A Lightweight Vision Transformer with Competitive Blocks for Finger Vein Recognition

Chi-An Chen
Department of Computer Science and Information Engineering
National Ilan University
Yilan, Taiwan
Email: b1043003@ems.niu.edu.tw

Tzu-Hung Chien
Taipei American School
Taipei, Taiwan
Email: 27henryc1@students.tas.tw

Liang-Ying Ke
Department of Engineering Science
National Cheng Kung University
Tainan, Taiwan
Email: n98131041@gs.ncku.edu.tw

Chih-Hsien Hsia, *Senior Member*, *IEEE*
Department of Computer Science and Information Engineering
National Ilan University
Yilan, Taiwan
Email: hsiach@niu.edu.tw

***Abstract*— With the widespread adoption of online payment systems, concerns over the security of financial transactions have grown significantly. Finger vein recognition has emerged as a promising biometric solution, offering enhanced security and reliability. However, existing deep learning-based finger vein recognition methods often suffer from high computational complexity and limited feature representation capabilities. To address these limitations, this study proposes a novel Lightweight Vision Transformer with Competitive Blocks (LViT-CB) architecture for finger vein recognition. The LViT-CB model enhances feature representation through a competitive mechanism, thereby improving the Correct Identification Rate (CIR). Experimental results demonstrate that the proposed model achieves CIRs of 99.93%, 99.17%, and 98.92% on the FV-USM, PLUSVein-FV3 (LED), and PLUSVein-FV3 (Laser) public datasets, respectively. In terms of Equal Error Rate (EER), the model attains 0.06%, 1.19%, and 1.09% on the corresponding datasets. Moreover, the LViT-CB model contains only 1.06 million parameters and operates with a computational cost of just 0.27 GFLOPs, making it highly suitable for deployment on resource-constrained platforms, such as those used in embedded systems for financial applications.**

Keywords*— Computer Vision, Vision Transformer, Biometrics, Finger Vein.

I. INTRODUCTION

The development of biometric recognition technologies has resulted in it being used in various fields for security and authentication purposes. This technology is able to identify the unique physical traits possessed by individuals, these characteristics are defined by the fundamental properties of universality, distinctiveness, permanence and collectability [1]. Common physical features used for biometric recognition include fingerprints, irises, palm prints, faces and finger veins. In contrast to other biometric traits, finger vein features provide higher security, anti-counterfeiting capabilities and the ability to detect whether an individual is living. Furthermore, since finger vein features are internal characteristics, they are not easily susceptible to environmental influence, therefore making them difficult to steal and thus, highly reliable for identity recognition [2]. Many studies have been able to successfully integrate Deep Learning (DL) and Computer Vision (CV) technologies into finger vein recognition. Despite this, Deep Neural Networks (DNNs) often have a large number of model parameters, which makes uses DNN-based finger vein recognition techniques on embedded devices and low-cost contexts more difficult. To solve this problem, some studies have introduced finger vein recognition models based on Federated Learning (FL) or Decentralised Learning, allowing DNN models to be trained on edge devices while synchronising trained model weights to a cloud platform to improve performance. However, since FL requires frequent synchronization of model weights between devices and servers, it is not suitable for affine access. To add on, the frequent communications result in increased computational and transmission costs affecting the efficiency of real-time recognition. Although decentralised learning does not rely on a central server, it requires devices to exchange model weights, which also thereby increases the computational costs. Therefore, designing low computational complexity and efficient finger vein recognition methods suitable for embedded devices has become a vital problem that must be addressed.

In order to effectively extra vein features while simultaneously reducing model complexity, previous studies have applied models such as Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Vision Transformers (ViTs) to finger vein recognition, achieving good results [3-5]. However, according to experimental results from these studies, merely increasing model depth does not necessarily improve recognition performance. Rather, it may increase model complexity and affect efficiency in execution. As a result, lightweight finger vein recognition models have become a research focus. These models, with low complexity and faster speeds in reasoning, are often more suited to usage of low-cost embedded platforms. Lu *et al*. [6] proposed a lightweight T2T-ViT model for finger vein recognition. This method reduced the channel dimension size within the model to lower computational complexity and reduce resource requirements. However, as the model's complexity decreased, a drop in Correct Identification Rate (CIR) was observed. In another study, Chai *et al*. [7] used a series of convolutional layers to extract spatial features from images of finger veins, followed by a global average pooling to average the spatial information of features and convert them into feature vectors. This avoids substantial increase in fully connected layer parameters that would result from flattening feature maps directly into one-dimensional vectors. However, since this model relies on convolutional layers alone to extract features, it lacks the ability to capture global features, making

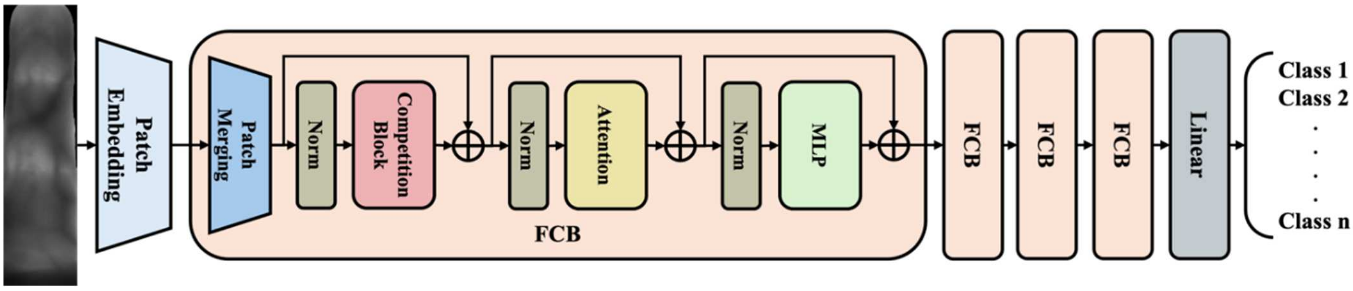

Fig. 1. LViT-CB model mentioned in this study.

it prone to missing Long-Range Dependencies (LRD), thus, ultimately affecting the CIR of identity recognition.

In view of this, this study proposes the use of Competition Blocks (CB) alongside a Lightweight Vision Transformer for finger vein recognition. The LViT-CB model determines competitive relationships among observed features, thereby enabling the finger vein recognition model to enhance its ability to represent finger vein features so that it can achieve a relatively higher CIR under limited computational resources.

II. Proposed Method

A. Overall Architecture

In order to effectively enhance the model's feature representation capability for finger vein images, this study proposes the LViT-CB model architecture, which integrates Competitive Blocks (CBs) with lightweight vision transformers for finger vein recognition, as illustrated in Figure 1. Firstly, the LViT-CB model uses convolution operations in a patch embedding layer to aggregate local features from the finger vein image and divide them into multiple patches. This process effectively suppresses redundant information in the image while significantly reducing the model's computational complexity that may slow it down. Next, the model uses Feature Competition Blocks (FCBs) – which include patch merging layers, CB modules, attention mechanisms and Multilayer Perceptrons (MLPs) – to extract finger vein features from patches in greater depth, improving, the model's ability to represent finger vein features. By repeatedly stacking FCB modules, the LViT-CB model progressively compresses the spatial dimensions of the feature maps while simultaneously expanding the number of channels. Finally, the model uses a linear layer to map the features extracted by the FCB modules to a classification dimension, thereby performing identity recognition on the finger vein image.

B. Competition Block

To effectively enhance the model's capabilities in extracting key features from finger vein images and to evaluate the importance of features across different dimensions in the recognition model, this study designs and proposes a CB that performs feature extraction based on a competitive mechanism. The model is shown in Fig. 2, where F represents the input features and F' represents the output features. The CB extracts finger vein features from the input feature map using two branches.

First, through the competition mechanism within the competition module, softmax operations are applied along the channel dimensions of the C-axis as well as the spatial dimensions of the X-axis and Y-axis. This allows the model to perceive critical features within each dimension. In Fig. 3, the extracted features are then averaged, where again F is the input and F' is the output. Second, the other branch uses convolutional layers to transform the feature map, extracting finer finger vein features, which are further refined by the competition module to emphasis important finger vein characteristics. Finally, the features extracted from both branches are concatenated, and a convolutional layer is used to restore the channel dimension to its original size for further processing by subsequent modules. Through this design, the CB module not only effectively captures the importance of features across both channel and spatial dimensions, but also significantly enhances the feature representation capability of the model, thereby improving its overall performance in identification for finger vein recognition. Moreover, this competition mechanism enables the overall model to perceive and identify critical information in multi-scale feature maps across different stages of the Feature Competition Blocks (FCBs). As a result, it can effectively learn the appearance textures and structural features of finger vein images, thereby significantly improving the model's ability to differentiate between identities in recognition tasks.

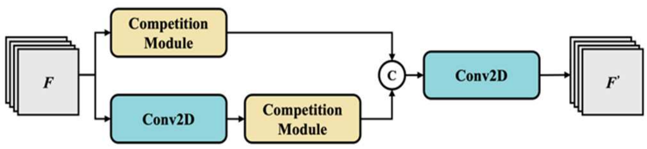

Fig. 2. Diagram of Competition Block.

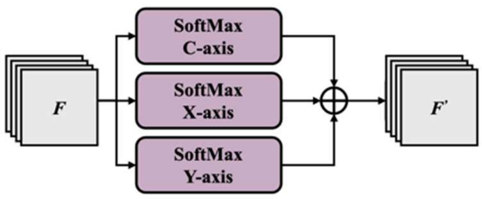

Fig. 3. Diagram of Compete Module.

III. Experimental Results And Analysis

This section provides a detailed overview of the experimental designs and related specifics of this study. First, the datasets used in the experiment are introduced in Section III-A. Then, the settings for model training and implementation environment are described in Section III-B. Finally, the experimental results and analysis are presented and discussed in Section III-C.

A. Public Datasets

To evaluate the identification and generalization capabilities of the proposed finger vein recognition model, the study utilizes two publicly available datasets namely FV-

USM [8] and PLUSVein-FV3 [9] that will train and evaluate the performance of the LViT-CB model. The content and characteristics of each dataset are described in detail below.

1) FV-USM Dataset

This dataset contains the finger vein images of 123 subjects, ranging from the age of 20 to 52, focusing on the left index finger, left middle finger, right index finger, and right middle finger. Each finger is treated as a separate class, resulting in a total of 492 classes. Each subject underwent two imaging sessions with a two-week interval, and six images per finger were captured during each session. In total, the dataset contains 5,904 near-infrared (NIR) images with resolutions of 640x480.

2) PLUSVein-FV3 Dataset

This data set uses NIR LEDs and laser sensors to collect the finger vein images of 60 subjects. Each sensor collects the images from the palmar (palm side) and dorsal (back side) of the hands of the subjects, forming two types of subsets. Each subset contains 1,800 images, and all four subsets combined include 7,200 images. The dataset collects images from six fingers per subject: left index, middle, and ring fingers, and right index, middle, and ring fingers. Each finger is treated as a distinct class, giving the dataset a total of 360 classes. The original images have a resolution of 1280 x 1024. To assist the model in effectively extracting features, regions of interest (ROI) are cropped from the images and resized to 192 x 736 resolution. For fair comparison with other models, only palmar images are used in this study. In order to effectively evaluate the model's CIR on these datasets, they are divided into training, validation and testing sets based on the number of samples per class, ensuring equal distribution across sets. The FV-USM dataset is split into a 4:1:1 ratio for training, validation, and testing, while the PLUSVein-FV3 dataset is split into a 3:1:1 ratio.

B. Environment Setting

In this study, the image size is set to 112, the batch size is set to 32, and the model is trained for 500 epochs. In order to ensure effective weight updates, the AdamW optimizer was used in this study, with an initial learning rate of 0.0002 and weight decay of 0.0001. All model architectures are implemented using PyTorch deep learning frameworks, and training is conducted on the hardware environments provided by the Intel ® Core(™) i7-12700 CPU and the Nvidia RTX 4080 graphics processing unit GPU.

C. Data Analysis

To evaluate the overall performance of the proposed finger vein recognition model, the public datasets of FV-USM, PLUSVein-FV3 (LED), and PLUSVein-FV3 (Laser) are used for training and testing. The model's recognition ability is assessed using two criterias: CIR and Equal Error Rate (EER). The CIR metric measures the accuracy of the model. A higher CIR indicates better security and overall performance of the model, thus making it suitable for various recognition tasks. The EER provides a more holistic evaluation of both accuracy and usability of the model. It is determined via adjusting thresholds until the False Accept Rate (FAR) equals the False Rejection Rate (FRR). A lower EER represents greater security and stability of the model.

Based on experimental results, the proposed LViT-CB model achieves CIR values of 99.93%, 99.17%, and 98.92% on the FV-USM, PLUSVein-FV3, and PLUSVein-FV3 datasets, respectively. In terms of EER, the model achieves 0.06%, 1.19%, and 1.09% on the same three datasets, as shown in Tables 1 and 2. Compared with previously superior finger vein recognition methods, the LViT-CB model demonstrates higher recognition accuracy and lower EER across all datasets. This further highlights the model's robust generalization ability and stability under various sensors and environments.

Moreover, in terms of model lightweighting, the proposed LViT-CB model contains only 1.06 million parameters and has a computational complexity of 0.27 GFLOPs, as shown in Table 3. When compared to previous methods, the proposed model not only achieves a lower parameter count but also demonstrates better CIR and EER results. These findings confirm that the LViT-CB model is suitable for deployment in edge computing and low-cost embedded platforms.

TABLE I. A COMPARISON IN CIR AND EER WITH PREVIOUS FINGER VEIN RECOGNITION METHODS ON THE FV-USM DATASET.

Methods	CIR (%)	EER (%)
ViT-Cap [5]	98.68	0.28
AGCNN [10]	93.80	3.89
lightweight CNN [11]	97.95	1.07
JAFVNet [12]	99.36	0.42
LCAModel [13]	99.42	0.19
VC-Gabor+MBS [14]	99.88	0.13
This work	**99.93**	**0.07**

TABLE II. A COMPARISON OF CIR AND EER WITH PREVIOUS FINGER VEIN RECOGNITION METHODS ON THE PLUSVEIN-FV3 DATASET.

Methods	CIR (%)		EER (%)	
	LED	Laser	LED	Laser
LDA-FV [15]	97.50	97.22	-	-
MultiFVNet [16]	-	-	1.27	-
ILCNN [17]	95.90	93.52	1.28	1.73
This work	**99.17**	**98.92**	**1.19**	**1.09**

TABLE III. A COMPARISON OF MODEL PARAMETERS AND COMPUTATIONAL COMPLEXITY WITH PREVIOUS FINGER VEIN RECOGNITION METHODS.

Methods	Params (M)	FLOPs (G)
VGG-16	136.28	7.94
EfficientNet	21.47	2.91
DenseNet-121	7.46	1.37
MobileNetV2	3.50	0.32
This work	**1.06**	**0.27**

To further evaluate the impact that the proposed CB block has on finger vein identification, this study uses the FV-USM, PLUSVein-FV3, and PLUSVein-FV3 datasets to conduct ablation studies, as demonstrated in Tables 4 and 5. According to experimental results, this study's proposed usage of CB blocks allowed for the improvement in the values of the previously mentioned CIR and EER metrics in all datasets. These results verify that the CB module can effectively capture the degree to which channel and spatial dimensions are important in detecting finger vein features, thereby enhancing the model's feature representation capability and improving performance in identity recognition.

TABLE IV. ABLATION STUDY OF THE CB MODULE IN THE LVIT-CB MODEL ON THE FV-USM DATASET.

Methods	CB	CIR (%)	EER (%)
This work	w/o	99.90	0.13
	w/	**99.93**	**0.06**

TABLE V. ABLATION STUDY OF THE CB MODULE IN THE LVIT-CB MODEL ON THE PLUSVEIN-FV3 DATASET.

Methods	CB	CIR (%)		EER (%)	
		LED	Laser	LED	Laser
This work	w/o	98.57	98.40	1.74	1.83
	w/	**99.17**	**98.92**	**1.19**	**1.09**

IV. CONCLUSION

To address the issues of high computational complexity and insufficient feature representation in existing finger vein recognition models, this study proposes a novel model, LViT-CB, which integrates a CB module with lightweight ViT architecture. The proposed CB module enables the LViT-CB model to effectively extract and filter salient features from finger vein images while significantly reducing both the number of parameters and overall computational complexity. Experimental results demonstrate that the proposed method achieves CIR of 99.93%, 99.17%, and 98.92% on the FV-USM, PLUSVein-FV3, and PLUSVein-FV3 public datasets, respectively. In terms of EER, the model attains 0.06%, 1.19%, and 1.09%, respectively. Furthermore, the LViT-CB model consists of only 1.06 million parameters and operates with a computational complexity of merely 0.27 GFLOPs. Compared to existing methods, the LViT-CB model offers substantial advantages in both model compactness and computational efficiency. These results underscore its potential for deployment on embedded platforms, particularly in applications within the financial payment sector.

REFERENCES

[1] C.-H. Hsia and C.-H. Liu, "New hierarchical finger-vein feature extraction method for iVehicles," *IEEE Sensors Journal*, vol. 22, no. 13, pp. 13612-13621, 2022.

[2] Y.-Y. Chen, C.-H. Hsia and P.-H. Chen, "Contactless multispectral palm-vein recognition with lightweight convolutional neural network," *IEEE Access*, vol. 9, pp. 149796-149806, 2021.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[4] H. Qin, M. A. E. Yacoubi, J. Lin, and B. Liu, "An iterative deep neural network for hand-vein verification," *IEEE Access*, vol. 7, pp. 34823-34837, 2019.

[5] Y. Li, H. Lu, Y. Wang, R. Gao, and C. Zhao, "ViT-Cap: a novel vision transformer-based capsule network model for finger vein recognition," *Applied Sciences*, vol. 12, no. 20, pp. 10364, 2022.

[6] Z. Lu, R. Wu, and J. Zhang, "Finger-vein feature extraction method based on vision transformer," *Journal of Electronic Imaging*, vol. 31, no. 4, pp. 043010, 2022.

[7] T. Chai, J. Li, S. Prasad, Q. Lu, and Z. Zhang, "Shape-driven lightweight CNN for finger-vein biometrics," *Journal of Information Security and Applications*, vol. 67, pp. 103211, 2022.

[8] C. Kauba, B. Prommegger, and A. Uhl, "The two sides of the finger - an evaluation on the recognition performance of dorsal vs. palmar finger-veins," *International Conference of the Biometrics Special Interest Group,* pp. 1-5, 2018.

[9] M. S. M. Asaari, S. A. Suandi, and B. A. Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3367-3382, 2014.

[10] Y. Zhang, W. Li, L. Zhang, X. Ning, L. Sun, and Y. Lu, "Adaptive learning Gabor filter for finger-vein recognition," *IEEE Access*, vol. 7, pp. 159821-159830, 2019.

[11] D. Zhao, H. Ma, Z. Yang, J. Li, and W. Tian, "Finger vein recognition based on lightweight CNN combining center loss and dynamic regularization," *Infrared Physics & Technology*, vol. 105, pp. 103221, 2020.

[12] J. Huang, M. Tu, W. Yang, and W. Kang, "Joint attention network for finger vein authentication," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021.

[13] Z. Zhang and M. Wang, "Finger vein recognition based on lightweight convolutional attention model," *IET Image Processing*, vol. 17, no. 6, pp. 1864-1873, 2023.

[14] J. Li, H. Wang, S. Wei, J. Zhou, Y. Shen, and L. Tao, "Variable curvature gabor convolution and multibranch structures for finger vein recognition," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 9, pp. 4753-4764, 2024.

[15] L.-Y. Ke and C.-H. Hsia, "Finger-vein secure access based on lightweight dual-attention convolutional neural network for quality distance education," *Sensors and Materials*, vol. 36, no. 3, pp. 945-955, 2024.

[16] X. Wang and P. Fang, "Enhancing multi-finger vein recognition based on multi-task learning," *International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics*, pp. 198-203, 2023.

[17] C.-H. Hsia, L.-Y. Ke, and S.-T. Chen, "Improved lightweight convolutional neural network for finger vein recognition system," *Bioengineering*, vol. 10, no. 8, pp. 919, 2023.

[18] L.-Y. Ke, Y.-C. Lin, and C.-H. Hsia, "Finger vein recognition based on vision transformer with feature decoupling for online payment applications," *IEEE Access*, vol. 13, pp. 54636-54647, 2025.