calculation was performed by combining image processing and deep learning techniques. Deep learning is a scientific subfield of machine learning, while image segmentation is a step in image processing. This study employed the U-Net segmentation method to identify different stages of heart area calculation. U-Net could conduct image segmentation with the small training dataset accurately. This study's population consisted of 800 chest X-ray images obtained from the Kaggle website, with human hearts as the sample. The findings revealed that the training data with the U-Net architecture model acquired an accuracy of 09.98. However, the testing data accuracy was still determined manually. In this work, the U-Net model employed an input shape measuring 256x256, a kernel size of 3x3, and 50 epochs.

Keywords: Chest X-ray, Deep Learning, Image Segmentation, U-Net Segmentation

1. Introduction

Every person's most fundamental human right is to have a healthy heart, which is also a critical component of global health justice. Health is a pillar of economic sustainability and community prosperity enhancement. Accordingly, the heart must be kept healthy, which can be accomplished by simple measures such as adopting a healthy lifestyle, eating nutritious meals on a regular basis, obtaining sufficient rest, quitting or avoiding smoking, and consulting a physician about heart health. A heart check, one of which utilizes an X-ray, is performed to prevent heart disease or determine the probability of a heart disorder. This X-ray examination is described as a medical imaging technique that employs electromagnetic radiation to capture images of the inside of the human body.

Living in a highly technological society, completing any endeavor is a breeze, attributed to the expansion of technological and informational knowledge. Every task necessitates accuracy and swiftness in acquiring information. Indeed, information management requires modern and adequate technology. Hence, creating a computerized system is one way to achieve the goal. Some implementation efforts are still being conducted conventionally. Implementing a design for determining the area of the human heart using chest X-rays is an example of how innovation in computerized system design is projected to contribute to health-related activities.

Combining image processing with deep learning techniques within machine learning allows for heart area calculation in the human body. Deep learning applies

the basic concept of machine learning through artificial neural network algorithms with additional layers. Due to the inclusion of several hidden layers between the input and output layers, this network is sometimes referred to as a deep neural network[1].

Image segmentation, which divides an image into similar areas based on the similarity of pixels, is one of the image processing methods[2]. This image is divided into sections of varying brightness to clearly distinguish the object and background. This segmentation is performed in alignment with the problem. It must be terminated if the objects have been highly visible or separated. The segmentation will generate an extremely high degree of accuracy, which is dependent on the amount of success of the applied analysis method.

This study utilized the U-Net segmentation method to identify various stages of heart area calculation. U-Net's architecture is a fully convolutional network (FCN)[3]. FCN consists of several artificial neural networks that have been modified and enhanced, allowing this architecture to function with a small number of training images and generate more accurate segmentation. The fundamental concept is to augment the conventional contracting network with successive layers in which the pooling operator is substituted with the upsampling operator. Consequently, this screen improves output resolution. The contracting path's high resolution is coupled with upsampled output for the location. Therefore, successive convolution screens can subsequently learn to gather more precise outputs. U-Net can accurately conduct image segmentation with minimal training data.

U-Net is one kind of encoder-decoder neural network to solve end-to-end problems, turning input image into feature representation and recovering feature representation into output image, which is very popular for semantic segmentation[4].

There are research results which state that the U-Net is more suitable for detecting skin diseases[5]. As for other architectural model methods, such as Mask−RCNN can be used to perform image segmentation, however these methods require longer processing time, have a complex design and are difficult to train compared to U-Net[6]. To measure how well segmentation predictions with ground truth annotations (segmentation by humans) can use the Intersection over Union (IoU) evaluation metric[7].

X-rays are a diagnostic method using electromagnetic waves in the form of X-rays. X-rays are emission of electromagnetic waves similar to ultraviolet light, heat, radio waves, which can penetrate objects because the waves are very short. In 1985, W.C. Rontgen, a German physicist, discovered X-Rays. X-rays exit and are directed from the tube through a window (diaphragm). The cooling radiator will dissipate the heat generated[8].

To make a medical diagnosis using X-rays, an X-ray machine or X-ray machine is needed. When an X-Ray machine produces X-rays, there will be an interaction of X-rays. X-rays will be generated by photons on the X-Ray machine. Decreasing the amount of intensity which is affected by the attenuation coefficient of the material (μ) and thickness (t) will result in different types of tissue. After going through the thickness of the material (t), the intensity (I) of the X-Ray irradiation will be related to the exponential equation with the initial intensity (Io)[9].

Digital image is a representative image that is taken using a machine based on quantization and sampling. Quantization states the number of colors in the image which is denoted by the brightness level value stated in the gray level value (grayscale) based on the amount of binary bit used by the machine. Sampling is the size of the box arranged in columns and rows. In other words, sampling is the size of the point (pixel) in the image[10].

2. Method

This research utilized secondary data gathered from the Kaggle website. The population consisted of 800 chest X-ray images, with the heart serving as the research sample. Spyder 4.1.5 with Python 3.8.3 and U-Net semantic segmentation were employed to discover the heart in Chest X-ray images (python reference).
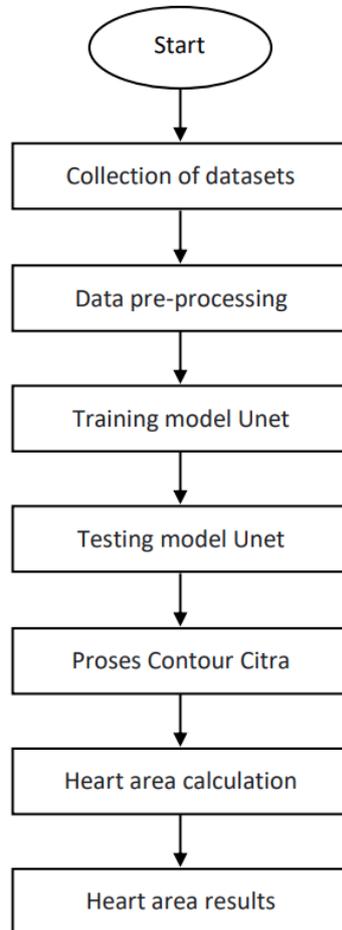

Figure 1 Research Stages

3. Results

The U-Net segmentation algorithm separated heart images from chest X-ray images. The initial step in constructing this model was data training, aiming to create a model for testing. The accuracy value was deployed to measure the model's success. The testing was then conducted, generating an output mask analyzed during training. Subsequently, the data were computed using a scale and converted to centimeters (cm).

3.1 Dataset Collection Results

This research utilized secondary data in the form of 800 chest X-ray images collected from the Kaggle website.

3.2 Image Mask-Making Results

Image mask creation was performed manually using the Krita program, and the resulting mask was stored in.png format.
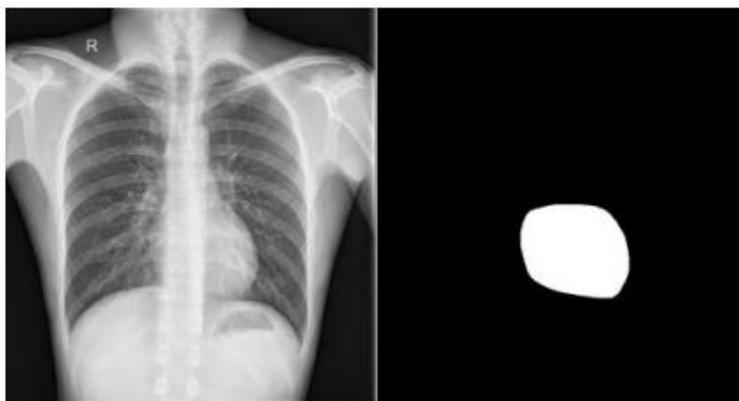
Figure 2 Mask-Making Result

Figure 2 illustrates a mask designed for training. The chest X-ray and mask images were separated into the Ori and Mask folders. Moreover, 586 of 800 images for training masks and 214 of 800 images for testing masks were identified.

3.3 Result of Resizing Dataset

The resizing of the dataset was performed as a result of adjusting the input of the U-Net model. The chest X-ray and mask images were resized to 256x256 pixels.
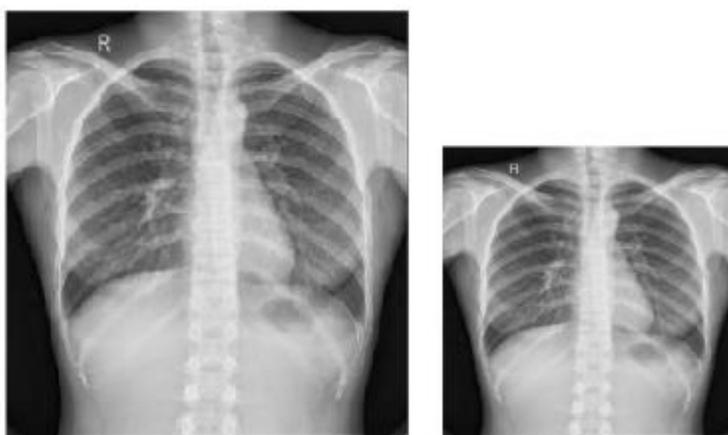

Figure 3 Result of Resizing a Chest X-ray Image

3.4 Results of U-Net Architecture Training Model

Figure 4 U-Net Architecture Training Model

Figure 4 depicts the training model of the U-Net architecture. The input images were 256x256 pixels.

1) Conv1

In this phase, the input was convoluted with 32 channels using a 3x3 kernel. Following convolution, ReLU activation was implemented. Then, the max pooling was run.

2) Conv2

Max pooling results of Conv1 were convolved again using 64 channels and a 3x3 kernel at this step. Immediately following convolution, the ReLU function was introduced. The max pooling was then utilized again.

3) Conv3

Max pooling results of Conv2 were then convoluted using 128 channels and a 3x3 kernel. After being convoluted, the ReLU function and max pooling were implemented.

4) Conv4

Max pooling results of Conv3 were convolved once more with 256 channels and a 3x3 kernel. Subsequently, the ReLU function and max pooling were applied.

5) Conv5

The Conv4 convolution results were convoluted again using 512 channels and a 3x3 kernel, according to the ReLU function. In this stage's convolution results, the model identified the mask in the image, but its actual existence was uncertain. The true existence information was lost due to the downsampling of Conv1 through Conv4 phases, decreasing the data size.

6) Up-Conv6

Due to these limitations, upsampling was finally implemented in this phase, with the results convoluted with 256 channels using a 3x3 kernel and the results of the fourth convolution stage. Concatenation connected two layers, producing information on the mask and its actual presence in the data. The method of upsampling was then repeated for the succeeding stage.

7) Up-Conv7

The previous stage's upsampling results were convoluted again using 128 channels, a 3x3 kernel, and the Conv3 results. The upsampling was then utilized again.

8) Up-Conv8

The upsampling results from the Up-Conv7 stage were then convoluted using 64 channels, a 3x3 kernel, and the Conv2 results. Then, the upsampling was implemented.

9) Up-Conv9

The upsampling results from the Up-Conv8 stage were convoluted using 32 channels, a 3x3 kernel, and the Conv1 results. Later, this stage's findings were transmitted to the fully connected layer stage.

10)   Conv 10

In this process, the data were reprocessed to be convoluted with two channels using a 1x1 kernel, and its mask was set based on its presence.

3.5 Training Results of the U–Net Model

The training aimed to train the model to detect the heart object on the chest X-ray images. The training utilized a model with the Seresnet34 architecture.
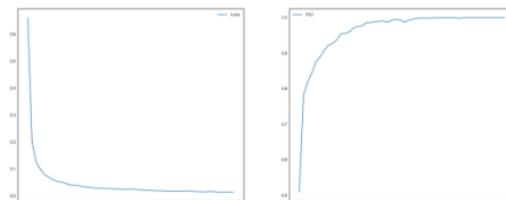

Figure 5 Training Results

Figure 5 displays that the training model's accuracy achieved 98% with a loss value of 0. This training employed 50 epochs with a batch size of four. The more epochs utilized, the better the findings, but the longer the process takes.

3.6 Testing Results of the U-Net Model

The testing employed 214 test data in which the output results were categorized into inaccurate and accurate predictions. The inaccurate predictions were then continued in the image contour process to rectify the incorrect portion.
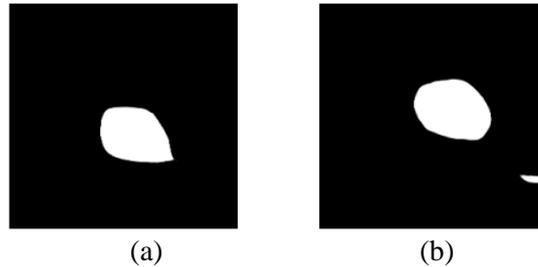

(a)   (b)

Figure 6 Testing Results: (a) Accurate and (b) Inaccurate

Of the 214 data generated in testing, 36 (17%) data fell into the inaccurate prediction category, while 178 (83%) belonged to the category of accurate prediction.

3.7 Image Contour Results

Contouring aimed to eliminate non-fitting objects, significantly influencing the heart area calculation. It must rectify two errors: the incorrect markings on the left and right sides and the not heart images.
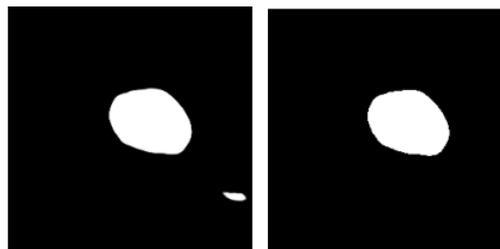

Figure 7 Contour Result

Figure 7 denotes the image contouring result. Unmatched objects were eliminated. However, the image matrix transformed into a binary matrix owing to the threshold function.

3.8 Heart Area Calculation Results with Scaling

The pixel scale was converted to a centimeter scale to determine the heart size in centimeters.

$$\text{Scale} = \frac{Specified\ actual\ image\ size}{Pixel\ size\ utilized}$$

The formula calculated the centimeter value of one pixel. The actual image was sized 30x25, and the pixel size was 256x256, resulting in a value of 0.011444091796875 cm per pixel.

3.9 Heart Area Calculation Results

After determining the scale, the heart area was calculated. The heart area calculated was the testing prediction results. To assess the accuracy of the area prediction results, the mask area constructed manually using Krita was compared to the predicted mask area acquired through testing.

Table 1 Difference in Heart Area Calculation

No	Name		Difference	Percentage
1	MCUCXR_03 62_1_mask	MCUCXR_03 62_1_predict	0.4799 484878	0.8%
2	MCUCXR_03 67_1_mask	MCUCXR_03 67_1_predict	3.7989 569663	7.5%
3	MCUCXR_03 69_1_mask	MCUCXR_03 69_1_predict	1.5529 11804	2.8%
4	MCUCXR_03 72_1_mask	MCUCXR_03 72_1_predict	2.9826 438958	7.9%
5	MCUCXR_03 75_1_mask	MCUCXR_03 75_1_predict	0.7262 198501	2%
6	MCUCXR_03 83_1_mask	MCUCXR_03 83_1_predict	1.6037 045074	3.6%
7	MCUCXR_03 87_1_mask	MCUCXR_03 87_1_predict	0.8804 4806	1.3%
8	MCUCXR_03 90_1_mask	MCUCXR_03 90_1_predict	0.2232 46315	0.5%
9	MCUCXR_03 93_1_mask	MCUCXR_03 93_1_predict	1.9969 612086	2.4%
10	MCUCXR_03 99_1_mask	MCUCXR_03 99_1_predict	2.5908 195687	6.7%

The data in Table 1 infer a negligible discrepancy between the prediction results and the mask results obtained manually. Once the chest X-ray image reached a certain threshold, it became a binary image, and the corresponding area could be obtained. Subsequently, the pixel area was calculated by adding the image's binary value. Hence, the pixel area was transformed into cm^2.

4. Conclusion

The following conclusions were drawn from the analysis results of implementing U-Net segmentation in classifying X-ray images.

1. The U-Net segmentation method could be utilized in the medical field. In this study, the U-Net could predict the heart organ on chest X-ray images, with the prediction result being determined in centimeters.
2. The training data from the U-Net architecture model generated an accuracy of 0.98. However, the testing results were still calculated manually. In this study, the input shape for the U-Net model was 256x256, and the number of epochs was 50, with 64 batches each.

References

[1] Goodfellow, I., Bengio, Y., & Courville, A., (2016). Deep Learning. The MIT Press Cambridge, Massachusetts, London, England.

[2] Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah. Emitor: Jurnal Teknik Elektro, 18(01), 15–21. https://doi.org/10.23917/emitor.v18i0 1.6235.

[3] biomedical image segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351, 234–241. https://doi.org/10.1007/978-3-319- 24574-4_28.

[4] Salumaa, S. O. (2018). Convolutional Neural Networks For Cellular Segmentation.Estonia: University Of Tartu.

[5] Listyalina, L. (2017). Peningkatan Kualitas Citra Foto Rontgen Sebagai Media Deteksi Kanker Paru. Jurnal Teknologi Informasi, XII(34), 110– 118. http://jti.respati.ac.id/index.php/jurnal jti/article/viewFile/1/1.

[6] Salumaa, S. O. (2018). Convolutional Neural Networks For Cellular Segmentation.Estonia: University Of Tartu.

[7] Falk, T., Mai, D., Bensch, R., Cicek, O., Abdulkadir, A., Marrakchi, Y, Ronneberger.,O. (2019). U-Net – Deep Learning For Cell Counting, Detection, And Morphometry .

[8] Souisa, F., Ratnawati, & Sudarsana, B. (2014). PENGARUH PERUBAHAN JARAK OBYEK KE FILM TERHADAP PEMBESARAN OBYEK I . PENDAHULUAN Radiografi

penunjang merupakan yang sarana sudah dengan tujuan untuk Pemanfaatan sinar-X dalam radiodiagnostik oleh fisika medis sangat menunjang untuk memperkuat diagnosa . 15(2), 15– 21.

[9]   Listyalina, L. (2017). Peningkatan Kualitas Citra Foto Rontgen Sebagai Media Deteksi Kanker Paru. Jurnal Teknologi Informasi, XII(34), 110– 118. http://jti.respati.ac.id/index.php/jurnal jti/article/viewFile/1/1.

[10]   Basuki, Achmad. 2005. Pengolahan Citra Digital Menggunakan Visual Basic. Graha Ilmu. Yogyakarta.