DOI: 10.24996/ijs.2022.63.4.37

ISSN: 0067-2904

Biometric Identification System Based on Contactless Palm-Vein Using Residual Attention Network

Husam Imad Abdulrazzaq[1]*, Rawaa Dawoud Al-Dabbagh[2]
[1]*Ministry of Higher Education and Scientific Research, Baghdad, Iraq*
[2]*Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq*

Received: 22/2/2021 Accepted: 11/6/2021 Published: 30/4/2022

Abstract

Palm vein recognition technology is a one of the most effective biometric technologies for personal identification. Palm acquisition techniques are either contact-based or contactless-based. The contactless-based palm vein system is considered more accurate and efficient when used in modern applications, but it may suffer from problems like pose variations and the delay in the matching process. This paper proposes a contactless-based identification system for palm vein that involves two main steps; First, the central region of the palm is cropped using fast extract region of interest algorithm, then the features are extracted and classified using altered structure of Residual Attention Network, which is a developed version of convolutional neural network that uses an attention mechanism. The altered structure is constructed by stacking multiple attention modules and pre-activation residual unit with additional down sampling layers in between. The proposed system was tested on contactless CASIA multispectral palm vein databases that contains palm images with obvious pose variations taken from 100 persons. The results show that our system has outperformed other state-of-the-art systems with 95.55% accuracy and fast identification process of 0.06 second per person.

Keywords: biometrical identification, contactless palm vein, residual attention network, convolutional neural network, deep learning.

نظام تحديد الهوية البايومتري بناء على اوردة اليد اللاتلامسية باستخدام شبكة الانتباه المتبقي

حسام عماد عبد الرزاق[1*]، رواء داود الدباغ[2]

[1]وزارة التعليم العالي والبحث العلمي، بغداد، العراق

[2]قسم علوم الحاسوب، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

تعد تقنية التعرف على وريد الكف واحدة من أكثر تقنيات القياسات الحيوية فعالية لتحديد الهوية الشخصية للأفراد. تقنيات اكتساب راحة اليد هي إما قائمة على اللمس أو بدون تلامس. يعتبر نظام التعرف على الوريد القائم على عدم التلامس أكثر كفاءة ودقة لاستخدامه بشكل جيد في التطبيقات الحديثة، ولكنه قد يعاني من مشاكل مثل اختلافات الوضع للكف وتأخر عملية المطابقة. اقترح هذا البحث نظام تحديد قائم على عدم التلامس لوريد الكف يتضمن خطوتين رئيسيتين. أولاً، يتم اقتصاص المنطقة الوسطى من الكف باستخدام

*Email: Hussam.i1986@gmail.com

خوارزمية الاستخراج السريع لمنطقة الاهتمام، ثم يتم استخراج الميزات وتصنيفها باستخدام بنية متغيرة لشبكة الانتباه المتبقي وهي نسخة مطورة من الشبكة العصبية التلافيفية التي تستخدم آلية الانتباه. يتم إنشاء الهيكل المعدَّل عن طريق تكديس وحدات الانتباه المتعددة والوحدة المتبقية قبل التنشيط مع طبقات أخذ عينات إضافية بينهما. تم اختبار النظام المقترح على قواعد بيانات وريد الكف متعدد الأطياف بدون تلامس CASIA التي تحتوي على صور كف مع اختلافات واضحة في الوضع مأخوذة من 100 شخص. تظهر النتائج أن النظام المقترح قد تفوق على أحدث الأنظمة بدقة 95.55٪ وعملية تحديد سريعة تبلغ 0.06 ثانية لكل شخص.

1. Introduction

Biometric identification has become an important field of scientific research in recent years. The contactless palm vein is an interesting type of biometric technologies. It is like palm print, but it needs Near Infra-Red (NIR) illumination to capture the veins that are hidden under the skin rather than the use of visible spectrum light to capture the palm print. Compared to other biometric techniques, palm vein is difficult to fake or change because each person has their unique patterns of veins that can efficiently be used in verification or identification. Moreover, they are less costly to obtain and easy to acquire [1].

Contactless based palm vein capturing is more complicated than contact based. However, it is more convenient and safer to users [2]. The main problem of contactless based technique is that the image of the palm is usually affected by the movement of the hand. This is because the palm is captured without a peg to guide the hand so there will be pose variations such as scaling, rotation, and translation transformations. Another problem in contactless is the matching speed especially in systems with large number of enrolled persons. It may not be convenient for a person to keep their hand raised waiting for identification. So, the system should be fast enough to handle this issue [2, 3].

Convolutional neural network (CNN) is one of the most powerful deep learning models. It has widely been used in image classification and pattern recognition because it can learn distinctive features of the input multimedia and, also, it is invariant to deformations such as rotation, scaling and translation. The most important features of CNN are local connectivity and shared parameters. Typically, CNN structure consists of several types of layers: 1) convolutional layer, which is 2-dimensional neurons 2) pooling layers, which is used to reduce the spatial size of the feature maps and 3) fully connected layer, which is a multi-layer perceptron.

In recent years, several structures and models inspired by CNN have already been proposed in [3, 4]. Wang et al. in 2017 [5] proposed Residual Attention Network (RAN), which is a developed CNN model with attention mechanism. It was built by stacking Attention Modules that generate attention-aware features. The bottom-up and top-down feedforward structure is used to clarify the feed forward and feedback attention process into a single feed forward process. The RAN model has been tested on CIFAR-10 and CIFAR-100 datasets and it achieved state-of-the-art performance in object recognition with best results.

In this paper, we propose a new person identification system using contactless palm-vein biometric trait. This system can deal with palm images captured with various pose variations such as rotation, scale, and translation. At the first stage, the system extracts the region of interest using the fast ROI algorithm [2] by applying it on the palm images for cropping palm central region. The features are then extracted and classified by special altered structure of RAN. Finally, the proposed system is evaluated using CASIA multispectral palm vein database, which is divided into training and testing portions. The last portion is used to compute the validation accuracy of the proposed system.

This paper is structured as follows. Section 2 presents a brief review of the related palm-vein research published for verification and identification. Section 3 presents the materials and methods on which the proposed system has been constructed. Section 4 shows the experimental results and comparisons to the proposed system. Finally, conclusions have been stated in Section 5.

2. Related Work

There are several researchers involved in developing personal identification systems using contact-based or contactless based palm-vein traits. For example, Abbas and George in [6] proposed a system of palm-vein recognition using spatial energy distribution of wavelet sub-bands. They first applied the discrete Haar wavelet (DHW) to the image and calculated the average energy distribution for each sub-band, then they obtained the feature vector by concatenating these sub-bands. A contact-based palm vein database was used for evaluating this system; where, the minimum equal error rate (EER) achieved was 0.068% and the recognition time was about 0.159 second.

Another research conducted by Wang et al. [7] that presented a palm-vein identification system that depends on Gabor wavelet. In this research, the contrast limited adaptive histogram equalization (CLAHE) was used for improving the image contrast in the first place, then image skeletonization has been applied for vein thinning. The Gabor wavelet transform method was then used for feature extraction. The system was evaluated using the authors' own collected palm vein database which was acquired in controlled conditions. They achieved a success rate of 98.88% and recognition time of 0.324 second.

Perwira et al. in [8] proposed personal palm-vein identification system using Principal Component Analysis (PCA) for the feature extraction and feature dimension size reduction, and Probabilistic Neural Network (PNN) as classification method. The system was evaluated with only 50 persons samples taken from CASIA multispectral palm vein databases and achieved a final accuracy of 84 %.

Jalali et al. in [9] presented a contactless palmprint biometric system using features of palm texture from an image captured by normal digital camera. Then, CNN was applied for the task of palmprint classification. The system has been evaluated using the authors' own contactless palm vein database that was collected from 10 persons only. The system achieved accuracy of 93.4%.

Dian and Dongmei in [3] proposed a palmprint recognition method using CNN. It is based on applying fuzzy enhancement algorithm for the pre-processing stage, then features were extracted using the AlexNet (a well-known structure of CNN). For the matching stage, the Hausdorff distance measure was used. The system achieved 0.044% EER when evaluated using contact-based database and 0.1113% EER on contactless database.

Another palm-vein recognition system was proposed by Holle et al. in [10] using the local line binary pattern (LLBP) method. The main steps of their system include, pre-processing, feature extraction using LLBP method and matching using Fuzzy KNN classifier. The system was evaluated on CASIA Multi-Spectral Image Database through conducting multiple experiments and achieved 93.2% accuracy. Finally, Hassan and Abdulrazzaq in [2] proposed an identification system for contactless palm-vein, which is mainly composed of three steps; Firstly, data is augmented by performing out-of-plane rotation. Second, palm region is cropped using the fast extract Region of Interest (ROI) algorithm. At the third and final stage, CNN is applied for features extraction and classification. The system evaluation result using the original CASIA Multi-Spectral Image Database achieved accuracy of 93%.

In this work, palm-vein personal identification system depending on a modified version of CNN named RAN is proposed. The main objective is to gain relatively high accuracy and high-speed identification time in comparison with other works when applied on contactless palm vein images, which could contain pose variations such as scaling, rotation, and translation transformations. For this reason, the system was tested on contactless CASIA multispectral palm vein database.

3. Materials and Methods

3.1 Residual Attention Network

Residual Attention Network (RAN) was proposed by Wang et al. [5] in 2017 as a developed version of CNN using Attention mechanism. RAN is constructed by the stacking of multiple Attention Modules where each module constitutes of two branches: trunk branch and mask branch. Trunk branch is where the feature processing is performed. In this work, as a basic unit, the pre-activation Residual Unit [11] is used to construct the Attention Module. To express it formally, assume the output of the trunk branch is $T(x)$ where x is the input, the bottom-up top-down structure [12] is then used by the mask branch to learn $M(x)$ mask with equal size. This adds soft weights to the required features (lines, edges, or points of interest) of the image in $T(x)$ feature map. The fast feedforward and feedback attention process is simulated by the bottom-up top-down structure. Like High Network, the $M(x)$ is used as by the trunk branch as control gates for the neurons. The output of Attention Module H is

$$H_{i,c}(x) = M_{i,c}(x) \cdot T_{i,c}(x) \quad (1)$$

where i indicates overall spatial positions and $c \epsilon \{1, \dots, C\}$ denotes the channel's index.

End-to-end (E2E) learning can be used to train the complete structure. The Residual Model proposed by He et al. [13] for training very deep networks since deeper neural networks are more difficult to train. They explicitly reformulate the layers with reference to the layer inputs. Figure 1 shows the residual block diagram.

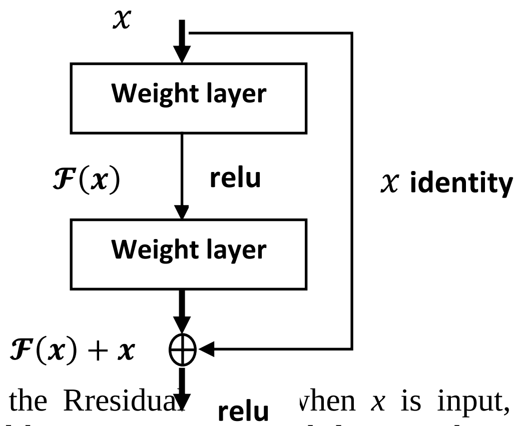


Figure-1 This diagram shows the Rresidual block when x is input, *relu* is an activation function and *weight layer* could be any neural network layer such as a convolutional layer [13].

During the back propagation and besides its function as a feature selector during forward propagation, the attention mask can also serve as a gradient update filter in the Attention Module. For the input feature in the soft mask branch, the gradient of mask can be expressed as

$$\frac{\partial M(x,\theta)\, T(x,\phi)}{\partial \phi} = M(x,\theta) \frac{\partial T(x,\phi)}{\partial \phi} \quad (2)$$

where ϕ are the parameters of the trunk branch and θ are the parameters of the mask branch. Figure 2 shows the full structure of the RAN module as proposed in [5]. This figure illustrates the numbers and types of the neural network layers that have been used for constructing RAN. In this module, the design of Attention Module has three hyper-parameters that are all related to the number of Residual Units (RUs) in different stages: 1) p is the number of RUs pre-processed before dividing into mask branch and trunk branch 2) t is the number of RUs in trunk branch. 3) r is the number of RUs in the mask branch between adjacent pooling layer. In this study, the following hyper-parameters setting: $\{p = 1\,,\ t = 2\,,\ R = 1\}$ were used. An equal number of trunk branches and corresponding number of channels in the soft mask RU was used [5].

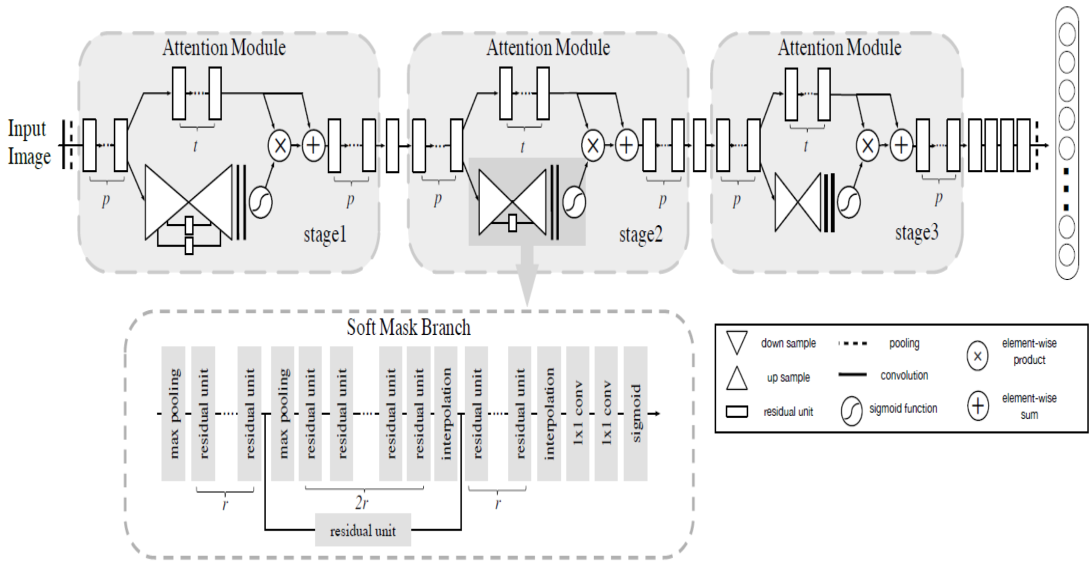


Figure 2-The full structure of RAN module as suggested by Wang et al. [5].

3.2 Contactless Palm-Vein Identification System using RAN

The proposed contactless palm-vein identification system includes two major phases: enrollment phase and identification phase. Enrollment phase uses the training portion of palm vein database, where all palm vein images of all persons are simultaneously passed into the system for training the RAN model for several epochs. In Identification phase, the trained system can directly identify any palm vein image of the enrolled persons. The test portion of database is used for conducting the experiments.

The overall system is composed of two major steps: first step is to extract the region of interest using a fast extract ROI method [2], which crops the most interesting region of the palm image. This method is invariant to translation and scale variations. The second step is the feature extraction and classification using RAN. In this work, the structure of RAN has been altered as shown in Figure 3. The altered structure has additional down sampling max pooling layers between each Residual Units and attention blocks for further dimensionality reduction. The last two residual blocks have been eliminated. This is for the purpose of decreasing the computational complexity as, mainly, the RAN structure consists of convolutional layer followed by three sequent groups of Attention and Residual blocks then another Residual block and the last fully connected layer, which size is equal to the number of enrolled persons. Finally, the last fully connected layer use SoftMax as an activation function that assigns a probability for each person class for matching.

Table 1 presents the layers of the proposed structure and the input and output size of each layer.

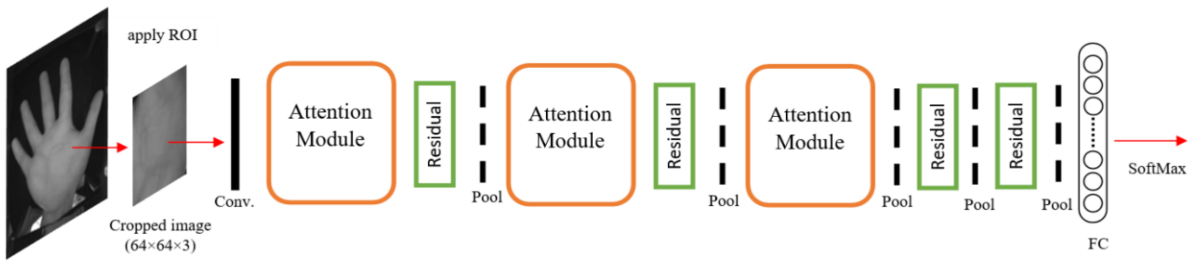


Figure 3-The main structure of proposed system, the black line represents a Convolutional layer, the dashed line represents Max Pooling layer, the orange rectangle represents Attention Module, the green rectangle represents Residual blocks.

Table 1-The structure of RAN. The input and output sizes are described in (rows×cols×#filters). The kernel for single layers only is specified as rows×cols; stride.

no	Layers	Input size	Output size	kernel
1	Convolutional layer	64×64×3	64×64×32	3×3; 1
2	Attention Module 1	64×64×32	64×64×32	
3	Residual block 1	64×64×32	64×64×**64**	
4	max pooling	64×64×64	**32×32**×64	2×2; 2
5	Attention Module 2	32×32×64	32×32×64	
6	Residual block 2	64×64×32	64×64×**128**	
7	max pooling	64×64×128	**16×16**×128	2×2; 2
8	Attention Module 3	16×16×128	16×16×128	
9	max pooling	16×16×128	**8×8**×128	2×2; 2
10	Residual block 3	8×8×128	8×8×**256**	
11	max pooling	8×8×256	**4×4**×256	2×2; 2
12	Residual block 4	4×4×256	4×4×256	
13	max pooling	4×4×256	**2×2**×256	2×2; 2
14	Fully connected + SoftMax	1024	**number of classes** (individuals)	

4. Experimental Results and Discussion

The proposed system was implemented using TensorFlow open-source framework. The classification accuracy metric was used to evaluate the proposed system. A well-known multi-spectral palm veins database, which is a CASIA multi-spectral palmprint image database V1.0, was used to verify this system [14]. It was composed of palm vein images of 100 person, captured using a contactless multiple spectral imaging device without pegs, so there is a certain degree of variations of different hand poses. In this experiment, 12 image samples of the right hand of each person were used. These images were captured under 850 nm and 940 nm near-infrared (NIR) illuminations, as shown in Figure 4.

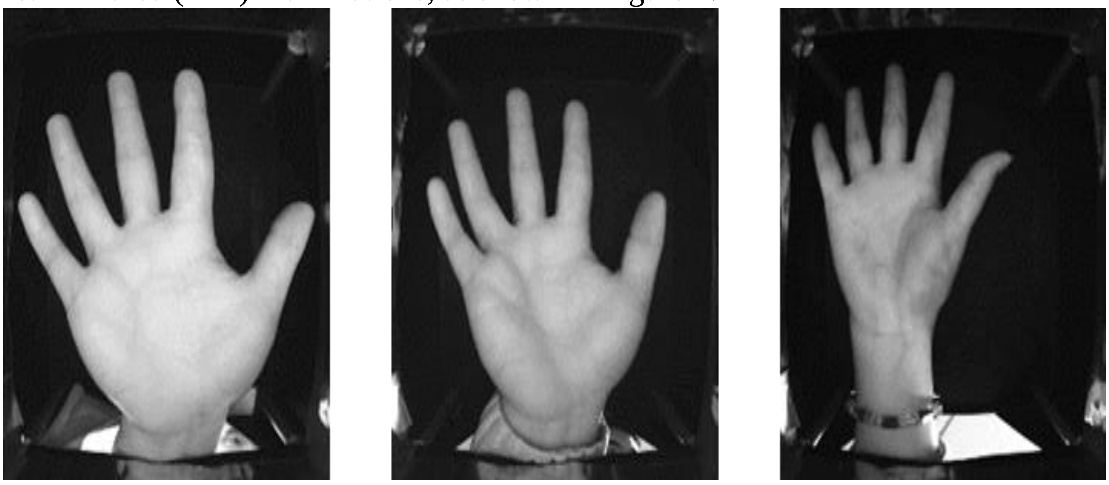


Figure 4-This figure shows some samples from CASIA database.

Throughout this experiment, $3:1$ was considered as the $training:testing$ ratio. The flow of the system process started by cropping all input palm vein images using the fast ROI algorithm. The cropped images were resized to a specific input image size then passed to RAN structure simultaneously for training. The system was trained for 120 epochs which takes about 8 hours on a personal computer with GPU of NVIDIA Quadro K1000M and 2 GB dedicated memory. The initial learning rate was set to 0.001 and the training batch size was set to 16.

The proposed system was tested with three different input image sizes, the results are presented in Table 2 which contains 5 columns: The first one represents the triple (Width, Hight, Channel) of each palm vein image input to the system while training or testing. The second column shows the complete training time in minutes for each corresponding input image size. The third column shows identification time in portions of second for one palm image. The fourth and fifth columns are for validation loss and validation accuracy.

Table 2-The results of testing the proposed system with three deferent input image size.

Input Image Size	Training Time (Minutes)	Identification Time (Second)	Loss	Accuracy
$32 \times 32 \times 3$	95	0.0178	0.438310	90.54 %
$64 \times 64 \times 3$	362	0.0570	**0.156274**	**95.55 %**
$96 \times 96 \times 3$	524	0.1012	0.832259	86.93 %

Table 2 shows that the best results were obtained with input image size of $(64 \times 64 \times 3)$. This proves the optimality of $(64 \times 64 \times 3)$ input size for training the proposed altered RAN structure. The testing portion of database is the same used for validation.

As shown in Table 3, the proposed system scores a relatively high validation accuracy of (95.55%) in comparison to other state-of-the-art works that have used the CASIA palm vein database to validate their systems. In this experiment, validation accuracy keeps stable for the last 20 epochs.

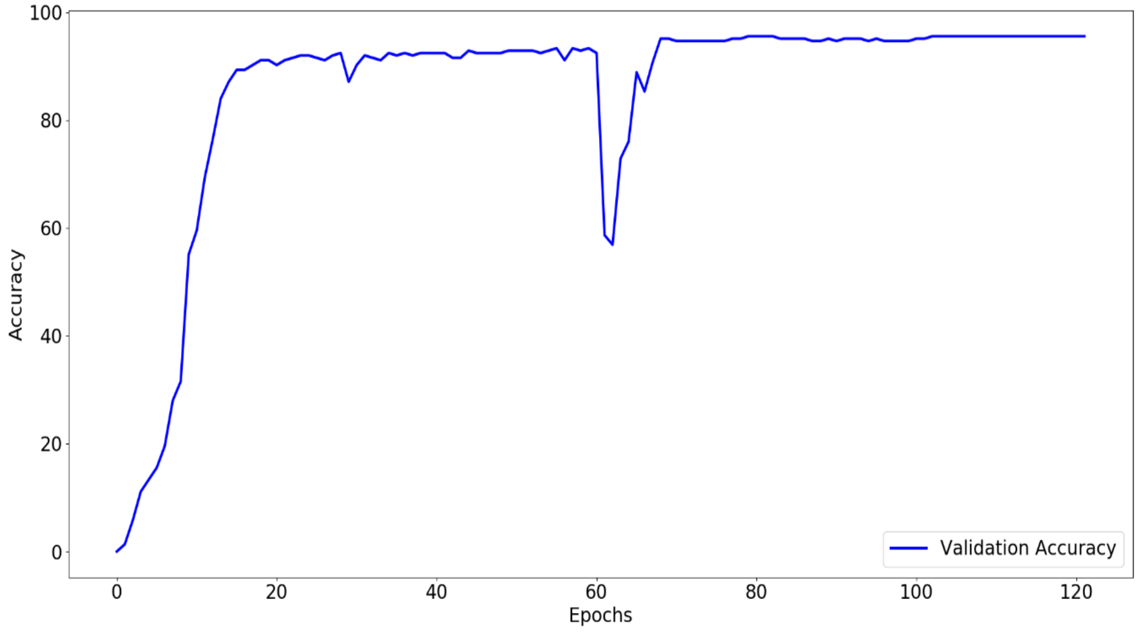
This system has also been tested using the original structure of RAN [5] which is illustrated in Figure 2, but the results were low since the obtained accuracy remains stable at 1.33%, which means that the neural network did not achieve the desired learning on palm veins.

Table 3-The proposed system in ccomparison with other works on CASIA database

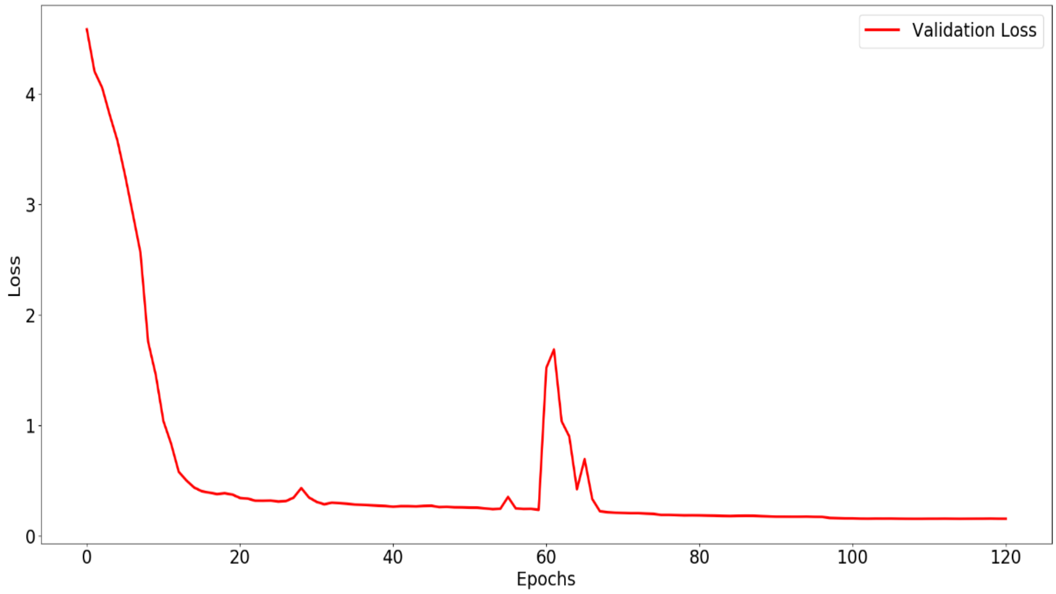
No.	System	Accuracy
1	Perwira et al. [8]	84.00 %
2	Hassan and Abdulrazzaq [2]	93.00 %
3	Holle et al. [10]	93.20 %
4	Rizki et al. [15]	94.67 %
5	The Proposed System	**95.55%**

This could be because of the relatively few features of veins image with high complexity of that structure. On the other hand, this proves the superiority of the proposed altered RAN structure on palm vein identification. This is mainly due to the additional successful down sampling of layers between each Residual Units and Attention blocks which puts more focus on most interesting feature points of the veins. Table 3 shows that the proposed system has achieved highest accuracy in comparison with other works that have used the same CASIA multi-spectral palm vein database. This also proves that the system is best at dealing with pose variation such as scaling, rotation, and translation transformations, which might be formed during capturing palm vein images in contactless mode. This is because of the characteristic of deep learning technique represented by the proposed altered RAN.

Furthermore, even though the system spends a relatively long time in training its identification process, it is fast because the time needed for identifying one palm is about (0.06) second which is fast enough for real time applications.



(a)



(b)

Figure 5-Results plots during 120 epochs of training, (a) Validation Accuracy, (b) Validation Loss

This high speed in identification process is achieved because the system does not use the preprocessing step for enhancing the visual quality of images since it will slow down the identification process without any improvement in the accuracy.

Figure 5 shows the validation accuracy and the validation loss plot which displays the changing of the results during the training process. In the same figure, the flat lines in (a) and (b) proves that the proposed structure is stable and has reached a good learning performance after 100 epochs.

5. Conclusions

In this paper, an altered structure of RAN has been applied to the identification of a person using their palm vein as contactless based system. The obtained results have proved the efficiency of this proposed system when tested on contactless CASIA multispectral palm vein database containing palm images with different pose variations. This modification to the RAN has helped the learning process to focus more on the interested feature points of the veins, which in turn has clearly affected the accuracy to attain 95.55% in comparison with other works. Also, it contributed to speeding up the identification process of one person to less than 0.06 seconds, which is another important factor that must be considered when designing such systems.

References

[1] Z. M. Noh, A. R. Ramli, M. I. Saripan and M. Hanafi, "Overview and challenges of palm vein biometric system," *International Journal of Biometrics,* vol. 8, no. 1, pp. 2-18, 2016.

[2] N. F. Hassan and H. I. Abdulrazzaq, "Pose Invariant Palm Vein Identification System using Convolutional Neural Network," *Baghdad Science Journal,* vol. 15, no. 4, pp. 503-510, 2018.

[3] L. Dian and S. Dongmei, "Contactless palmprint recognition based on convolutional neural network," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, Chengdu, China, 2016.

[4] M. Thoma, "Analysis and Optimization of Convolutional Neural Network Architectures," Karlsruhe Institute of Technology, Germany, 2017.

[5] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang and X. Tang, "Residual Attention Network for Image Classification," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.

[6] A. M. J. Abbas and L. E. George, "Palm Vein Identification and Verification System Based on Spatial Energy Distribution of Wavelet Sub-Bands," *International Journal of Emerging Technology and Advanced Engineering,* vol. 4, no. 5, pp. 727- 734, 2014.

[7] R. Wang, G. Wang, Z. Chen, Z. Zeng and Y. Wang, "A palm vein identification system based on Gabor wavelet features," *Neural Computing and Applications,* vol. 24, no. 2014, p. 161–168, 2014.

[8] D. Y. Perwira, . T. B. W. Agung and M. D. Sulistiyo, "Personal Palm Vein Identification Using Principal Component Analysis and Probabilistic Neural Network," in *2014 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2016.

[9] A. Jalali, R. Mallipeddi and M. Lee, "Deformation Invariant and Contactless Palmprint Recognition Using Convolutional Neural Network," in *HAI '15: Proceedings of the 3rd International Conference on Human-Agent Interaction*, New York ,NY, United States, 2015.

[10] K. F. H. Holle, J. Y. Sari and Y. P. Pasrun, "Local Line Binary Pattern and Fuzzy K-NN for Palm Vein Recognition," *Journal of Theoretical and Applied Information Technology,* vol. 95, no. 13, pp. 2906-2912, 2017.

[11] K. He, X. Zhang, S. Ren and J. Sun, "Identity Mappings in Deep Residual Networks," in *European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016.

[12] A. Newell, K. Yang and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," in *European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

[14] N. L. o. P. R. (NLPR), "Biometrics Ideal Test," Institute of Automation, Chinese Academy of Sciences(CASIA), 2020. [Online]. Available: http://biometrics.idealtest.org.

[15] F. Rizki, T. A. B. Wirayuda and K. N. Ramadhani, "Identity recognition based on palm vein feature using two-dimensional linear discriminant analysis," in *2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, 2016.