생체 인식 인식 시스템을 위한 주의 인식 잔차 분할

앤디 [1], 이효종 [1]
[1] 전북대학교 공과대학
andy@jbnu.ac.kr, hlee@jbnu.ac.kr

Attention Aware Residual U-Net for Biometrics Segmentation

Aung Si Min Htet[1], Hyo Jong Lee[1*]
[1]Division of Computer Science and Engineering, Jeonbuk National University, Korea
*Corresponding author

Abstract

Palm vein identification has attracted attention due to its distinct characteristics and excellent recognition accuracy. However, many contactless palm vein identification systems suffer from the issue of having low-quality palm images, resulting in degradation of recognition accuracy. This paper proposes the use of U-Net architecture to correctly segment the vascular blood vessel from palm images. Attention gate mechanism and residual block are also utilized to effectively learn the crucial features of a specific segmentation task. The experiments were conducted on CASIA dataset. Hessian-based Jerman filtering method is applied to label the palm vein patterns from the original images, then the network is trained to segment the palm vein features from the background noise. The proposed method has obtained 96.24 IoU coefficient and 98.09 dice coefficient.

1. Introduction

Biometric identification is a technique for authenticating identity that employs physical or behavioral characteristics of the human body. In the past decade, biometric identification systems have been increasingly employed for personal identification in many applications. However, extrinsic biometrics such as fingerprint, iris, face, and palmprint has limitations, including the fact that they are sensitive to spoofing and are heavily influenced by age, health, injuries, and skin condition. In contrast, intrinsic biometrics like palm vein recognition have obtained significant attention. The palm vessels, which are concealed beneath the human skin, provide better privacy and security advantages as it becomes difficult to forge under visible light. Furthermore, palm vascular structures remain constant throughout one's life and vanish when there is no blood flow. Thus, the intricate and persistent palm veins are considered to be more reliable for biometric identification than palmprints which can be easily forged and affected by external factors.

Because contactless palm images are acquired using near-inferred light, the contrast between the veins and background is generally low and the visibility of the blood vessel is affected by various external factors such as temperature, light conditions, and illuminations. Therefore, enhancing the vein patterns is crucial to improve the robustness of palm vein features for the convolutional neural network to learn effectively. In this research, we proposed attention aware residual U-Net to segment the palm vein patterns for learn domain specific features.

The remaining paper is systematized as follows. Section 2 discusses the related work. Section 3 presents the proposed attention aware residual U-Net segmentation method. Section 4 contains the experimental result and method evaluation. The conclusion is given in Section 5.

2. Related Work

Recent studies for palm vein biometrics have shown that palm vein enhancement can improve the model accuracy significantly by boosting the intensity values of elongated structures like palm veins. However, traditional image processing and filtering methods for vein enhancement can be difficult to obtain the accurate vessel shape of palm vein, thus, degrading the accuracy and affecting model performance. Handcrafted palm vein enhancement methods such as by Frangi-based and Laplacian filters [5], and Jerman filter [6] are also proposed, which can be error-prone in real-world applications. Furthermore, deep-learning based methods are also proposed for segmentation. PVSNet [2] proposed the novel Encoder-Decoder network to learn the domain-specific features, which afterward trained the encoder separately to identify the identities using Hard Triplet Loss. [3] and [4] proposed U-Net based palm vein segmentation architecture. However, with the limitation of U-Net architecture, the segmentation results are not promising.

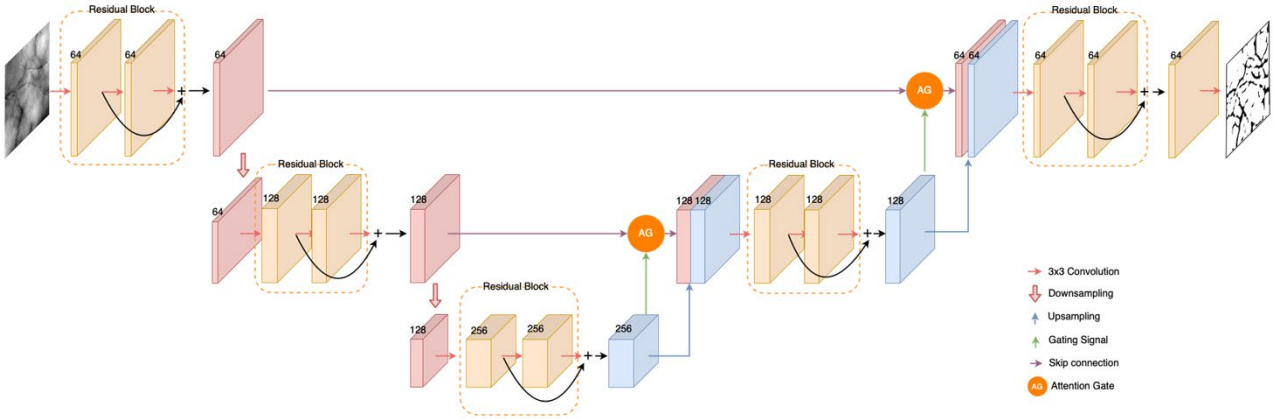
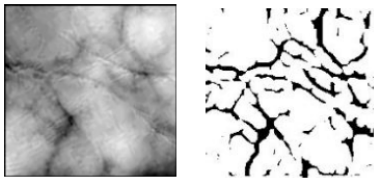


Figure 1. Proposed Attention Aware Residual U-Net Architecture

3. Methodology

3.1 Preprocessing

First, we label the palm vein pixels by applying a Hessian-based traditional handcraft method called Jerman filtering method [8] which extract the vascular vein structures from the background pixels. The Gaussian size and τ value of the method are carefully selected as too small kernel size can become sensitive to the noise and produce inconsistent vein patterns. In our research, we intent to obtain the ground truth label mask to be as much accurate to the actual vein pixels that is in the original image. To reduce unclear and blurry result, each pixel is filtered with a specific threshold value and weak pixels are eliminated. The vein pixels are then set to 0 and background pixels to 255 for proper labelling.



(a) (b)

Figure 2. (a) Original Grayscale Image (b) Labeled using Jerman Filtering Method

3.2 Attention Gate

The attention gate mechanism was introduced as an additional convolution module to suppress irrelevant regions and highlighting salient features useful for a given task. Thus, attention gate can be combined with U-Net architecture in the skip-connections to obtain more precise segmentation result with minimal computational overhead. The attention gate mechanism is illustrated as Fig. 4.

To define focus areas for a given feature map x_l, a gating vector g_i is used for each pixel i. The gating vector contains the contextual information from the lower-level feature responses. Attention coefficients, α, detect the salient areas and prune feature responses to maintain only the activations relevant to the given task. The output is the element-wise multiplication of x_l and α, which can be defined as follows:

$$x_{out} = x_l * \alpha$$

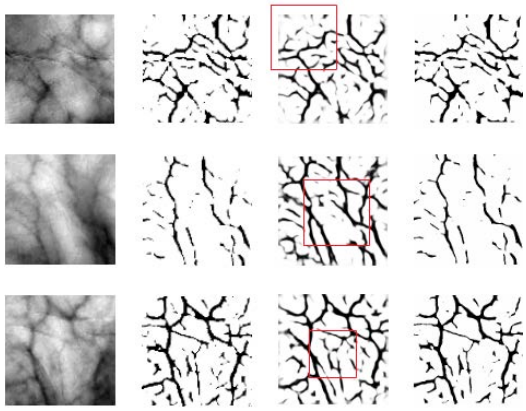
The additive attention [9] is used to obtain the gating coefficient, which requires a high computational cost but with the trade-off for obtaining more accurate segmentation results. Additive attention is formulated as follows:

$$\alpha = \sigma_2(\psi^T\left(\sigma_1\left(W_x^T x_l + W_g^T g_l + b_g\right)\right) + b_\psi)$$

where σ_1 and σ_2 are ReLU and Sigmoid functions, W_x, W_g and ψ are linear transformations, and b_g and b_ψ are bias.

3.3 Proposed Attention Aware Residual U-Net

The Fig. 1 shows the network architecture of proposed attention aware residual U-Net model. U-Net has achieved a promising result in the field of medical image analysis. It contains the down-sampling encoder and up-sampling decoder and the skip-connection between them. At deeper stage of the encoder path, the network can learn low-level context information. However, because U-Net model uses a pooling layer during the down-sampling process, it may lose many image features. In addition, skip connections in U-Net model is used to concatenate the low-level features with the high-level features. Due to lack of spatial information in the low-level features, the spatial details tend to be lost while preserving along the up-sampling process.



(a) (b) (c) (d)

Figure 3. (a) Original Grayscale Image (b) Label (c) Result from U-Net Model (d) Result from Proposed Method

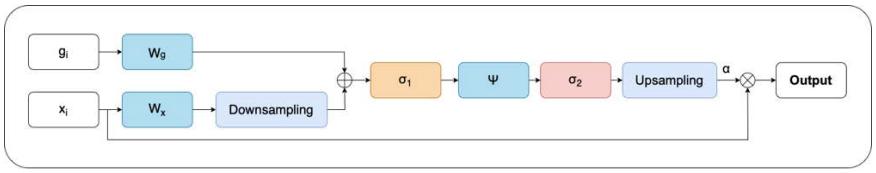


Figure 4. Attention Gate Module

To solve this problem, we integrate attention gate to the skip-connections of U-Net model to eliminate the irrelevant features and restore more accurate high-resolution image. Another problem with deep neural networks is the vanishing gradient problem, which occurs when the gradient is repeatedly multiplied in the backpropagation and causes the gradient to be infinitely small. To overcome this problem, we used the residual block to replace the convolution process at each level of the encoder and decoder networks. By combining the strength of residual blocks and attention gate in the skip-connection, the segmentation result of our proposed method is higher than the baseline methods.

3.4 Loss Functions

The Dice coefficient is widely used metric in computer vision tasks to measure the distance between the segmentation result and the label. In this study, we used Dice Loss for training the segmentation model, which is defined as follows:

$$DL(p,g) = 1 - \frac{2pg + 1}{p + g + 1}$$

where p and g represent pairs of corresponding pixel values of prediction and ground truth, respectively.

4. Experimental Result

The proposed method is experimented on the CASIA Multi-Spectral Palm-print Image Database [7]. The database contains palmprint and palm vein images of 100 different volunteers using multiple spectral wavelengths between 460nm and 940nm. Left-hand and right-hand images are considered as different identities. The palm vein images at 850nm and 940nm were selected for the experiment as blood vessel appears vividly at these wavelengths.

As illustrated in Fig. 3, the U-Net model (c) generates irrelevant vein patterns that are not present in the original images or labels due to a lack of spatial information. Our proposed model (d), on the other hand, can accurately execute the segmentation while suppressing the unwanted background pixels. As the ablation study, the comparison with other state-of-the-art methods on palm vein segmentation is evaluated in Table 1.

Table 1. Methods Comparison of Palm Vein Segmentation

Method	IoU Coef.	Dice Coef.
U-Net	95.59	97.75
Attention U-Net	95.83	97.87
Residual U-Net	96.16	98.04
Proposed Method	**96.24**	**98.09**

5. Conclusion

In this research, we have experimented the effectiveness of attention gate and residual block in a single U-Net architecture to distinguish between salient and noisy feature responses while highlighting important features. The proposed method outperformed the baseline U-Net model, attention U-Net and residual U-Net, respectively.

Acknowledgment

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2019R1D1A3A03103736 and in part by project for Joint Demand Technology R&D of Regional SMEs funded by Korea Ministry of SMEs and Startups in 2021 (No. S3035805).

Reference

[1] Synced G, Shaoyou (Victor) L, Baorui (Alex) C, Qingyan T, Chenchen (Chain) Z, Chen (Robert) T, Meghan H. (2018). Year of AI: How Global Public Company Adapted to the Wave of AI Transformation: A 2018 Report about Fortune Global 500 Public Company Artificial Intelligence Adaptivity (Kindle Edition). ASIN: B07K91RKVK.

[2] D.Thapar, G.Jaswal, A.Nigam, and V.Kanhangad, "PVSNet:Palmvein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features," in Proc. IEEE 5th Int. Conf. Identity, Secur., Behav. Anal. (ISBA), Jan. 2019, pp. 1–8, doi: 10.1109/ISBA.2019.8778623.

[3] Wang, P., Qin, H.: Palm-vein verification based on u-net, vol. 806. Institute of Physics Publishing, (2020). https://doi.org/10.1088/ 1757-899X/806/1/012043

[4] Marattukalam, F., Abdulla, W.H.: Segmentation of Palm Vein Images Using U-Net

[5] Noh, Z.M., Ramli, A.R., Hanafi, M., Saripan, M.I.,Ramlee, R.A.: Palm vein pattern visual interpretation using laplacian and frangi-based filter. Indonesian Journal of Electrical Engineering and Computer Science 10 (2018). https://doi.org/10.11591/ijeecs.v10.i2.pp578-586

[6] Obayya, M.I., El-Ghandour, M., Alrowais, F.: Contactless palm vein authentication using deep learning with bayesian optimization. IEEE Access 9, 1940–1957 (2021). https://doi.org/10.1109/ACCESS.2020. 3045424

[7] Casia-ms-palmprintv1 http://biometrics.idealtest.org/

[8] Jerman, T., Pernuˇs, F., Likar, B., ˇZiga ˇSpiclin: Beyond Frangi: an im- proved multiscale vesselness filter. In: Ourselin, S., Styner, M.A. (eds.) Med- ical Imaging 2015: Image Processing, vol. 9413, pp. 623–633. SPIE, (2015). https://doi.org/10.1117/12.2081147.

[9] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pan- creas," 2018, arXiv:1804.03999.