Development of A Near-Infrared (NIR) Forearm Subcutaneous Vein Extraction Using Deep Residual U-Net

A K Nuraini Huda[1*], C M Goh[2], C H Lim[1], S A Z Sayed Aluwee[2], M N Bajuri[3], N H Abdul Wahab[4]

[1]Department of Industrial Engineering, Universiti Tunku Abdul Rahman, Malaysia
[2]Department of Computer Science, Universiti Tunku Abdul Rahman, Malaysia
[3]Citra Centre for Liberal Study, Universiti Kebangsaan Malaysia
[4]School of Computing, Universiti Teknologi Malaysia

*nurainihuda@gmail.com

Abstract. Impotence to locate the forearm subcutaneous vein leads to multiple intravenous (IV) attempts causing pain and injuries to patients such as bruise or vein damages. Various technologies and techniques were proposed and developed to overcome the multiple IV access problems. The standard techniques used in research and hospitals are Transillumination, Ultrasound Imaging, and Near-Infrared (NIR). Among those techniques, NIR is the most optimal way of locating the subcutaneous vein because of its non-invasive properties, low-cost implementation. The device can be assembled in a small size product. Nevertheless, the NIR forearm images contain noises that cause difficulties in extracting the vein features. Hence, the performance of NIR vein extraction is having the bottleneck of detecting the vein pixel accurately. Many research studies have been conducted to work on the NIR forearm subcutaneous vein detection due to such a limitation. Artificial intelligence is one of the powerful technology that would benefit this study. However, a limited number of articles were found on the patentability search, and thus we propose an automatic vein extraction algorithm using Deep Residual U-Net architecture. Our algorithm shows 75 percent of the accuracy in extracting the NIR vein from the experiments that tested. These results show the evidence that the Deep Residual U-Net can be applied to extract the NIR vein.

1. Introduction

In the medical industry, venipuncture is one of the essential procedures. The procedure can either draw blood samples (phlebotomy) or intravenous (IV) treatment that infuses intravenous solutions, medications, nutrients, supplements, or blood directly into a vein. The fundamental procedure of accessing the vein is by applying a tourniquet mentioned in [1].

The conventional method is not suitable for all patients due to physiological factors such as infants, aged individuals, obese, dark skin, hairy forearm, scar and more. The difficulties of finding the vein leading to the multiple trials for IV insertion and high probability causing the patient to suffer from internal bleeding, vein damage and infiltration, and extravasation [2],[3].

A survey done by [4] finds that the estimation of 150-200 million times of IV is performed annually. The survey [4] also mentioned that up to 8-23% of patients experience difficult peripheral IV placement, narrowing down to the emergency department. These patients are more likely to require central venous access, including significantly higher associated morbidity. Some existing technologies and researches

developed to solve such problems in locating the forearm vein such as Color Vision, Pressure Sensor, Transillumination, Multispectral Imaging, Infrared Thermal Imaging, Ultrasound Imaging, and Near-Infrared (NIR) Imaging.

Transillumination is widely used in the industry as it offers a low-cost solution and is small in size [5]. Although it provides a low-cost product, it might induce negative consequences to patients, such as skin and surrounding cells burning due to the LEDs' heat. Compared to Transillumination, Ultrasound Imaging used to be the safest and accurate technique in the medical industry, but it is costly and bulk in size [1]. The ultrasound can penetrate deep underneath the skin and return broad information of the imaging. Thus, the procedure requires an expert known as a sonographer to allocate the right blood vessel to use during the course.

The researchers are done in such as [6], [7] widely uses NIR imaging for IV procedure and claims that NIR is the best technique that offers non-invasive, low cost, and small in size. The light can penetrate up to 3mm depth under the skin, and only the blood vessel that contains the oxygenated blood is visible. Even though with the NIR imaging assistant technology, multiple attempts of vein access still occurs.

The reason is that NIR image enhancement is necessary to extract the vein for the clinician references. Some researchers such as [8] and [9] applied some image filter applications to enhance vein structure and provide better contrast between the surrounding images. With technology advancement [10] especially intelligent system field, researchers begin to use deep learning as a tool for extracting the vein automatically. Deep learning is knowingly a powerful tool in the automation industry, and automatic object recognition and segmentation are part of the applications [11]. Image segmentation is a widespread practice to extract the necessary object in the medical imaging industry because it can skip some processes such as image pre-processing and image enhancement.

In [12], the researchers use deep learning to enhance the NIR images, improving the robustness of vein recognition and feature extraction used in biometric identification, using the parameters named SparsityProportion, L2WeightRegularisation and SparsityRegularisation. Based on the research [12] outcome, the proposed method significantly improves the feature extraction procedure compared with the original model.

A group researchers from [13] uses a recurrent fully convolutional network (Rec-FCN) architecture to extract the forearm subcutaneous vein automatically NIR stereo image. The FCN has the advantage of feeding the rectangular image to the model. Also, [14] uses U-Net architecture for the vein segmentation task, with Frangi vesselness filter applied to the image for vein visibility enhancement. The images are going through the training with U-net based architecture.

2. Methodology

The methodology of automatically extract the forearm subcutaneous vein using Deep Residual U-Net are as follows. This section is discussing the project platform, image preparation, and the most important is the model structure. The architecture of the model is consists of three neural networks; deep residual block, common convolutional neural network block and U-Net as a whole.

2.1. Project Platform

This project builds the deep learning model using Keras API and TensorFlow 2 as the machine learning tools and coded in Python. The model selected in this project is Deep Residual U-Net, and it is a combination of Residual block and U-Net architecture. The elaboration of the model is described in the next section.

2.2. Image Preparation

The input data is the most crucial part of a successful neural network model to work as the model will be learning based on the given dataset. This project is different from other segmentation projects because the input training images don't have any class to differentiate the image objects but only depending on extracting the vein. The extracted vein image from the original image is called the ground truth image.

The images (original images and ground truth images, respectively) have been through some processes, such as cropped, resized, flipped and rotated because the medical image is known very limited sources and quantity. The purpose of images being resized to the fixed size of training images; 224x224 pixels and 128x128 pixels is to have a result comparison.

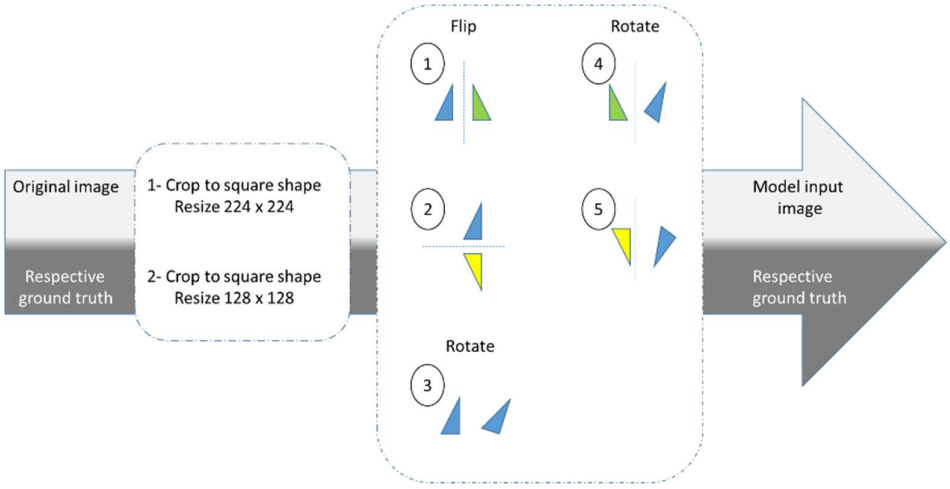


Figure 1. Image preparation flow diagram.

Referring to Figure 1, there are two groups of images. The first group is the original image while the second group is the respective ground truth image. Both groups of images going through the crop and resize process. It is because the model architecture built is accepting the square image only. The images being resized into two sizes, 128x128 and 224x224. The resized images are then required to go through another five processes; flip vertically, flip horizontally, rotate, flip vertically and rotate, flip horizontally and rotate. From Figure 1, the images in the bullet number 1 to 3 are the original images, while the images in the bullet number 4 and 5 are the flipped images. Even though the images are the same, but the processes (flip and rotate) make the image numbers different with each other. Thus, the neural network model views them as different images.

2.3. *Network Architecture: Deep Residual U-Net*

The model selected is the Deep Residual U-Net model, inspired by the research from [15] that combines residual neural networks and U-Net architecture, and Figure 2 showed the model structure used in this project. There are five main blocks in the model architecture as shown in Figure 2: Input, Encoder, Bridge, Decoder and Output. The Encoder and Decoder blocks contain residual blocks, while the bridge consists of the Convolution Neural Network (CNN) blocks. The residual block and CNN block are built from the combinations of the Convolutional layer (conv2D), Batch Normalization (BN) and Rectified Linear Unit (ReLU).

The automated forearm subcutaneous vein algorithm begins with feeding the dataset as the Input to the Encoder. The images are being shuffled before going to each process (training, validation and test). The Encoder block is a down-sampling feature map where the number of feature channel is doubled, and the image been resized in half. For example, in Figure 2, in the beginning of the architecture, the image size (I) is 224x224 and the feature channel (C) is 64, thus, the minimum I is 14 and the maximum C is 1024. The Encoder block consists of four residual blocks with the iteration of two BN, ReLU and 3x3 Convolutional layers with stride 2. The residual block illustration is showed in figure 3.

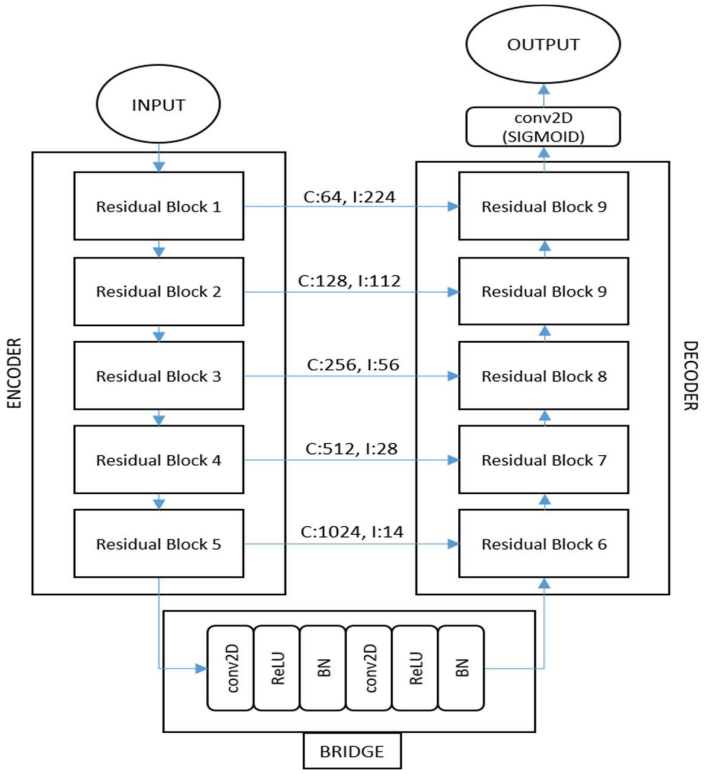
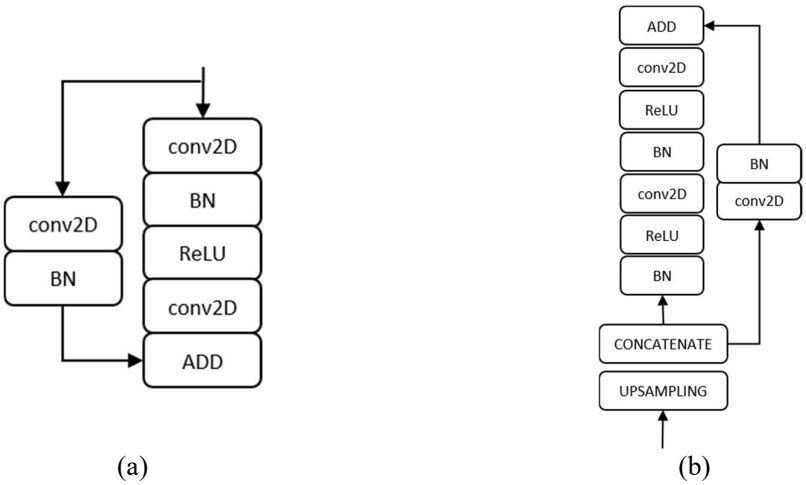


Figure 2. The architecture of the deep Residual U-Net algorithm.

The Encoder's output is fed to the bridge block, and the Decoder block organized in parallel. The output from each of the Bridge block also being fed to the Decoder block, respectively. The Decoder block works contradict the Encoder block. It is an up-sampling feature map where the number of feature channels is halved, and the image size is doubled, thus, the layers are connected with the same values. The decoder block's output going through the convolution with sigmoid activation, and the desired segmentation is mapped from the process. The blocks that being assembled is symmetrical and forming a "U" shape. Thus, this architecture is called U-Net architecture.



(a) (b)

Figure 3. A residual architecture used in the project. (a) is the down-sampling residual architecture within the Encoder. (b) is the up-sampling residual architecture within the Decoder.

Figure 3 showed the architecture of the residual block that being used within the Encoder and Decoder blocks. By referring to Figure 2, the residual architecture in Figure 3(a) is the representative of each Residual Block 1 to 5 while the residual architecture in Figure 3(b) is the representative of each Residual Block 6 to 7. The Encoder and Decoder blocks have interconnected each other thru the ADD layer (down-sampling architecture in Figure 3(a)) and the CONCATENATE layer (up-sampling architecture in Figure 3(b)).

2.4. *Residual Block*

The traditional neural networks work by feeding each layer into the next layer. While for the network with residual blocks, each layer into the next layer and their output is added to the next jumped layer [15]. Thus, this method could solve the degradation problem, which causes saturated accuracy [15]. According to [16], the residual network is easy to optimize and the accuracy gained from considerably increased depth.

2.5. *U-Net Architecture*

U-Net is a combination of convolutional neural network architecture; developed by Olaf Ronneberger et al. [17]; that is most suitable for biomedical image segmentation due to the low quantity of dataset. According to [16], U-Net architecture consists of two paths: contracting path (Encoder) and expansive path (Decoder), has been explained previously and shown in Figure 2. As the architecture is symmetry and resulting a U-shape architecture, thus, the model is being call U-Net.

2.6. *Metrics Accuracy*

In this work, the Dice coefficient metric is used to calculate the similarity between two images and widely used in medical image segmentation [18]. The Dice coefficient will measure the similarity of two samples as measured in Equation (1).

$$D = \frac{2*|A \cap B|}{|A| + |B|} \tag{1}$$

A and B represent the pixel value of ground truth and predicted image, respectively. In the numerator, if the pixel value of A and B is the same, it will become 1; otherwise, it is 0. The total value of the intersection is then amplified up twice. The denominator is the total pixel value of both predicted and ground truth images. The value for the Dice; D is between 0 and 1. Consequently, the closer the number to 1, the higher the accuracy.

2.7. *Other Parameters*

In this architecture, there are 29 convolutional operations in total. By referring to Figure 1, there are three big blocks; Encoding, Bridge and Decoding. The Residual Block is used during the Encoding and Decoding stages with strides of 2 and resulting in the dimension of the input image is doubled.

The ordinary convolution convolutional operations with BN, ReLU activation, kernel size [3, 3], "same" padding and stride (1, 1) applied to the Bridge block. The Bridge block acted as the bridge that connects the Encoding and Decoding block, making the architecture symmetric as a U-shape.

At the end of the architecture, 1 X 1 convolution and a sigmoid activation layer are used to project the multi-channel feature maps into the desired segmentation.

3. Preliminary Results and Discussion

Some hyperparameters of gradient descent are being tuned during the training and validation process. Eight sets of training were done repeatedly, and the experiment results are summarized in Table 1. The batch size divided the training samples into the batches to run in parallel and updating the model's internal parameters. The optimum epoch value are selected to manage the number of complete passes

through the training dataset and balanced with the learning rate value to manage the size of the weights that are updated during training in response to the estimated error without disrupting the training process.

The experiments begin with the smallest value of batch size, epoch and learning rate. The numbers for each variable (image size, batch size, epoch and learning rate) are selected and matched randomly. The test for image size 224x224 also begins at the best hyperparameters for image size 124x124 resolution. The reason is because it took high time, average between two to three days, depending on the computer and dataset, to train the model. According to [19], the best training stability and generalization performance could be achieve using small size of batch size. To date, there is no findings on analytically measure the optimal hyperparameters value on a given dataset for a given model. Therefore, the trial and error are the only way to found out the best learning rate.

Table 1. The hyperparameters used during model training and results, respectively.

Image Size	Batch Size	Epoch	Learning Rate	Dice Score
128x128	5	10	1e-5	0.4868
128x128	5	20	1e-5	0.6227
128x128	5	30	1e-5	0.6512
128x128	5	40	1e-5	0.6628
128x128	5	50	1e-5	0.6675
128x128	5	50	1e-6	0.1537
128x128	3	50	1e-5	0.7008
224x224	3	50	1e-5	0.7145
224x224	5	50	1e-5	0.7599

From the results shown in the table 1, the model trained with the hyperparameters with the batch size sets to 5, epoch at 50 and learning rate at 1e-5 gives the highest dice score value. This is shown that the higher the image resolution, the more the dice score resulted. The hyperparameters for the model optimization also depending onto the image's resolution.

Some of the images being presented in Figure 4. Images in Figure 4(a) column are the input images fed to the model, images in Figure 4(b) column are the ground truth of the images, while images in Figure 3(c) are the predicted images results from the model trained respectively.

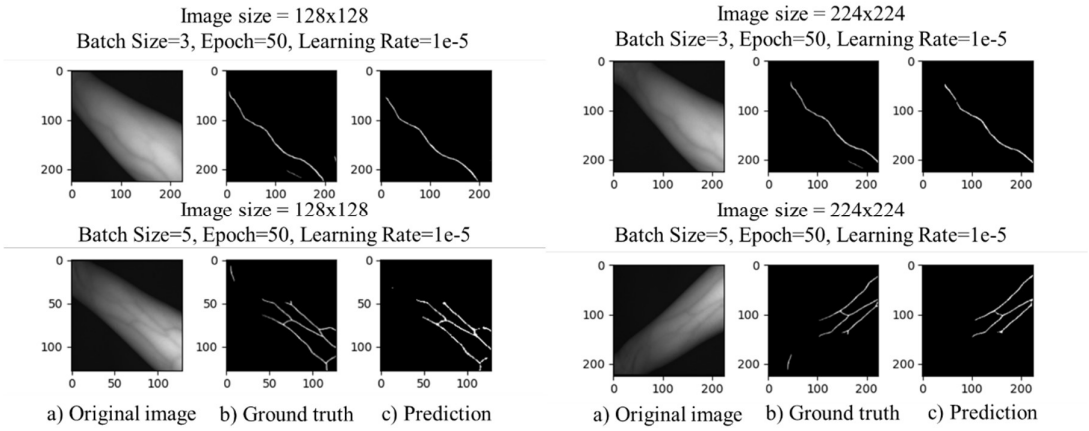


Figure 4. The result of the vein extraction in (c) from the input image in (a) using the last two parameters in Table 1. Image (b) is the corresponding ground truth image to the input image.

As mentioned in section 2.3 before, the images are being shuffled before the process, thus, the output also return different images. Based on the vision, the results have equivalency image drawn with the ground truth image.

4. Conclusions

This project implements a Deep Residual U-Net, which combines residual block and convolutional neural network (CNN) in a U-Net architecture to extract the forearm subcutaneous vein automatically. The models trained in this research have a significant performance from the results achieved and show a promising result that motivates the further development for subcutaneous vein extraction using deep learning, specifically on deep residual U-Net architecture.

The project should be further developed with more experiments to set the value for the variables in hyperparameters of the neural network optimization to validate the best metric accuracy score of the system.

Acknowledgements

This research was supported by a UTAR Research Fund (UTARRF) with the project number of IPSR/RMC/UTARRF/2019-C2/L06. Thank you for the funding.

References

[1] Cantor-peled G, Ovadia-Blechman, and Zehava MH 2016 Peripheral vein locating techniques *Imaging Med* **8** 83–88.

[2] Cooke M, Ullman AJ, Ray-Barruel G, Wallis M, Corley A, and Rickard CM 2018 Not 'just' an intravenous line: Consumer perspectives on peripheral intravenous cannulation (PIVC). An international cross-sectional survey of 25 countries *PLoS One* **13** 1–18.

[3] Marsh N, Webster J, Larsen E, Cooke M, Mihala G, and Rickard CM 2018 Observational study of peripheral intravenous catheter outcomes in adult hospitalized patients: A multivariable analysis of peripheral intravenous catheter failure *J. Hosp. Med.* **13** 83–89.

[4] Kaur P, Rickard C, Domer GS, and Glover KR 2019 Dangers of Peripheral Intravenous Catheterization: The Forgotten Tourniquet and Other Patient Safety Considerations *Vignettes Patient Saf* **4** 116-136.

[5] Agnalt SK, Canfield DM, Perreault KM, Legris JD, and McPheron BD 2016 Vein detection using vein transillumination and contrast differentiation for practitioner aid *2016 IEEE MIT Undergrad. Res. Technol. Conf. URTC* 1–4.

[6] Marathe M, Bhatt NS, and Sundararajan R 2014 A novel wireless vein finder *Proc. Int. Conf. Circuits, Commun. Control Comput. I4C 2014* 277–280.

[7] Al Ghozali HK, Setiawardhana, and Sigit R 2016 Vein detection system using infrared camera *Proc. - 2016 Int. Electron. Symp. IES 2016* 122–127.

[8] Ahmed KI, Habaebi MH, Islam MR, and Zainal NAB 2017 Enhanced vision based vein detection system 2*017 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2017* 1–6.

[9] Lu CY, Jing BZ, Chan PPK, Xiang D, Xie W, Wang J, Yeung DS 2016 Vessel enhancement of low quality fundus image using mathematical morphology and combination of Gabor and matched filter *Int. Conf. Wavelet Anal. Pattern Recognit.* 168–173.

[10] Zaleha SH, Haliza NAW, Ithnin N, and Ahmad J, Hidayah NZ, Chinoso O and Huda NAK 2021 Microsleep Accident Prevention for SMART Vehicle via Image Processing Integrated with Artificial Intelligent.

[11] Yang T, Yoshimura Y, Morita A, Namiki T and Nakaguchi T 2018 Fully automatic segmentation of sublingual veins from retrained u-net model for few near infrared images *Sensors (Switzerland)* **10** 1–7.

[12] Varastehpour S, Sharifzadeh H, Ardekani I, Francis X, and Baghaei N 2019 An Adaptive Method for Vein Recognition Enhancement Using Deep Learning *2019 IEEE 19th International Symposium on Signal Processing and Information Technology* 1-6.

[13] Chen AI, Balter ML, Maguire TJ, and Yarmush ML 2020 Deep learning robotic guidance for autonomous vascular access *Nat. Mach. Intell.*, **2** 104–115.
[14] Leli VM, A. Rubashevskii, Sarachakov A, Rogov O, and Dylov DV 2020 Near-Infrared-to-Visible Vein Imaging via Convolutional Neural Networks and Reinforcement Learning *16th IEEE Int. Conf. Control. Autom. Robot. Vision, ICARCV 2020* 434–441.
[15] Zhang Z, Liu Q, and Wang Y 2018 Road Extraction by Deep Residual U-Net *IEEE Geoscience and Remote Sensing Letters* **15** 749–753.
[16] He K, Zhang X, Ren S, and Sun J 2 016 Deep residual learning for image recognition *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 770–778.
[17] Navab N, Hornegger J, Wells WM, and Frangi AF 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation *Springer International Publishing Switzerland 2015* **9351** 12–20.
[18] Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, and Blaschko MB 2020 Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index," *IEEE Trans. Med. Imaging*, vol. **39** 3679–3690.
[19] Masters D and Luschi C 2018 Revisiting Small Batch Training for Deep Neural Networks 1–18 arXiv:1804.07612.