Thesis for the Degree of Master of Engineering

Convolutional Neural Network for Photoacoustic Imaging Segmentation Applied in Human Study

by

Vu Thi Thu Ha

Department Industry 4.0 Convergence Bionics Engineering

The Graduate School

Pukyong National University

August 2022

Convolutional Neural Network for Photoacoustic Imaging Segmentation Applied in Human Study

인간 대상 연구에 적용된 광음향 영상 분할을 위한 컨볼루션 신경망

Advisor: Prof. Junghwan Oh

by

Vu Thi Thu Ha

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Engineering

in Department of Industry 4.0 Convergence Bionics Engineering,

Pukyong National University

August 2022

Convolutional Neural Network for Photoacoustic Imaging Segmentation Applied in Human Study

A dissertation

by

Vu Thi Thu Ha

Approved by:

Prof. HaeGyun Lim, Ph.D.
(Chairman)

Prof. Byeong-il Lee, Ph.D.
(Member)

Prof. Junghwan Oh, Ph.D.
(Member)

August 26th, 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

인간 대상 연구에 적용된 광음향 영상 분할을 위한 컨볼루션 신경망

부티투하

부경대학교 4차산업융합바이오닉스공학과

요 약

광음향 현미경은 레이저에서 생성된 광음향 효과를 기반으로 하는 하이브리드 방식의 생물학적 이미징 기술로, 피부층, 정상 또는 손상된 혈관, 그리고 종양 등의 이미징이 가능함. 주요 부위의 시각화 및 분석을 용이하게 하기 위해 이미지 세분화 기술이 사용되고 있으며, 이는 당뇨병, 피부과, 종양학 및 신경과와 같은 질병의 보다 정확한 진단과 직접적으로 관련되어 있는 중요한 기술임. 그러나 전통적인 기법들을 기반으로 하는 기존의 연구에서 피부와 혈관에서 확보한 광음향 신호를 분할하는 데 여러 가지 어려움에 직면했으며 이들은 대부분 노이즈, 얕은 이미징 깊이 및 낮은 이미지 해상도와 관련이 있음.

이에, 본 연구에서는 광음향 이미지의 각 픽셀을 개별적으로 분할하기 위해 U2-Net 신경망을 제안함. 본 신경망은 트레이닝을 위해 원본 B-scan 의 데이터 세트와 레이블 데이터 세트를 기반으로하고 세그먼트 영역을 생성하기 위한 예측을 수행하는 역할을 함. 해당 연구에서는 인간 대상 연구를 통해 확보한 손의 데이터로부터 혈관과 피부층을 분할하고. 해당 실험의 결과, 제안된 U2-Net 신경망 모델이 모든 실험(인간 손)의 데이터 세트에 대해 높은 광음향 영상 분할 성능을 보였으며, 평균 픽셀 정확도는 인간 손의 경우 ~94%에 도달했음. 이를 통해 본 연구에서 수행된 방법을 적용함으로써 임상 수준의 광음향 영상 진단 응용에 있어 향상된 품질의 광음향 영상을 제공할 수 있음을 확인했으며, 이는 방법이 용이하고 성능 또한 우수한 유망한 이미징 처리 기술이 될 수 있음을 시사함.

CONVOLUTIONAL NEURAL NETWORK FOR PHOTOACOUSTIC SEGMENTATION APPLIED IN HUMAN STUDY

Vu Thi Thu Ha

Department of Industry 4.0 Convergence Bionics Engineering, The Graduate School
Pukyong National University

ABSTRACT

Photoacoustic microscopy (PAM) is a hybrid modality of biological imaging (skin layers, normal or damaged blood vessels, and tumors) technique based on the photoacoustic effect generated by Lasers. To easily visualize and analyze the significant regions, now a day's segmentation techniques are used more frequently. This technique is directly associated with the more accurate diagnosis of human diseases such as diabetes, dermatology, oncology, and neurology. However, many studies using classical segmentation methods have encountered several limitations in the segmentation of photoacoustic signals obtained from the targeted tissues, mostly associated with noises, low image depth, and poor image resolution.

This thesis proposes an U^2-Net neural network to segment each pixel in the PA image individually. This method is based on the original B-scan dataset and label dataset for training, followed by making predictions to generate segment regions. The study was performed on human hands to segment blood vessels and skin layers, whereas mice models were used to segment the tumor and healthy blood vessels. The following experimental results show that the U^2-net model could achieve a good PAI segmentation performance for all the experimental (human hand) datasets, where the average pixel accuracy reached ~94%. Therefore, this method could be a promising easy-to-process technique for PAM imaging with enhanced quality for clinical-level photoacoustic imaging diagnosis application.

CHAPTER 1: INTRODUCTION

Photoacoustic (PA) microscopy (PAM) is a hybrid modality of biological imaging (skin layers, normal or damaged blood vessels, and tumors) technique based on the photoacoustic effect generated through the photoacoustic effect [1]. It means a short-pulsed laser (a few nanoseconds) is absorbed by biological tissues making an ultrasonic emission phenomenon. PAM is known as a technique that detects ultrasonic emission when cells within tissues absorb a short-pulsed laser, which is called PA [1][2]. PAM provides high spatial resolution by using raster scanning of PA [1][3]. In recent years, the fast scanning of PAM becomes more popular in biomedical research. Especially, PAM has a good optical absorption contrast and high (diffraction-limited) spatial resolution, which also has an important role in diagnosing and studying diseases such as diabetes, dermatology, and neurology.

Besides the development of PAM, photoacoustic imaging (PAI) has particular attention because of the advances in reconstructing images with high resolution in deep tissue. In the PAI of humans, after using a laser for scanning the skin, the internal tissue is hidden. Therefore, the segmentation of blood vessels and skin structure is an important stage of subcutaneous microvasculature. Each B-scan image has complete data on the skin and blood vessels [5]. However, manual identification is difficult and takes a long time if the number of B-scan images is up to hundreds or thousands. Many results of studies have proved that manual segmentation between blood and skin is not effective [6][7]. Therefore, this thesis applies a method named deep learning (DL) in skin and blood vessel segmentation.

Deep learning is a part of ML - machine learning, that is essentially a neural network with multiple layers. Recently, DL was rapidly developing and widely applied to image processing to get the higher-level features. DL approaches such as convolutional neural networks also brought positive strides to medical image processing in general, including PA images. Geert Litjens et al, had an analysis of the application of AI, and DL in medical images [8]. With PA imaging, Yang et al, and Deng et al, already summarize some DL methods and their applications in PA image segmentation, PA image detection, and PA image reconstruction [9][10][11][12].

Here, we use a neural network (U^2-net technique) that can simply segment each pixel in the PA image individually. Our method will be based on the original B-scan dataset and label dataset for training, then make predictions to segment blood vessels

and skin. When testing the result of our model, we use only one dataset to compare our method with the U-net method that Ly et al, used in their paper [2]. In addition, Segnet network and PSPNet network also used to evaluate the accuracy of U^2-net. The method we proposed gives some outstanding results such as higher accuracy and reducing training time, which is described in detail in the Result part.

CHAPTER 2: MATERIALS AND METHODS

1. Network architecture

1.1. Residual U-blocks

Unlike CNN architectures (ResNet or DenseNet, and so on), the size of convolutional filters such as 3×3 or 1×1 is usually used in for feature extraction. However, the output of the field of 1×1 or 3×3 filters is too small to capture global information. To get high-resolution information from shallow layers, we need to enlarge the receptive field. However, this solution needs too many memory resources.

Therefore, to solve the problem, a novel ReSidual U-block (RSU) is proposed by Xuebin Qin et al. [13]. The structure of RSU is shown in Figure 1.

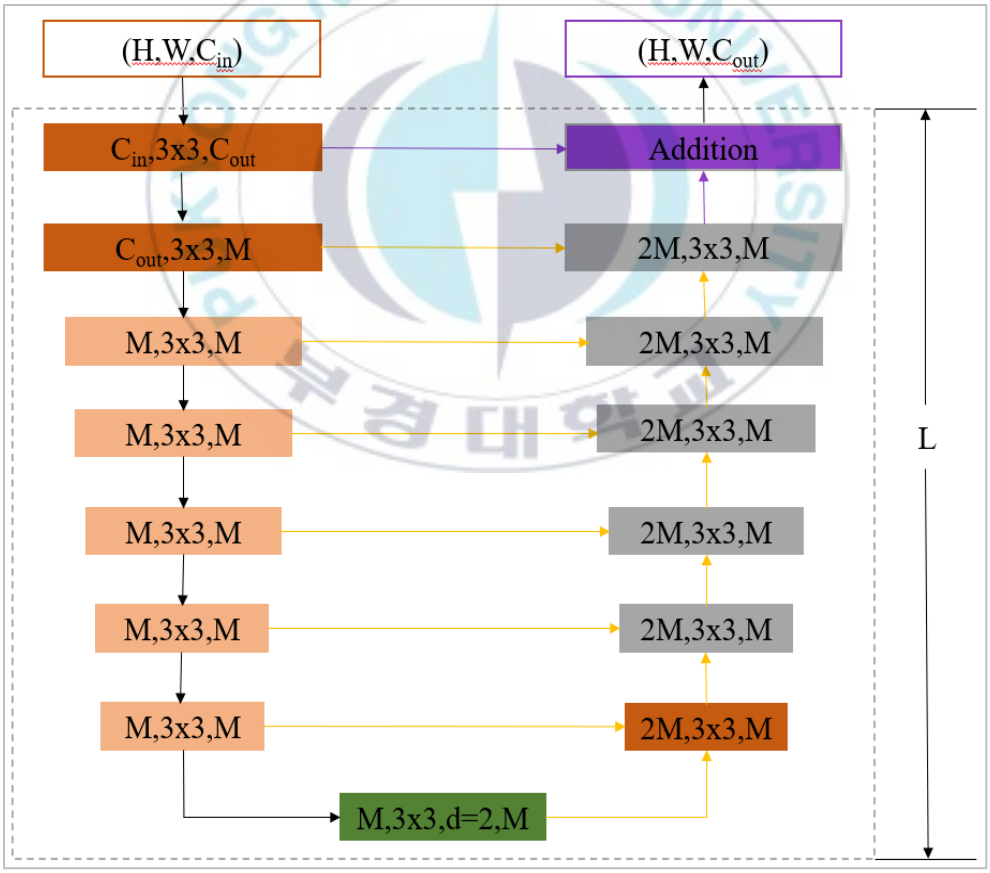


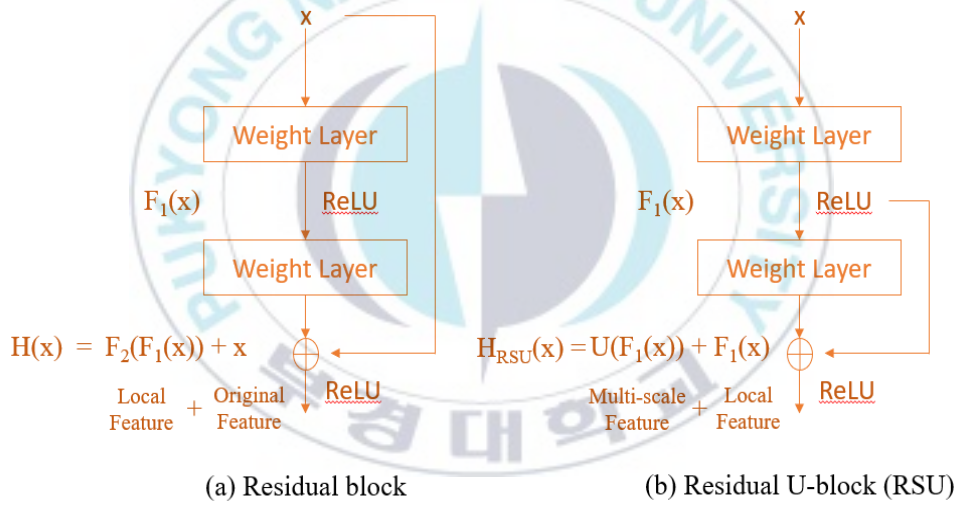
Figure 1. The structure of ReSidual U-block (RSU)

With each RSU, a U-net structure is used likes an encoder-decoder. The structure of RSU-L (C_{in}, M, C_{out}) where L is the number of layers in the encoder, C_{in} is input channels, C_{out} is output channels, and M is the number of channels in the internal layers of RSU. There are three main parts in RSU:

(1) H x W x C_{in}: an input convolution layer will be fitted with an intermediate map F1(x) of C_{out}.

(2) a symmetric unit like a U-shaped structure with height L will take a map F1 as input and then learn and extract and encode for U(F1(x)). When L is larger, the pooling operations and the range of receptive fields are larger and richer local and global features.

(3) the last connection for multi-scale features and fuses local features: F1(x) + U(F1(x)).



(a) Residual block (b) Residual U-block (RSU)

Figure 2. The difference between the residual block and RSU

The original residual block: $H(x) = F_2(F_1(x)) + x$

Here,

- H(x): the desired mapping of the input features x.
- F1, F2: the weight layers.

The residual U-block (RSU): HRSU(x) = U(F1(x)) + F1(x)

Here:

- U: the multi-layer U-structure.
- F1: the weight layers.

From each residual block of multiple scales, this network can extract many features.

1.2. U^2-net

Many variations of U-net were developed for a while such as CU-Net [13] and DocUNet [13]. Most networks will use the structures stacked U-block with n as the number of U-block repeats. However, U×n-net is a different formulation: instead of using stacking n layer, it will use nested U-structure. That is the reason why it has exponential notation. Theoretically, we can set n as an arbitrary positive integer. But in real applications, when n is getting bigger, the network will get more complicated and difficult.

U^2-net means that n is set to 2. There are 11 stages of the U-structure in this U^2-net network. The illustration of U^2-net architecture is shown in Figure 3. This architecture including of three parts: 6 stages of the encoder layer, 5 stages of the decoder layer, and a composite module.

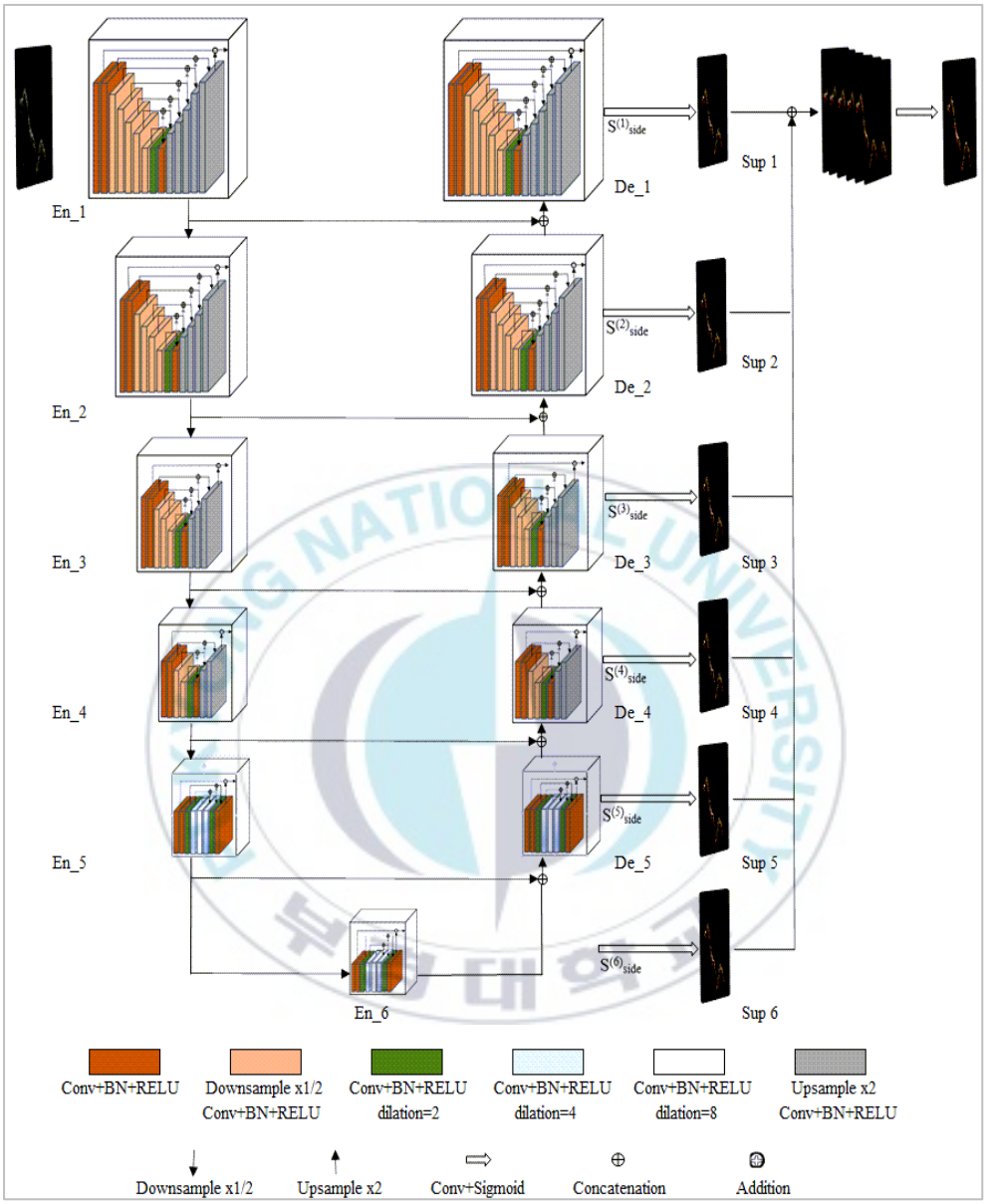


Figure 3. The illustration of U2-Net architecture (Each stage consists of RSU)

Figure 4 to Figure 9 show the details of each Encoder, Decoder from 1 layer to 6 layer:

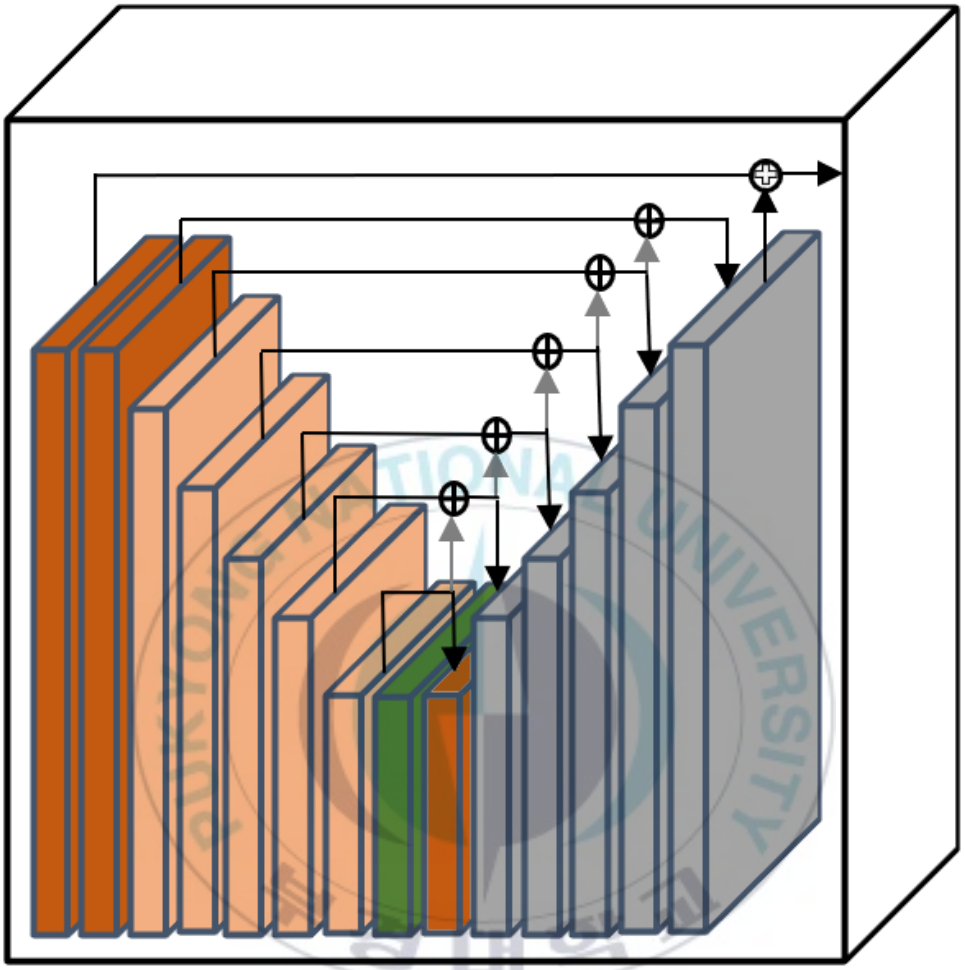


Figure 4. The architecture of Encoder_1

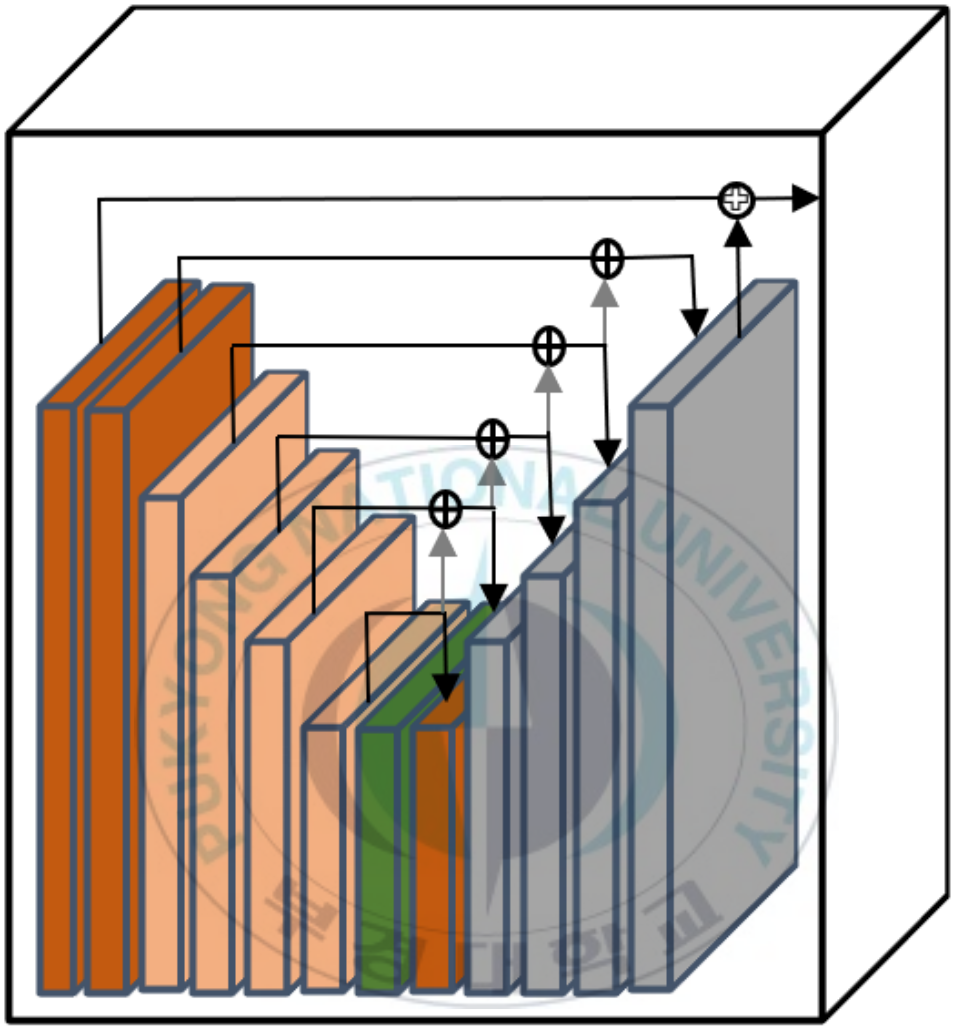


Figure 5. The architecture of Encoder_2

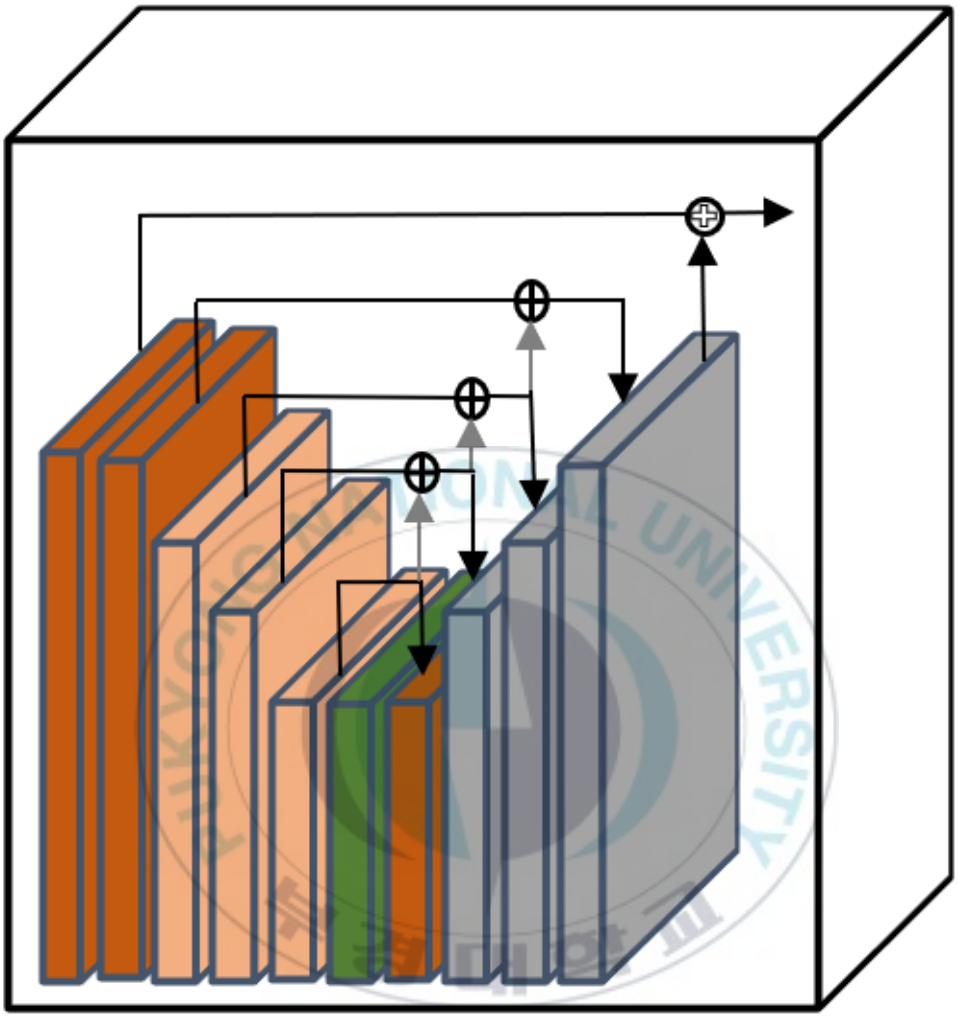


Figure 6. The architecture of Encoder_3

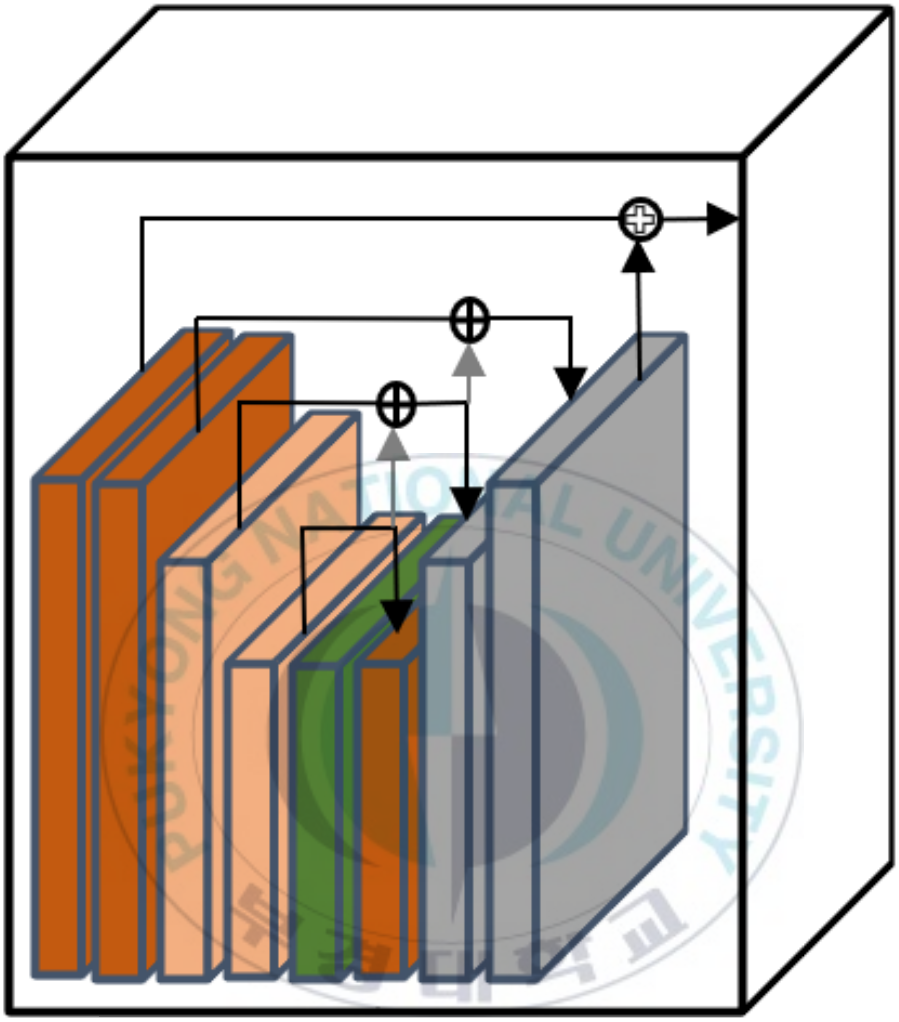


Figure 7. The architecture of Encoder_4

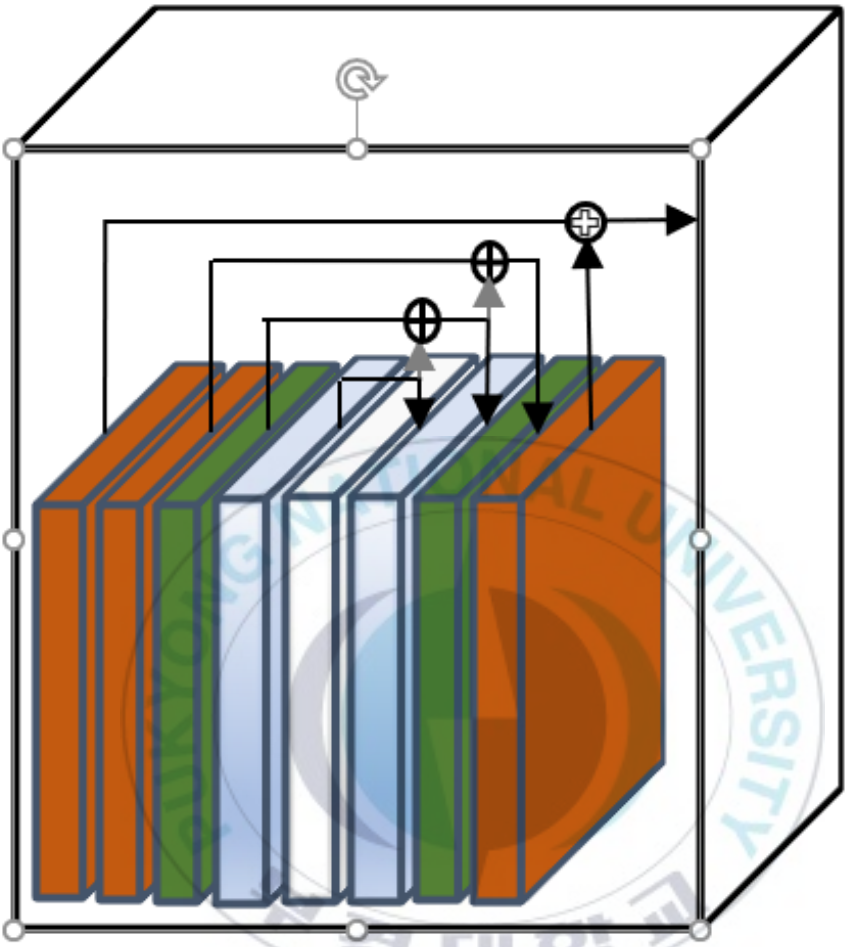


Figure 8. The architecture of Encoder_5

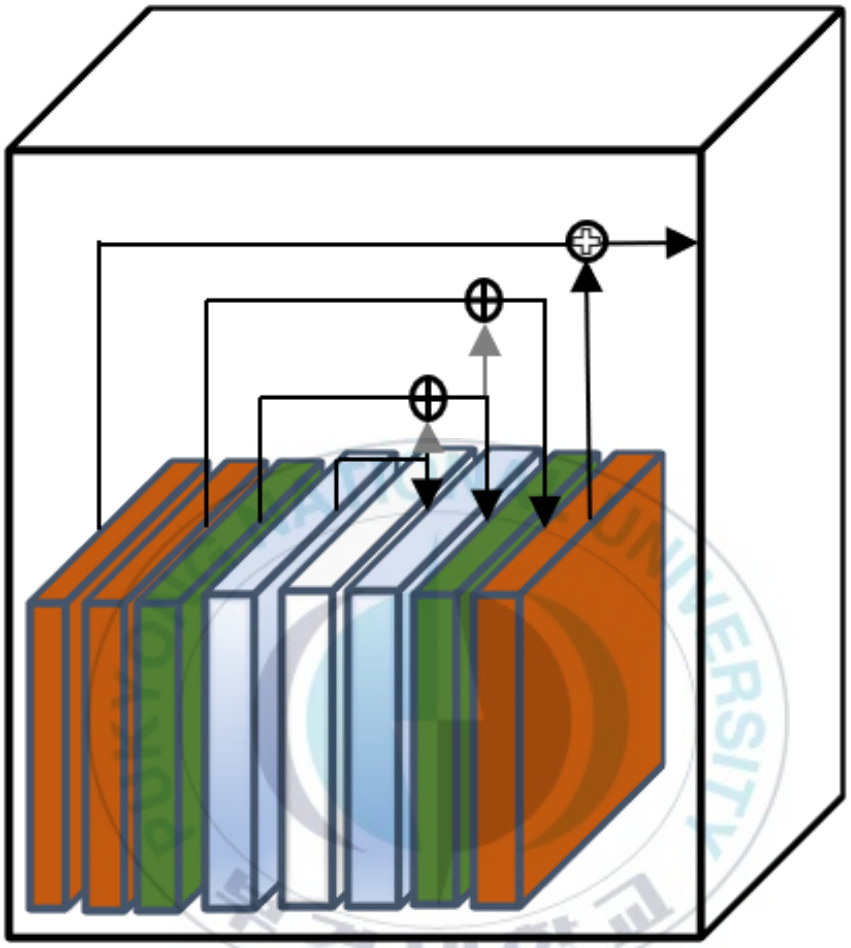


Figure 9. The architecture of Encoder_6

Both encoder and decoder stages have the same structures. Encoder stages use the residual U-block (RSU) (inner layer) [14]. From stage one to stage four, RSUs are applied to increase large-scale information [13]. In the Encoder 5 and Encoder 6 stages, max pooling and up-sampling are changed to the dilated convolutions. These two steps are important in the case the receptive field is expanded and input feature maps stay the same. With the Decoder 5 stage, it is equal to Encoder 5 and Encoder 6 stages. The Decoder 5 stage is the combination of up-sampled feature maps from its previous stage (Encoder 6) and those from its symmetrical encoder stage (Encoder 5) [13]. This process is similar to other decoder stages Decoder 4, Decoder 3, Decoder 2, and Decoder 1.

The final feature map is the fusion connected from the six-side output on the model

body S^6side, S^5side, S^4side, S^3side, S^2side, and S^1side. After up sampling the input size, the last output result is a combination of the input image and a concatenation operation [13].
In summary, the U^2-net structure has a good architecture with outstanding features as well as a memory architecture for a low cost [13]. In addition, U^2-net is only using RSU blocks instead of using pre-trained backbones, therefore it allows the algorithm to prune the unrelated background volumes and make it easy to apply in different environments.

2. Preparing dataset

2.1. Experimental setup

We used an acoustic-resolution PAM system for this experiment as illustrated in Figure 10. During all periods of experimenting, we were complying with The Institutional Review Board of Pukyong National University. In the process of doing the experiment, the hand of a volunteer is scanned by fast scanning of the PA probe.

Figure 10. Experimental setup

The scanning system used the 532 nm laser under the American National Standards Institute standards safety limit (20 mJ/cm2 for 532 nm wavelength) [15]. The frequency 50 MHz of the Olympus transducer is used to receive the signal [2]. The sampling rate of the

ultrasound signal was 200 MHz and was directly fed into A-scan (a time series data line). Lastly, the LabVIEW internal software is used to control the motion of the transducer, turn on/turn off the laser, and receive data.

2.2. Data preparation

The data that we get after experimenting including of 800 *tdms* files. Because we will process on Two-dimensional image, therefore the format of the root data is converted from *tdms* to *png* file for easy handling. We manually delineated the outer boundaries for the blood with the ground truth dataset, which we developed in python. To get the skin dataset, we use the original image and the manual label image through the math on the 2D matrix.

This ground truth dataset is a grayscale image and is normalized to (0-1). The training dataset preparation is divided into n classes corresponding to n input datasets. Therefore, our dataset includes original data, vascular segmentation images, and skin segmentation images. All images were resized into 512 x 512 pixels, then we used data augmentation such as flipping, rotational transform, and lateral and vertical shift to expand our dataset. The dataset was divided into the training set (70 images), the validation set (30 images), and the testing set (10 images).

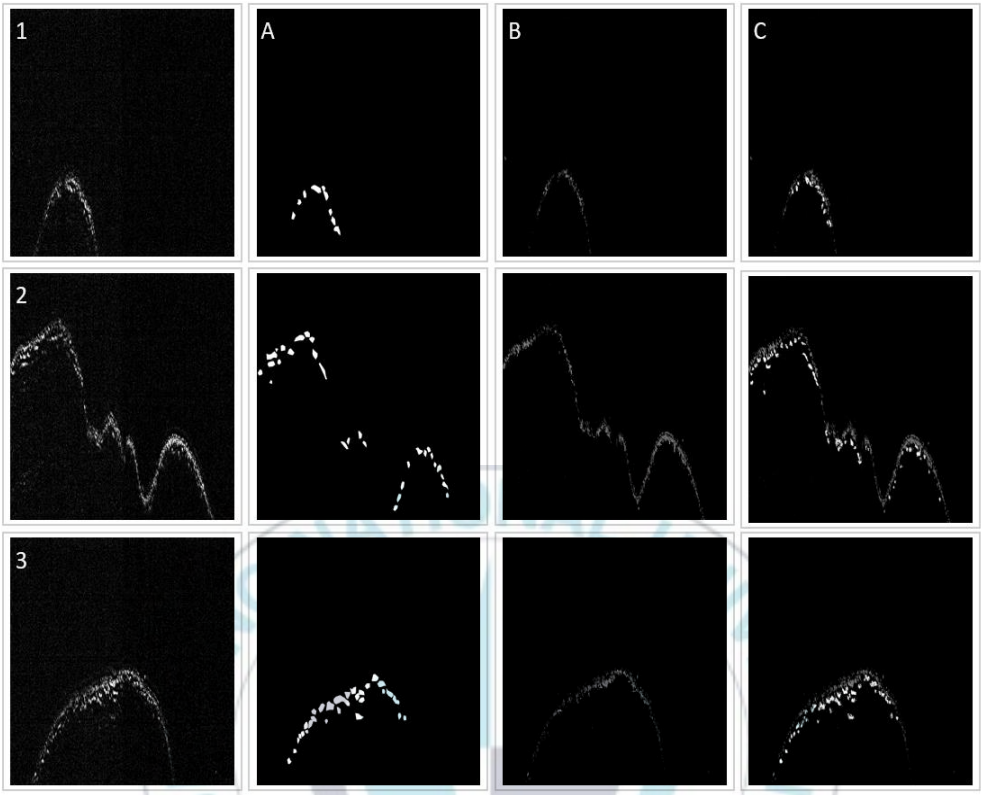


Figure 11. Th sample of the dataset. (1) Original image; (A) Label of blood vessels; (B) Label of skin layer; (C) Merger label of blood and skin

The details of the dataset shown in the pictures below:

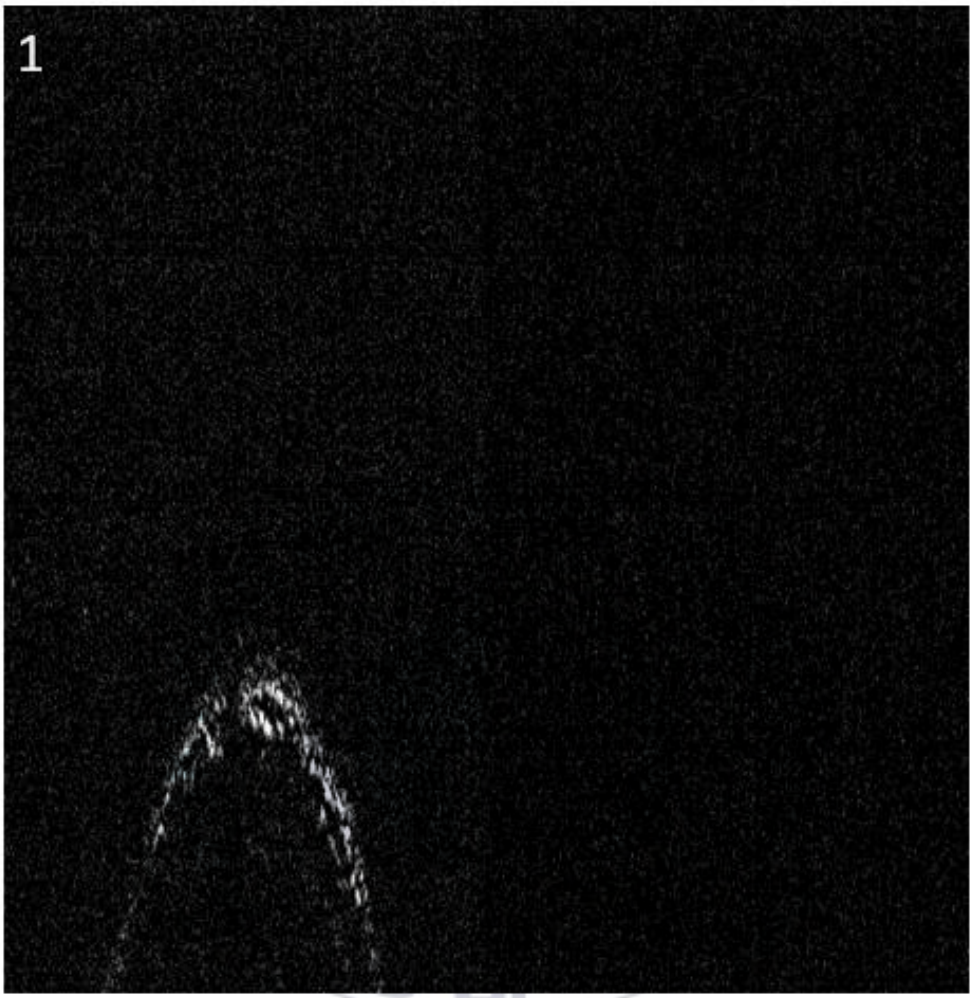


Figure 12. Image of the original B-scan image

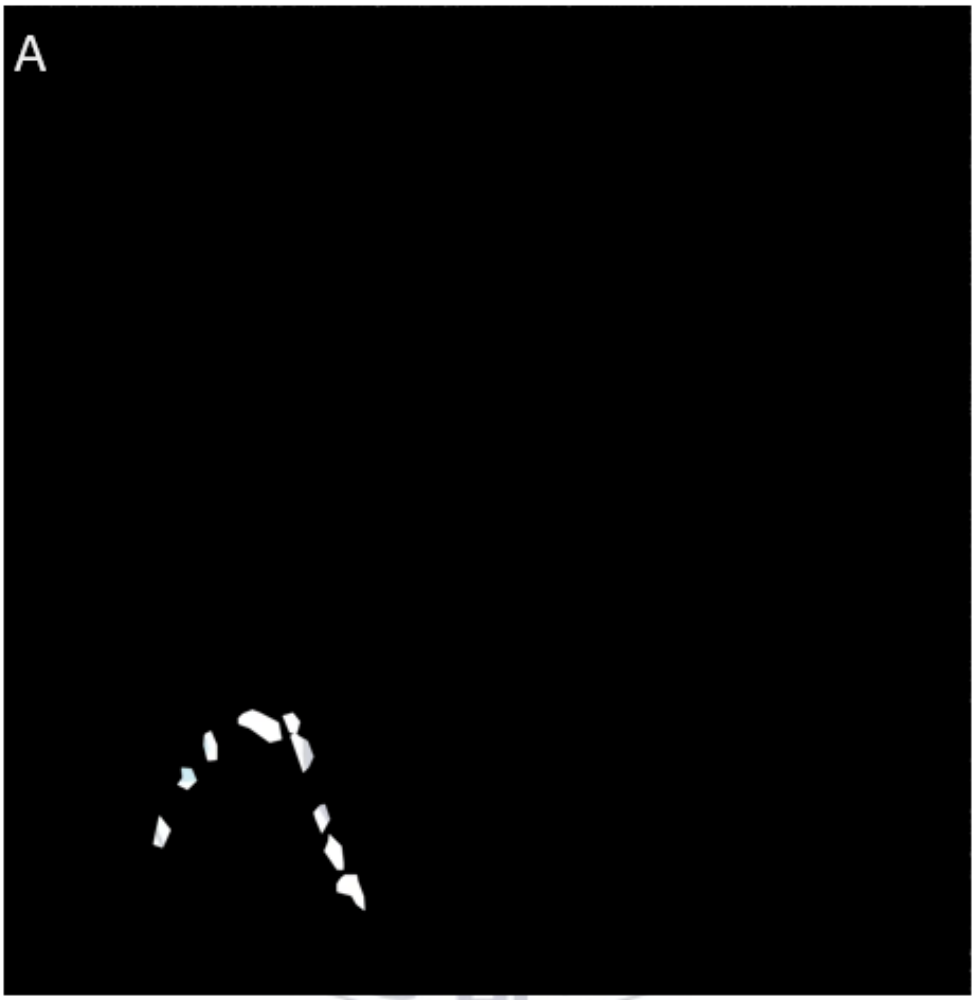


Figure 13. Image of blood vessels labeled

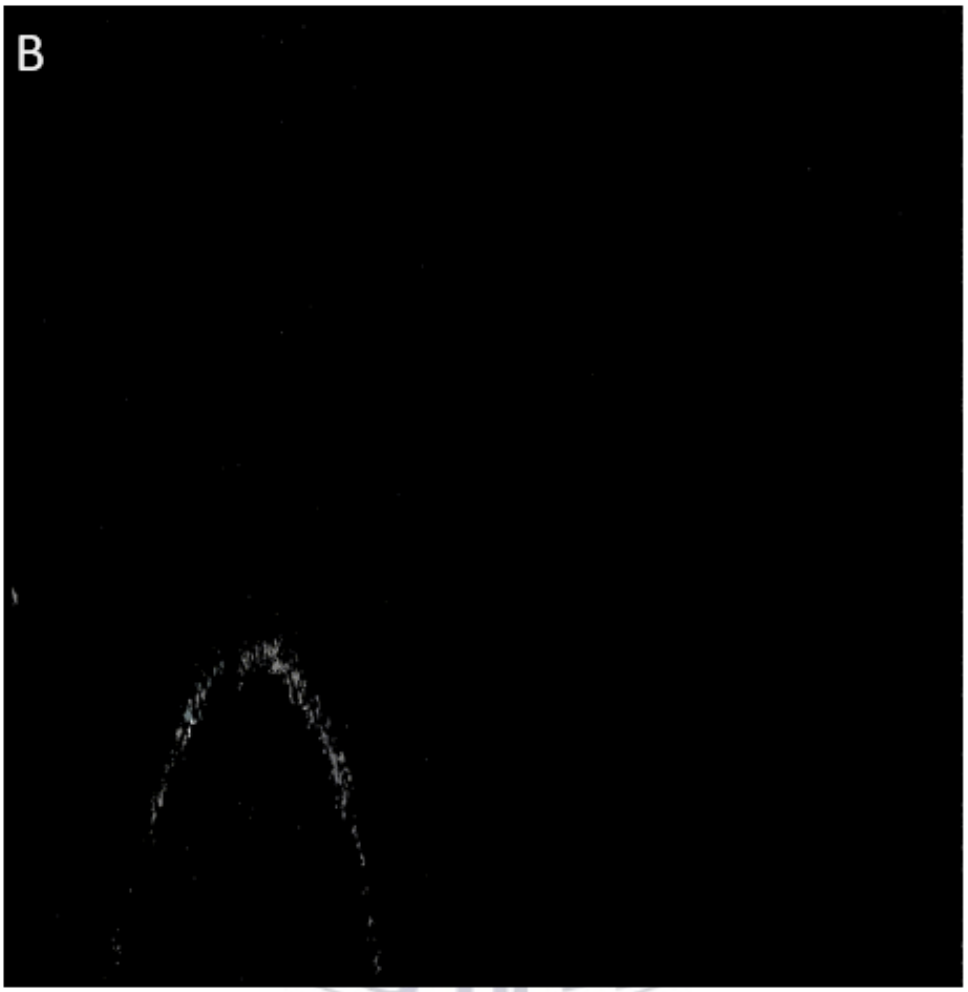


Figure 14. Image of skin layers labeled

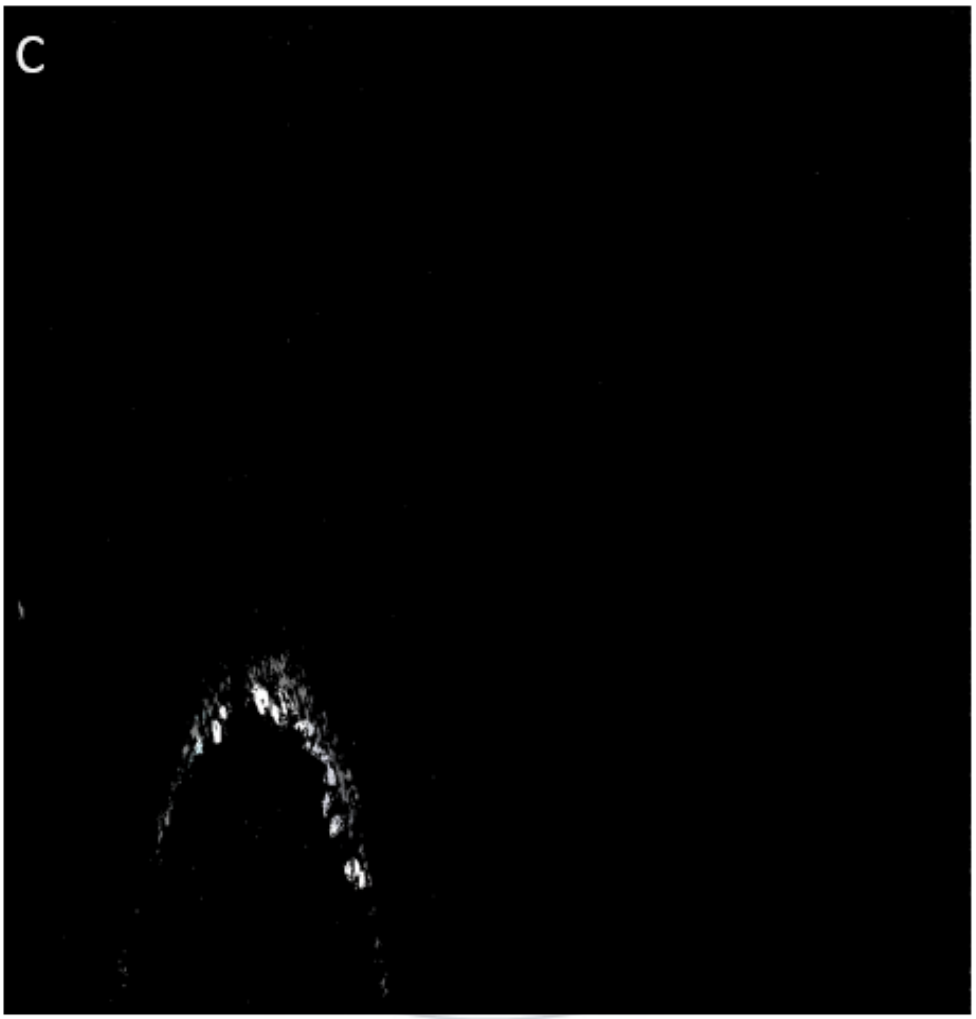


Figure 15. Image of the merger between blood vessels and skin layers labeled

3. Training and Testing

To measure the accuracy of the model, we use the method training set and testing set. Training the model means creating the model while testing means testing the accuracy of the model. We use 70% train dataset and 30% test data set with our model. Here, we also used validation data because it provides the estimate of model skill while tuning model's hyperparameters, which training data cannot do it.

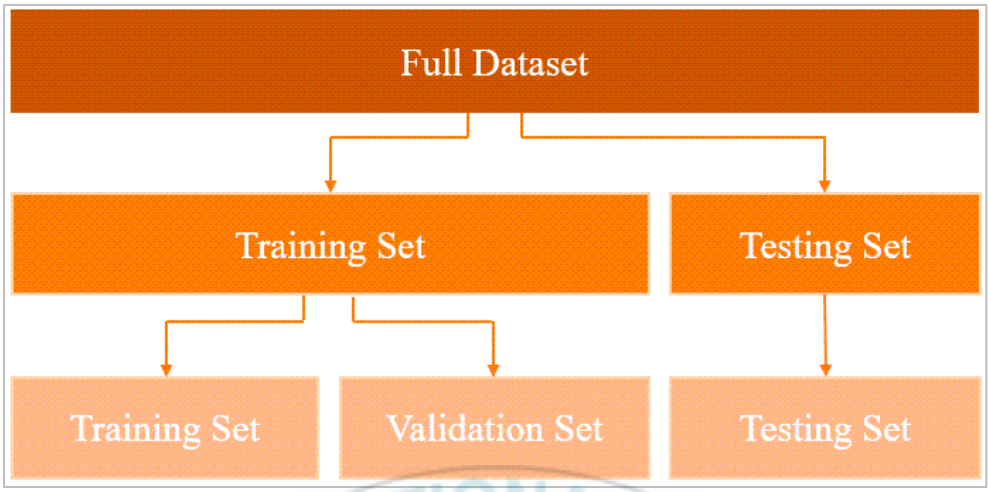


Figure 16. Data set division

Because we do not have any pre-trained model, therefore we started training from scratch. The hyperparameter settings of the U^2-net model are the input image size (512x512 pixels), batch size (4), learning rate (0.0001), and the number of epochs is 50. However, in real-time training, we got the overfitting problem. Overfitting is Overfitting is a mistake that happened when a training function is too fit with data points. Therefore, the model just gets high accuracy in its initial data set, with other data sets it may not good enough.

Figure 17. Sample of overfitting

To resolve the overfitting problem, we use the early stopping technique. Early stopping means that we set the number of epochs as *a* and if the loss does not decrease

in *a* epochs, the training will stop. All the processes are implemented with PyTorch [16]. The training model requires 45 minutes on the computer Intel Core i7-6700, RAM (Random Access Memory) 8GB, and NVIDIA 1060 with 32 GB memory.

4. Loss and Evaluation methods

The loss function defines an important method that is used to determine the error between the output of the algorithm and the given target value. It plays what a good prediction is and isn't. In this paper, we focus on a common loss function named Dice Loss. Dice loss is proposed by Milletari in 2016, which is a measure of overlap used to evaluate segmentation performance when having the ground truth.

$$\textbf{Dice loss} = 1 - \frac{\Sigma_{n=1}^{N}(p_n r_n + \epsilon)}{\Sigma_{n=1}^{N}(p_n + r_n + \epsilon)} - \frac{\Sigma_{n=1}^{N}(1-p_n)(1-r_n) + \epsilon}{\Sigma_{n=1}^{N}(2 - p_n - r_n + \epsilon)} \quad (1)$$

Here, r is the reference foreground segmentation (gold standard) with voxel values r_n; p is the predicted probabilistic map for the foreground label over N image elements p_n; and ϵ is used here to ensure the stability and avoids the emptiness of r and p.

The segmentation challenge is evaluated through parameters, which are defined in formulas below. First is the average accuracy is defined as follows:

$$\textbf{Accuracy (\%)} = \frac{\Sigma_{i=0}^{k}(P_{ii})}{\Sigma_{i=0}^{k}\Sigma_{j=0}^{k}(P_{ij})} \quad (2)$$

Here, k are types of classification, i and j are different classification; p_{ii} indicates correctly classified pixels in the image; and p_{ij} represents misclassified pixels in the image.

$$\textbf{IoU} = \frac{T_P}{T_P + F_P + F_N} \quad (3)$$

$$\textbf{Precision} = \frac{T_P}{T_P + F_P} \quad (4)$$

$$\textbf{Recall} = \frac{T_P}{T_P + F_N} \quad (5)$$

$$\textbf{F1-score} = \frac{2T_P}{2T_P + F_P + F_N} \quad (6)$$

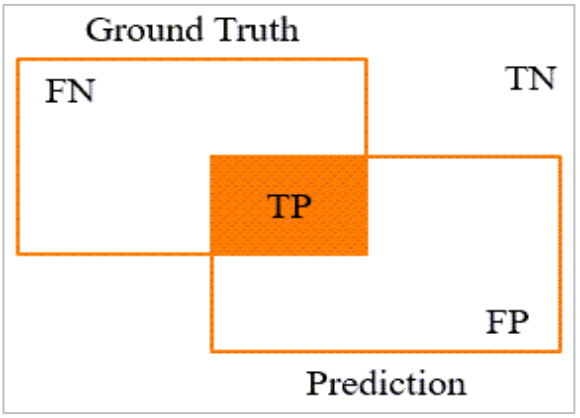


Figure 18. Intersection over Union (IoU)

Here, TP is true positive samples that is the overlap between target mask and the prediction output; TN is true negative samples in which the model predicted the negative class correctly in the output; FP is false positive samples and indicates the actual result should be false, but the predictive result is true; FN is false negative and indicates the actual result should be true, but the predictive result is false.

CHAPTER 3: RESULT

1. Model architecture result

The segmentation performance of the U^2-net model is show below.

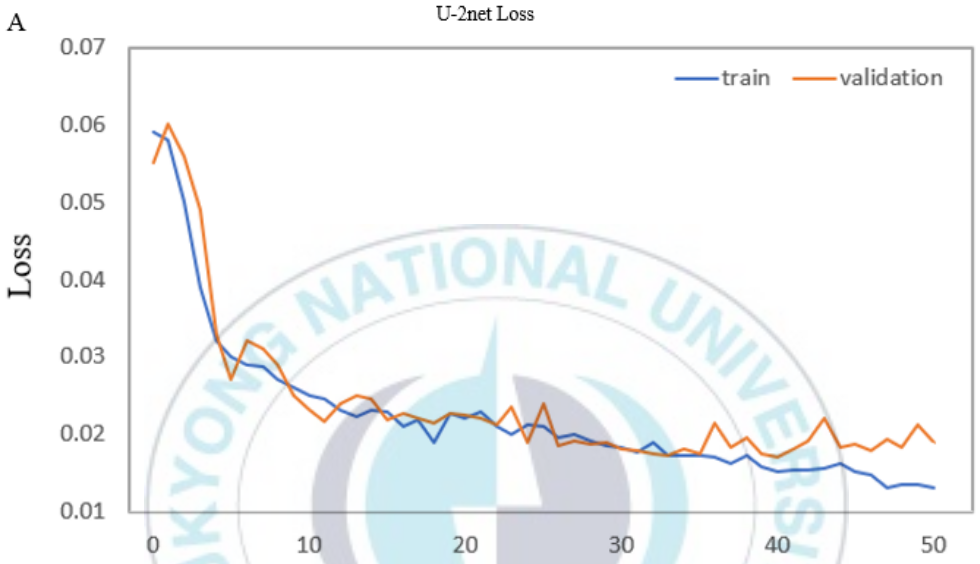


Figure 19. U^2-net Loss

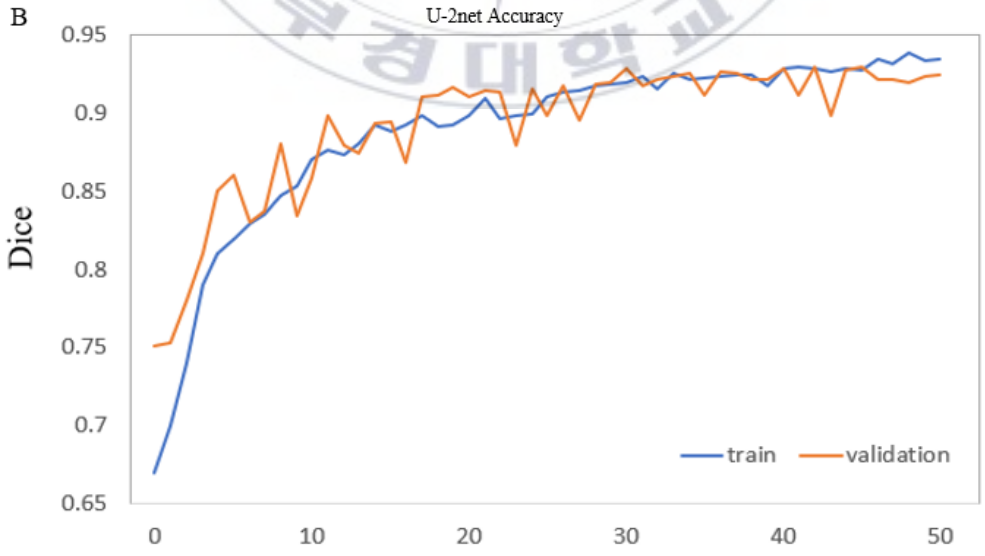


Figure 20. U^2-net Accuracy

Observing the U^2-net model proposed in this thesis, it is superior to the other models in stability and convergence speed. Loss can be fast and stable in the early training period, with a very small fluctuation range and quite good robustness. The visualization of Figure 20 and Figure 21 shows the training model result. With 50 epochs, the Loss of the training process reaches 0.0131 while the Loss of the validation process is 0.189. The average accuracy of the training process is 0.9349~0.94% while the average accuracy of the validation process is 0.9242.

To assess the segmentation quality results, we also build a U-net model following the research of Ly [2], SegNet model [21], and PSPNet model [22]. Then we trained and tested with the same PAM dataset. Table 1 shows the experimental results between our model and other models on PAM dataset. As the most effective and popular models for segmentation purposes, the training and validation accuracy of U^2-net is 94% while U-net, SegNet, and PSPNet are 88%, 62%, and 51%, respectively. Moreover, the training time of U^2-net is better (requirement of 280 minutes for 50 epochs while U-net, SegNet, PSPNet needs 320 minutes, 270 minutes, and 295 minutes, respectively). U^2-net also outperforms other networks in metrics such as IoU, Recall, and F1-score (as show in Table 1).

Table 1. The experimental result on the testing dataset

	U^2-net	U-net	SegNet	PSPNet
Average Accuracy (%)	94	88	62	51
IoU	64.8	61.1	48	43
Recall	82.5	76.8	57	55
F1-score	78.9	75.6	50	47
Training time (min)	280	320	270	295

As seen in Table 1, with all indicators, the U^2-net approach is better than U-net, SegNet, and PSPNet models. It can prove that U^2-net network has certain advantages in the segmentation of blood vessels and skin layers in PAM images.

Figure 21 shows the predicted result of four models on the B-scan dataset.

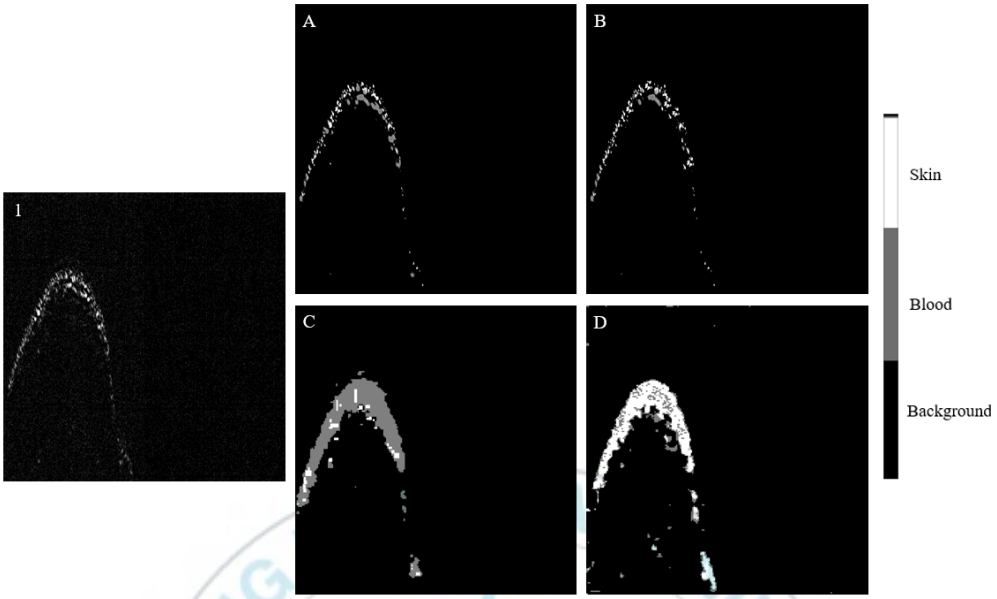


Figure 21. Visual comparison of different segmentation models on our B-scan dataset

The 1st image is our original B-scan image. Image A shows the result for U^2-net model. Images B, C, and D show the predicted images of U-net, SegNet, and PSPNet models, respectively. As seen, U^2-net model is suitable for the segmentation of blood vessels and skin layers while other models misinterpret the blood margins and skin boundaries.

Here, the detail of the original image and predicted images of models are shown:

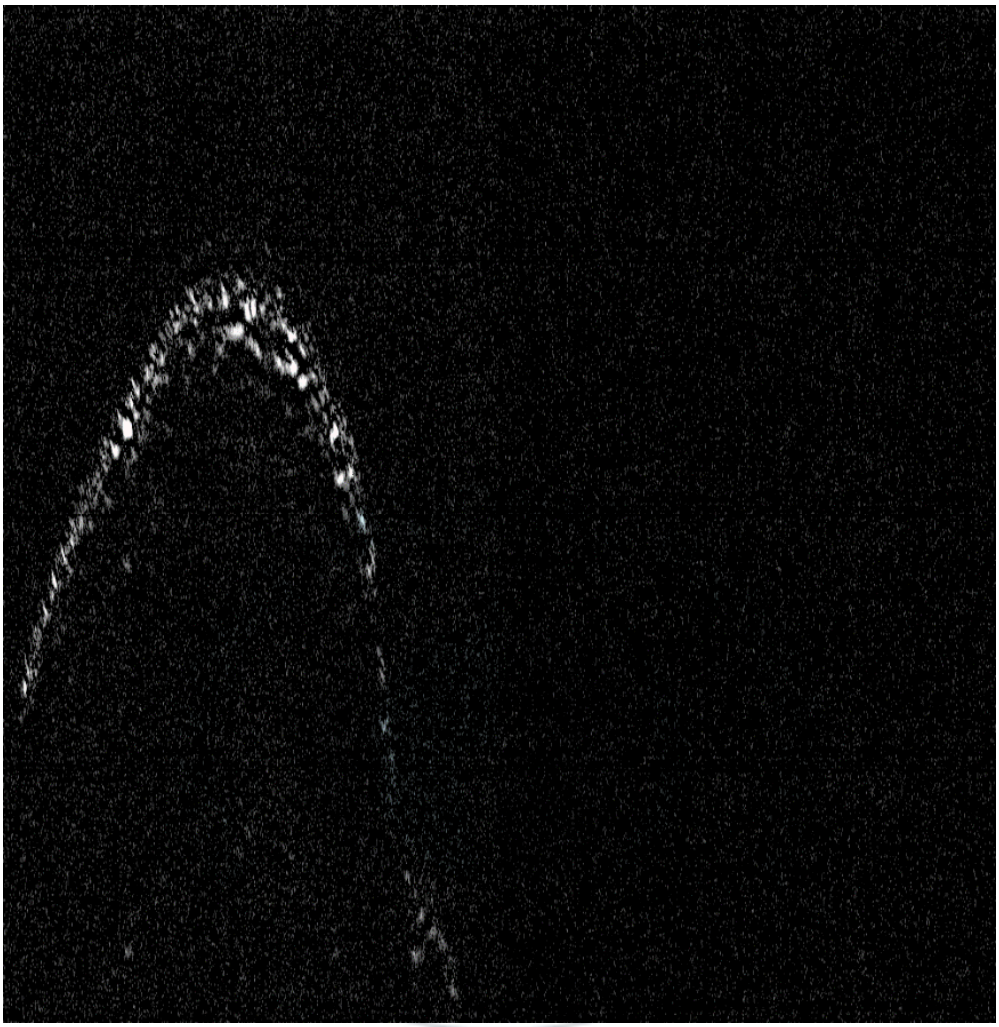


Figure 22. Original B-scan image

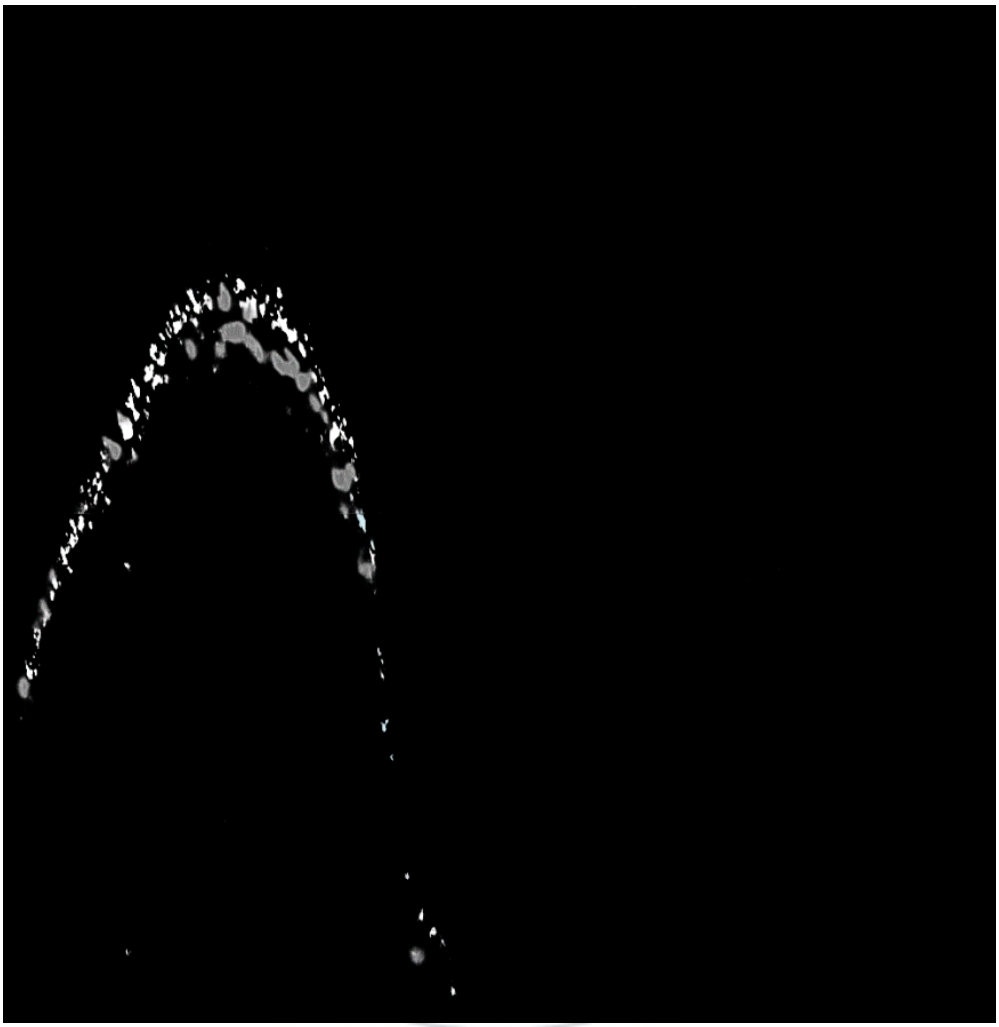


Figure 23. Predicted image of U^2-net model

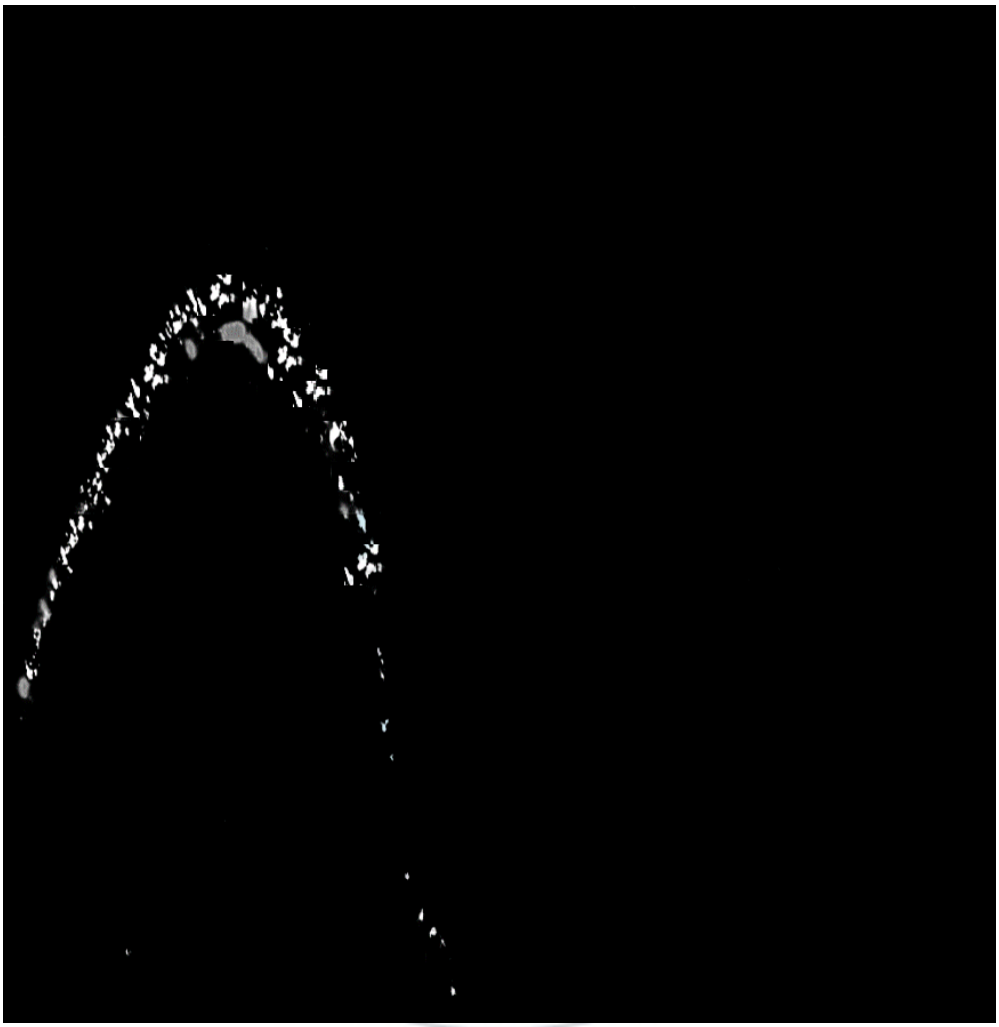


Figure 24. Predicted image of U-net model

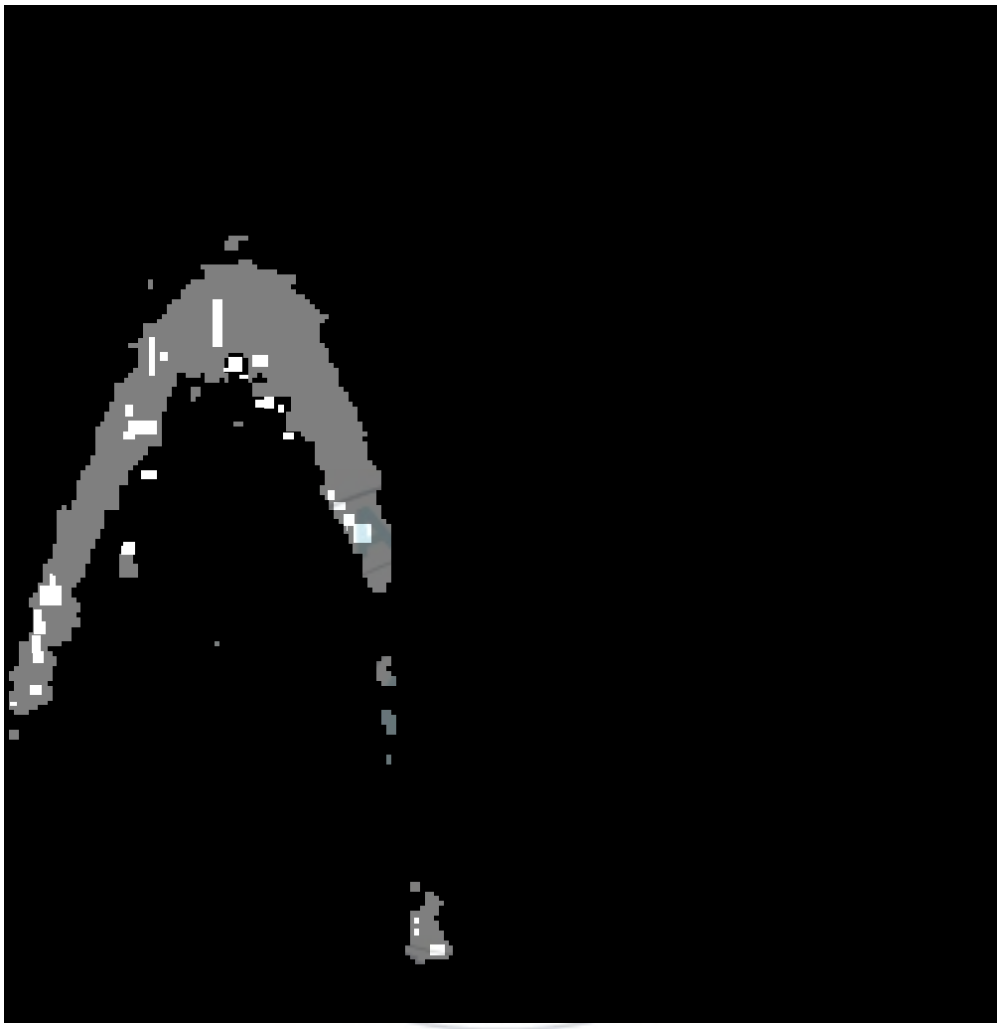


Figure 25. Predicted image of SegNet model

Figure 26. Predicted image of PSPNet model

2. A framework for visualization B-scan images

In recent years, much research has been developed to improve the quality of PAM images such as imaging speed, imaging reconstruction, or imaging segmentation [20]. We also want to support researchers who have a deep understanding of animal and human anatomy in vivo to easy visualization B-scan images between skin and blood vessel. Therefore, we developed a framework for processing and visualization B-scan

images of PAM data. Our software was built on QT5 and ran on Linux operating systems. Python programming languages are chosen because it's easy to use and supports strong libraries.

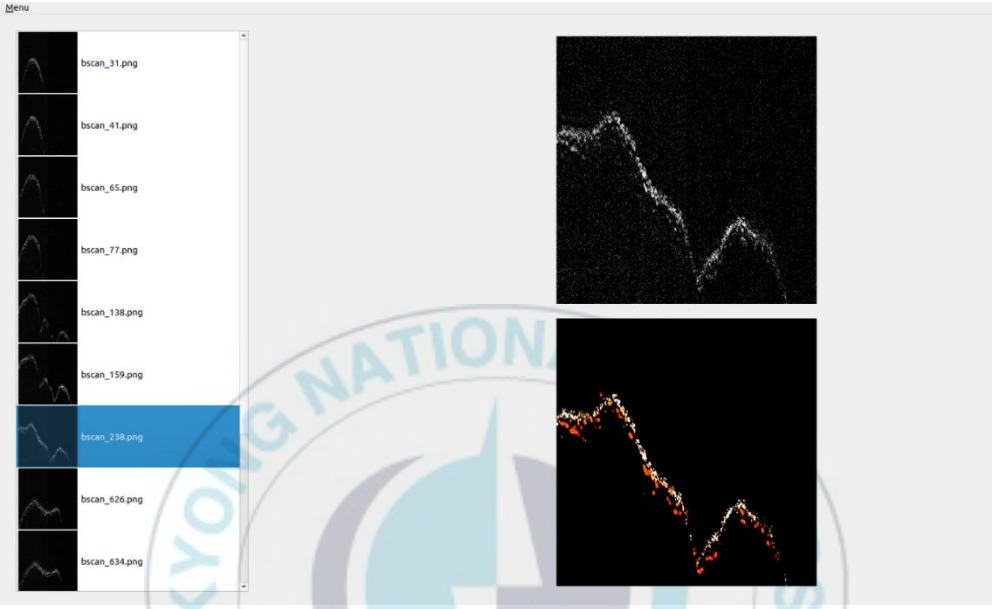


Figure 27. GUI of the framework for visualization B-scan images prediction

The user interacts with the framework through Graphical User Interfaces (GUI). It is understood that software includes graphical controls that the user can use a keyboard or mouse to handle. The software can open multiple data files through the "Open" visual icon. Then click on the button "Predict", the software will automatically call the model, which we already trained. The corresponding predict feature will be processed then shown the predicted image on the other display frame.

We believe our software is a valuable framework for researchers, promoting development in a wide range of studies. The framework was incorporated into the QT creator with an impressively user-friendly interface and fast data processing. It helps researchers have a better overview of the prediction between blood and skin. Furthermore, our products have some advantages like it is completely free, and easy to install and use. We hope our framework will become more popular for researchers

in the near future.

3. 3D Framework for photoacoustic image rendering

Because PAM has become an interesting framework that can support researchers who have a deep understanding of animal and human anatomy in vivo. Therefore, much research has been developed to improve the quality of PAM images such as imaging speed, imaging reconstruction, or imaging segmentation. However, almost all results are viewed by 2D computer graphics. With 3D PAM images, researchers need to use several frameworks that support the visualization of the illusion of depth in an image. A few companies build software for processing and rendering data for 3D images, such as Amira (Thermo Fisher Scientific, United States of America) and 3D Slicer (Free and open-source software package). Researchers usually have paid a usage fee when using these frameworks or some software features are limited in the case it is free. Therefore, we developed a 3D framework for processing and visualization of PAM data. Our software was built on QT creator, a cross-platform integrated development environment, and ran on Windows, Linux, and macOS desktop operating systems. This software includes applications: processing data and visualization, rotational the acquired image, zoom in/out, image capture, and frame recording. We hope this software will be known and used by more researchers, especially in medical image processing.

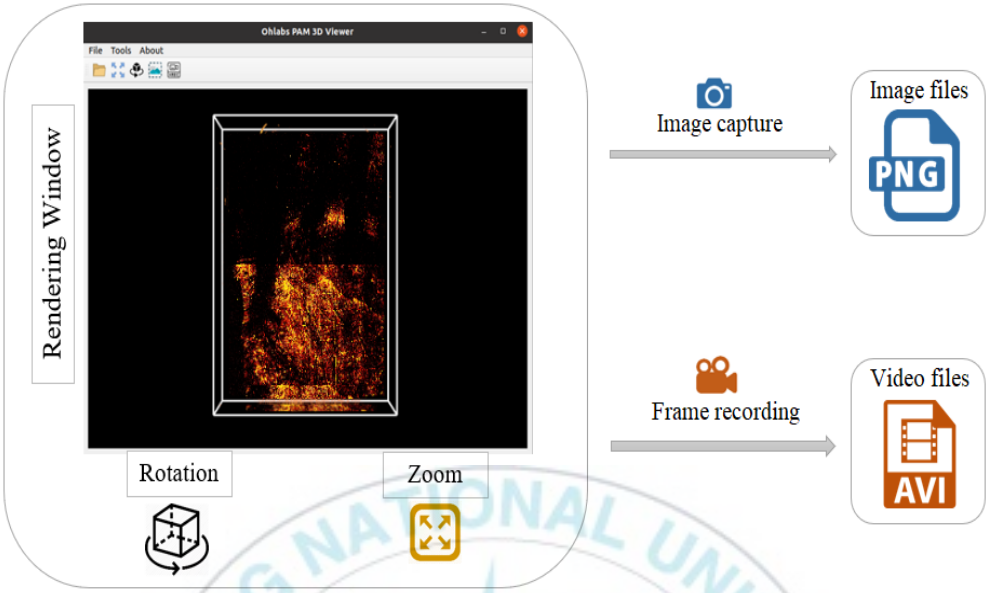


Figure 28. The user interface of the 3D framework applications

The data that we use is the same as shown on *2.2 Data preparation*. After scanning, the data will be in the form of the *.tdms* format. We processed the data to convert it to *.nrrd* format before doing visualization. This processing uses the segmentation algorithm to clearly distinguish the classes in the data (with our data segmentation for the blood vessel and skin).

The user interacts with the 3D rendering engine through the Graphical User Interfaces (GUI). It is understood that software includes graphical controls that the user can use a keyboard or mouse to handle. The interface of the program encloses a menu bar, windows, and applications buttons. Users interact with software through manipulating by visual icon, and then the corresponding features will be displayed.

The applications are presented below, acting as a guideline for the user in 3d visualization and 3D rending for photoacoustic microscopy images:

- **Processing data**: The software can open and rendering multiple data files at the same time. Users can also customize the display for each different data such as opacity and color.

- **Zoom**: The user clicks on the icon Zoom on the menu bar, the window can

zoom in/out. When zooming, it is necessary that context can be retained.

- **Rotation**: It provides a precise way to rotate the 3D view. Users can use the mouse to rotate the image in the direction which they want to view. Especially, rotation direction and rotation direction are unlimited.

- **Image Capture**: The software can take a screenshot with the press of a key. The user can choose the path to save the image and the supported format image is .png. The screen captures are automatically named and timestamped.

- **Frame Recording**: The framework can record high-quality videos through modern algorithms, which allow us to record any time with one click mouse.

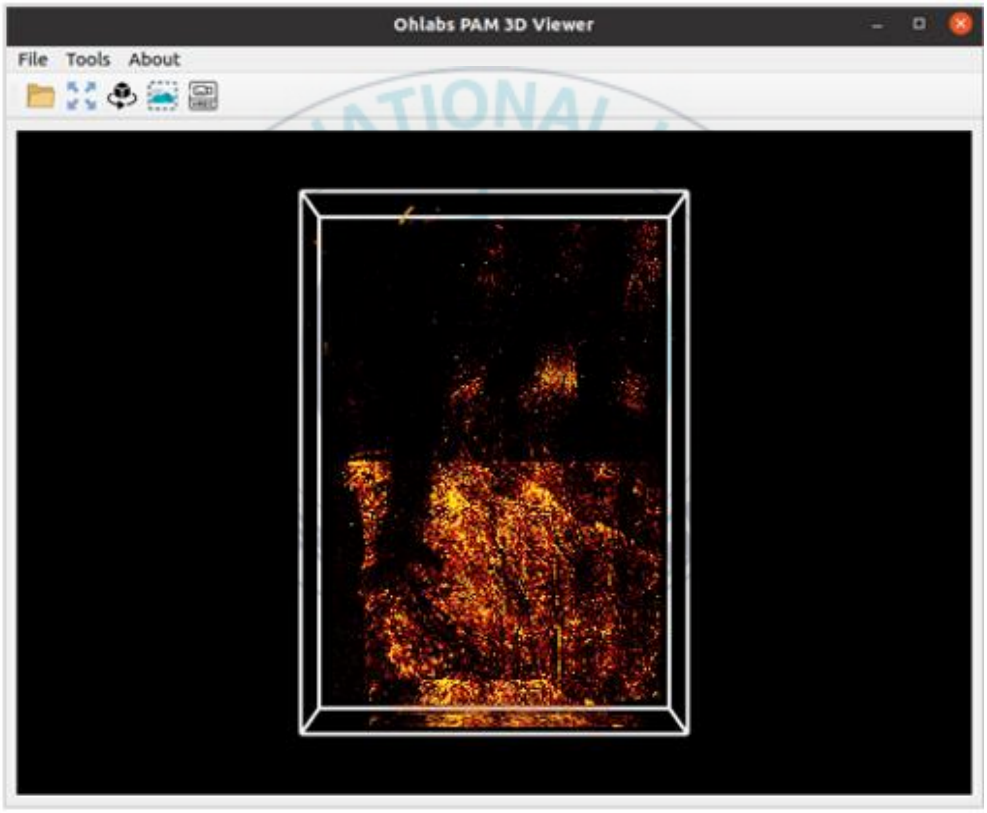


Figure 29. GUI of the framework for visualization of 3D photoacoustic image

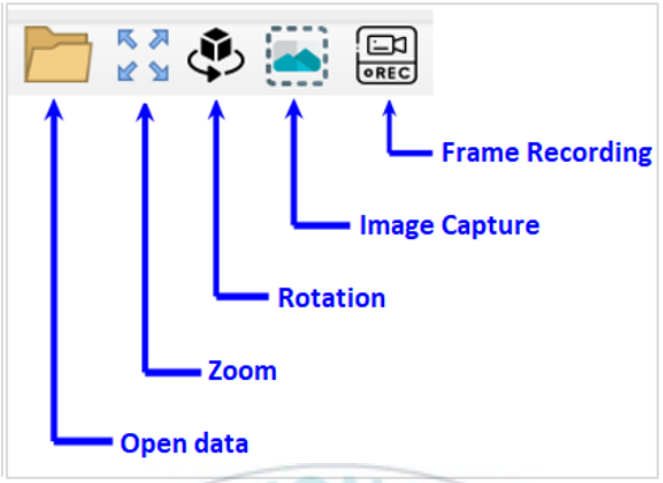


Figure 30. Application buttons

The trend of visualization data is getting developed rapidly in the photoacoustic microscopy image field. Recently, there are many new algorithms to improve the quality of visualization software. Therefore, many new techniques have been applied in our 3D software to enhance the visualization of PAM images. We believe our software is a valuable framework for researchers, promoting development in a wide range of studies. The framework was incorporated into the QT creator with an impressively user-friendly interface and fast data processing. it helps researchers have a better overview of the 3D data, especially photoacoustic data [1]. Furthermore, our products have some advantages like it is completely free, easy to install and use, many applications are integrated for PA imaging. To improve the quality of 3D Viewer framework, we plan to add new features such as saving images captured with more formats (*jpg* or *bmp*), developing medical image format (*dicom*), and customized color for each color channel.

CHAPTER 4: DISCUSSION

In this thesis, a systematic analysis of semantic segmentation of photoacoustic images using U^2-net, U-net, SegNet, and PSPNet has been performed. We anticipated that U^2-net could better reconstruct high-frequency information than other networks because U^2-net structure only uses RSU blocks. As expected, we could see the more detailed structures in the U^2-net -PA images than in another model. As the result, U^2-net achieves good results which segmentation well in blood vessels, skin, and background.

However, there are several restrictions in our algorithms. Firstly, the number of datasets in our research is still limited although it is still enough for the training of the network. Moreover, in the process of collecting data, we need to focus more on PAM imaging systems such as scanning methods, optical wavelength, and ultrasonic frequency. In the next stage, we will do more experiments to get more data types such as scanning on the human foot or on the body animal. Secondly, in our progress, we process the raw data immediately after scanning without reducing the noise of the signal. In fact, there is much research that has applied filter techniques to get a better signal. if the original signal is not good enough, the image quality is also not good. Hence, signal processing needs to be considered in the future study. Thirdly, at present, we prepare datasets by manual technique. it means that the label image for the training model is done by a human. Moving forward, when the number of datasets is increasing, it may take much time to ground the truth image. Finally, to be suitable with network architecture, our method needs to resize the input image for the training process. It will affect the quality of the input photoacoustic imaging in the training process. In the future study, we will try to find and apply some algorithms to solve this problem.

CHAPTER 5: CONCLUSION

In conclusion, this thesis applied a deep learning method for automated semantic segmentation of blood vessels and skin layers in PAM images. U^2-net architecture demonstrated the best performance compared to other networks such as U-net, SegNet, PSPNet (as described in Table 1). The results showed that this method brings the better accurate in determining blood, skin, and background (average accuracy reaches 94%). Through the experimental results, this method could be an easy-to-process technique for PAM imaging with enhanced quality for clinical-level photoacoustic imaging diagnosis application. In the future, to improve the accuracy of U^2-net model, we will extend the dataset with diverse data types. In addition, we are looking for new research topics such as improvements in the image quality of other modalities of the PAM imaging (lipids, tumor cells, oxygen saturation, melanoma, and organs). The purposed image segmentation techniques could help an important role in diagnosing and studying diseases in humans such as diabetes, dermatology, and neurology. Lastly, our research on reproducing and enhancing the image quality of PAM images on animals is also planned such as segmentation of the tumor and healthy blood in mice.

REFERENCES

1. Nguyen, V., et al. *Ultra-widefield photoacoustic microscopy with a dual-channel slider-crank laser-scanning apparatus for in vivo biomedical study*. Photoacoustics, Volume 23, September 2021, 100274.
2. Ly, C., et al., *Full-view in vivo skin and blood vessels profile segmentation in photoacoustic imaging based on deep learning*. Photoacoustics, Volume 25, March 2022, 100310.
3. JiashengZhou., et al., *Photoacoustic microscopy with sparse data by convolutional neural networks*. Photoacoustics, Volume 22, June 2021, 100242.
4. Seungwan Jeon., et al., *A Deep Learning-Based Model That Reduces Speed of Sound Aberrations for Improved In Vivo Photoacoustic Imaging*. IEEE Trans Image Process. 2021;30:8773-8784.
5. Maslov, K., et al. *In vivo dark-field reflection-mode photoacoustic microscopy*. Opt. Lett. 30 (2005) 625–627.
6. Khodaverdi, A., et al., *Automatic threshold selection algorithm to distinguish a tissue chromophore from the background in photoacoustic imaging*. Biomed. Opt. Express 12 (2021) 3836–3850.
7. Baik, J.W., et al., *Super wide-field photoacoustic microscopy of animals and humans in vivo*. IEEE Trans. Med. Imaging 39 (2020)975–984.
8. GeertLitjens, et al., *A survey on deep learning in medical image analysis*. Medical Image Analysis Volume 42, December 2017, Pages 60-88.
9. Yang, C., et al., *Review of deep learning for photoacoustic imaging*. Photoacoustics 21 (2021), 100215.
10. Deng, H., et al., *Deep learning in photoacoustic imaging: a review*. Biomed. Opt. (2021), 040901.
11. Zhang, J., et al., *Photoacoustic image classification and segmentation of breast cancer: a feasibility study*. IEEE Access 7 (2019) 5457–5466.
12. Chlis, N.K, et al., *A sparse deep learning approach for automatic segmentation of human vasculature in multispectral optoacoustic tomography*. Photoacoustics 20 (2020), 100203.
13. Xuebin Qin, et al., *U^2-Net: Going Deeper with Nested U-Structure for Salient Object Detection*. Computer Vision and Pattern Recognition, 2020.

14. Xin-Yi Tong, et al., *Land-Cover Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models*, Remote Sensing of Environment Volume 237, February 2020, 111322.

15. Zabinski, J, et al., Vardoulakis, *Encyclopedia of Computer Science and Technology*. 2017, pp. 130–132.

16. Paszke, A. et al., *PyTorch: an imperative style, high-performance deep learning library*. Adv Neur In. 2019; arXiv:1912.01703:12.

17. Yuwei Cai, et al., *A Comparative Study of Deep Learning Approaches to Rooftop Detection in Aerial Images*. Canadian Journal of Remote Sensing, Volume 47, 2021 - Issue 3.

18. Bischke, B., et al., *Multi-task learning for segmentation of building footprints with deep neural networks*. In Proceedings of 2019 IEEE International Conference on Image Processing (ICIP), pp. 1480–1484.

19. Chen, Q., et al., *Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings*. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 147: pp. 42–55.

20. Changchun, Y., et al., *Review of deep learning for photoacoustic imaging*. Photoacoustics, Volume 21, March 2021, 100215.

21. Vijay Badrinarayanan, et al., *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*. IEEE.

22. Hengshuang Zhao, et al., *Pyramid Scene Parsing Network* . IEEE, 2017.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude from the bottom of my heart to my advisor, Prof. Junghwan Oh, who gave me the opportunity to study and research in The NBM (Nano Bio Medicine) Lab, Department of Industry 4.0 Convergence Bionics Engineering, at Pukyong National University (PKNU). I appreciate a lot and would like to give my sincere thanks to him for his dedication, enthusiasm, invaluable guidance, support, patience, and encouragement toward the completion of my master's course.

Secondly, I am grateful to my labmates, my foreign friends, and all my Vietnamese friends at PKNU, who helped me a lot in my study and research. They are my big motivation to overcome the hardness in my abroad lifetime and this study would not have been possible without their advice.

Thirdly, I wish to thank Grandma 김옥희, a kind-hearted person from Korea. Thank you for being my second family in Korea.

And finally, my sincere gratitude goes to my family for always supporting, encouraging, and trusting me. They are always by my side to cheer up me and be a great source of motivation for me to strive every day. Thanks for the love and support you all have given me on this journey and forever.

Busan, August 2022

Vu Thi Thu Ha