The Visual Computer
https://doi.org/10.1007/s00371-024-03286-6

RESEARCH

Veintr: robust end-to-end full-hand vein identification with transformer

Shenglin Lu[1] · Sheldon Fung[2] · Wei Pan[1] · Nilmini Wickramasinghe[2] · Xuequan Lu[2]

Accepted: 20 January 2024

Abstract
Hand vein identification stands out to be an increasingly popular approach for biometric identification due to its distinctiveness and convenience. While state-of-the-art techniques are able to achieve good performance, they share two common drawbacks: (1) complex preprocessing procedures, e.g., vein enhancement and Region of Interest (ROI) extraction, and (2) vein information loss due to hand ROI partition. To address these issues, we propose VeinTr, an end-to-end full-hand vein identification approach. In particular, our VeinTr consists of three components: a local feature extractor, a lightweight transformer, and a global feature decoder. We first obtain local features via convolution-based ResNet-like blocks. Then the attention mechanism is employed to aggregate global features from local features, which can be then decoded as global hand vein features. Finally, a global feature decoder is applied to generate robust hand features. By doing so, VeinTr is capable of directly extracting robust hand vein features from raw hand vein images. We evaluate our method on CASIA, TPV, and PLUSVein hand vein datasets. Experimental results show that our approach outperforms the state-of-the-art methods and has strong inter-dataset generalization abilities.

Keywords Palm vein identification · Biometrics security · Transformer

Mathematics Subject Classification 0000 · 1111

1 Introduction

Secure personal identification has become increasingly significant nowadays due to the rapid evolution of digital technology. The consequences of erroneously granting access to impostors within critical systems, including but not limited to banking systems and social networks, are potentially catastrophic. Such errors can lead to not only losses at the financial or reputational level but also other immeasurable ramifications. Token-based authentication methods such as passwords and personal information are feasible in certain scenarios; however, they require users with good memory and can be possibly attacked. Physical identifiers, for instance, bank cards, compromise convenience and meanwhile can potentially be forged. In recent years, biometric identification has come to the horizon and swiftly become a dominant alternative in the field primarily due to its distinctiveness and convenience.

Among the human physiological features, the face [1, 2] and hand stand out as the two most widely used components in contemporary applications. However, the face-based recognition approach can be potentially attacked [3, 4]. As the awareness of privacy grows among the public, hand-based methods have gained preference over face-based methods. Such preference is driven by the fact that hand-based methods are inherently less likely to be exposed to the public and subsequently subject to malicious data collection. In the domain of hand feature (e.g., finger and palm) identification, hand vein features have taken precedence over hand print features due to similar reasons [5].

Earlier research attempts to leverage the handcrafted-based geometrical [6–10] or statistical [11–14] information exhibited in hand vein image under near-infrared light (NIR). These methods are feasible in certain application scenarios and, however, are still subpar under perturbations such as

Shenglin Lu and Sheldon Fung contributed equally.

✉ Xuequan Lu
b.lu@latrobe.edu.au

[1] OPT MV, Dongguan, China

[2] La Trobe University, Melbourne, VIC, Australia

Published online: 20 March 2024

illumination and noise. In the last decade, advanced deep learning techniques have been exploited and combined with handcrafted features to extract robust vein features [15, 16]. It merely alleviates the drawbacks of handcrafted features but is still vulnerable to real-world application. More recent works [17, 18] fully abandoned the handcrafted features and employed convolutional neural networks (CNNs) to extract features from vein images.

Nonetheless, current deep learning methods typically focus on feature extraction from the Region of Interest (ROI). This requires a series of preprocessing steps involving hand localization and ROI extraction. Consequently, the performance of those methods still relies on the robustness of ROI extraction, which is challenging in real-world applications. On the other hand, the vein pattern throughout the hand is unique for each individual. Therefore, using only the ROI for feature representation is potentially inferior in comparison to leveraging the full-hand vein pattern.

To this end, we propose VeinTr, an end-to-end full-hand vein identification/matching approach. Without ROI extraction, it directly takes a full-hand vein image as input. In particular, the local feature extractor employs CNNs to simultaneously reduce the pixel-wise feature size and extract prior local features. These features are then fed to a series of residual blocks [19] to enhance the local feature representation. Subsequently, a lightweight transformer [20, 21] is adopted to enlarge the perceptive field and aggregate the global feature. In the end, the hand vein features can be obtained from the global feature decoder.

The contributions of this work can be listed as follows:

- We propose VeinTr, a hand vein identification framework that consists of CNNs and a lightweight Transformer.
- Compared to previous methods, our designed network directly takes the full-hand image as input, eliminating complex preprocessing, e.g., ROI extraction.
- We conduct extensive experiments, including intra- and inter-dataset validation on the proposed method, and compare the results with the state-of-the-art approaches, demonstrating the effectiveness of our method.

The rest of the paper is organized as follows. Section 2 reviews previous research work. Section 3 presents the proposed approach. In Sect. 4, we evaluate our method using three publicly available datasets and analyze the experimental results. Section 5 concludes this work.

2 Related work

Typically, state-of-the-art techniques for hand vein identification follow a four-step procedural framework: (1) hand localization, (2) ROI extraction, (3) feature extraction, and (4) probe-to-database matching. Among these, extracting robust vein feature representations stands out as the decisive procedure that leads to a reliable hand vein biometric identification system. Vein feature extraction methods can be roughly categorized into two groups: handcrafted and learning methods.

Handcrafted methods Fingerprints [22–25] and palmprints [26–29] for biometric recognition are widely studied. Similar to them, hand vein image consists of rich topological geometry information such as the principal line, branches, and minutiae points. Building stable feature representations from these visual cues experienced a longstanding investigation. Early attempts employed explicit geometry-based detectors such as wide line detector [6] and neighborhood matching radon transform [7]. In line with this approach, Wu et al. [8] proposed to use a directional filter bank to extract line-based features, complemented by a minimum directional code for feature encoding. Wirayuda et al. [9] proposed to exploit the vein topology and utilize the minutiae feature for vein representation.

Another line of research focuses on leveraging statistical information in the vein image. Techniques such as invariant moments [30], local binary pattern (LBP) [11–14], and local derivative pattern (LDP) [11, 14, 31] have found widespread application for vein feature extraction. Although these methods exhibit high performance on public datasets, they are still vulnerable to perturbations in real-world applications. To address vein distortion in practical scenarios, Kang et al. [32] proposed to adopt RootSIFT to overcome the projection transformation in contact-free settings. Rahul et al. [33] proposed the mutual foreground local directional texture pattern (MF-LDTP) for feature extraction, aimed at noise suppression. Pratiwi et al. [13] employed a local binary pattern rotation invariant (LBPROT) to enhance the robustness against rotation. Wu et al. [34] proposed Haar-wavelet decomposition and partial least square (HDPLS) to effectively extract the main subspace feature while mitigating the non-significant noise.

Learning-based methods Although the above-mentioned approaches alleviate the shortcomings of handcrafted methods to an extent, their performances are still conditioned to imaging factors such as noise and lighting. Recently, researchers have resorted to deep learning techniques to extract robust vein features. Perwira et al. [35] adopted principal component analysis (PCA) for feature extraction and a probabilistic neural network (PNN) to perform classification. Building upon this work, Fronitasari et al. [15] proposed to improve the performance of PNN by incorporating LBP features. Similarly, Bhilare et al. [16] employed center-symmetric LBP for vein feature extraction, subsequently applying DeepMatching [36] for probe-to-database matching.

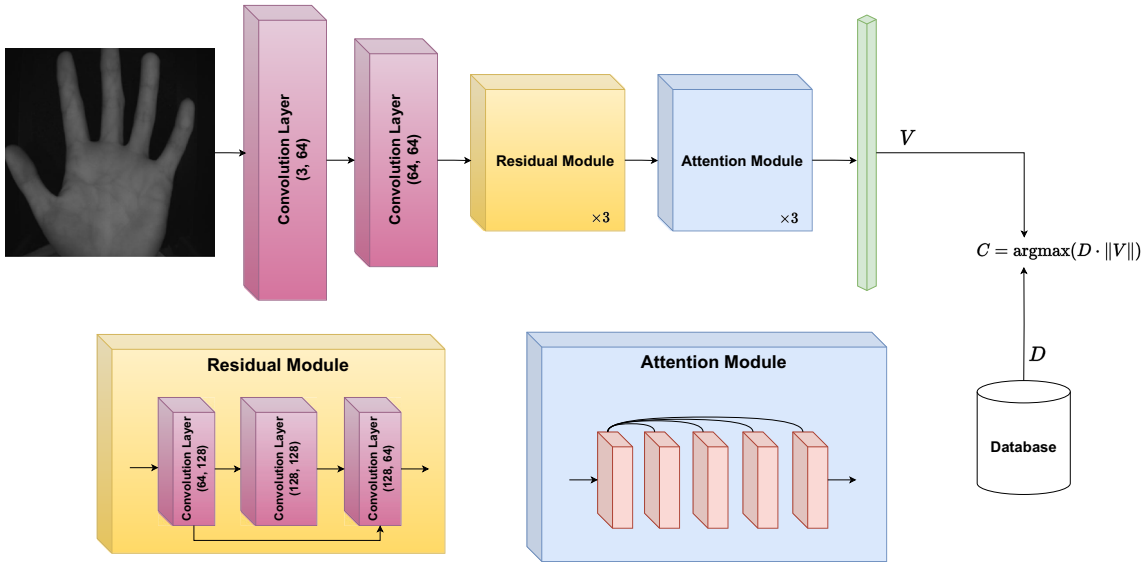


Fig. 1 The overview of our proposed VeinTr pipeline. The full-hand image is first fed into a local feature encoder to extract local features and meanwhile reduce its spatial size. Then the attention blocks are employed to aggregate global features from the extracted local features. Finally, the hand vein features can be obtained through the MLP-based global feature decoder. Notably, our method pioneers the end-to-end extraction of full-hand features

However, the performance of the aforementioned methods is still constrained by the initial handcrafted features. To address this issue, Thapar et al. [17] proposed an end-to-end PVSNet, leveraging a deep convolutional neural network (DCNN) for feature extraction. Chen et al. [18] addressed challenges related to insufficient training samples and high computation costs by adopting symmetric discrete wavelet transform (SMDWT-PCA) for vein image augmentation and depth separable convolution (DSC) for model parameters reduction, respectively.

Unlike the previous methods, our proposed method does not require an ROI extraction step and can be directly fed with unprocessed full-hand image data.

3 Method

In previous works, ROI extraction is a crucial step in the context of biometric identification based on hand attributes, e.g., palm vein [17, 18] or finger vein [37, 38]. Consequently, the identification accuracy exhibits a positive correlation with the quality of the extracted ROI. On the other hand, the unique patterns present in both palm and finger areas contribute to the extraction of distinctive feature representations from a specific hand. In the neighboring research domain, the fusion of palmprint and fingerprint has been investigated in pioneering work [28]. Nevertheless, it extracts the features from those two areas individually, potentially compromising the topology structure of the hand pattern as a whole.

To address these issues, we propose VeinTr, an end-to-end full-hand vein identification/matching approach. Our method consists of only two steps: (1) feature extraction and (2) probe-to-database matching. Figure 1 illustrates our overall framework.

3.1 Vein transformer

Given a full-hand image $X \in \mathbb{R}^{H\times W\times C}$, where H and W are the height and width of the image and C is the number of image channels, a local feature encoder that consists of a convolutional neural network and residual blocks is employed to extract the local features $F_l \in \mathbb{R}^{H'\times W'\times D}$. Then a lightweight transformer is adopted to aggregate global contextual features by enlarging the perceptive field via the attention mechanism. The global features $F_g \in \mathbb{R}^{L\times M}$ outputted from the transformer are then fed to a global feature encoder to obtain the final feature representation $F^h \in \mathbb{R}^O$ for probe-to-database matching.

Local feature encoder We adopt a series of convolution blocks to simultaneously extract the prior local feature and reduce the size of the input data. A convolution block consists of a convolutional layer, a batch normalization operation, and an activation function. We employed PReLU [39] as the activation function in our experiments. The outputted prior local features $F_p \in \mathbb{R}^{H'\times W'\times D}$ from the convolution blocks are then fed into a three-layer residual block [19] to extract local feature $F_l \in \mathbb{R}^{H'\times W'\times D}$. In our experiments, given an image $X \in \mathbb{R}^{256\times 256\times 3}$, $H' = W' = D = 64$.

Attention layer To leverage the attention mechanism for global feature aggregation, we first reshape the local feature $F_l \in \mathbb{R}^{H' \times W' \times D}$ into a patch-based 2D sequence $F_s \in \mathbb{R}^{N \times (P^2 \times D)}$, where P is the edge length of each patch (e.g., 16 in our experiments) and $N = H'W'/P^2$ is the resulting number of patches. Each patch is flattened and projected into a lower dimension d. Subsequently, the projected features $F_d \in \mathbb{R}^{N \times d}$ are fed into a L-layer multi-head attention blocks. Each multi-head attention block is defined as follows:

$$\mathcal{A}(Q, K, V) = (Head_1 \oplus \cdots \oplus Head_H)W^O, \tag{1}$$

where Q, K, and V are the input query, key, and value, respectively. $\oplus$ denotes the concatenation operation. $W_h^Q \in \mathbb{R}^{Hd_{head} \times d}$ is a learnable projection matrix. The number of heads H is set to 8 in our experiments and $d_{head} = d/H$. $Head_h$ is defined as follows:

$$Head_h = \text{softmax}\left(\frac{QW_h^Q(KW_h^K)^T}{\sqrt{d_{head}}}\right)VW_h^V, \tag{2}$$

where $W^O, W_h^Q, W_h^Q \in \mathbb{R}^{d \times d_{head}}$ are learnable projection matrices. In the case of $L = 1$, the global feature $F_{g'}$ can be obtained by setting F_d to be query, key, and value as follows:

$$F_{g'} = \text{LNorm}(\mathcal{A}(F_d, F_d, F_d)), \tag{3}$$

where $LNorm(\cdot)$ denotes the layer normalization operation. In our experiments, we set the number of attention layers L to 3.

Global feature decoder The extracted global feature F_g can be used as a robust hand vein representation. To this end, we use a two-layer MLP to reduce the dimensionality of the output vein feature. F_g is first flattened to $\hat{F}_g \in \mathbb{R}^{1 \times LM}$ and then the final hand vein feature $F^h \in \mathbb{R}^{1 \times O}$ can be obtained as follows:

$$F^h = \text{BNorm}(\text{BNorm}(\hat{F}_g W_1 + b_1)W_2 + b_2), \tag{4}$$

where BNorm$(\cdot)$ denotes the batch normalization operation. W_1, W_2, b_1, and b_2 are learnable weights and biases in MLP layers, respectively. The output feature dimension O is set to 512 in our experiments.

3.2 Loss functions

Our proposed network can be trained in an end-to-end manner as a metric learning task. Concretely, the training should optimize the vein feature F^h to enforce higher similarity among intra-class samples, and conversely, distinctiveness for inter-class samples. In our experiments, we adopt ArcFace loss [40] as our objective function, which is defined as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1, j\neq y_i}^{n} e^{s\cdot\cos\theta_j}}, \tag{5}$$

where N and n are the batch size and the number of classes, respectively. $s = \|F_i^h\|$, where $\|F_i^h\|$ is the i-th sample, belonging to the y_i-th class and $\|\cdot\|$ is the l_2 normalization.

$$\cos\theta_j = \frac{W_j^T F_i^h}{\|W_j\|\|F_i^h\|}, \tag{6}$$

where $W_j \in \mathbb{R}^O$ denotes the j-th column of the learnable weight $W \in \mathbb{R}^{O \times n}$.

3.3 Probe-to-database matching

Given a database comprising data samples $G = \{g_i | g_i \in \mathbb{R}^{H \times W \times C}, i = 1, ..., n\}$, our initial step involves the creation of a database matrix denoted as $D \in \mathbb{R}^{n \times O}$. This matrix is formed by assigning $\|VeinTr(g_i)\|$ to the i-th column of D, where $g_i \in G$ and $VeinTr(\cdot)$ refers to our proposed method, as elaborated in Sect. 3.1. Then given a set of probe data samples $P = \{p_i | p_i \in \mathbb{R}^{H \times W \times C}, i = 1, ..., n\}$, the predicted class of a sample p_i can be calculated as follows:

$$C = \text{argmax}(D \cdot \|VeinTr(p_i)\|), \tag{7}$$

where C denotes the predicted class.

4 Experiment results

4.1 Implementation

Our method is implemented in PyTorch and we use an SGD optimizer during the training stage with an initial learning rate of $1e-2$. The momentum and the weight decay rate are $9e-1$ and $5e-4$, respectively. The batch size is set to 128, and the total number epochs is 1000. All training is conducted on an NVIDIA 2080Ti GPU.

4.2 Datasets

We employ three hand vein datasets that are publicly available to validate our method.

CASIA dataset [41] is a multi-spectral palmprint image repository containing a total of 7200 hand images captured from 100 different people using a custom-designed multiple spectral imaging device. For each hand, 6 samples are included, collected across two sessions at one-month

intervals. Each sample contains 6 images of different wavelengths. We only select images with a wavelength of 850 and 940 nm for our experiments. Image samples of the CAISA dataset are shown in Fig. 2a.

Tongji palm vein dataset (TPV) [42] is a large-scale contactless palm vein repository containing a total of 12000 hand images captured from 300 volunteers. For each hand, 10 images are captured in each session with a total of two sessions. Similar to the CASIA dataset, the average time interval between sessions was about two months. Image samples of the Tongji dataset are shown in Fig. 2b.

PLUSVein dataset [43] is a relatively small palm vein dataset containing a total of 420 hand images captured from 42 volunteers using 850 nm wavelength illumination. For each hand, five images are captured. Image samples of the PLUSVein dataset are shown in Fig. 2c.

In our experimental setup, we employ the CASIA dataset specifically for the intra-dataset vein matching task, given its extensive utilization in prior research as a widely recognized benchmark. We incorporate all three datasets in the inter-dataset vein matching task to facilitate comprehensive comparisons and assessments.

The uniqueness of hand vein patterns extends to both hands of an individual. Therefore, treating the right and left hands as a single identity leads to a mislabeling scenario where distinct classes share the same label. We follow [17, 18], the right and left hands of the same person are regarded as hands from two independent identities during training and testing. Hand vein images from each person in all datasets are evenly separated as the training set (i.e., the gallery set) and the testing set (i.e., the probe set). For the PLUSVein dataset, we assign 3 images to the training set and the rest to the test set since they cannot be evenly separated.

4.3 Evaluation metrics

We follow Chen et al. [18] and report the correct recognition rate (CRR) and equal error rate (EER). CRR is defined as follows:

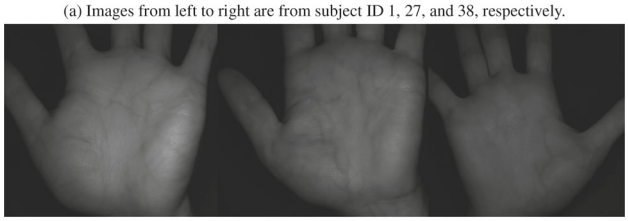
$$CRR = \frac{T_p}{N_g} \times 100\%, \tag{8}$$

where T_p is the correctly matched samples and N_g is the number of samples in the gallery. EER is the point at which the false acceptance rate (FAR) and false rejection rate (FRR) are equal.

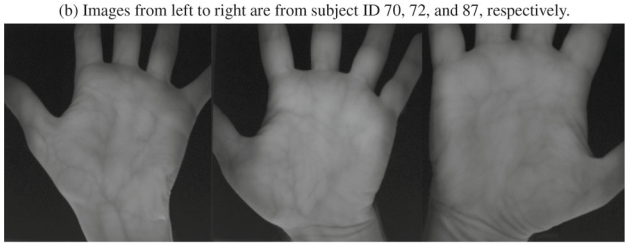
4.4 Intra-dataset vein matching

In the intra-dataset matching experiment, our proposed VeinTr is trained and tested with the CASIA dataset. The results have been compared with state-of-the-art methods, as summarized in Table 1. Notably, when considering correct recognition rate (CRR), all handcrafted methods have consistently achieved CRR exceeding 90%, and the performance of NAF [18] is nearly saturated, reaching over 99%. Despite that, our proposed method still exhibits robust capabilities, outweighing NAF [18] by 0.02% and 0.42% for 850 and 940 nm wavelengths, respectively.

(a) Images from left to right are from subject ID 1, 27, and 38, respectively.



(b) Images from left to right are from subject ID 70, 72, and 87, respectively.



(c) Images from left to right are from subject ID 5, 9, and 28, respectively.

Fig. 2 **a**–**c** are sample images from CASIA, Tongji, and PLUSVein, respectively

Although the equal error rate (ERR) metric has been presented by only a few methods, our proposed method significantly surpasses the performance of PVSNet [17] by a substantial margin of 3.61%, achieving an ERR of 0.10%. Among handcrafted methods, our proposed method outperforms the second-best method, NMRT [7], by 0.41%.

4.5 Inter-dataset vein matching

In the inter-dataset matching experiment, our proposed VeinTr is individually trained on each dataset and subsequently tested with all three datasets. The results are shown in Table 2. Notably, when the model is trained on the CASIA dataset [41] and the TPV dataset [42], the test results across three datasets consistently achieve CRR exceeding 90%, demonstrating the powerful generalization capabilities of our method.

In particular, when the model is trained on the CASIA dataset [41], it achieves remarkable test results on the TPV

Table 1 Intra-dataset vein matching evaluation on the CASIA dataset

Methods	ROI	Wavelength (nm)	CRR%	ERR%
Handcrafted Methods				
WLD [6]	Finger	850	97.50	6.08
NMRT [7]	Palm	850	99.17	0.51
MFFM [9]	Palm	850	90.87	N/A
MF-LDTP [33]	Palm	850	95.00	N/A
LBPROT [13]	Palm	940	96.00	11.70
LDP [31]	Palm	940	98.30	N/A
Deep learning methods				
PCA-PNN [35]	Palm	850	84.00	N/A
MLPB [15]	Palm	850	92.75	N/A
MDP [16]	Palm	850	89.72	N/A
FaceNet [17]	Palm	850	77.16	5.77
PVSNet [17]	Palm	850	85.16	3.71
NAF [18]	Palm	850	99.83	N/A
NAF [18]	Palm	940	99.50	N/A
VeinTr (ours)	Full hand	850	**99.85**	**0.10**
VeinTr (ours)	Full hand	940	**99.92**	**0.09**

Bold values indicate top results. A higher CRR value represents stronger performance. A lower ERR value represents stronger performance
Note that N/A stands for non-applicable

Table 2 Inter-dataset vein matching evaluation on the multiple datasets

Train set	Test set CRR%		
	CASIA [41]	TPV [42]	PLUSVein [43]
CASIA [41]	99.85	98.17	97.62
TPV [42]	94.68	99.45	96.12
PLUSVein [43]	89.76	81.43	99.05

dataset [42] and the PLUSVein dataset [43], reaching CRR of 98.17 and 97.62%, respectively.

The PLUSVein dataset is a relatively smaller dataset as described in Sect. 4.2. When the model is trained on the PLUSVein dataset, the learned feature representations are relatively weaker than those of the model trained on the large-scale dataset. Although the overall generalization ability is somewhat weaker for the model trained on the PLUSVein dataset [43], it still delivers promising results, achieving CRR of 89.76 and 82.43% when tested on the CASIA dataset [41] and the TPV dataset [42], respectively.

4.6 Ablation study

To demonstrate how each component contributes to the overall performance, we further conduct experiments specifically on the two principle components of our proposed method: the residual blocks and attention layers. In the experiments, all networks are trained only on the CASIA dataset [41] from scratch and we test it on the CASIA [41], TPV [42], and PLUSVein [43] datasets. We report the final correct recognition rate (CRR) to demonstrate the corresponding performance.

How effective is each module? For a better understanding of each module in our proposed method, we first ablate the residual blocks and the attention layers. We report the experimental results in Table 3. When we remove both the For the model without attention layers, a substantial drop of 16.42% in CRR can be observed when tested with the CASIA dataset [41]. More notably, the CRR performances on the other two datasets show a more dramatic decrease of over 30%. We notice a similar trend when the residual blocks are eliminated. Concretely, the test results on the CASIA dataset [41] experience a significant drop, reaching a CRR of 95.12%. Moreover, test results on the other two datasets also exhibit a performance decrease over 10% in CRR.

Attention layer number We conduct experiments by varying the number of attention layers while keeping other components at their default settings. The experiment results are presented in Fig. 3. For CASIA [41], CRR is saturated at the 3rd layers of attention blocks. Similarly, the CRR for TPV [42] and PLUSVein [43] also reach the optimal performance at the 3rd attention layers as well. However, it saw a notable performance drop at both the 4th and the 5th attention layers due to overfitting. We thus choose 3 layers of attention blocks by default.

Residual block number In our experiments, we manipulate the number of residual blocks while maintaining other com-

Table 3 Ablation experiments of each module

Residual	Attention	Test set CRR%		
		CASIA [41]	TPV [42]	PLUSVein [43]
–	–	56.28	29.63	34.44
✓	–	83.43	65.14	55.10
–	✓	95.12	75.07	83.67
✓	✓	99.85	98.17	97.62

✓ means the corresponding module is used in the ablation study. "Residual" and "Attention" represent the local feature encoder attention layer, respectively

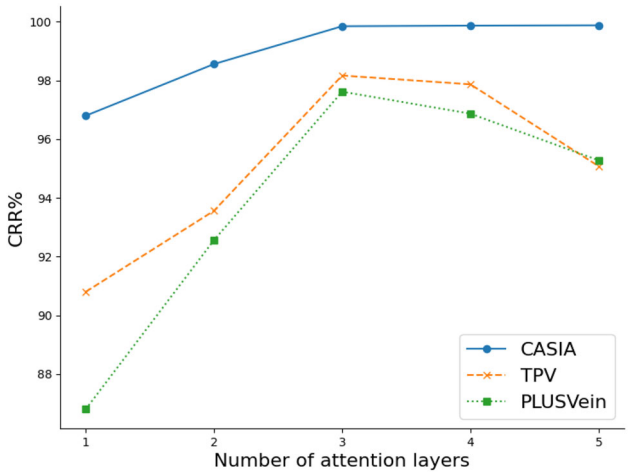


Fig. 3 CRR performance of various number of attention layers

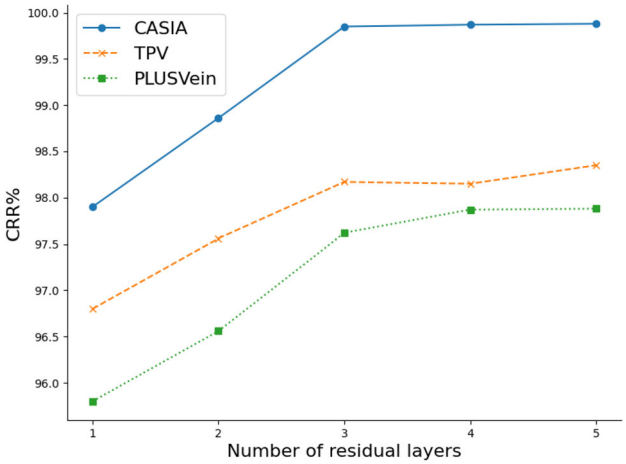


Fig. 4 CRR performance of various number of residual layers

ponents at their default settings. The experimental results are displayed in Fig. 4. Notably, similar to the trend observed with attention layers, the testing CRR of the CASIA [41] dataset saturates at the 3rd layers of residual blocks. Likewise, a steep increase in testing CRR can also be observed in both TPV [42] and PLUSVein [43] datasets when the layer number is below 4. However, unlike the behavior of attention layers, the overall testing CRR of PLUSVein [43] still saw a rising trend at the 4th and the 5th layer. For the TPV dataset [42], we can observe an increase of CRR at the 5th layer, despite a small fluctuation at the 4th layer. Thus, 3 layers of residual block is a good choice to maintain the trade-off between the CRR performance and the computational expense.

5 Conclusion

In this study, we have presented VeinTr, a novel hand vein biometric identification framework that leverages residual blocks and attention mechanisms to extract robust full-hand vein features. It addresses the two common shortcomings associated with learning-based techniques: complex preprocessing procedures and information loss from partial ROI. We evaluate our proposed method with three publicly available hand vein datasets in both intra- and inter-dataset validation manner. Experimental results show our approach outperforms the state-of-the-art methods. Additionally, we conduct ablation studies to systematically analyze and verify the effectiveness of the key components in VeinTr. Our experiments demonstrate that incorporating full-hand vein patterns contributes to the overall robustness, particularly in inter-dataset validation. Therefore, in future work, we would like to investigate the fusion of full-hand vein information with full-hand print data, further advancing the capabilities of biometric identification systems.

Acknowledgements This work is supported in part by the investigator fund (3.2501.11.47) and the industry fund (3.6267.01).

Author Contributions Shenglin Lu and Sheldon Fung did the experiments and wrote the manuscript text. Wei Pan, Nilmini Wickramasinghe, and Xuequan Lu helped with the experiment and refining the text and figures. All authors reviewed the manuscript.

Data Availability No datasets were generated during the current study.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

References

1. Jing, Y., Xuequan L., Shang G.: 3D face recognition: a comprehensive survey in 2022. Comput. Vis. Media **9**(4), 657–685 (2023)
2. Zeng, S., Xiong, Y.: Weighted average integration of sparse representation and collaborative representation for robust face recognition, Computational Visual. Media **2**, 357–365 (2016)
3. Feng, D., Lu, X., Lin, X.: Deep detection for face manipulation. In: Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18–22, 2020, Proceedings, Part V, vol. 27, pp. 316–323. Springer (2020)
4. Fung, S., Lu, X., Zhang, C., Li, C.-T., Deepfakeucl: deepfake detection via unsupervised contrastive learning. In: International Joint Conference on Neural Networks (IJCNN), vol. 2021, pp. 1–8. IEEE (2021)
5. Wu, W., Elliott, S.J., Lin, S., Sun, S., Tang, Y.: Review of palm vein recognition. IET Biometr. **9**, 1–10 (2020)
6. Huang, B., Dai, Y., Li, R., Tang, D., Li, W.: Finger-vein authentication based on wide line detector and pattern normalization. In: 20th International Conference on Pattern Recognition, vol. 2010, pp. 1269–1272. IEEE (2010)
7. Zhou, Y., Kumar, A.: Human identification using palm-vein images. IEEE Trans. Inf. Forens. Secur. **6**, 1259–1274 (2011)
8. Wu, K.-S., Lee, J.-C., Lo, T.-M., Chang, K.-C., Chang, C.-P.: A secure palm vein recognition system. J. Syst. Softw. **86**, 2870–2876 (2013)
9. Wirayuda, T.A.B.: Palm vein recognition based-on minutiae feature and feature matching. In: 2015 International Conference on Electrical Engineering and Informatics (ICEEI), pp. 350–355. IEEE (2015)
10. Ananthi, G., Raja Sekar, J., Arivazhagan, S.: Human palm vein authentication using curvelet multiresolution features and score level fusion. Vis. Comput. 1–14 (2022)
11. Mirmohamadsadeghi, L., Drygajlo, A.: Palm vein recognition with local texture patterns. Iet Biometr. **3**, 198–206 (2014)
12. Kang, W., Wu, Q.: Contactless palm vein recognition using a mutual foreground-based local binary pattern. IEEE Trans. Inf. Forens. Secur. **9**, 1974–1985 (2014)
13. Pratiwi, A.Y., Budi, W.T.A., Ramadhani, K.N.: Identity recognition with palm vein feature using local binary pattern rotation invariant. In: 2016 4th International Conference on Information and Communication Technology (ICoICT), pp. 1–6. IEEE (2016)
14. Piciucco, E., Maiorana, E., Campisi, P.: Palm vein recognition using a high dynamic range approach. Iet Biometr. **7**, 439–446 (2018)
15. Fronitasari, D., Gunawan, D.: Palm vein recognition by using modified of local binary pattern (LBP) for extraction feature. In: 2017 15th International Conference on Quality in Research (QiR): International Symposium on Electrical and Computer Engineering, pp. 18–22. IEEE (2017)
16. Bhilare, S., Jaswal, G., Kanhangad, V., Nigam, A.: Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach. Mach. Vis. Appl. **29**, 1269–1286 (2018)
17. Thapar, D., Jaswal, G., Nigam, A., Kanhangad, V., Pvsnet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features. In: IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA), vol. 2019, pp. 1–8. IEEE (2019)
18. Chen, Y.-Y., Jhong, S.-Y., Hsia, C.-H., Hua, K.-L.: Explainable AI: a multispectral palm-vein identification system with new augmentation features. ACM Trans. Multimedia Comput., Commun., Appl. (TOMM) **17**, 1–21 (2021)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv. Neural Inf. Process. Syst. **30** (2017)
21. Lin, X., Sun, S., Huang, W., Sheng, B., Li, P., Feng, D.D.: EAPT: efficient attention pyramid transformer for image processing. IEEE Trans. Multimedia (2021)
22. Öztürk, H.İ., Selbes, B., Artan, Y.: Minnet: minutia patch embedding network for automated latent fingerprint recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1627–1635 (2022)
23. Zhang, Y., Zhao, R., Zhao, Z., Ramakrishnan, N., Aggarwal, M., Medioni, G., Ji, Q.: Robust partial fingerprint recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1011–1020 (2023)
24. Kolberg, J., Priesnitz, J., Rathgeb, C., Busch, C.: Colfispoof: a new database for contactless fingerprint presentation attack detection research. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 653–661 (2023)
25. Johnson, J., Chitra, R.: Multimodal biometric identification based on overlapped fingerprints, palm prints, and finger knuckles using BM-KMA and CS-RBFNN techniques in forensic applications. Vis. Comput. 1–15 (2023)
26. Ito, K., Sato, T., Aoyama, S., Sakai, S., Yusa, S., Aoki, T.: Palm region extraction for contactless palmprint recognition. In: 2015 International Conference on Biometrics (ICB), pp. 334–340. IEEE (2015)
27. Gumaei, A., Sammouda, R., Al-Salman, A.M., Alsanad, A.: An effective palmprint recognition approach for visible and multispectral sensor images. Sensors **18**, 1575 (2018)
28. Genovese, A., Piuri, V., Scotti, F., Vishwakarma, S.: Touchless palmprint and finger texture recognition: a deep learning fusion approach. In: 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 1–6. IEEE (2019)
29. Chai, T., Prasad, S., Yan, J., Zhang, Z.: Contactless palmprint biometrics using DeepNet with dedicated assistant layers. Vis. Comput. **39**(9), 4029–4047 (2023)
30. Li, X., Guo, S., Gao, F., Li, Y.: Vein pattern recognitions by moment invariants. In: 2007 1st International Conference on Bioinformatics and Biomedical Engineering, pp. 612–615. IEEE (2007)
31. Akbar, A.F., Wirayudha, T.A.B., Sulistiyo, M.D.: Palm vein biometric identification system using local derivative pattern. In: 2016 4th International Conference on Information and Communication Technology (ICoICT), pp. 1–6. IEEE (2016)
32. Kang, W., Liu, Y., Wu, Q., Yue, X.: Contact-free palm-vein recognition based on local invariant features. PLoS One **9**, e97548 (2014)
33. Rahul, R.C., Cherian, M., Mohan, M.: A novel MF-LDTP approach for contactless palm vein recognition. In: 2015 International Conference on Computing and Network Communications (CoCoNet), pp. 793–798. IEEE (2015)
34. Wu, W., Elliott, S.J., Lin, S., Yuan, W.: Low-cost biometric recognition system based on NIR palm vein image. IET Biometr. **8**, 206–214 (2019)
35. Perwira, D.Y., Agung, B.T., Sulistiyo, M.D.: Personal palm vein identification using principal component analysis and probabilistic neural network. In: 2014 International Conference on Information Technology Systems and Innovation (ICITSI), pp. 99–104. IEEE (2014)
36. Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C.: Deepmatching: hierarchical deformable dense matching. Int. J. Comput. Vis. **120**, 300–323 (2016)
37. Qin, H., El-Yacoubi, M.A.: Deep representation-based feature extraction and recovering for finger-vein verification. IEEE Trans. Inf. Forens. Secur. **12**, 1816–1829 (2017)

38. Xie, C., Kumar, A.: Finger vein identification using convolutional neural network and supervised discrete hashing. Pattern Recogn. Lett. **119**, 148–156 (2019)
39. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
40. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)
41. Hao, Y., Sun, Z., Tan, T., Ren, C.: Multispectral palm image fusion for accurate contact-free palmprint recognition. In: 2008 15th IEEE International Conference on Image Processing, pp. 281–284. IEEE (2008)
42. Zhang, L., Cheng, Z., Shen, Y., Wang, D.: Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset. Symmetry **10**, 78 (2018)
43. Kauba, C., Prommegger, B., Uhl, A.: Combined fully contactless finger and hand vein capturing device with a corresponding dataset. Sensors **19**, 5014 (2019)

Shenglin Lu is an associate professor in mechanical engineering and the Chairman/R&D Director of OPT Machine Vision Co., Ltd. With over a decade in machine vision, he has successfully leaded and completed various national projects including "Micro-level Real-time Visual Inspection" and "Robot 3D Vision Intelligent Grasping System." He has published over 300 patents and 10 papers. His research interests include machine vision system and image processing.

Sheldon Fung is a PhD student at La Trobe University in Melbourne, Australia. His research focuses on point cloud data processing, particularly in areas such as point cloud registration and 2D-3D cross-modal learning.

Wei Pan is currently working in OPT Machine Vision Corp. as a research lead in 3D algorithm development. Prior to that, he worked as a research fellow at Shenzhen University and South China University of Technology after he got his PhD degree from Singapore University of Technology and Design. His research interests include 3D imaging, 3D Data processing, computer vision, machine learning, and computer graphics.

Nilmini Wickramasinghe, PhD, MBA, is the Professor and Optus chair of Digital Health at La Trobe University within the School of Computing, Engineering and Mathematical Sciences. She also holds honorary research professor positions at the Peter MacCallum Cancer Centre, MCRI, Epworth HealthCare and Northern Health.

Xuequan Lu received the PhD degree from Zhejiang University, China, in 2016. He spent more than two years as a Research Fellow in Singapore. He is currently a Senior Lecturer with the School of Information Technology, Deakin University, Australia. His research interests include visual computing, for example, geometry modeling, processing, and analysis, animation/simulation, and 2D data processing and analysis.