

Artificial Neural Networks for Finger Vein Recognition: A Survey

Yimin Yin^{1†}, Renye Zhang^{2†}, Pengfei Liu³, Wanxia Deng⁴, Siliang He¹, Chen Li⁵ and Jinghua Zhang^{6*}

¹School of Mathematics and Statistics, Hunan First Normal University, Changsha, 410205, Hunan, China.

²School of Computer Science, Hunan First Normal University, Changsha, 410205, China.

³College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha, 410073, Hunan, China.

⁴College of Meteorology and Oceanography, National University of Defense Technology, 410073, Changsha, China.

⁵College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, 110016, Liaoning, China.

^{6*}College of Intelligence Science and Technology, National University of Defense Technology, Changsha, 410073, Hunan, China.

*Corresponding author(s). E-mail(s): zhangjingh@foxmail.com;

[†]These authors contributed equally to this work.

Abstract

Finger vein recognition is an emerging biometric recognition technology. Different from the other biometric features on the body surface, the venous vascular tissue of the fingers is buried deep inside the skin. Due to this advantage, finger vein recognition is highly stable and private. They are almost impossible to be stolen and difficult to interfere with by external conditions. Unlike the finger vein recognition methods based on traditional machine learning, the artificial neural network technique, especially deep learning, it without relying on feature engineering and have superior performance. To summarize the development of finger vein recognition based on artificial neural networks, this paper collects 149

related papers. First, we introduce the background of finger vein recognition and the motivation of this survey. Then, the development history of artificial neural networks and the representative networks on finger vein recognition tasks are introduced. The public datasets that are widely used in finger vein recognition are then described. After that, we summarize the related finger vein recognition tasks based on classical neural networks and deep neural networks, respectively. Finally, the challenges and potential development directions in finger vein recognition are discussed. To our best knowledge, this paper is the first comprehensive survey focusing on finger vein recognition based on artificial neural networks.

Keywords: Finger vein recognition, Artificial neural networks, Deep learning, Convolutional neural networks, Image analysis

1 Introduction

1.1 Biometric recognition

The identity verification system is essential in many fields, such as account login, online payment, and automated teller machines. This technology aims to guarantee the security of user privacy information. The most widely used security mechanism is the classical password. Due to its time-consuming input process, potential leakage risk, and weak anti-attack ability, the classical password is inefficient. However, with the rapid development of information technology, biometric recognition systems are commonly applied in many authentication scenarios [1]. Biometric recognition aims at identifying a person based on physical and behavioral features, such as face [2], voice [3], fingerprint [4], etc. The general workflow of the biometric identity verification system is shown in Fig. 1. It usually contains the register and match parts. In the register part, the first step is encoding the original biometric data into feature representations through pre-processing operations, such as image processing techniques [5]. Then, these features will be registered as prototypes for their corresponding labels in the database. In the match process, the original data is encoded into the feature representation using the same or similar approach as the register part. The feature will be compared with the prototypes in the database to perform the identification task.

Biometric recognition technology is more efficient than the traditional secure identification process due to its convenience and steady security. With the increasing demand for digital security identification systems, biometric recognition demonstrates many advantages. For instance, the biometric recognition system, especially the fingerprint-based system, is widely used and efficient in many home security systems because it can make people released from having to remember passwords or carry keys [6]. Besides, biometric recognition technology, such as the face, handwriting, and voice recognition, also

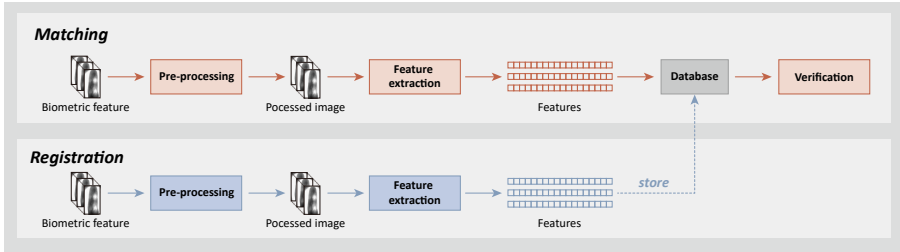


Fig. 1 The general workflow of biometric identity verification system.

plays a vital role in crime investigation [7]. Furthermore, in some financial services, the biometric features, which include finger vein, face, fingerprint, etc., can guarantee that the individual attempting to access financial data is only an authorized user [8]. In addition to above mentioned biometric features, iris [9], retina, and gait [10] are also widely used. The details of these biometric recognition technologies are provided in Tab. 1 to understand their characteristics better. The example image of these biometric features is given in Fig. 2.

Table 1 Several biometric recognition technologies [10–12]. **N** represents non-contact. **C** represents contact. **RF** represents radio frequency. **NIR** represents near-infrared.

Feature	Security	Obstruction	Data	Device	Contact	Cost
Face	Normal	Illumination	Image	Camera	N	Low
Voice	Normal	Noise	Audio	Microphone	N	Low
Fingerprint	Good	Skin surface	Image	Optical sensor Thermal sensor RF sensor	C	Low
Iris	Superior	Glasses	Image	Special camera	N	High
Retina	Good	Glasses	Image	Special camera	N	Middle
Gait	Normal	Personal appearance Filming angle	Video Foot pressure Velocity Frequency	Camera Floor sensor Accelerometer Radar	N	Low
Signature	Normal	Randomness of writing	Image Writing pressure Writing posture	Scanner Electronic tablet & Electronic pen	N	Low
Finger vein	Superior	Few	Image	NIR sensor & NIR sensitive camera	N	Low



Fig. 2 The images of biometric features. These images from left to right are face, voice, fingerprint, iris, retina, gait, signature, and finger vein [13–15].

Although all the biometric features in Tab. 1 have particular applications in people's lives, they have different characteristics. Unfortunately, several representative biometric identification technologies are struck by some bottlenecks. For instance, the fingerprint recognition rate is significantly affected by the finger surface, and the fingerprint faces the risk of being forged. Voice recognition usually requires a relatively quiet environment. The recognition rate of iris systems is outstanding, but it requires expensive sensors and may be affected by device location and the environment around the iris. The face data obtained on the user side usually differs from registration in the face recognition system, which may hinder the recognition process [15].

1.2 Finger vein recognition

Different from the above biometric features, *Finger Vein Recognition* (FVR) has many advantages since it utilizes the feature extracted from the intrinsic physiological structure of organisms. It matches the vascular feature extracted from the human finger with the previously registered prototypes to perform the recognition task. A Japanese scientist introduced the FVR technology in 2002 [16]. FVR was initially used for medical diagnosis rather than biometric recognition, so early FVR researchers excessively focused on improving the quality of finger vein images, neglecting the practicality of FVR devices in daily life [16]. In 2006, Hitachi invented the first FVR device for identification [17]. FVR has been widely used in the biometric recognition field and created a boom in Japan, with financial institutions in hundreds of cities applying FVR to their security systems [18, 19]. Since the finger veins are hidden deep beneath the skin surface, and it is usually observed by the *Near-Infrared* (NIR) light [20] instead of the visible light [18], this characteristic makes FVR-based security systems more private than other biometric-based recognition methods. To better understand the principle of FVR based on NIR, the details are illustrated in Fig. 3.

In addition to privacy, FVR also has the characteristic of uniqueness and stability. Even between identical twins, their finger vein structures are different from each other [22]. Besides, finger veins can maintain their structures with age [22]. Due to the above-maintained advantages, FVR is challenging to be affected by external factors. Compared with other biometric recognition approaches, finger veins are virtually impossible to be stolen, which brings FVR strong security. In addition, The FVR is more hygienic because its acquisition is non-contact, avoiding public health infections [22], and more efficient due to the small size image it processes.

AI technology, especially DL technology, has developed rapidly in recent years. Compared with traditional image processing methods, DL achieves overwhelming performance in many tasks of computer vision, such as biometric recognition [23], biomedical image analysis [24], and autonomous driving [25]. The traditional FVR process usually includes image capture, image data pre-processing, feature extraction, and matching or other analysis tasks. The application of DL-based methods, especially *Convolutional Neural Networks*

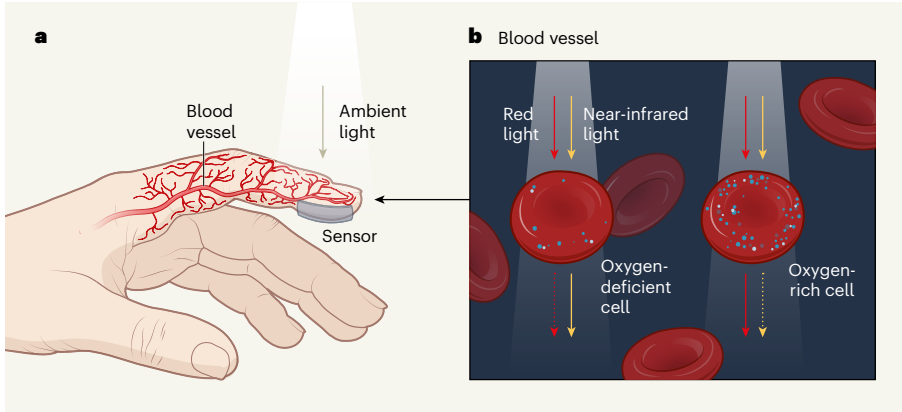


Fig. 3 The specific principle of finger vein image capture [21]: The oxygenated and de-oxygenated Hemoglobin in the venous vessel absorbs the NIR light passing through the finger.

(CNNs), dramatically changes the manual feature extraction process. The performance of conventional *Machine Learning* (ML) approaches is significantly influenced by feature engineering, in which the feature selection is based on human domain knowledge. Nevertheless, CNNs can extract abstract but efficient features by supervised or semi-supervised learning. DL-based methods have highly simplified the recognition process. Due to this significant advantage of DL, DL-based methods are widely used in FVR tasks. For instance, [26] uses GAN to generate data in the FVR task to suppress overfitting. [27] employs a deep *Convolutional Auto-Encoder* (CAE) structure and bihashing algorithm to encrypt the finger vein feature templates. These DL-based approaches have worked innovatively on the FVR tasks and achieved outstanding performance.

1.3 The motivation about this survey

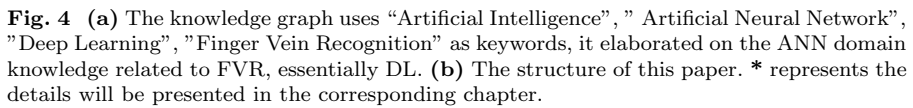
The traditional feature extraction methods depend on prior knowledge, and designing a manual feature extraction algorithm for FVR usually requires the knowledge of finger vein anatomy, information coding, and computer vision [28]. These traditional feature extraction methods are complex and gradually bottlenecked due to the requirement of prior knowledge. Since the widespread use of ANN technology, especially DL, the traditional image feature extraction process has been dramatically changed. ANN-based FVR is attracting attention as a high-performance second-generation biometric technology [29]. To comprehensively describe the application of ANNs on FVR, this paper reviews classical neural networks and deep neural networks used in FVR. Although there are some existing surveys on FVR, none provide a comprehensive view of the application of ANNs. To figure out our contribution and the difference between our paper and other surveys, we discuss recent surveys on FVR [19, 28, 30] in the following parts.

In [19], the technology involved in each step of the traditional FVR workflow, such as pre-processing, feature extraction, and matching, is presented. Some traditional ML methods and DL methods for FVR are also discussed. [28] summarizes the feature extraction methods commonly used in FVR, with vein patterns based-methods, dimensionality reduction-based methods, LBP based-methods, image transformations based-methods, and other feature extraction methods. Besides, this survey compares the characteristics between traditional feature extraction methods and feature learning methods. However, the ANN technique mentioned in these papers is not systematic and comprehensive enough, and the amount of literature presented is insufficient. [30] focuses on the software development design, and the hardware development design of FVR is presented. Besides, this paper summarizes the challenges of FVR in several aspects. The solutions to these challenges are also processed based on existing research results. Nevertheless, this paper lacks a summary of related technical papers. These surveys are comprehensive and novel, providing summaries of FVR from different aspects. However, these surveys present a non-negligible drawback. ANN, a critical technology in FVR, is not comprehensively elaborated in these surveys, and the existing FVR surveys lack an overall summary of ANN's wide range of applications in FVR. Inspired by these papers, we have conducted this comprehensive survey of ANN-based FVR. From classical neural networks to deep neural networks, our survey provides a comprehensive summary of ANN applications in FVR.

To conduct this paper, we summarized 149 papers in the field of ANN-related FVR from 2004 to 2022, covering tasks such as verification, image enhancement, segmentation, *Presentation Attack Detection* (PAD), Template protection of finger vein images. These papers are collected from mainstream academic datasets or search engines, including IEEE Xplore, Springer, Elsevier, ACM, MDPI, World Scientific, and Google Scholar. We use “**finger vein image analysis**” AND (“**deep learning**” OR “**neural network**” OR “**ANN**” OR “**CNN**” OR “**GAN**” OR “**RNN**” OR “**LSTM**”) as the searching keywords. FVR is intimately connected with ANN technology, and the related keyword knowledge graph is shown in Fig. 4a.

The contributions of this paper are as follows:

- To the best of our knowledge, this is the first comprehensive survey summarizing the application of ANNs, including classical neural networks and deep neural networks in FVR. To conduct this survey, we discuss 149 relevant papers. In addition, we briefly describe the history of ANN development and summarize the commonly used public datasets in the FVR domain.
- We divide the involved ANNs into two types, classical neural networks and deep neural networks. In classical neural networks, we follow the traditional image process of the biometric recognition paradigm for analysis, including pre-processing, feature extraction, and matching.
- In deep neural networks, we follow the different image analysis tasks for summary, including verification, image enhancement, segmentation,



- We have compiled some of the challenges encountered in FVR and provided potential directions for the development of FVR.

The structure of this survey is as follows: In Sec. 2, we introduce the development history of ANN and highlight its groundbreaking achievements. We also introduce representative deep neural networks and transfer learning. The public datasets widely used for FVR are illustrated in Sec. 3. In Sec. 4, we summarize the application of classical neural networks on FVR according to the recognition workflow. Sec. 5 presents the summary of tasks of deep neural networks on FVR according to the different tasks. In Sec. 6, we outline the potential directions of FVR. Sec. 7 summarized the entire paper. The specific content of each section in this survey is shown in Fig. 4b.

Since our work focuses on applying ANN to FVR, we briefly describe the development history and breakthrough discoveries of ANN and introduce the representative deep neural network structures widely used in FVR. At the same time, transfer learning, which is widely used in FVR, is also included in our discussion. The specific structure of this section is shown in Fig. 5.

ANN is a computational model designed to mimic biological neural networks. The structure of basic ANN, MLP, is shown in Fig. 6. ANN consists of multiple nodes interconnected with each other, and nodes between two adjacent layers are given different connection weights to extract features from the original information. Each node in the ANN performs forward propagation by

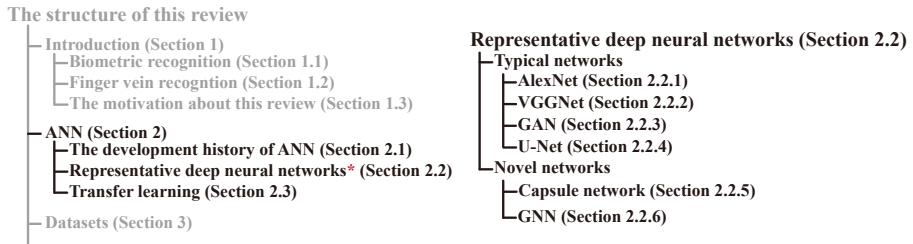


Fig. 5 This section structure.

receiving the output from one or more nodes of the previous layer. ANN uses the interconnection between nodes to perform mathematical modeling to solve complex problems.

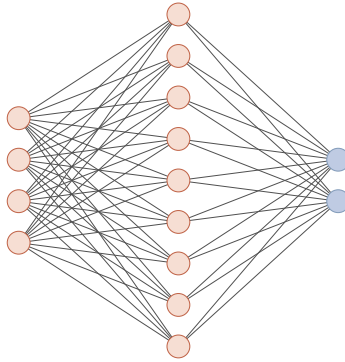


Fig. 6 Typical ANN structure represented by MLP.

Nowadays, ANN has achieved great success on many tasks. However, these achievements have been obtained by many AI scientists through continuous research for up to half a century. To finger out the developing process of ANNs, we provide an overview of it in Fig. 7.

The starting date of ANN research can be traced back to the 1940s. In 1943, McCulloch and Pitts were inspired by biological neurons to propose the first artificial neuron model, the M-P model, [31]. More details are shown in Fig. 8. This model is implemented through a physical network, consisting of resistors and other elements. It can be used to perform simple logical operations. Based on the M-P neuron, Roseblatt proposed the perception in 1958 [32]. This model is capable of determining the weights between neuronal connections after training. Perception led to the first boom in ANN research. However, in 1969, Minsky and Papert pointed out that perception could not solve the linear indivisibility problem, XOR problem [33]. This finding brought ANN's research to a low ebb.

In 1983, Hopfield created a stir in the ANN field by proposing the *Hopfield Neural Network* (HNN) that used associative memory [34]. HNN achieved the

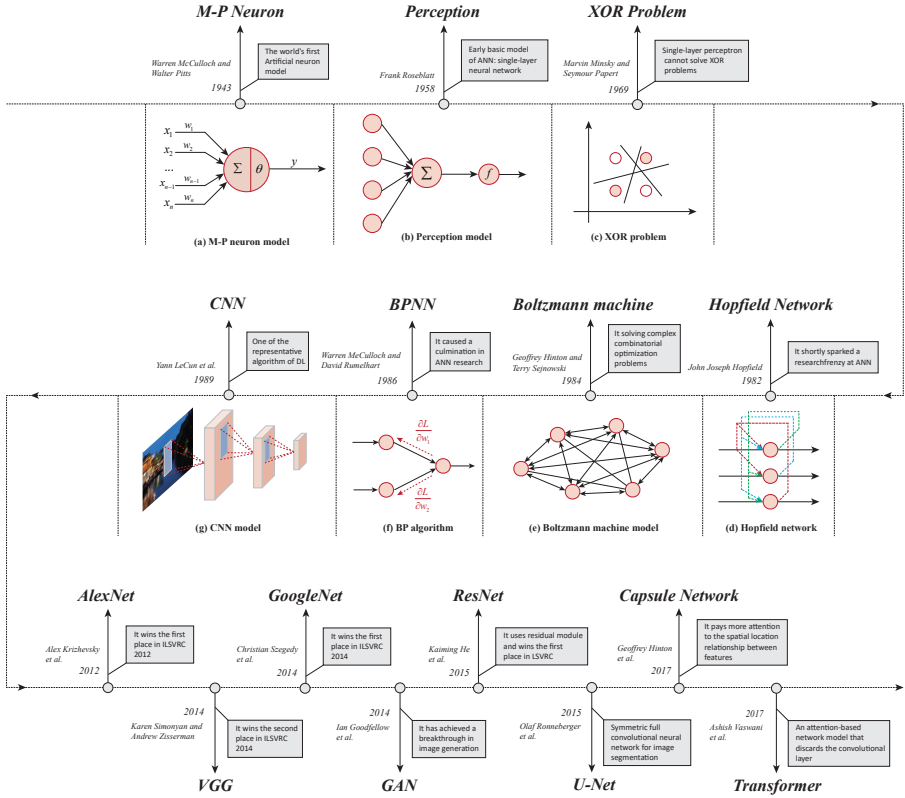


Fig. 7 Representative results of ANN's development history.

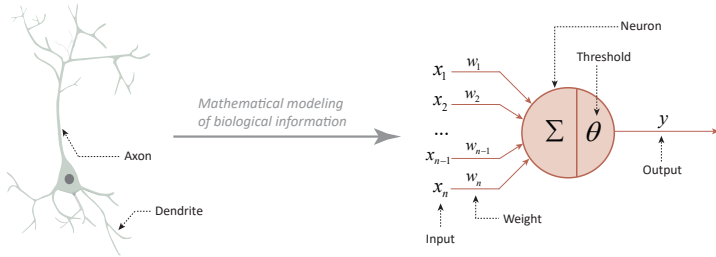


Fig. 8 M-P neurons are modeled according to the physiological structure of the biological neurons, especially the function of dendrites and axons on biological neurons. Dendrites can receive stimuli and transmit excitation to the cell body, and axons can transmit their excitation to other neurons. This constitutes the basic function of artificial neurons.

best result on the traveling salesman problem [35]. HNN has brought the previously lukewarm ANN back into the limelight of scientists. After that, Hinton proposed the famous Boltzmann machine, a randomized HNN [36]. However, the emergence of *Back Propagation* (BP) algorithms led to the second wave of ANN research. In 1986, McCulloch and Rumelhart proposed the BP algorithm [37] and applied it to *Multi-layer Perceptron* (MLP) that can solve the

XOR problem. In 1989, LeCun et al. introduced convolutional layers into ANN by the biological primary visual cortex. They also introduced the BP algorithm into the network and achieved great success in *Handwritten Digit Recognition* (HDR) tasks [38, 39]. The BP algorithm is one of the most successful and fundamental ANN algorithms. Even now, it is still an essential element in the training process of ANN.

Although ANN can model complicated patterns and forecast issues by increasing the number of hidden layers and neurons, the computer performance at that time was so limited that it was almost impossible to train a large-scale ANN. Additionally, the overfitting problem also hindered the development of ANN. With the popularity of ML methods, especially *Support Vector Machine* (SVM) [40], the research related to ANN gradually fell into a depression for the second time.

Despite the dilemma of ANN research, Hinton and Bengio et al. still focused on ANN research [41–51]. In 2006, [43] proposed a *Deep Belief Network* (DBN), which employed the Layer-wise pre-training strategy. After the pre-training, the weights in DBN were fine-tuned with the BP algorithm. This approach made it possible to train a deep ANN at that time. Besides, The deep neural network based on the pre-training strategy achieved significant breakthroughs in speech recognition tasks [50]. Meanwhile, the remarkable AlexNet designed by Hinton and Krizhevsky [51] won first place in ImageNet ILSVRC-2012. Benefiting from these outstanding achievements, the ANN technologies represented by DL attracted researchers' attention again. Recently, with the significant improvement in computer performance, the training cost of deep neural networks has decreased. In addition, the amount of data available for network learning has increased greatly compared to before, which can well avoid the problem of overfitting caused by limited data. With the above progress, ANN has once again reached a climax.

2.2 Representative deep neural networks

As mentioned above, training a complex neural network is no longer difficult due to the development of computer hardware and ANN technology. To figure out the application trend of deep neural networks in FVR, we perform frequency statistics based on the literature we summarized. The details are provided in Fig. 9. To better understand these popular networks, their characteristics are briefly summarized. Besides, since U-Net is widely used in the finger vein image segmentation tasks, we also introduce the structure of U-Net. Some novel networks that have great potential to be applied to FVR tasks are also introduced in this section, including the capsule network and *Graph neural network* (GNN).

2.2.1 AlexNet

AlexNet is a milestone network because it is the first CNN that won first place in the ILSVRC 2012. Before it, the development of neural network technology

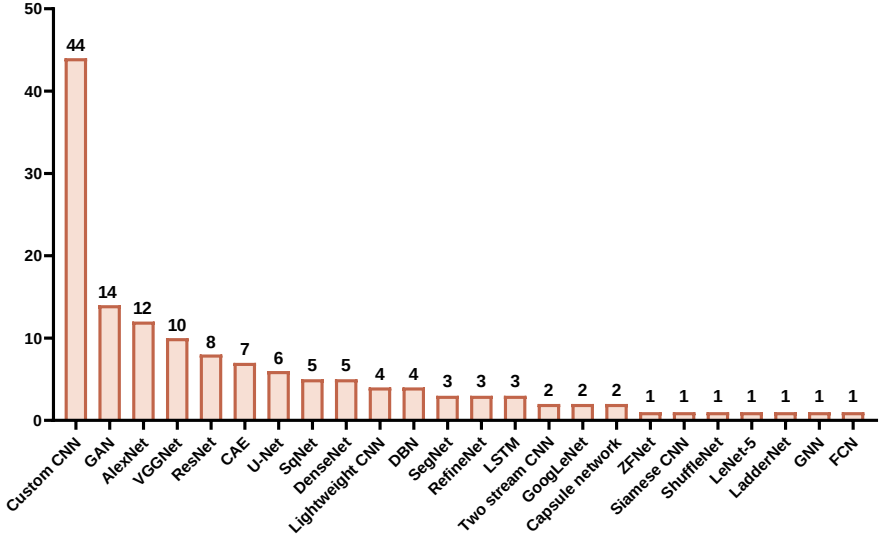


Fig. 9 The frequency of deep neural networks on FVR tasks.

was at a low ebb for many years. Since the success of AlexNet, deep CNNs have become the mainstream technology in many computer vision tasks [52]. AlexNet has 600 million parameters and 650,000 neurons. As the structure of AlexNet shown in Fig. 10, it contains five convolutional layers and three fully connected layers with 4096, 4096, and 1000 neurons, respectively [51]. AlexNet uses overlapping pooling instead of convolutional pooling and uses the *Rectified Linear Unit* (ReLU) as the activation function because it is faster in gradient descent than a saturated nonlinear function. The equation of ReLU is formulated in Eq.(1). Additionally, AlexNet was trained on multi-GPUs, since the performance of GPU was limited at that time [51]. To avoid overfitting, AlexNet utilized data augmentation and dropout operations.

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (1)$$

2.2.2 VGGNet

Compared with AlexNet, the significant contribution of VGGNet is only using 3×3 convolutional kernels instead of large convolutional kernels used in AlexNet to compose the network structure. This innovation reduces the parameters while enhancing the nonlinear fitting ability of the network. There are various structures of VGGNet in the earliest research of [53]. Among them, VGG-16 and VGG-19 are the most widely used. The detailed structures are shown in Fig. 10, VGG-16 consists of 13 convolutional layers, five max-pooling layers, and three fully connected layers. VGG-19 has three more convolutional

In theory, the deeper the network structure is, the more rewarding the performance has. This is because deep CNNs are capable of extracting more efficient features. However, a too deep network structure is prone to lead to the problem of gradient explosion or vanishing [54]. To overcome this problem, [55] proposed

Residual Network (ResNet) by introducing residual learning in 2015, which won first place in ILSVRC 2015. As shown in Fig. 10, ResNet improves the efficiency of information propagation by adding shortcut connections between the convolutional layers. The combination structure is called the residual unit, and ResNet is a deep neural network composed of several residual units.

In the residual unit, the input can be denoted as x , and the feature map learned for x by the residual unit can be represented by $H(x)$. As shown in Eq.(2), $H(x)$ is split into x and $H(x) - x$, which called identity mapping and residual mapping respectively. Since the x is invariant, the learning target of the network changes from $H(x)$ to the residual mapping $F(x) = H(x) - x$. Benefiting identity mapping, ResNet can build the network deeper than other networks without gradient vanishing during gradient backpropagation. This deep network structure allows ResNet to learn more useful features than shallow networks, which usually promote better performance.

$$H(x) = \underbrace{x}_{\text{Identity mapping}} + \underbrace{[H(x) - (x)]}_{\text{Residue mapping}} \quad (2)$$

2.2.4 Generative adversarial network

Generative adversarial network (GAN) proposed by [56] is the hottest DL algorithm in the image generation task. The specific workflow of the algorithm is shown in Fig. 10. Unlike the neural networks used for classification, GAN is composed of a generator and a discriminator. The purpose of the generator is to generate fake samples that can fool the discriminator. The task of the discriminator is to distinguish samples between true and fake.

In the training process of GAN, the generator is first initialized by random noise. Then, the samples obtained by the generator are combined with the real samples to train the discriminator to distinguish samples between true and fake. The predicted probability of the discriminator is then used to promote the updating of the generator by the loss function. After sufficient training rounds, when the generator can fool the discriminator, the network parameters will be fixed and used to generate the image. The generative adversarial process can be modeled in the form of Eq.(3). V is the objective function of the entire model. D is discriminator, G is generator, $E_{x \sim P_{data}}$ represents the true data distribution, $E_{z \sim P_z(z)}$ represents random noise distribution. The entire formula reveals the GAN optimization process.

$$\min_G \max_D V(G, D) = E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z(z)} [\log [1 - D(G(z))]] \quad (3)$$

2.2.5 U-Net

U-Net is a popular fully convolutional network for image segmentation, initially used for biomedical image segmentation. As shown in Fig. 10, the network structure of U-Net is symmetrical and includes a compressed path to extract

features and an extensive path to perform up-sampling. This network structure can achieve precise segmentation with few images since it takes advantage of data augmentation.

2.2.6 Capsule network

CNNs learn global information by deepening the number of layers. However, they cannot capture spatial information. This limitation hinders the network from considering the feature-to-feature interrelationships and positional relationships. Meanwhile, pooling operations widely used in CNNs are susceptible to information loss. Based on these shortcomings of CNNs, [57] proposed the capsule network that is a more effective image processing network. The capsule network uses capsules instead of ordinary artificial neurons to store feature information. A capsule is a carrier that contains multiple neurons and is expressed as a vector. The mode value of the capsule vector indicates the degree to which the capsule recognizes the features, and the direction indicates the posture information of the feature. The parameters of the capsule vector can reflect the variation of feature locations. The capsule network can store spatial information of the features. In addition, the capsule network uses the dynamic routing algorithm to train the weights among the capsules rather than the BP algorithm. The algorithm difference between capsule network and traditional CNN is shown in Tab. 2.

Table 2 Difference between capsule network and CNN [57].

		Capsule network	CNN
Input from low-level		vector $\mathbf{u}_i$	scalar x_i
Operation	Transform	$\hat{\mathbf{u}}_{j i} = \mathbf{W}_{ij}\mathbf{u}_i$	-
	Weighted summation	$\mathbf{s}_j = \sum_i c_{ij}\hat{\mathbf{u}}_{j i}$	$a_j = \sum_i w_i x_i + b$
	Activation	$\mathbf{v}_j = \frac{\ \mathbf{s}_j\ ^2}{1 + \ \mathbf{s}_j\ ^2} \frac{\mathbf{s}_j}{\ \mathbf{s}_j\ }$	$h_j = f(a_j)$
Output		vector $\mathbf{v}_i$	scalar h_i

2.2.7 Graph neural network

As a data structure consisting of vertices and edges, the graph can store rich information and represent relationships between information [58]. To learn the graph structure, GNN is proposed as a variant of CNN [59]. The structure of GNN is defined by the node and connection matrix. GNN represents a new node by aggregating adjacent feature vectors and extends the traditional convolutional layers to the non-Euclidean space to capture the information in a graph structure. Finger vein features can be considered a graph structure because of the interlocking venous vascular lines embedded in images.

2.3 Transfer learning

Since most ML methods require the training and testing data from the same label space [60], it usually needs to reconstruct a proprietary dataset when the learning model needs to perform different tasks. However, re-collecting and labeling the data in many real application scenarios is time-consuming and costly. However, Semi-supervised learning [61] can partly solve this problem by relaxing the annotation requirements of the data. It is sometimes challenging to collect enabled raw samples in large quantities. To reduce the cost of reconstructing datasets from scratch, transfer learning is widely used in various tasks, including FVR.

In transfer learning, the training and test samples are called the source and target domain, and transfer learning resolves target tasks by transferring knowledge from the source domain to the target domain, thereby reducing the cost of reconstructing data for new tasks. Transfer learning in the ML domain is proposed by the knowledge transfer in the human knowledge domain. For instance, someone who has studied piano may have an easier time learning violin, and someone who can ride a bicycle will seem more proficient on a motorcycle. As shown in Fig. 11, the goal of transfer learning is to utilize knowledge from source domains to increase the learning performance of the target task by minimizing the number of labeled examples required for the target domain. Due to the insufficient number of samples on the public datasets in the FVR domain, transfer learning has been used extensively in the process of model training of FVR tasks. Pre-training network module such as AlexNet and VGGNet on ImageNet [62] before performing the FVR significantly improves the training process of the models.

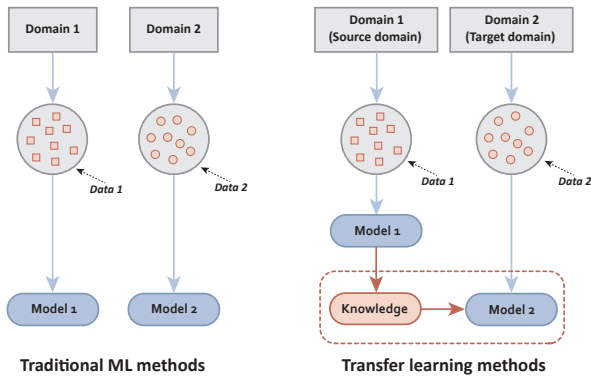


Fig. 11 Transfer learning method.

3 Datasets

As DL occupies an important position in the research field of AI, it is also widely used in FVR task [63]. DL is a data-driven learning paradigm that

aims to learn effective features based on abundant training data to perform the analysis task. Therefore, the dataset plays an essential role in the development of DL. In this section, we introduce the commonly used datasets in FVR. By investigating relevant papers, we conduct a statistic to summarize the usage frequency of various FVR datasets. The details are demonstrated in Fig. 12. It can be found that the popular datasets for FVR mainly include SDUMLA-HMT [64], FV-USM [65], HKPU [66], MMCBNU-6000 [67], UTFVP [68, 69], THU-FVFDT [70, 71], SCUT [72], and IDIAP [73]. The fundamental information of these datasets is provided as follows:

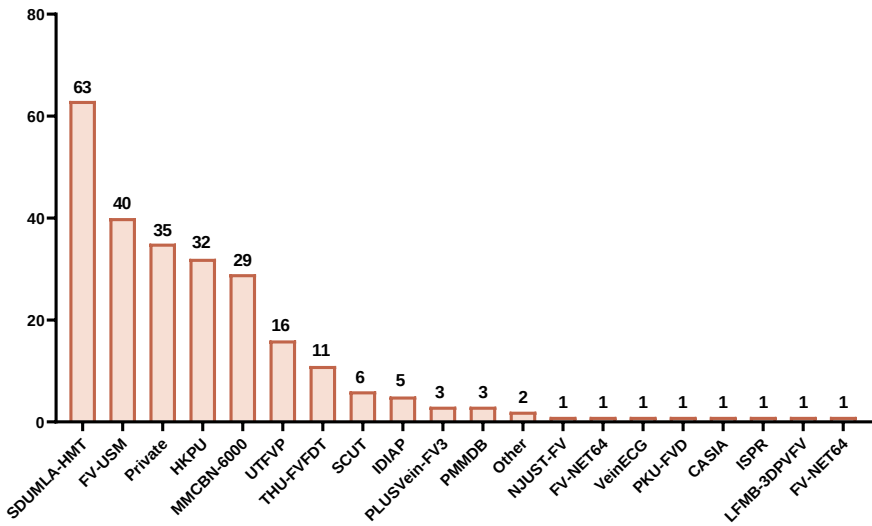


Fig. 12 The frequency of all datasets on FVR tasks.

SDUMLA-HMT is a homologous multi-modal traits database containing multiple biometric features such as face, finger veins, gait, iris, and fingerprints. The finger vein part of SDUMLA-HMT is the first publicly available finger vein dataset, consisting of 3816 images. These images were collected from each of the six fingers of 106 people, and six images were collected from each finger.

FV-USM contains 5904 images obtained from 123 volunteers, including 93 males and 40 females, ranging in age from 20 to 50. The image collection process was divided into two stages. The time gap between these two stages is more than two weeks. Each person provided four fingers for image capture. For each image collection stage, six images were taken for each finger.

HKPU contains 6264 images acquired from 156 subjects. Half of these images are finger vein images, and the rest are finger texture images. 93% of the subjects are younger than 30 years old. Images were acquired in two separate sessions with a minimum interval of one month and a maximum interval of six months. The average interval is 66.8 days. In each session, every subject

provided six samples. Each sample contains one vein image and one finger texture image.

MMCBNU-6000 contains 6000 finger vein images collected from 100 volunteers from 20 different countries. These volunteers have different skin tones. Each subject provided their index finger, middle finger, and ring finger, and each finger was photographed ten times in an office environment (rather than a dark environment).

UTFVP contains 1440 vascular pattern images obtained from 60 volunteers. These images were captured in two sessions. The average time gap between these sessions is 15 days. The vascular pattern of the six fingers from each subject was taken two times.

THU-FVFD1 contains two versions. The first version, THU-FVFD1, contains 440 finger vein images from 220 subjects. The second version, THU-FVFD2, contains 2440 finger vein and finger dorsal texture images from 610 subjects. Both datasets were acquired with only one finger of each subject, and their image acquisition process was finished in two sessions.

SCUT contains 10800 images acquired from 100 subjects. Each subject provided six fingers, and each finger was photographed 18 times. For each finger, the first six images were taken in a normal posture, and the last 12 images were taken at a rotational angle of less than 20°.

IDIAP consists of 880 cropped and full versions of real and faked images from 110 subjects, and these subjects were from different races. Half of these images are real acquisitions, and half are fake. The fake images are created based on some images of the VERA dataset after the simple pre-processing. This dataset is mainly used for PAD.

To summarize the above-mentioned datasets, the crucial information of these datasets is provided in Tab. 3.

Table 3 Detail of public datasets that widely used in FVR. **SN** represents subject number. **IN** represents image number. **M** represents middle finger. **I** represents index finger. **R** represents ring finger. **No.** **FS** represents number of finger for each subject.

Dataset	SN	IN	No. FS	Resolution	URL
SDUMLA-HMT [64]	106	3816	6 (both M, I, R)	320 × 240	http://mla.sdu.edu.cn/info/1006/1195.htm
FV-USM [65]	123	5904	4 (both M, I)	640 × 480	http://drfendi.com/fv.usm.database/
HKPU [66]	156	6264	2 (left M, R)	513 × 256	http://www4.comp.polyu.edu.hk/~csajaykr/fvdatabase.htm
MMCBNU-6000 [67]	100	6000	6 (both M, I, R)	640 × 480	http://multilab.jbnu.ac.kr/MMCBNU_6000
UTFVP [69]	60	1440	6 (both M, I, R)	672 × 380	https://pythonhosted.org/bob.db.utfvp/
THU-FVFD1 [70]	220	440	1 (left I)	200 × 100	https://www.sigs.tsinghua.edu.cn/labs/vipl/thu-fvfdt.html
THU-FVFD2 [71]	610	2440	1 (left I)	200 × 100	https://www.sigs.tsinghua.edu.cn/labs/vipl/thu-fvfdt.html
SCUT [72]	100	10800	6 (both M, I, R)	640 × 480	https://github.com/SCUT-BIP-Lab/SCUT-RIFV
IDIAP [73]	110	440	4	665 × 250	https://www.idiap.ch/dataset/vera-fingervein/index.html

4 Classical neural network for finger vein recognition

In this section, we follow the biometric recognition process to summarize the application of classical neural networks on FVR, including pre-processing, feature extraction, and matching. These ANNs transmit information through weighted connections between artificial neurons instead of convolutional layers. The specific structure of this section is shown in Fig. 13.

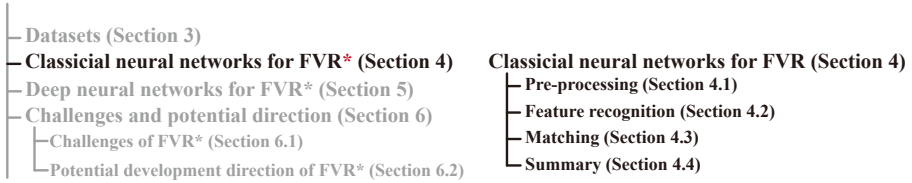


Fig. 13 This section structure.

4.1 Pre-processing

The purpose of image pre-processing is to enhance the effectiveness of the features of the object and suppress the features of other factors. Typical pre-processing methods include image enhancement, filtering, segmentation, etc. In the FVR tasks using classical neural networks, it is vital to pre-process the finger vein images to make the finger vein features more clear. [74] uses multi-scale self-adaptive enhancement transform based on wavelet to denoise and enhance the finger vein feature in the finger vein images. The wavelet transform is good at extracting point features. However, the finger vein images exhibit more significant linear features than point features. To improve the previous work, [75] proposes the multi-scale self-adaptive enhancement transform based on curvelets. The method applies curvelets decomposition to finger vein images making the pre-processing strategy not only have the local time-frequency analysis capability of wavelets but also the differentiation capability of direction selection and the identification capability of linear features. The pre-processing steps in [76] are divided into four parts: vein-region division, Gamma enhancement, Gary-scale, and contrast enhancement. Firstly, the vein region is divided in the original image. After the feature of the vein region is Gamma enhanced to obtain better image tones, followed by the conversation of RGB image to grayscale using the Gary-scale method, and finally, the contract of the image is enhanced. The above pre-processing method is complex, but the pre-processing of some studies is straightforward, and it also helps improve the network's recognition performance. In [77], the range of pixel values is normalized, followed by ROI extraction and image enhancement. [78] uses image cropping to highlight the finger vein region.

4.2 Feature extraction

Extracting venous features from finger vein images is a critical step in finger vein verification. Although pre-processing techniques can improve the quality of images, the processed images still contain some irrelevant information, which is inefficient to be used for model training directly. Feature extraction techniques focus on extracting the required information from a large amount of information and removing the irrelevant information to the maximum extent possible. [74, 75] construct a local interconnection neural network to extract features from pre-processed finger vein images. As shown in Fig. 14, this network has seven columns with seven nodes in each column in the input layer, seven nodes in the hidden layer, and one node in the output layer. Nodes in one column of the input layer form a local interconnection structure with a node in the hidden layer. This network reduces the computational effort of fully connected networks, and both local and global information can be considered.

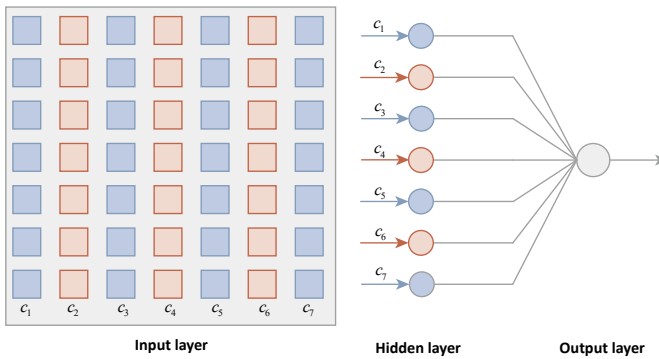


Fig. 14 The structure of the local interconnection neural network [74, 75].

[76] uses Radon transform to extract the feature. Radon transform is a mathematical projection method that condenses the image's information in a few high-value coefficients on the transformed domain. In [77], repeated line tracking, Gabor filter, and image segmentation are used to extract the feature. The repeated line tracking method clears the irrelevant information by removing irregular shadows generated by the thickness of the finger bones. The Gabor filter is mainly used to detect the finger vein images' length and width; subsequently, the finger vein contour is segmented from the entire image.

In addition to these feature extraction methods mentioned above, PCA and *Linear Discriminant Analysis* (LDA) are also applied to FVR [78–82]. PCA is a common data dimensionality reduction method, which finds the most significant features from the high-dimensional features for retention, thus realizing the dimensionality reduction of features and simplifying the computation. [81] introduces LDA to extracting features. Analogous to PCA, LDA is a data dimensionality reduction method that separates two or more classes by finding a linear combination of features.

4.3 Matching

In the matching phase, the feature extracted from the row data is compared with the finger vein information stored in the registration database for identity verification. Since the finger vein structure contains feature points formed by the intersection of vein lines, template matching based on pixel value is suitable for FVR. [74, 75] address the lack of robustness of traditional template matching methods by identifying the blurred areas around the vein vessels and ignoring the slick misalignments between vein patterns. The method is implemented by relabeling the vein track space in the blurred region with pixel values between 46 and 170.

ANNs show a powerful performance in matching step [76–82]. [79, 80] use *Adaptive Neuro-Fuzzy Inference Systems* (ANFIS) to match the finger vein features. The structure of ANFIS is a merger of an adaptive network and a fuzzy inference system, which inherits the interpretability of the fuzzy inference system and the learning ability of the adaptive network. The ANFIS can change the system parameters based on prior knowledge to make the output closer to optimization [83]. These papers also conduct comparative experiments using MLP based on the BP algorithm, and the experimental results show that ANFIS significantly outperforms MLP. [76] uses *Radial Basis Function Neural Network* (RBFNN) and *Probabilistic Neural Network* (PNN) to perform the classification of finger vein images. The structures of RBFNN and PNN are shown in Fig. 15. In RBFNN, the hidden layers provide the clustering ability in the classification process because its nodes consist of RBF that realize the nonlinear transformation from the input space to the hidden space. PNN is a supervised feed-forward neural network whose network structure is built directly through Parzen nonparametric probability density function. The experimental results show that both RBFNN and PNN have rewarding performance, achieving identification rates of 98.3% and 99.2%, respectively. Additionally, the training time of PNN is shorter.

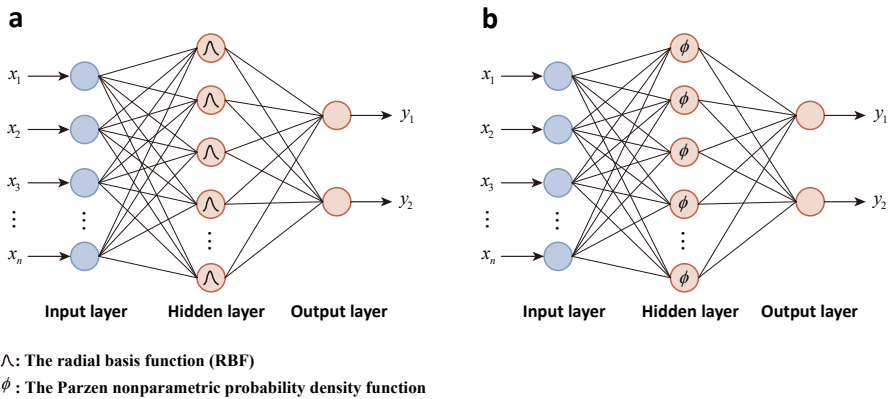


Fig. 15 The structure of RBFNN and PNN [76]. (a) RBFNN. (b) PNN.

4.4 Summary

In the papers we summarized, the classical neural networks for FVR are most used for classification, and only [74, 75] are used for feature extraction. This may be because that early ANNs had limited ability to extract venous features since the networks were too shallow. Although their ANN-based feature extraction methods achieve great experimental results, the samples used in their experiments are deficient. Experiments conducted on the small samples fail to demonstrate the practical application of the model. Unlike classical neural networks, the PCA-based feature extraction method requires few pre-processing steps and eventually achieves excellent performance, indicating that PCA has a solid ability to abstract complex features. In the classification process, the MLP based on the BP algorithm that is widely used in other tasks performs poorly on FVR tasks, with *Accuracy* (ACC) rates lower than 50% in [79, 80]. Conversely, ANFIS achieves satisfactory performance. We have provided a summary table for these papers in Tab. 4.

5 Deep neural network for finger vein recognition

This section introduces the FVR task based on deep neural networks. Unlike classical neural networks, deep neural networks have been used in a wide range of applications on FVR. We will discuss the literature based on the tasks of the paper, which are verification, image enhancement, segmentation, PAD, template protection, and other tasks. In addition, multi-biometric recognition, including finger vein, is also discussed. The specific structure of this section is shown in Fig. 16

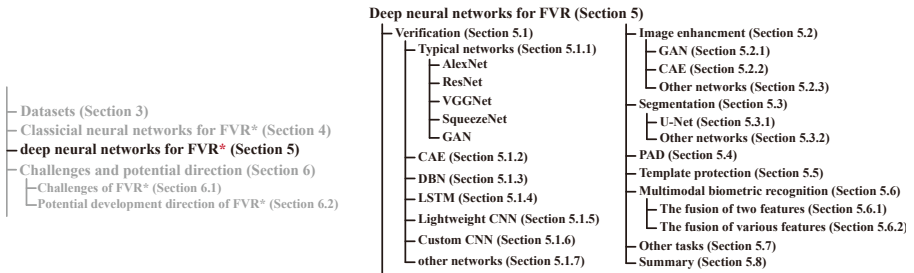


Table 4 Summary of the FVR based on classical neural networks.

Year	Reference	Task	Method		Dataset	Results
			Per-processing	Feature extraction		
2005	[74]	Verification	Multiscale self-adaptive enhancement transform based on wavelet	Local interconnection NN	Template	Private EER = 0.130%
2006	[75]	Verification	Multiscale self-adaptive enhancement transform based on curvelets	Local interconnection NN	Template	Private EER = 0.128%
2009	[76]	Verification	Vein-region Segmentation Image enhancement Gray-scale	Random transform	RBFNN PNN	Private ACC = 98.3% ACC = 99.2%
2010	[79]	Verification	\	PCA	MLP ANFIS	Private ACC = 48% ACC = 99%
2011	[80]	Verification	\	PCA	MLP ANFIS	Private ACC = 48% ACC = 99%
2011	[81]	Verification	\	PCA LDA	SVM ANFIS	Private ACC = 98% ACC = 98%
2015	[77]	Verification	Image normalization ROI extraction Image enhancement	Repeated line tracking Gabor filter Image segmentation	ANN	\
2017	[78]	Verification	Image cropping	PCA	MLP	Private ACC = 81%
2017	[82]	Verification	\	PCA	ANN	Ptivate FAR = 6.53% FRR = 0.71% ACC = 93.27%

5.1 Verification

5.1.1 Verification based on typical CNN

AlexNet

Since the pre-processing process of the traditional ML-based FVR method is too complicated, [84] proposes an AlexNet-based finger vein verification system to solve the problem. The AlexNet is used to extract feature vectors, and Euclidean distance is used to calculate the distance between two feature vectors for verification. The experimental results show that the method achieves an *Equal Error Rate* (EER) of 0.21% on the private dataset, indicating that the network can effectively discriminate the distance between intra-class and inter-class. In [85], the finger vein texture features are first extracted by a local coding method constructed by a set of fixed sparse predefined binary convolution filters. The features extracted by the local encoding method are robust to rotation and illumination changes. These features are fed to pre-trained modified AlexNet for further learning. Finally, the SVM is used for classification. The experimental results show that this method obtains an average ACC of 98.78% on the SDUMLA-HMT dataset.

[86, 87] design the modified AlexNet network for end-to-end learning. In [86], The ROI extracted images are directly fed into AlexNet for classification. The experiment achieves a correct rate of 99.53% on the SDUMLA-HMT dataset. Compared with [86], the network in [87] is more lightweight and can be deployed on Android platforms. Meanwhile, this system incorporates the ResNet module and SENet module to enhance the ability of feature extraction. This system reaches a recognition rate of 94.53%.

[88] proposes a network structure inspired by AlexNet. The research introduces the densely connected layers to the base structure of AlexNet. The study also compares the performance of two loss functions, Softmax and triplet-loss. Based on the network structure designed in this research, Softmax and triplet-loss achieve the ACCs of 97.00% and 97.35% on FV-USM, respectively. As AlexNet was an early successful CNN for image classification, there are related transfer learning methods on FVR using pre-trained AlexNet networks. [89, 90] use a pre-trained and fine-tuned AlexNet network for classification. Using transfer learning instead of initial tuning of network weights makes network training faster.

ResNet

[91] proposes a real-time FVR system that employs a fusion loss to learn more robust features by combining classification loss and metric learning loss, and an inter-class data augmentation technique is used to solve the lack problem of training data. In this system, ROI extraction and alignment are performed by the method of [92], and the ResNet-18 based on the transfer learning technique is used for feature extraction. This network employs cosine similarity as the metric for matching. The specific flow of the whole system is shown in Fig. 17. The experimental results show that the method achieves an EER of 0.48% on

FV-USM, which is significantly lower than other DL methods. [93] proposes a FVR method based on ResNet and U-Net. Both network models are trained by the end-to-end approach. The proposed method introduces bias field correction and spatial attention mechanisms. In this system, the contrast of the original images is first adjusted by the bias field correction model, followed by the inversion of the pixel values. The processed images are fed into the U-Net-based spatial attention model for enhancing the selective region information. Finally, ResNet-50 classifies the enhanced images. The method reduces the impact of low-quality images and classifies finger veins using global features, and makes the network can extract more significant features. The experimental results show that the method has a rank-one verification rate of 99.53% on SDUMLA-HMT and 98.20% on THU-FVFD2. [94–96] propose improving CNN structures with the help of the residual idea of ResNet rather than just using the original ResNet for image analysis tasks. [94] proposes *Efficient Channel Attention Residual Network* (ECA-ResNet) to enhance the practical application ability on FVR tasks. The ECA [97] can bring significant benefits to the model with a small number of parameters, breaking the paradox between performance and complexity. [95] proposes a new ResNet-based architecture, called *FV2021*. The network is compact and suitable for installation on mobile devices. FV2021 uses separable convolutional layers instead of normal convolution layers to reduce the model complexity. In [96], a novel ResNet-based network architecture, ResNext, is proposed. This network is employed for classification, it has a homogeneous multi-branch structure with only a few hyperparameters to tune, and this structure uses the split-transform-merge strategy for scaling any large number of transformations without the need for specialized design. Meanwhile, this model uses *cutout* [98] as the data augmentation strategy. The specific structure of ResNext is shown in Fig. 18. The model still outperforms the original ResNet even with the same model complexity. As the extension of this method, [99] uses a neural architecture search network for FVR. This network uses a controller neural network to sample subnetworks with different structures. This approach is used to update the parameters of the controller network to generate a better architecture. The performance of this network is better than the previous ResNext.

VGGNet

[100, 101] use VGG-16-based CNN for finger vein verification. [100] resizes the detected finger vein ROI to a 224×224 pixel image and then obtains the difference image between the input finger vein ROI image and the registered finger vein ROI image. The difference image is fed to the VGG-16-based CNN model to directly obtain the matching results without the redundant image pre-processing step. The experimental results show that the adjusted VGG-16 network achieves EERs of 0.396%, 1.275%, and 3.906% on three datasets with different image quality (from high to low), respectively. VGG-16 also inspires the network of [101], and this study uses a wide line detector [92] to extract

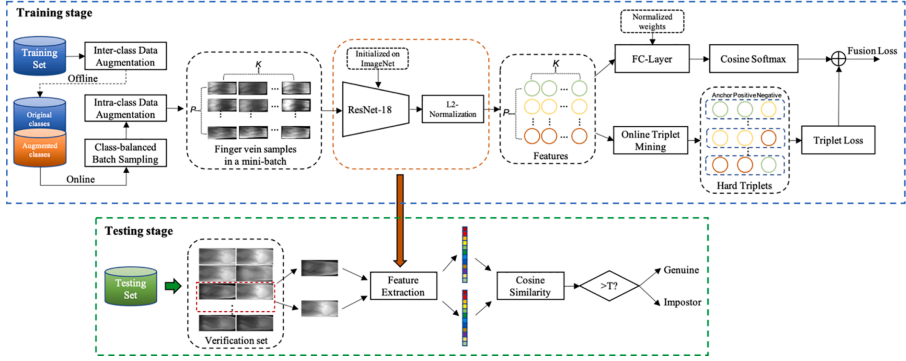


Fig. 17 The workflow of the real-time FVR system [91].

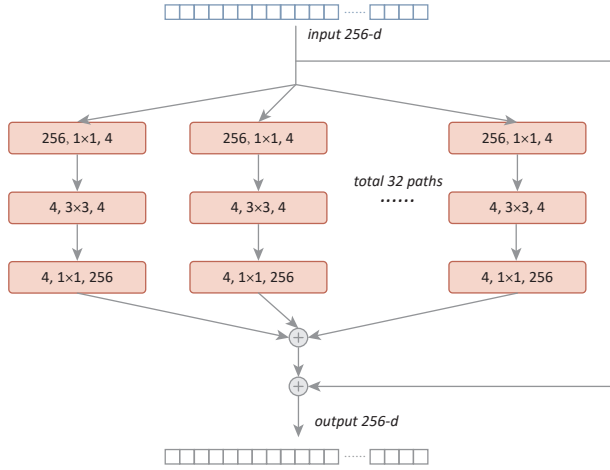


Fig. 18 The split-transform-merge strategy of ResNext [98].

finger vein features from the normalized images in addition to simple ROI extraction. The method achieves an EER of 0.42% on the private dataset.

SqNet

SqNet has fewer parameters without losing accuracy than other deep neural networks and is also used in some FVR tasks. [102] proposes a lightweight SqNet that can be deployed on hardware platforms with limited computational power and memory. The study uses 3 *Channel* (3C) images as the input of the SqNet. The 3C images are obtained by the different operation between the input images and registered images. The SqNet can be fine-tuned to achieve the most accurate performance by feeding 3C images into a network that has been pre-trained on ImageNet. The experimental results fully demonstrate that the method achieves high recognition rates while simplifying the network structure, whose EERs of 1.889% and 4.906% on MMCBNU-6000

and SDUMLA-HMT, respectively. [103] uses pre-trained SqNet to extract the features of the left finger vein and right finger vein and then uses SVM for classification. This method significantly decreases the feature dimension. The experimental results show that this feature fusion method achieves an ACC of 99.81% on SDUMLA-HMT and an ACC of 99.36% in one section of FV-USM. The longitudinal rotation of fingers in the acquisition process can affect recognition performance. To solve this problem, [104] proposes a neural network with robustness to longitudinal finger rotation by training the neural network using finger vein images with different angles. These images from different angles are from two places. One is captured from different angles, and another uses a data augmentation strategy to simulate the longitudinal rotation of the fingers. The study uses two networks, Triplet-SqNet and DenseNet-161, to train these images separately. Experimental results on the PLUSVein dataset show that training with images acquired from different angles leads to a more robust model, especially Triplet-SqNet.

GAN

[105] proposes a novel structure termed FV-GAN based on *Cycle-consistent Adversarial Network* (CycleGAN) to extract features and perform the verification of finger veins. The generator in the FV-GAN consists of an image generator based on U-Net and a pattern generator based on an encoder-decoder network. The pattern generator extracts the finger vein patterns from the image and outputs the probability of each pixel belonging to the vein pattern, and a binary discriminator is used for verification. In addition, FV-GAN uses a fully convolutional structure to reduce the cost of computation. The experimental results show that the EER of FV-GAN is 0.94% on SDUMLA-HMT and 1.12% on THU-FVFD2. [26] proposes a GAN-based structure termed triplet-classifier GAN, which combines a conditional generator and an angular triple loss-based classifier. As shown in Fig. 19, the triplet-classifier GAN is used for data augmentation and classification, and the data augmentation strategy can enhance the training effect of the network. In addition, the cosine similarity is used to replace the Euclidean distance in the designed angular triple loss to improve the feature extraction ability. The experimental results show that this model achieves the EERs of 0.05%, 0.14%, and 0.15% on SDUMLA-HMT, FV-USM, and HKPU, respectively. [106] employs a *Deep Convolutional Generative Adversarial Network* (DCGAN) to identify genuine and fake finger vein images. The generator generates fake images from genuine images, and the discriminator is used to distinguish between genuine and fake images. The experimental results show that DCGAN significantly improves the performance of finger vein verification. The method can be used for forensic identification and biometric verification of the finger vein. Traditional FVR systems are trained using one type of data and have serious performance degradation problems when the trained model is applied to different types of data. To improve the recognition performance of the network on heterogeneous datasets, [107] proposes a FVR system incorporating the domain adaptation technique based on CycleGAN.

All input samples are pre-processed and fed into CycleGAN to generate composite images so that data from different domains have some similarity. Due to the advantage of this method, the model generality is enhanced. The composite image is fed to DenseNet-161 for classification. This experiment is trained on SDUMLA-HMT and tested on HKPU, and an EER of 0.85% is obtained.

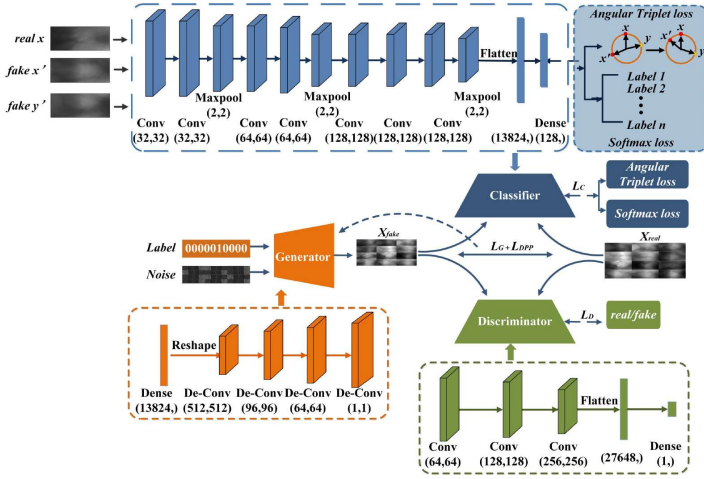


Fig. 19 Architecture of the triplet-classifier GAN [26].

5.1.2 Verification based on CAE

[108] proposes a finger vein verification system that integrates CAE-based feature learning methods with CNN. This system first extracts features from the finger vein images using CAE. Then, these features are fed into the CNN to further depth feature extraction and classification. Using this method, the recognition rate for the FV-USM is 99.16%, and the EER is 0.16%. The workflow of [109] is similar to [108]. CAE is used to learn features of a particular distribution from original finger vein images, and SVM is used for classification. The experimental results show that the method achieves an EER of 0.12% on FV-USM and 0.21% on SDUMLA-HMT. [110] uses CAE to extract preliminary features from the image, and then the parameters of the CAE are used to initialize the parameters of the convolutional layers of a deep CNN. This deep CNN is used to classify the finger vein images. Meanwhile, the *Extreme Learning Machine* (ELM) layers are used to replace the fully connected layers after the training is completed. Because the ELM has faster convergence and better generalization ability than the traditional BP algorithm. The experimental results show that this method achieves EERs of 98.88% and 98.58% on FV-USM and SDUMLA-HMT.

5.1.3 Verification based on DBN

[111, 112] proposes a feature fusion FVR algorithm based on DBN and CNN. It uses the features extracted from DBN and CNN and then uses the similarity measure for matching. To reduce the learning and detection time of the model, feature points based on endpoint and intersection are extracted using the feature extraction method based on eight neighborhoods. These feature point sets are then used as inputs to the network, simplifying the computation of the network and making it adaptable to mobile terminals. The ACC of the method on the private dataset reaches 99.6%. [113] proposes a FVR algorithm based on DBN and uniform LBP operator. The texture features are extracted from the sub-blocks of finger vein curvature gray images using the uniform LBP operator. This makes the learned feature contain more vein information. Then, the histogram of the sub-block features is computed and integrated into an overall histogram for training the DBN. The experimental results show that the recognition rate of this method is 97.4% on FV-USM.

5.1.4 Verification based on LSTM

[114] proposes a network model for feature extraction by combining CNN and LSTM. CNN represents the vein texture features in the local region, and LSTM is used to capture the spatial location relationship within the region. This way of extracting features considers the spatial location relationship between features and makes the model more robust. Finally, the Hamming distance is used for matching. The proposed LSTM-based network achieves an EER of 0.95% on HKPU. [115] designs a bidirectional LSTM-based verification system. The system uses ROI extraction and Gaussian filtering for pre-processing, followed by feature extraction using bidirectional LSTM and shark smell optimization algorithm to optimize the hyperparameters. Finally, Euclidean distance is used for matching. The method surpassed the earlier methods, and the maximum ACC is 99.93%. The traditional FVR system requires the users to hold their finger for a few seconds to complete the verification process. [116] conducts a real-time verification system of finger veins. The system can obtain the user's finger vein feature dynamically. The system extracts finger vein image sequences from recorded videos and uses CNN to extract features from the pre-processed sequences. Then, LSTM is used to find and track the temporal dependencies within the input feature sequences for verification. The system achieves an ACC of 99.13% on the collected dataset. The images in this dataset were taken from different exposure times.

5.1.5 Verification based on Lightweight CNN

Some deep neural networks achieve excellent performance by stacking layers but lose the ability to be applied well because the models are too complex. To let the networks perform verification tasks on mobile terminals, it is necessary to explore the development of lightweight CNN on FVR tasks. [117, 118]

use lightweight CNN for FVR, and these network structures have a few convolutional and fully connected layers. The network of [117] is end-to-end, and it is trained by the joint supervise based on the center loss function and the Softmax loss function, which can obtain highly discriminative features for FVR. [118] proposes a lightweight CNN to perform the classification task along with feature extraction and optimization method of maximum curvature finger vein features based on Gaussian filter, which reduces the influence of image noise on recognition. The lightweight CNN proposed in [119] has two structures by using different loss functions, and these two architectures are *Closed-set* (CS) architecture and *Open-set* (OS) architecture. The specific process is shown in Fig. 20. The CS architecture uses Softmax to predict the class of the input samples. The OS architecture in this study outputs the feature vectors of the input samples and registered samples. The experimental results show that the network achieves EERs of 2.29% and 0.47% on SDUMLA-HMT and MMCBNU-6000. [120] designs a lightweight CNN model using a partially pre-trained MobileNetV2 [121] as a backbone. The model achieves high performance on FVR tasks by using pre-trained auxiliary blocks and customized auxiliary blocks while simplifying the training process. This network achieves the CIR of 96.98% on public datasets.

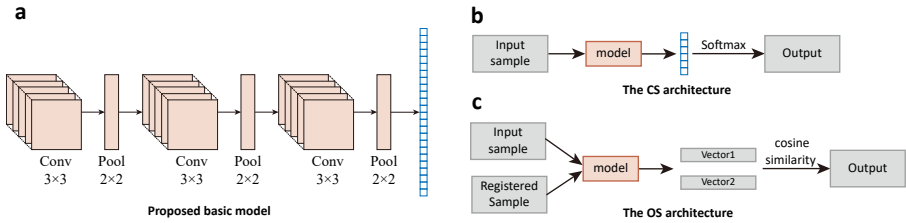


Fig. 20 The details of the framework [119]. (a) The Proposed lightweight CNN. (b) CS architecture. (c) OS architecture.

5.1.6 Verification based on custom CNN

To perform FVR tasks in a targeted manner, many studies use customized deep network structures to analyze finger vein images. The researchers design the number of layers and parameters of these networks according to the FVR task without the help of representative deep neural networks such as AlexNet, VGGNet, and ResNet.

[122, 123] use *Curvature Gray Images* (CGIs) instead of original finger vein images as the input of the network. Their CGIs are obtained from 2D Gaussian templates, and CGIs are directly fed into customized CNN for classification. These studies employ the improved activation function instead of ReLU. The improved activation function has both the ReLU function's sparsity and the Softplus function's smoothness. The ReLU and Softplus functions are shown in Eq.(1) and Eq.(4). Eq.(5) illustrate this improved activation function. This

method is effective to resist noise interference and improve the recognition rate. [124] uses two relatively independent sub-convolutional networks with different granularity for FVR and LeakyReLU as the activation function. The experimental results show that the two sub-convolutional networks can extract feature more effectively and achieves an ACC of 95.1% on the small dataset.

$$S(t) = \ln(1 + e^t) \quad (4)$$

$$S(t) = \begin{cases} 0 & t < 0 \\ \ln(1 + e^t) - \ln 2 & t \geq 0 \end{cases} \quad (5)$$

Researchers have designed some neural networks to perform only one part of feature extraction or classification. [125] removes the bottom structure from a pre-trained CNN model and retains the front structure to extract features from finger vein images. These images are matched using a template matching strategy. [126] uses PACNet [127] instead of CNN for feature extraction and ridge regression classifier [128] for classification. The PCANet filter is generated based on the correlation between the original image and the grayscale image. The best performance of these two methods on the public dataset reaches an EER of 0.04% and an ACC of 100%, respectively. [129] uses PCA for feature extraction and then uses a custom CNN model for classification, achieving a recognition rate of 98.53% on FV-USM.

To solve the problem that traditional 2D finger vein images are easily affected by finger position and posture changes during acquisition, [130] constructs 3D finger vein images and uses these images for detection. This study uses three cameras to perform the imaging process. The 3D images usually include more sufficient vein information than 2D. This research uses a 3D reconstruction algorithm and a corresponding texture mapping algorithm to create a 3D image based on three 2D finger vein images, and 3D finger vein images are shown in Fig. 21. Finally, the lightweight CNN with depth-wise separable convolution is used for feature extraction and matching. The experimental results are shown that this method achieves EERs of 0.94%, 1.69%, and 2.40% on FV-USM, SDUMLA-HMT, and HKPU. As an extension of previous work, [131] uses a contour-based optimization model for 3D FVR and a corresponding acceleration strategy to obtain 3D point clouds of finger vein structures. A custom CNN structure, 3DFVSNet, is used to extract rotationally invariant features, and the specific network structure is shown in Fig. 22. Cosine similarity distance is used for verification. The Experimental results show that 3DFVSNet has powerful robustness to axial rotation.

Some structures use Gabor filters [132] instead of convolutional layers. Gabor filter is a wavelet with good transform properties of time and frequency domains. [133, 134] use Gabor convolutional layers instead of the normal convolutional layers in the CNN structure, and the whole model is end-to-end. The method maintains the advantages of the Gabor filter in FVR and has the excellent feature extraction ability of CNNs. They all achieve favorable recognition results on public datasets. [135] proposes a novel *Trilateral Filter-ative Hermitian Feature Transformation based Deep Perceptive Fuzzy Neural*

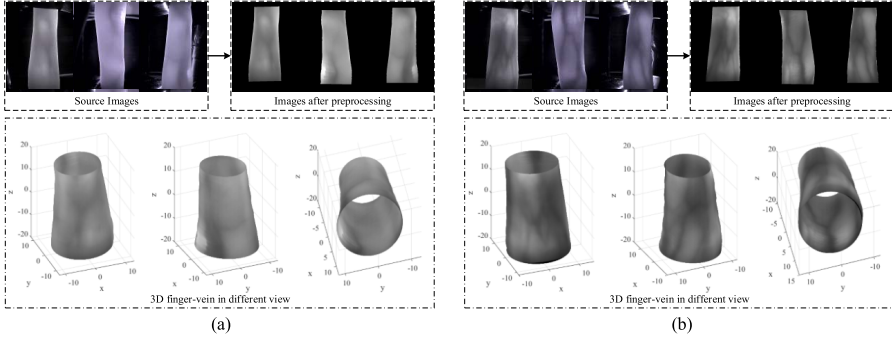


Fig. 21 The 3D finger vein image [130].

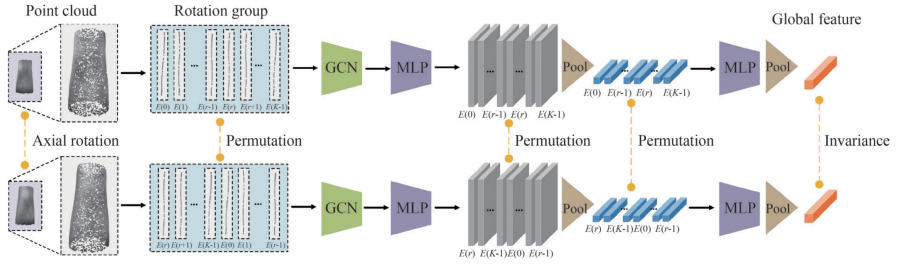


Fig. 22 The structure of 3DFVSNet [131]. **GCN** represent graph convolutional neural network.

Network (TFHFT-DPFNN) model for improving the verification accuracy of FVR. In this system, the Hermitian Hat wavelet is first used to decompose and extract features from the noise-reduced finger image, and then the features are transferred to the next hidden layer. These features are matched by the Jaccard similarity index. Finally, the fuzzy membership function of the output layer is applied to output the validation results. The TFHFT-DPFNN achieves an ACC of 98% on SDUMLA-HMT.

The attention mechanism has been introduced into some custom CNN structures. [136] designs a network structure consisting of a convolutional layer and three *Joint Attention* (JA) models. The JA model can improve the model's discrimination of low-contrast finger vein images by exploiting the interdependence between the spatial locations of channels and feature maps. In addition, this network introduces GeM pooling layers to enhance the feature. Numerous experiments demonstrate the validity of each network component, and the network has an EER of only 0.34% on FV-USM. [137] proposes an improved FVR model based on the residual attention mechanism. The main part of this network is the residual attention blocks, and this network is divided into the trunk branch and the soft mask branch. The trunk branch extracts finger vein features from feature maps that are generated by the previous layers, and the soft mask branch learns global vein information using the hourglass network [138]. Benefiting the advantages of this structure, this network can learn

more abundant finger vein information than other simple networks without branching structures. This method achieves the ACCs of 98.58% and 97.54% on FV-USM and MMCBNU-6000. [139] designs a network based on bilinear pooling to extract the second-order features of finger veins, and the complexity of this network is reduced by replacing conventional convolution with DSC. Meanwhile, this method designs a dimensional interactive attention mechanism to enhance the correlation between channels and space, further increasing the model's recognition accuracy. This method reaches an ACC of 100% on FV-USM.

In [140], The vein areas are extracted from original finger vein images using Sobel operator and polygon ROI, and then the dark vein lines are generated using double *Contrast Limited Adaptive Histogram Equalization* (CLAHE). Finally, these processed images are fed into a 20-layer CNN for feature extraction and verification. The application of CLAHE enhances the finger vein image quality. The experimental results show that this method achieves an ACC of 94.88% on SDUMLA-HMT.

[141] proposes a new FVR method that uses two different fingers instead of a single finger to perform verification, and the method is called *Re-enforced DL* (RDL). The CNN model of RDL stores the weights of both the index and middle finger veins. In the RDL, the index finger is used for the first verification, and the middle finger for enhanced verification. The RDL method effectively utilizes the venous features of both fingers and achieves an ACC of 91.19% on HKPU. [142] performs finger vein verification with the ensemble DL system. In this system, several CNNs are first trained as weak classifiers, and then these classifiers are ensembled to obtain an integrated classifier for recognition. The input of the ensemble DL system is the feature maps from other CNNs, Gabor filters, and LBP. The ensemble DL system obtains ACCs of 92.11% and 94.17% on HKPU and FV-USM.

5.1.7 Verification based on other networks

LeNet-5 is one of the most representative structures of early CNNs. It defined the main structure of the CNN, convolutional layer, pooling layer, and fully connected layer. [143] designs a MATLAB-based FVR system. This system uses a modified LeNet-5 as the backbone. In this system, the image pre-processing and other steps are performed on different platforms, and these two parts are connected using MATLAB's MEX files. The experimental results show that this FVR system recognizes ten subjects within less than ten seconds and achieves an ACC of 96%. [144, 145] use the network proposed by [146] for FVR. This network is more lightweight than LeNet-5, with only two convolutional layers and two fully connected layers. To reduce the number of parameters, [144] cuts off the connections between some neurons in the fully connected layers. In addition, this network uses the stochastic diagonal Levenburg-Marquardt algorithm to ensure fast convergence. The network achieves recognition rates of 100% and 99.38% on 50 and 81 subjects, respectively. ZFNet is also one of the classical CNNs. In [147], the pre-processed

images were fed to ZFNet for classification. The first convolutional layer of ZFNet uses small convolutional kernels and strides, which can retain more features of the original information. This network achieves an ACC of 86% on one section of SDUMLA-HMT.

[148] compares the performance of three well-known deep CNN structures, AlexNet, SqNet, and GoogLeNet, on FVR tasks. These networks have been pre-trained on ImageNet. The image noises are removed using a wiener filter, and these networks are trained with these enhanced images. The experiment results show that GoogLeNet outperforms other networks in terms of classification. [149] introduces the DSC into a pre-trained Xception model, replaying the traditional convolutional layers with DSC. The DSC enables Xception to learn more robust features from images and achieves superior classification performance than normal convolutional layers. This method achieves an ACC of 99% and an F1-score of 98% on the SDUMLA-HMT dataset, and an ACC of 90% and an F1-score of 88% on the THU-FVFD2. [150] uses discrete supervised hashing sequences and triplet loss function to train the network. Discrete supervised hashing can reduce the size of stored feature templates, and this method can improve the matching speed. This paper compares a lightweight CNN with the improved VGG-16, and the experimental results show that lightweight CNN has a superior performance.

DenseNet [151] has a more dense connection mechanism than ResNet. Each layer of DenseNet accepts the output of all its preceding layers as its additional input. This design greatly suppresses the gradient vanishing. [152, 153] use DenseNet-based network architecture to perform FVR tasks. In [152], three images are simultaneously fed into a pre-trained DenseNet-161. The input image, the registered image, and their composite image. This method solves the problem that the normal different image is susceptible to noise, and the composite input of three images makes the recognition performance more accurate. The method achieves EERs of 0.33% and 2.35% on HKPU and SDUMLA-HMT, respectively. [153] is the first study to use both venous texture and finger shape features for FVR. The texture image and the segmented shape image are fed into the DenseNet-161 and then output their respective matching scores. Finally, the verification result is presented by score-level fusion. The method effectively utilizes multiple features of finger vein images and enhances the noise resistance of the model.

ShuffleNet [154, 155] maintains a balance between speed and accuracy by introducing pointwise group convolution and channel shuffle. [156] uses ShuffleNet V2 as the backbone and removes the first pooling layer to generate larger feature maps, which helps to retain fine-grained features. In addition, this model uses Triplet and Softmax-based fusion loss function instead of the original Softmax loss function. The network achieves an EER of 0.05% on the public dataset.

[72] transfers the knowledge of a pre-trained CNN model to a more lightweight Siamese CNN by knowledge distillation. This Siamese CNN uses a new modified contrastive loss function to improve the discriminative ability for

features. The experimental results show that the EERs of this Siamese CNN on MMCBNU-6000, FV-USM, and SDUMLA-HMT are 0.08%, 0.11%, and 0.75%. [157] proposes a FVR system based on a two-branch network incorporating joint Bayesian loss. In this system, finger shape images and ROI vein images are fed into the two-branch network to extract more effective features from finger vein images. Joint Bayesian loss is used to train the network. The experimental results show that the system obtains the EERs of 0.17% and 0.94% on MMCBNU-6000 and SDUMLA-HMT. In [158], two finger vein images from the same finger are fed into two CNNs that have parallel structures, and the two outputs are combined using the CLAHE method and the Gabor filtering method. This method utilizes the vein information from both images instead of a single image, and achieves an ACC of 99.56% on THU-FVFD2T.

Different from the traditional CNNs, some novel network structures are used in FVR. [159] uses a capsule network for FVR. Capsule network extract features of finger vein in a more reasonable way than CNNs by their translational and rotational invariance. In addition, the special structure of GNN determines that it can effectively learn the graph structure features of finger veins. [160] applies GNN to perform FVR task for the first time. The method without pre-processing steps and data augmentation strategy. In this study, the images are first fed into a small CNN to extract feature vectors, and then an edge features learning network is used to model the relationships between all node pairs. Finally, the results are fed into a GNN to perform the classification. The experimental results show that GNN achieves high performance on FVR tasks with fewer parameters and faster model convergence.

5.2 Image enhancement

5.2.1 Image enhancement based on GAN

Finger vein images may be blurred due to low temperature, vein stretching, and illumination change. These factors seriously affect the quality of the images. Some GAN-based methods can repair these defects of impaired images to make these images greatly recognized by ANN. [161] proposes a GAN based on *Neighbors-based Binary Patterns* (NBP) to recover the finger vein images. This model uses NBP texture loss between the input image and the generated image to train the generator network, and the loss function is able to enhance the deblurring ability of the network. Meanwhile, residual connections are added to the generator network to prevent overfitting. The *Peak Signal-to-Noise Ratio* (PSNR) of this GAN based on NBP reaches 30.42. dB. [162] proposes a novel finger vein image restoration method, and this method also uses the NBP texture loss function. Meanwhile, the method uses Possion fusion in the input process to reconstruct the finger vein images to make the image boundary connection of the image more natural. The discriminator network consists of two Wasserstein GAN with Gradient Penalty modules to ensure consistency between the global and local information of the restored image. The experimental results show that adding texture loss can better recover

the veins. [163] and [164] are focus on the optical blurring and the motion blurring of finger vein images, respectively. [163] uses a modified conditional GAN to restore optically blur red finger vein images. Previous conditional GANs use random noise in the form of dropout to ensure the randomness of noise. However, this method may change the information of the restored finger vein image. This study proposes a modified conditional GAN without dropout because of the need for deterministic output. A comparison with the original conditional GAN shows that the approach performs better image restoration. As an extension of the previous work, [164] reduces the number of residual modules in the generator network to speed up the model inference and reduce the network parameters.

To overcome the obstacle of illumination, [165] designs a *GAN for The Illumination Normalization of Finger Vein Images* (INF-GAN). The structure of INF-GAN is designed based on Pix2Pix-HD [166]. The residual model is used as the generator and PatchGAN [167] as the discriminator. The residual image generation block can highlight the vein textures distorted by serving uneven illumination. [168] proposes a finger vein images denoise method based on GAN, which called *Custom Sample Texture Conditional GAN* (CS-TCGAN). This approach designs a joint loss function that combines adversarial loss, content loss, and texture loss to obtain more abundant vein information from the image than Softmax loss. Meanwhile, the CS-TCGAN describes the rough finger vein structures by the de-convolutional layers operation and then fills the details to generate the image. To make the training samples simulate the real noise distribution, this study also designs a dataset with a mixture of Gaussian, Poisson, and speckle noise. The CS-TCGAN shows a great denoising performance on a private dataset.

5.2.2 Image enhancement based on CAE

[169, 170] use the CAE network for finger vein image enhancement. [169] proposes a finger vein image with spots and stains repair solution. The scheme first removes the effect of illumination changes on the image by the Gabor filter and Weber's low descriptor. Then, the images are fed into an encoder network structured as AlexNet to mark the smeared pixels and learn shallow features. Finally, the marked images are fed into the decoder network to recover the image. The workflow of [170] is similar like [169], which uses the adaptive thresholding approach to detect contaminated areas in finger vein images before feeding them into CAE for image recovery. These two CAE-based image enhancement methods have great performance on private datasets.

5.2.3 Image enhancement based on other networks

[171] pre-trains a lightweight VGG-16, and the pre-processed images are fed into this network for image enhancement. This modified VGG-16 structure includes 13 convolutional layers and five max-pooling layers. Meanwhile, The VGG-16 removes the fully connected layer. The ACC of more than 99% is achieved using enhanced images by this network for recognition. [172, 173]

uses custom CNN structures for finger vein image enhancement. In [172], the captured images are directly fed into the CNN for image enhancement. This network contains three convolutional layers and two deconvolutional layers. The convolutional layers are used to learn the distribution of noise features and generate feature maps, and the deconvolutional layers use the feature maps to reconstruct noise-free finger vein images. The model achieves a PSNR of 29.638 dB on HKPU. [173] builds an end-to-end *Finger Vein Image Scattering Removal Network* (FVSR-Net) by combining an optical scattering model and a multi-scale CNN named E-Net. The specific workflow of this method is shown in Fig. 23, and the theory of the optical scattering model is shown in Eq.(6). I and I_0 represent the original and restored images. The $E(x)$ and a represent the output of E-Net and bias. The FVSR-Net achieves the PSNR of 13.5929 dB on SDUMLA-HMT.

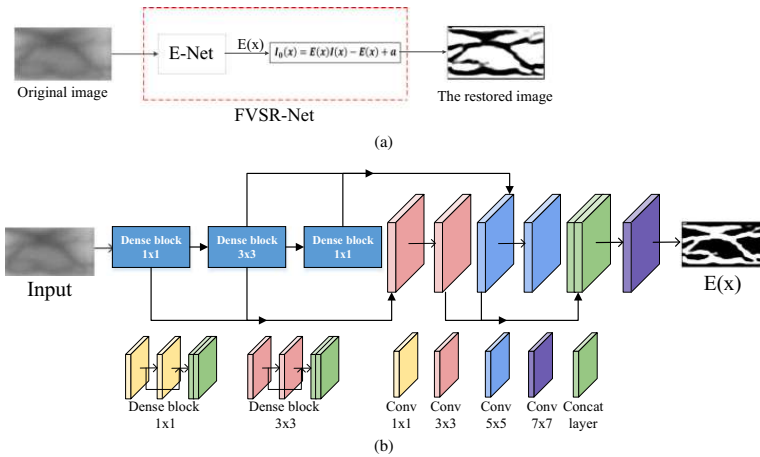


Fig. 23 The framework of the proposed scattering removal model [173]. (a) The structure of the FVSR-Net. (b) The structure of the E-Net.

$$I_0(x) = E(x)I(x) - E(x) + a \quad (6)$$

In addition to these CNN-based structures mentioned above, a novel neural network is used to enhance the finger vein images. [174] uses a modified *Pulse Coupled Neural Network* (PCNN) for finger vein image enhancement. PCNN has the characteristics of synchronous pulse release and global coupling to extract sufficient information from complex backgrounds. This study proposes a parameters tuning scheme based on the original PCNN to adjust the parameters without any empirical correlation automatically. The experimental results show that the PCNN-enhanced images can produce a rewarding recognition performance.

5.3 Segmentation

5.3.1 Segmentation based on U-Net

[175–177] compare the segmentation performance of three fully CNNs: U-Net, RefineNet, and SegNet. These studies use the manual labeling method and automatic labeling method to generate the ground truth images from the original datasets, and these ground truth images are used to train these networks. Finally, the segmentation performance is evaluated by calculating the correlation between the input images and the ground truth images. Numerous experiments illustrate that labeled images generated by automatic labeling methods can improve the segmentation accuracy of the network, and U-Net is more sensitive to the input image quality, while RefineNet and SegNet are more stable.

The existing large finger vein segmentation networks are not suitable for implementation in mobile terminals since they are too deep. To solve this problem, [178] proposes a lightweight network for finger vein segmentation. First, the DSC is introduced to the original U-Net to reduce the model's parameters, and the Ghost model [179] is introduced to the network to compress the network further. In addition, channel shuffling is introduced to the model to shuffle and reorganize all feature channels uniformly. Finally, this network obtains better segmentation performance and shorter segmentation time by using filter pruning via geometric median. Experimental results demonstrate that the network achieves great segmentation performance.

[180] proposes a LadderNet-based segmentation method for finger vein images. The LadderNet is an improved network based on the conventional U-Net, which can fuse multi-path transmission information to obtain complex vein features. In this study, the venous features are extracted by the local maximum curvature detection method. These features are used to train LadderNet. The experimental results show that the LadderNet obtains a AUC of 91.56% and 92.91% on SDUMLA-HMT and MMCBNU-6000.

5.3.2 Segmentation based on other networks

[181] evaluates three *State-Of-The Art* (SOTA) semantic segmentation networks on finger vein image segmentation task, and they are Mask RCNN [182], CCNet [183], and HRNet [184]. In this research, the pre-processing steps only contain the resizing and normalization of the input images. The experimental results show that the performance of these networks is unstable when tested on different public datasets. [185] designs a CNN-based multi-scale feature representation method for finger vein image segmentation, and the specific flow is shown in Fig. 24. The *Global Guiding Feature* (GGF) model is used to extract multi-scale feature information and enables the vein features to be better separated from the background. The GGF model comprises four *Local Similarity Pyramid* (LSP) models based on different scales. In addition, the network uses the *Pyramid Fusion Module* (PFM) to enhance the multi-scale features, and this approach avoids the contextual information loss between

different sub-regions. In the feature aggregation phase, this network uses the *Feature Aggregation Module with Channel Attention* (FAMCA) to retain the important feature mappings and ignore the irrelevant ones. The network can automatically exploit multi-scale features in finger vein images to improve the segmentation performance by fusing GGF, PFM, and FAMCA models.

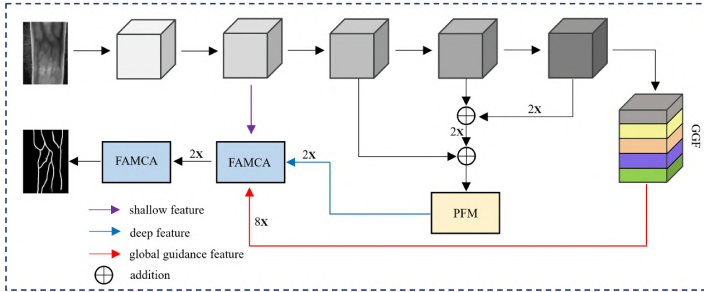


Fig. 24 The framework of the proposed multi-scale feature representation method [185].

5.4 Presentation attack detection

[186, 187] use CNN structure based on AlexNet for finger vein PAD. [186] uses a pre-trained CNN model to extract features from the pre-processed finger vein images and uses PAC for feature extraction. Finally, the SVM is used to verify the authenticity of the images. This PAD method achieves zero error on IDIAP. [187] extends the original AlexNet with seven layers to enrich the robustness of the model. Meanwhile, the study uses patches corresponding to real and fake images to fine-tune the network to suppress overfitting. This method performs well for inkjet printed artifacts with 3.48% APCER and 0% BPCER.

[188] proposes a DSC with residual structure and Linear SVM for finger vein PAD tasks. This is the first time DSC has been used for feature extraction in the finger vein PAD tasks. This feature extraction method allows more comprehensive processing of real-time scenes and makes the network more lightweight. In addition, residual modules are added to the DSC to prevent the gradient vanishing. Linear SVM is used for classification. The experimental results show that the error rate of this model on both IDIAP and SCUT is 0.00%.

[189, 190] perform PAD using a custom CNN structure. [189] designs a lightweight CNN that consists of two convolutional layers and pooling layers followed by two fully-connected layers. This network is not pre-trained in any way, but the training samples of the network are enriched by data augmentation directly on the original images. The method achieves an ACC of 100% on both public datasets. [190] uses a multi-task learning approach to integrate the recognition task and the PAD task into a united CNN model and designs

an FVR system. In this system, the image with the most obvious vein information is selected by applying a multi-intensity illumination strategy, and if it passes the anti-spoofing detection, the features of that image are used for subsequent registration and recognition. The experiments illustrate the excellent performance of the system even on challenging databases with images depicting axial rotation.

5.5 Template protection

[191] presents a template generation framework based on random projection and DBN, which is termed FVR-DLRP. In the FVR-DLRP, the features extracted from the original finger are transform the high-dimensional space to the low-dimensional space by random projection while generating a protected template by combining randomly generated keys. These templates are trained on the DBN, and the complexity of the DBN structure ensures the safety of these templates. The experimental results show that the FVR-DLRP achieves a recognition rate of 96.9% and a *False Acceptance Rate* (FAR) of 1.5% on FV-NET64. [192, 193] use CAE structure to learn deep features from the feature map generated by conventional FVR methods. Biohash generates a protection template for these features. This method performs better than the original biohash protection template in the stolen scenario. [194] encodes the finger vein images into one-dimensional vectors, then encrypts the vectors and re-encodes the vectors into images using the *Rivest–Shamir–Adleman* (RSP) algorithm, and finally feeds it into a modified ResNet containing the Squeeze-and-Excitation block for feature extraction and recognition. [195] proposes a biometric protection algorithm based on the *Binary Decision Diagram* (BDD) [196] and a *Multi-Layer ELM* (ML-ELM) [197]. This BDD-based secure template generation algorithm extracts features from the pre-processed finger vein images using Gabor filters with LDA. These features are transformed into binary-based features, and BDD generates the protected templates of these features. Finally, these templates are fed into the ML-ELM for training. ML-ELM is a multi-layer ANN structure with a faster learning speed than other deep CNNs. The BDD-based protected template generation method can be applied to any binary-valued feature vector, and this method requires only a little storage space. The experimental results show that the method achieves CIRs of 93.09%, 98.70%, and 98.61% on SDUMLA-HMT, MMCBNU-6000, and UTFVP.

5.6 Multimodal biometric recognition

5.6.1 The fusion of two biometric information

[198–201] use finger vein and another biometric information for recognition. [198] uses finger shape and finger vein for recognition, where the finger shape image is a two-dimensional spectrogram image that expresses the change in the frequency component of the finger thickness based on the horizontal position of the finger. The finger shape image and the finger vein image are fed into a

more lightweight ResNet with only four residual blocks to output the matching scores separately. Score level fusion is used for multimodal recognition of finger shape and finger vein. The experimental results show this multimodal recognition framework obtains EERs of 3.509% and 1.706% on SDUMLA-HMT and HKPU. [201] performs score level fusion recognition of face and finger vein. The CNN structures used here are based on AlexNet, and the CNN for faces has two more convolutional layers and one more pooling layer than that for finger veins. This method achieves an ACC of 99.78%.

[199] proposes two methods for multimodal recognition regarding finger knuckle and finger vein. The respective specific processes of the two methods are shown in Fig. 25. This study compares the recognition performance of three pre-trained CNN structures for finger knuckle print image and finger vein image: AlexNet, VGG-16, and ResNet-50. Among these networks, ResNet-50 achieves the best experimental results, with the ACC of 99.89% in the score level fusion method and 98.84% in the feature level fusion method. [200] presents a multimodal biometric recognition system for *Electrocardiogram* (ECG) and finger veins. A custom CNN containing nine convolutional layers is used to extract deep features, and the K-Nearest Neighbors is used for classification. In addition, this system uses multi-canonical correlation analysis to express deep features in low-dimensional space and accelerate the validation. The experimental results show this approach achieves EERs of 1.40% on the score level fusion and 0.12% on the feature level fusion.

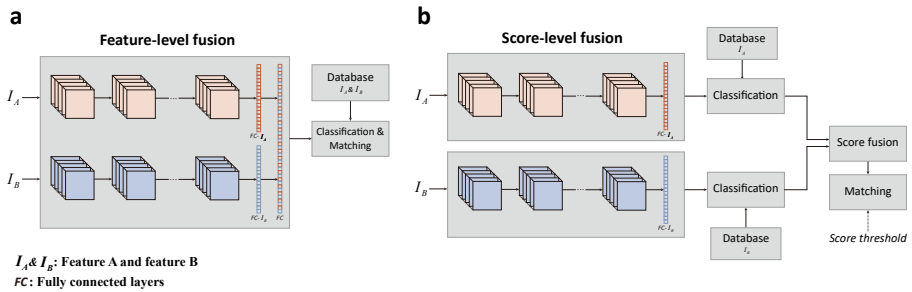


Fig. 25 Two fusion methods on multimodal biometric recognition [199]. (a) feature level fusion method. (b) score level fusion method.

5.6.2 The fusion of various biometric information

[202, 203] performs multimodal biometric recognition using multiple features, including finger veins. [202] aggregates fingerprints, finger veins, and faces for recognition using the score level fusion method. The CNN structure used for feature extraction and the three network structures used to detect the three biometric information are the same. The method related to finger vein uses random forest [204] as the classifier, and other biometric information uses Softmax. This multimodal recognition method achieves an ACC of 99.49% on

SDUMLA-HMT. [203] designs a score level fusion recognition approach with iris, face, and finger vein, and the experiments of this study are also performed on SDUMLA-HMT. It uses pre-trained VGG-16 as the CNN structure to perform the classification, and this method achieves an ACC of 99.39%.

5.7 Other tasks of FVR

Finger vein image quality assessment is an essential part of FVR, and the finger vein images of unsatisfactory quality may lead to verification failure. [205–208] performs the quality assessment of finger vein images based on deep neural networks. [205, 206] extracts binary finger vein patterns from grayscale images, feeds them into a deep CNN for training, and outputs the features using the last fully connected layer in this deep CNN. This deep feature is fed into the SVM to generate image quality scores. Experiments on two public datasets show that the CNN + SVM scheme can accurately discriminate between high and low-quality finger vein images. [207] uses a lightweight CNN with only two convolutional layers and two fully connected layers for image quality assessment. In this approach, the segmented image sub-blocks are fed into the network instead of the entire image, and the network generates the respective quality scores. Finally, the quality score of the whole image is the average of all sub-blocks scores. [208] uses the histogram of competitive Gabor responses [209] to label the training samples. Meanwhile, to compensate for the insufficient number of low-quality images in the training data, it uses the SMOTH method to perform the data augmentation. A custom CNN model trained by these samples achieves an ACC of 98.3% on MMCBNU-6000.

[210, 211] focus on the feature extraction step in FVR. [210] proposes a finger vein feature extraction model based on fully CNN and *Conditional Random Field* (CRF). DSC is added to fully CNN in this model to capture complex vein features by adaptively adjusting the received field. In addition, residual recurrent convolution is used to mine the deep features further. The CRF-RNN module is embedded in the model to output the feature maps. [211] proposes a new loss function. It can dynamically adjust margins of different types to obtain more representative features. Experiments prove that the features extracted by this method have stronger geometric interpretations. [212, 213] focus on ROI extraction of finger vein images. [212] uses VGG-16 for ROI extraction of finger vein images. This research compares the feature maps outputted by each convolutional layer, and the experiment results show that the feature map output by the third convolutional layer converges the fastest and has the best performance. [213] uses a capsule network to perform the ROI extraction task of FVR. The dynamic routing algorithm between capsules replaces the pooling layer in CNN, thus avoiding the loss of location information due to pooling operations. This method not only improves the efficiency of ROI extraction but also simplifies the network model compared to deep CNN.

[214] proposes a GAN-based image generation method to compensate for the lack of finger vein images in public datasets by generating finger vein images. This network generates new finger vein images from the segmented

finger vein image by convolution and deconvolution operations, adding residual blocks to suppress overfitting. The images generated by this method are clearer than the original images in the datasets. [215] uses a modified Pix2Pix model to generate grayscale finger vein images from binary templates. The generator network uses a U-Net structure to generate the vein images and adds residual blocks between the up-sampled and down-sampled convolution layers. This method can be used for data reconstruction across datasets rather than just intra-class data augmentation.

[216, 217] explore the similarities between finger veins from the same people. To investigate the vein similarity between the left and right hand symmetrical fingers, a pair of symmetrical fingers from the same subject is considered a class in [216]. DenseNet-201 is used to train the manually extracted features. This paper experimentally considers that although there is some similarity between symmetrical fingers of the same people, this similarity is not sufficient for applicants on FVR. As an extension work, [217] introduces the triple loss into the CNN model, enabling vein similarities between symmetric fingers to be successfully detected.

5.8 Summary

Compared with classical neural networks, deep neural networks perform more tasks in the field of FVR. DL is now the mainstream technology used in the FVR. To illustrate the application of deep neural networks on FVR, we provide a summary table of related papers in Tab. 5.

Verification tasks are the most frequent tasks involved in FVR. Typical structures, AlexNet, VGGNet, ResNet, etc. have achieved excellent results in the verification task. In addition, many custom CNN structures have also obtained satisfactory performance. The construction of 3D finger vein images provides a desirable prescription for finger posture change problems. GANs are the mainstream network in image enhancement tasks on FVR due to their powerful image generation ability. The CAEs are also performed image enhancement by the convolutional layers in the encoder and de-convolutional layers in the decoder. Like many other computer vision tasks, U-Net is the most commonly used network for performing segmentation tasks in FVR. However, the number of papers related to segmentation tasks is scarce, and this research direction needs further exploration. The main task of the PAD on FVR is to identify the printed finger vein images and the real collected finger vein images. Many research works have achieved complete success on this task. The study of the template protection tasks helps prevent data leakage from the FVR system and fully safeguards users' privacy.

6 Challenges and potential directions

Although the finger vein is stable and challenging in forgery due to the finger vein being deep inside the skin, the recognition process of FVR still faces some challenges. In this section, we summarize the challenges related to FVR and

Table 5 Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2015	[111]	Verification	DBN	Private (IN:6000 SN:64)	ACC = 96.9% EER = 1.5%
2015	[205]	Image Q-Assessment	Custom	HKPU FV-USM	ACC = 88.89% (HI) ACC = 88.18% (LI) ACC = 74.98% (HI) ACC = 70.07% (LI)
2016	[143]	Verification	LeNet-5	Private	ACC = 96.78%
2016	[144]	Verification	Constom	UTFVP	ACC = 100%
2017	[84]	Verification	AlexNet	Private	ACC = 99.4% EER = 0.21%
2017	[100]	Verification	VGG-16	SDUMLA-HMT Private1 (IN:1200 SN:20) Private2 (IN:1980 SN:33)	EER = 0.396% EER = 1.275% EER = 3.906%
2017	[101]	Verification	VGG-16	Private1 (Outdoor continuous acquisition) Private2 (Indoor intermittent acquisition) Private3 (Indoor continuous acquisition)	EER = 0.42% EER = 1.41% EER = 2.14%
2017	[189]	PAD	Custom	IDIAP SCUT	ACER = 0.00% ACER = 0.00%
2017	[86]	Verification	AlexNet	SDUMLA-HMT Private (IN:2970 SN:198)	EER = 0.80% EER = 0.079%
2017	[112]	Verification	DBN Custom	Private (IN:960 SN:15)	ACC = 99.6%
2017	[206]	Image Q-Assessment	Custom	FV-USM HKPU	ACC = 71.01% (HI) ACC = 73.57% (LI) ACC = 87.08% (HI) ACC = 86.36%
2017	[218]	Image enhancement	Custom	HKPU FV-USM	EER = 2.70% EER = 1.42%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2017	[186]	PAD	AlexNet VGG-16	IDIAP ISPR	APCER = 0.2018% BPCER = 0.1863% ACER = 0.1940% APCER = 0.0000% BPCER = 0.1240% ACER = 0.0620%
2017	[187]	PAD	AlexNet	Private	APCER = 3.48% BPCER = 0% (Inkjet printed) APCER = 0% BPCER = 0% (Laserjet printed)
2018	[169]	Image enhancement	CAE	Private (IN:5850)	PSNR = 22.86 dB (Single irregular-region) PSNR = 21.65 dB (Multiple irregular-region) PSNR = 24.01 dB (Square-region)
2018	[219]	Verification	Two stream CNN	MMCBNU-6000 SDUMLA-HMT	EER = 0.10% EER = 0.47%
2018	[220]	Verification	Custom	HKPU FV-USM SDUMLA-HMT UTFVP	ACC = 95.32% ACC = 97.53% ACC = 97.48% ACC = 95.56%
2018	[221]	Verification	Custom	Private (IN:1200)	ACC = 100%
2018	[198]	Verification	ResNet	SDUMLA-HMT HKPU	EER = 2.4088% EER = 0.8255%
2018	[175]	Segmentation	U-Net RefineNet SegNet	UTFVP SDUMLA-HMT	EER = 0.46% FMR = 0.60% ZFMR = 1.75% (U-Net) EER = 0.27% FMR = 0.27% ZFMR = 0.46% (RefineNet) EER = 1.20% FMR = 2.87% ZFMR = 4.44% (SegNet) EER = 6.15% FMR = 13.37% ZFMR = 16.80% (U-Net) EER = 2.45% FMR = 5.87% ZFMR = 8.65% (RefineNet) EER = 5.50% FMR = 12.63% ZFMR = 16.34% (SegNet)
2018	[191]	Template protection	DBN	FV-NET64	GAR = 96.9% FAR = 1.5%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2018	[108]	Verification	CAE	FV-USM	ACC = 99.49% EER = 0.16%
2018	[89]	Verification	AlexNet	Private	ACC = 95%
2018	[125]	Verification	Custom	MMCBNU-6000	EER = 0.04% (Closed-set) EER = 0.30% (Open-set)
				FV-USM	EER = 0.06% (Closed-set) EER = 0.76% (Open-set)
				SDUMLA-HMT	EER = 0.46% (Closed-set) EER = 1.20% (Open-set)
2018	[207]	Image Q-Assessment	Lightweight CNN	MMCBNU-6000	ACC = 71.95%
				SDUMLA-HMT	ACC = 74.63%
2019	[222]	Verification	Custom	SDUMLA-HMT	ACC = 99.21%
				HKPU	ACC = 99.61%
				FV-USM	ACC = 98.44%
				MMCBNU-6000	ACC = 98.58%
2019	[172]	Image enhancement	Custom	HKPU	PSNR = 29.638 dB
2019	[150]	Verification	Lightweight CNN	HKPU	EER = 0.1497%
			Custom1 (Triplet loss)		EER = 0.1316%
			Custom2 (Joint Bayesian)		EER = 0.1327%
			VGG-16		EER = 0.1223%
2019	[159]	Verification	Capsule network	SDUMLA-HMT	100%
				UTFVP	94%
				HKPU	88%
				MMCBNU-6000	100%
2019	[122]	Verification	Custom	Private (IN:3000 SN:1000)	ACC = 96.8%
2019	[223]	Verification	Custom	UTFVP	EER = 0.0614%
				SDUMLA-HMT	EER = 0.0395%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2019	[174]	Image enhancement	PCNN	NJUST-FV	Gradient in spatial = 0.6087 Grey contrast = 31.5743
				SDUMLA-HMT	Information capacity = 9.2709 Deep evaluator = 6.4869
				THU-FVFDTr	Gradient in spatial = 0.6855 Grey contrast = 23.9111
				HKPU	Information capacity = 6.1768 Deep evaluator = 6.1851
					Gradient in spatial = 0.576 Grey contrast = 23.8313
2019	[224]	Similarity explosion	Custom	MMCBNU-6000	Information capacity = 7.3566 Deep evaluator = 6.6593
				SDUMLA-HMT	Gradient in spatial = 23.1059 Grey contrast = 0.6294
					Information capacity = 3.6167 Deep evaluator = 6.0966
					EER = 0.74%
					EER = 2.37%
2019	[113]	Verification	DBN	Private (IN:3000 SN:1000)	ACC = 97.4%
2019	[114]	Verification	LSTM	HKPU	EER = 0.95%
2019	[133]	Verification	Custom	(Train) MMBNU-6000	ACC = 86.49% AUC = 0.9337
				(Test) Private (IN:9888 SN:103)	
2019	[72]	Verification	Siamese CNN	MMCBNU-6000	EER = 0.12%
				FV-USM	EER = 0.30%
				SDUMLA-HMT	EER = 0.66%
				SCUT	EER = 2.78%
2019	[90]	Verification	AlexNet	Private (IN:100 SN:12)	ACC = 91.67%
					AUC = 0.942
2019	[176]	Segmentation	U-Net	UTFVP	EER = 0.64% FMR = 1.85% ZFMR = 3.47%
			RefineNet		EER = 1.76% FMR = 4.12% ZFMR = 6.34%
			SegNet		EER = 2.21% FMR = 6.20% ZFMR = 11.25%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2019	[85]	Verification	Custom	Private (IN:500 SN:50)	ACC = 98.64% ACC = 98.95%
2019	[105]	Verification	GAN	SDUMLA-HMT THU-FVFD ² T	EER = 0.94% EER = 1.12%
2019	[225]	Image enhancement	GAN	HKPU	EER = 0.87%
2019	[160]	Verification	GNN	MMCBNU-6000 SDUMLA-HMT	ACC = 99.98% ACC = 99.98%
2019	[109]	Verification	CAE	FV-USM SDUMLA-HMT	ACC = 99.95% ACC = 99.78% EER = 0.12% EER = 0.21%
2019	[126]	Verification	Custom	FV-USM SDUMLA-HMT THU-FVFD ² T	ACC = 99.49% ACC = 98.19% ACC = 100%
2019	[102]	Verification	SqNet	MMCBNU-6000 SDUMLA-HMT	EER = 1.89% EER = 4.91%
2019	[170]	Image enhancement	CAE	Private (IN:5850 585)	EER = 0.16%
2019	[180]	Segmentation	LadderNet	SDUMLA-HMT MMCBNU-6000	ACC = 92.44% ACC = 93.93% AUC = 91.56% AUC = 92.91%
2019	[226]	Verification	Custom	FV-USM	ACC = 82.45% (T.S. 1) ACC = 99.52% (T.S. 2)

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance		
2019	[180]	Segmentation	LadderNet	SDUMLA-HMT MMCBNU-6000	ACC = 92.44% ACC = 93.93%	AUC = 91.56% AUC = 92.91%	
2019	[134]	Verification	Custom	MMCBNU-6000 FV-USM SDUMLA-HMT HKPU	EER = 0.11% EER = 0.57% EER = 1.09% EER = 1.67%		
2019	[216]	Verification	DenseNet	SDUMLA-HMT	EER = 0.54%		
2019	[152]	Verification	DenseNet	SDUMLA-HMT HKPU	EER = 2.35% EER = 0.33%		
2019	[130]	Verification	Custom	Private (3D Image IN:8526 SN:203) FV-USM SDUMLA-HMT HKPU	EER = 2.84% EER = 0.94% EER = 1.69% EER = 2.40%		
2020	[156]	Verification	ShuffleNet	SDUMLA-HMT FV-USM MMCBNU-6000	EER = 0.37% EER = 0.31% EER = 0.05%		
2020	[158]	Verification	Custom	FV-USM THU-FVFDt2	ACC = 96.15% ACC = 99.48% ACC = 99.56%		
2020	[110]	Verification	CAE	FV-USM SDUMLA-HMT	ACC = 98.02% (CAE) ACC = 97.86% (CAE)	ACC = 98.88% (CAE+ELM) ACC = 98.58% (CAE+ELM)	

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2020	[123]	Verification	Custom	Private (IN:3000 SN:1000) FV-USM	ACC = 98.4% ACC = 98.0%
2020	[124]	Verification	Custom	FV-USM	ACC = 95.1% EER = 0.0373%
2020	[116]	Verification	LSTM Custom CNN	Private (IN:9000 SN:100)	ACC = 99.13%
2020	[163]	Verification	GAN	HKPU SDUMLA-HMT	EER = 1.814% EER = 3.934%
2020	[210]	Segmentation	U-Net	SDUMLA-HMT MMCBNU-6000 HKPU	EER = 5.827% EER = 0.364% EER = 2.372%
2020	[178]	Segmentation	U-Net	SDUMLA-HMT MMCBNU-6000	EER = 0.1546% EER = 0.1556%
2020	[190]	PAD	Custom CNN	IDIA SCUT	EER = 5.61% (Recognition) EER = 2.18% (Recognition) HTER = 0.00% (PAD) HTER = 0.00% (PAD)
2020	[227]	Image enhancement	GAN	MMCBNU-6000 FV-USM	EER = 5.66% EER = 2.37%
2020	[208]	Image Q-Assessment	Custom	MMCBNU-6000	ACC = 98.3%
2020	[202]	M-Biometric recognition	Custom CNN	SDUMLA-HMT	ACC = 99.49%
2020	[228]	Verification	ZFNet	SDUMLA-HMT	ACC = 86%
2020	[212]	ROI extraction	VGG-16	THU-FVFDt1 UTFVP HKPU	EER = 0.1987% IOU = 94.95%
2020	-	Verification	Custom CNN	Private	ACC = 100%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2020	[203]	M-Biometric recognition	VGG-16	SDUMLA-HMT	ACC = 99.39% (FLF) ACC = 100% (SLF)
2020	[199]	M-Biometric recognition	AlexNet VGG-16 ResNet-50	SDUMLA-HMT	ACC = 79.03% (Softmax) EER = 0.2925% (Softmax) 0.322% (SVM) ACC = 85.04% (Softmax) EER = 0.1490% (Softmax) 0.204% (SVM) ACC = 98.58% (Softmax) EER = 0.0142% (Softmax) 0.066% (SVM)
2020	[153]	Verification	DenseNet	HKPU SDUMLA-HMT	EER = 0.05% EER = 1.65%
2020	[117]	Verification	Lightweight CNN	MMCBNU-6000 FV-USM	ACC = 99.05% EER = 0.503% ACC = 97.95% EER = 1.070%
2020	[177]	Segmentation	U-Net RefineNet SegNet	UTFVP	EER = 0.462% FMR = 0.787% ZFMR = 1.388% EER = 1.437% FMR = 2.870% ZFMR = 4.444% EER = 0.686% FMR = 1.388% ZFMR = 2.916%
2020	[93]	Verification	U-Net ResNet	SDUMLA-HMT THU-FVFD12	ACC = 99.53% ACC = 98.4%
2020	[214]	Image generation	GAN	MMCBNU-6000	EER = 0.24%
2021	[171]	Image enhancement	VGG-16	SDUMLA-HMT FV-USM UTFVP THU-FVFD11	ACC = 99.84% ACC = 98.6% ACC = 99.10% ACC = 98.9%
2021	[194]	Template protection	Custom CNN	SDUMLA-HMT MMCBNU-6000 HKPU FV-USM	EER = 2.451% CIR = 95.912% (LBP) EER = 2.137% CIR = 96.698% (One-D vector) EER = 0.232% CIR = 99.100% (LBP) EER = 0.090% CIR = 99.667% (One-D vector) EER = 0.277% CIR = 99.038% (LBP) EER = 0.312% CIR = 98.397% (One-D vector) EER = 0.091% CIR = 99.593% (LBP) EER = 0.214% CIR = 99.187% (One-D vector)
2021	[118]	Verification	Lightweight CNN	SDUMLA-HMT PKU-FVD	EER = 1.13% ACC = 99.3% EER = 0.67% ACC = 99.6%
2021	[119]	Verification	Lightweight CNN	SDUMLA-HMT MMCBNU-6000	EER = 2.29% EER = 0.47%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2021	[136]	Verification	Custom CNN	SDUMLA-HMT MMCBNU-6000 FV-USM SCUT	EER = 1.18% EER = 0.23% EER = 0.49% EER = 0.86%
2021	[139]	Verification	Custom CNN	FV-USM SDUMLA-HMT	ACC = 100% ACC = 99.82%
2021	[107]	Verification	GAN	SDUMLA-HMT HKPU	EER = 0.85% (Train: SDUMLA-HMT Test: HKPU) EER = 3.40% (Train: HKPU Test: SDUMLA-HMT)
2021	[213]	ROI extraction	Capsule network	FV-USM SDUMLA-HMT	Pixel recognition rate = 99.7% Pixel recognition rate = 97.5%
2021	[173]	Image enhancement	Custom CNN	Private (IN:5850 SN:585) SDUMLA-HMT	PSNR = 15.9569dB SSIM = 0.8896 Scoot = 0.7102 PSNR = 13.5929dB SSIM = 0.7947 Scoot = 0.8083
2021	[87]	Verification	AlexNet	FV-USM SDUMLA-HMT	ACC = 97.93% ACC = 97.71%
2021	[103]	Verification	SqNet	FV-USM SDUMLA-HMT	ACC = 89.35% ACC = 99.81%
2021	[229]	Verification	AlexNet	FV-USM SDUMLA-HMT	ACC = 97.35% ACC = 92.29%
2021	[201]	M-Biometric recognition	AlexNet	FV-USM SDUMLA-HMT	EER = 0.050% ACC = 99.85% EER = 2.610% ACC = 94.87%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2021	[115]	Verification	LSTM	Private	ACC = 99.93% EER = 0.07%
2021	[217]	Similarity explosion	SqNet	SDUMLA-HMT	EER = 2.7% (SqNet) 3.1% (ResNet) 4.9% (LCNN)
			LCNN	UTFVP	EER = 2.5% (SqNet) 3.6% (ResNet) 4.6% (LCNN)
			ResNet	PLUSVein-FV3	EER = 2.4% (SqNet) 3.2% (ResNet) 4.7% (LCNN)
				HKPU	EER = 3.7% (SqNet) 5.6% (ResNet) 10.0% (LCNN)
2021	[148]	Verification	AlexNet SqNet GoogleNet	SDUMLA-HMT	ACC = 82.17% ACC = 87.06% ACC = 92.22%
2021	[140]	Image enhancement	Custom CNN	SDUMLA-HMT	ACC = 94.87%
2021	[161]	Image enhancement	Custom CNN	(Train) Private1 (IN:427 SN:960) (Test) Private2 (IN:720 SN:72)	PSNR = 30.42dB SSIM = 0.9885
2021	[99]	Verification	Custom CNN	FV-USM	ACC = 98.89%
2021	[185]	Segmentation	VGG-16	(Train) Private1 (IN:950) (Test) Private2 (IN:150)	JS = 0.926 MBE = 0.077 MAE = 16.46
2021	[94]	Verification	ResNet		ACC = 99.99% EER = 0.03% (Closed-Set)
					ACC = 99.79% EER = 0.25% (Open-Set)
				FV-USM	ACC = 99.56% EER = 0.72% (Closed-Set)
				SDUMLA-HMT	ACC = 99.25% EER = 1.53% (Open-Set)
				HKPU	ACC = 99.60% EER = 0.55% (Closed-Set)
				private (IN:8316 SN:1386)	ACC = 99.07% EER = 1.30% (Open-Set)
					ACC = 99.20% EER = 1.59% (Closed-Set)
					ACC = 97.90% EER = 3.97% (Open-Set)

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2021	[96]	Verification	ResNet	FV-USM UTFVP FV-USM PALMAR SDUMLA-HMT THU-FVFD ^T 1 IDIAP MMCBNU-6000 HKPU	ACC = 98.10% AUC-ROC Score = 0.9992 AUC-ROC Score = 1.0000 AUC-ROC Score = 1.0000 AUC-ROC Score = 1.0000 AUC-ROC Score = 0.9993 AUC-ROC Score = 0.9990 AUC-ROC Score = 0.9997 AUC-ROC Score = 1.0000
2021	[95]	Verification	ResNet		
2021	[193]	Template protection	CAE	UTFVP	FMR = 0.0% FNM ^R = 0.0% EER = 0.0% (Nomal) FMR = 18.7% FNM ^R = 16.7% EER = 17.7%
2021	[104]	Verification	SqNet DenseNet	PLUSVein-FV3 PMMDB	
2021	[211]	Feature extraction	Custom CNN	FV-USM HKPU SDUMLA-HMT MMCBNU-6000 THU-FVFD ^T 3	EER = 0.81% EER = 1.59% EER = 0.31% EER = 1.43% EER = 0.39%
2021	[168]	Image enhancement	GAN	(Train) Private1 (IN:680 IN:68) × 10 (Test) Private2 (IN:600 SN:60) × 6	PSNR = 34.71dB
2021	[91]	Verification	ResNet	HKPU MMCBNU-6000 FV-USM	EER = 1.90% EER = 0.21% EER = 0.48%

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2021	[165]	Image enhancement	GAN	HKPU SDUMLA-HMT	EER = 1.65% EER = 3.17%
2021	[142]	Verification	Custom CNN	HKPU FV-USM	ACC = 92.11% ACC = 94.17%
2021	[192]	Template protection	CAE	UTFVP	FMR = 0.5% FNMR = 0.7% EER = 0.6% (Nomal) FMR = 8.5% FNMR = 10.3% EER = 9.4%
2021	[200]	M-Biometric recognition	Custom CNN	VeinECG	EER = 0.12% (FLF) EER = 1.40% (SLF)
2021	[120]	Verification	Lightweight CNN	HKPU FV-USM SDUMLA-HMT UTFVP	CIR = 96.98% CTR = 98.58% CIR = 97.75% 98.61%
2021	[137]	Verification	Custom CNN	FV-USM MMCBNU-6000	ACC = 95.59% ACC = 97.54%
2021	[157]	Verification	Two-stream CNN	MMCBN-6000 SDUMLA-HMT	EER = 0.17% EER = 0.94%
2021	[141]	Verification	RDL	HKPU	ACC = 91.19%
2021	[164]	Image enhancement	GAN	SDUMLA-HMT HKPU	PSNR = 32.64dB SNR = 22.75 SSIM = 0.85 PSNR = 27.70dB SNR = 20.17 SSIM = 0.90

Table 5 Continued: Summary of the FVR based on deep neural networks.

Year	Reference	Task	Network	Dataset	Performance
2022	[188]	PAD	DSCNN	IDIAP SCUT	APCER = 0.00% BPCER = 0.00% ACER = 0.00% ACPER = 0.00% BPCER = 0.00% ACER = 0.00%
2022	[26]	Verification	GAN	SDUMLA-HMT FV-USM HKPU Private (IN: 8316 SN:1396)	EER = 0.05% (Closed-set) 1.33% (Open-set) EER = 0.03% (Closed-set) 0.14% (Open-set) EER = 0.15% (Closed-set) 0.40% (Open-set) EER = 0.06% (Closed-set) 1.34% (Open-set)
2022	[149]	Verification	DSCNN	SDUMLA-HMT THU-FVFD _T 2	ACC = 98.5% ACC = 90%
2022	[129]	Verification	Custom CNN	FV-USM	AAC = 98.53%
2022	[131]	Verification	Custom CNN	SCUT LFMB-3DPV _{FV}	EER = 2.61% EER = 2.81%
2022	[181]	Segmentation	Mask RCNN CCNet HRNet	HKPU PLUS _{Vein} PMMDB-FR PMMDB-FV3 UTFVP	IOU = 0.92 (Mask RCNN) 0.67 (CCNet) 0.73 (HRNet) IOU = 0.59 (Mask RCNN) 0.57 (CCNet) 0.75 (HRNet) IOU = 0.95 (Mask RCNN) 0.86 (CCNet) 0.87 (HRNet) IOU = 0.92 (Mask RCNN) 0.91 (CCNet) 0.84 (HRNet) IOU = 0.96 (Mask RCNN) 0.96 (CCNet) 0.98 (HRNet) IOU = 0.86 (Mask RCNN) 0.79 (CCNet) 0.83 (HRNet)
2022	[106]	Verification	GAN	THU-FVFD _T 2	-
2022	[162]	Image enhancement	GAN	Private	PSNR = 50.173dB (P-oil dirt) 32.978dB (P-finger molting)

analyze existing research on these challenges. Meanwhile, the ANN technology represented by DL has been popular in recent years. In this section, we present the potential development directions of FVR based on cutting-edge knowledge in the field of ANN and the properties of finger veins as biological tissue.

6.1 Challenges of FVR

6.1.1 Finger posture changes

During the finger vein image acquisition, the user may place the finger on the finger vein collector in different positions, which may result in misaligned images. Meanwhile, the skin's thickness on the finger's surface is hardly perfectly uniform. Some light angular rotations of the finger during the capture process can make the image quality change, and severe rotation of the finger resulting in postural changes can lead to significant differences in the captured vein structure from the registered images. [159, 213] uses capsule networks for processing finger vein images to mitigate the effects of finger pose changes on FVR by taking advantage of the capsule network's ability to store information on feature location changes. However, the training time of the capsule network is too long, and too much redundant information will lead to the capsule shedding problem [230]. Moreover, the capsule network is still in development and not yet mature. [130, 131] construct 3D images using multiple 2D finger vein images, and then use the constructed 3D finger vein images for recognition instead of traditional 2D images. The 3D finger vein images are shown in Fig. 21. The 3D finger vein image can fully render the entire finger's vein structure instead of the localized structure captured on one side in the 2D image. The vein structure in the final 3D finger vein image is consistent regardless of finger posture variations. However, the 3D modeling process is computationally complex compared to the traditional FVR approach using 2D images. Meanwhile, the multiple 2D finger vein images used to construct the 3D model need to be captured from different angles using multiple NIR cameras. This method of capturing images is even more costly. There is also a lack of large public datasets of 3D finger veins for other network models to be trained. In conclusion, more cost-effective and convenient methods need to be developed to solve the challenge of image misalignment caused by finger posture changes in FVR.

6.1.2 Illumination blur

Illumination conditions can affect the finger vein image quality during the image acquisition phase [174]. Affected by uneven illumination condition, the brightness of the central part of the finger vein image is too high, and both sides are too low, resulting in the loss of some of the vein pattern information. Meanwhile, the obtained finger vein images are frequently damaged due to light attenuation in biological tissue [169]. To address the degradation phenomenon of image quality caused by uneven illumination, most research works employ cumbersome image pre-processing methods to improve image contrast.

However, the restoration effect of this method is hardly significant when the image degradation is severe. To enhance the quality of finger vein images damaged by uneven illumination, [165] restores the vein information damaged by light with the help of residual blocks. However, the robustness of this model is insufficient, and the recovery of uneven illumination is unsatisfactory for general scenes. To address this challenge, on the software side, the network model needs to improve robustness without increasing complexity as much as possible, and it also needs to optimize the image enhancement methods further. In terms of hardware, creating an image acquisition environment that is as free from light as possible is necessary.

6.1.3 Dynamic recognition

In a typical biometric recognition system, the user usually need some time to pick up biometric information from the body during the image acquisition process. It is undoubtedly more satisfying if the feature acquisition process can be completed without dwell. However, image capture while the user is in motion will cause image blurring, and these blurred images can affect the recognition performance of ANNs. [164] uses a DeblurGAN based on an improved loss function to focus the motion blurred finger vein image restoration. The approach performs image enhancement, but the matching process still has failed cases because of the difficulty in ROI extraction of motion blurred finger vein images. To more thoroughly address the challenge of dynamic verification for FVR, [116] uses an array of multiple cameras for image capture, and these cameras can collect multiple finger vein videos at different exposures during user movement. The LSTM and CNN are used to process the captured images. However, this system is designed with a non-end-to-end network structure, and the processing is complex. In addition, as with several other solutions to the challenge, the image acquisition method for this method is costly. FVR still needs further exploration in dynamic recognition.

6.1.4 Other challenges

FVR faces several challenges that haphazardly interfere with its performance and application capabilities of FVR. Temperature changes may impact blood flow in the veins, affecting the contrast of the vein picture. Furthermore, various thicknesses of finger bones and muscles will produce irregular shading, which can detect ambiguous or incomplete vessel structures. Besides, although the finger vein information is claimed to be only detected on aliveness, many recognition systems can be cheated by forged vein patterns printed on distinctive paper. The excellent performance, privacy, and stability of FVR are already widely appreciated, and if researchers can overcome these above challenges, the FVR technology will be more powerful.

6.2 Potential development direction of FVR

6.2.1 Knowledge distillation

In addition to training network models, how to deploy the models on terminals with limited hardware conditions is a critical issue for FVR. In general, the trained and deployed models are usually the same in FVR tasks, and networks with excellent recognition performance in FVR tasks usually have complex structures. These complex network models are difficult to deploy on mobile terminals due to their slow inference speed and high computational resource requirements. To enable the network models with excellent performance to be deployed on mobile terminals to enhance the application of capabilities of FVR, the model compression methods are used. Traditional compression methods mainly focus on designing lightweight network structures. The lightweight networks can achieve comparable performance to complex networks on easy tasks through a refined design, but their structure is shallow, resulting in limited feature extraction capability that fails to solve the tasks that have large data volumes [231]. Knowledge distillation [232] can enable lightweight network models to have the better recognition performance comparable to that of deep networks, thus improving the application capability of FVR.

Knowledge distillation is a model compression algorithm and employs a training method based on a “teacher-student” network. The schematic diagram is shown in Fig. 26. This approach uses the knowledge of larger networks with better performance to supervise the training process of the lightweight network to allow the lightweight network to achieve better performance. This large network is called the “teacher” model, and the lightweight model is called the “student” model. In the case of using knowledge distillation, the teacher model only serves as a guide, and it is the student model that is really deployed on terminals.

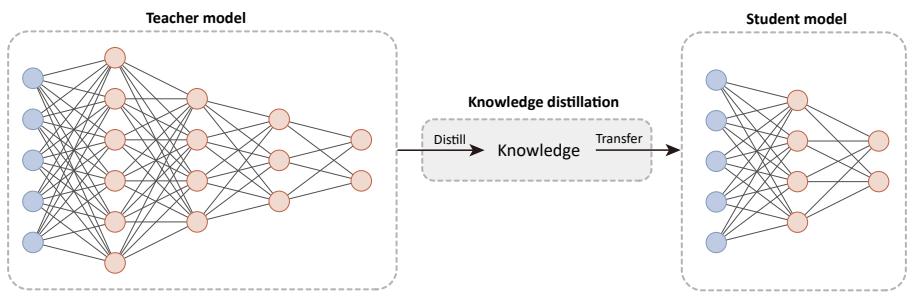


Fig. 26 The knowledge distillation based on “teacher-student” network [232].

Knowledge distillation can be divided into response-based distillation and feature-based distillation according to the different distillation methods. In the classification task, the network uses a Softmax function at the final layer to obtain the probabilities of the positive and the negative samples. The response-based distillation uses the output of Softmax from the teacher model to assist

the training of the student model, allowing the student model to learn the generalization ability of the teacher model by directly combining the labels of samples [233]. In this strategy, the output of Softmax is called soft target, and the true labels of samples are called hard target. Due to all data in the same distribution having similarities [60], and negative samples with higher scores have more similarities to positive samples, in addition to positive samples, a large number of negative samples contain abundant information inferred from the teacher model. However, using the output of Softmax directly as soft target makes its contribution to the loss function minor. As shown in Eq.(7), the response-based distillation amplifies the effect of information carried by negative samples on model training by setting the parameter T in the Softmax function. p_i^T represents the value of the Softmax output of the teacher model on class i at the temperature equal to T , and q_i^T represents the student model. N represents the total number of samples.

$$\begin{aligned} p_i^T &= \frac{\exp(v_i/T)}{\sum_k^N \exp(v_k/T)} \\ q_i^T &= \frac{\exp(z_i/T)}{\sum_k^N \exp(z_k/T)} \end{aligned} \quad (7)$$

In addition to the output of the final layer in the network, the feature maps output from the middle layers of the network in the teacher model can also be the knowledge for supervising the training of the student model [234]. The feature-based distillation is an extension of response-based distillation, and the difference between the two methods is shown in Fig. 27. The feature-based knowledge distillation leans not only on the output results in the teacher model but also on the features extracted from hidden layers in the teacher model.

According to the literature we surveyed, most of the model compression methods involved in the FVR are to design lightweight networks [83, 117–120, 235, 236], and knowledge distillation has never been widely used. Therefore, applying knowledge distillation in FVR is yet to be explored. The development of this technology will take FVR to a new level.

6.2.2 Transformer

Transformer is a landmark model proposed by Google in 2017 [237], which has created a revolution in the field of *Nature Language Process* (NLP). Transformer uses an Encoder-Decoder structure based on self-attention, and the specific structure is shown in Fig. 28. In NLP tasks, RNNs process data in a serial manner. In contrast to RNNs, The innovation of Transformer processes data in a parallel manner since the attention mechanism allows the model to consider the interrelationship between any two words regardless of their position in the text sequence. Relying on this advantage, Transformer has achieved great success in NLP tasks [238, 239]. At the same time, several studies have pioneered the introduction of Transformer into computer vision with excellent performance, including image classification [240], object detection [241], image

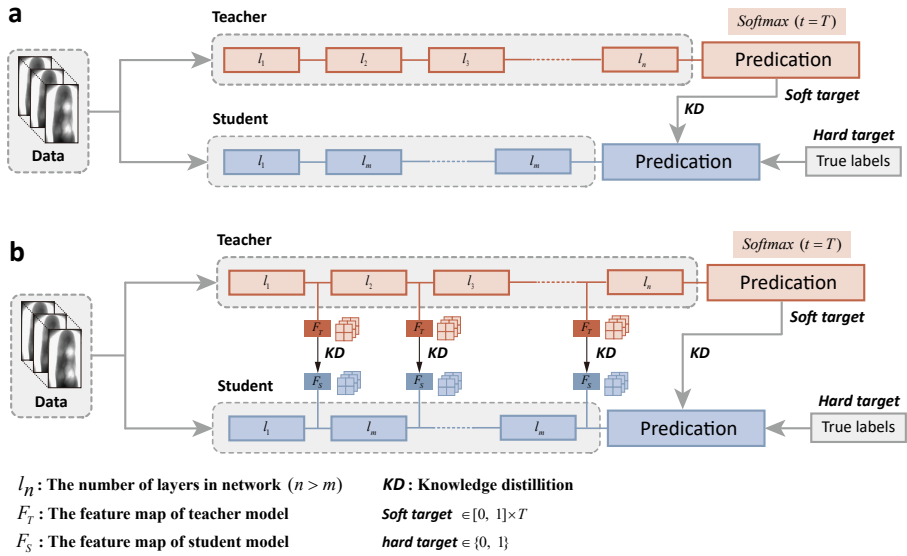


Fig. 27 The response-based distillation and the feature-based distillation. (a) The response-based distillation. (b) The feature-based distillation.

segmentation tasks [242], etc. The *Vision Transformer* (ViT) and the *Segmentation Transformer* (SETR) are two representative Transformer-based models in the CV field, based on which Transformer can make a breakthrough in the verification task and segmentation task of FVR.

6.2.3 The similarity of finger veins

One of the conveniences of FVR is that even if one finger is in an accident, the other fingers can still be used for identification. In addition, it is undoubtedly excellent to lock the identity of a suspect in forensic identification with the veins of any one finger. Nevertheless, registering ten fingers simultaneously in an identification system is a hassle for users. Therefore, it is necessary to explore whether the finger veins of the ten fingers of the same individual are similar in future FVR research work. If there are some connections between different finger veins of the same person and they can be identified by FVR systems, it will take the convenience of FVR systems to a new level. Although [216, 217] focus on this problem, they still have some limitations. [216] considers the connection between the veins of different fingers of the same person is too weak to perform the recognition. [217] utilizes the triplet loss with hard triplet online mining for FVR. This strategy successfully verified that symmetric fingers (the same sort of finger from opposite hands in the same individual) have enough similarities to be recognized. The similarities of other asymmetric fingers are also proved in [217], but the proposed recognition system can still not effectively identify these asymmetric finger veins. Therefore, related work can still be further explored in the future.

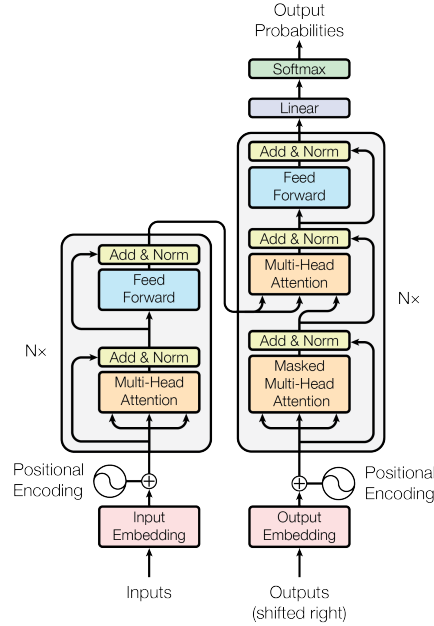


Fig. 28 The Encoder-Decoder structure based on self-attention [237].

7 Conclusion

This paper provides a comprehensive survey of the ANN-based FVR and compensates for the lack of a comprehensive survey related to ANN in the field of FVR. A total of 149 papers have been collected to support this work. The purpose of this paper is to discuss the FVR tasks based on ANNs. In this paper, some FVR-related information is first presented, including the background of FVR in Sec. 1, representative network structures in Sec. 2, and commonly used public datasets in Sec. 3. The literature review section follows this paper's most important work. In Sec. 4, we discuss the application of classical neural networks in FVR from the perspective of the biometric recognition process, including pre-processing, feature extraction, and matching. The classical neural networks are mainly used to perform the classification task in FVR. The MLP and ANFIS are the most widely used in these networks. In Sec. 5, we discuss the application of deep neural networks in FVR from tasks of the papers. In these papers, the typical networks such as AlexNet, VGGNet, ResNet, etc., and some custom networks demonstrate the advantages of DL methods on FVR. The GAN and U-Net show their excellent performance on finger vein image enhancement tasks and segmentation tasks, respectively. In addition, the PAD and template protection tasks on FVR are researched, and these researches achieve outstanding performance. The end of this survey summarizes some typical challenges in FVR and suggests potential directions for FVR. This content can inspire future research in the field of FVR.

8 Acknowledgements

This work is supported by the “Scientific Research Project of Education Department of Hunan Province” (No. 21C0839). We also thank Mr. Shanjun Chen and Mr. Rui Pan for their important discussion in this work. Renye Zhang is the co-first author of this paper.

References

- [1] Kaur, N., *et al.*: A study of biometric identification and verification system. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 60–64 (2021). IEEE
- [2] Ali, W., Tian, W., Din, S.U., Iradukunda, D., Khan, A.A.: Classical and modern face recognition approaches: a complete review. *Multimedia Tools and Applications* **80**(3), 4825–4880 (2021)
- [3] Antonova, V., Balakin, K., Grechishkina, N., Kuznetsov, N.: Development of an authentication system using voice verification. *Journal of Communications Technology and Electronics* **65**(12), 1460–1468 (2020)
- [4] Yang, W., Wang, S., Hu, J., Zheng, G., Valli, C.: Security and accuracy of fingerprint-based biometrics: A review. *Symmetry* **11**(2), 141 (2019)
- [5] Zhuang, X., Zhang, T.: Detection of sick broilers by digital image processing and deep learning. *Biosystems Engineering* **179**, 106–116 (2019)
- [6] Vinh, P.V., Dung, P.X., Tien, P.T., Hang, T.T.T., Duc, T.H., Nhat, T.D.: Smart home security system using biometric recognition. In: International Conference on Internet of Things as a Service, pp. 405–420 (2020). Springer
- [7] Alrahawe, E.A., Humbe, V.T., Shinde, G.: A biometric technology-based framework for tackling and preventing crimes. *Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications*, 133–160 (2021)
- [8] Deepa, N., Priya, J.S., *et al.*: A behavioral biometric system for recognition and authenticity via internet in detecting attacks to provide information security using multiple security interceptive biometric scanners (msibs). In: 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 911–915 (2021). IEEE

- [9] Lei, S., Dong, B., Shan, A., Li, Y., Zhang, W., Xiao, F.: Attention meta-transfer learning approach for few-shot iris recognition. *Computers and Electrical Engineering* **99**, 107848 (2022)
- [10] Wan, C., Wang, L., Phoha, V.V.: A survey on gait recognition. *ACM Computing Surveys (CSUR)* **51**(5), 1–35 (2018)
- [11] Al-Tamimi, A.-K., Qasaimeh, A., Qaddoum, K.: Offline signature recognition system using oriented fast and rotated brief. *International Journal of Electrical & Computer Engineering* (2088-8708) **11**(5) (2021)
- [12] Kulkarni, S., Raut, R., Dakhole, P.: A novel authentication system based on hidden biometric trait. *Procedia Computer Science* **85**, 255–262 (2016)
- [13] of Bologna, U.: Fingerprint verification competition 2004 dataset. <http://bias.csr.unibo.it/fvc2004/default.asp>
- [14] Fuhl, W., Kasneci, G., Kasneci, E.: Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types. In: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 367–375 (2021). IEEE
- [15] Goldbaum, M.: STructured Analysis of the Retina. <http://cecas.clemson.edu/~ahoover/stare/>
- [16] Kono, M.: A new method for the identification of individuals by using of vein pattern matching of a finger. In: Proc. Fifth Symposium on Pattern Measurement, Yamaguchi, Japan, 2000, pp. 9–12 (2000)
- [17] Uhl, A., Busch, C., Marcel, S., Veldhuis, R.: *Handbook of Vascular Biometrics*. Springer, ??? (2020)
- [18] Hashimoto, J.: Finger vein authentication technology and its future. In: 2006 Symposium on VLSI Circuits, 2006. Digest of Technical Papers., pp. 5–8 (2006). IEEE
- [19] Shaheed, K., Liu, H., Yang, G., Qureshi, I., Gou, J., Yin, Y.: A systematic review of finger vein recognition techniques. *Information* **9**(9), 213 (2018)
- [20] Kauba, C., Dražanský, M., Nováková, M., Uhl, A., Rydlo, Š.: Three-dimensional finger vein recognition: A novel mirror-based imaging device. *Journal of Imaging* **8**(5), 148 (2022)
- [21] Akinwande, D., Kireev, D.: Wearable graphene sensors use ambient light

- to monitor health. Nature Publishing Group (2019)
- [22] Al-Khafaji, R.S., Al-Tamimi, M.S.: Vein biometric recognition methods and systems: A review. *Advances in Science and Technology. Research Journal* **16**(1) (2022)
 - [23] Ghosh, U.B., Sharma, R., Kesharwani, A.: Symptoms-based biometric pattern detection and recognition. In: *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*, pp. 371–399. Springer, ??? (2022)
 - [24] Zhang, J., Li, C., Kosov, S., Grzegorzec, M., Shirahama, K., Jiang, T., Sun, C., Li, Z., Li, H.: Lcu-net: A novel low-cost u-net for environmental microorganism image segmentation. *Pattern Recognition* **115**, 107885 (2021)
 - [25] Chen, L., Lin, S., Lu, X., Cao, D., Wu, H., Guo, C., Liu, C., Wang, F.-Y.: Deep neural network based vehicle and pedestrian detection for autonomous driving: a survey. *IEEE Transactions on Intelligent Transportation Systems* **22**(6), 3234–3246 (2021)
 - [26] Hou, B., Yan, R.: Triplet-classifier gan for finger-vein verification. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–12 (2022)
 - [27] Madhusudhan, M., Udaya Rani, V., Hegde, C.: Finger vein recognition model for biometric authentication using intelligent deep learning. *International Journal of Image and Graphics*, 2240004 (2021)
 - [28] Sidiropoulos, G.K., Kiratsa, P., Chatzipetrou, P., Papakostas, G.A.: Feature extraction for finger-vein-based identity recognition. *Journal of Imaging* **7**(5), 89 (2021)
 - [29] Liu, W., Lu, H., Wang, Y., Li, Y., Qu, Z., Li, Y.: Mmran: A novel model for finger vein recognition based on a residual attention mechanism. *Applied Intelligence*, 1–18 (2022)
 - [30] Mohsin, A.H., Zaidan, A., Zaidan, B., Albahri, O., Ariffin, S.A.B., Alemran, A., Enaizan, O., Shareef, A.H., Jasim, A.N., Jalood, N., *et al.*: Finger vein biometrics: taxonomy analysis, open challenges, future directions, and recommended solution for decentralised network architectures. *Ieee Access* **8**, 9821–9845 (2020)
 - [31] McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* **5**(4), 115–133 (1943)
 - [32] Rosenblatt, F.: The perceptron: a probabilistic model for information

- storage and organization in the brain. *Psychological review* **65**(6), 386 (1958)
- [33] Minsky, M., Papert, S.: *Perceptrons: An Introduction to Computational.* MIT press, ??? (1988)
 - [34] Hopfield, J.J.: Hopfield network. *Scholarpedia* **2**(5), 1977 (2007)
 - [35] Flood, M.M.: The traveling-salesman problem. *Operations research* **4**(1), 61–75 (1956)
 - [36] Hinton, G.E., Sejnowski, T.J., Ackley, D.H.: *Boltzmann Machines: Constraint Satisfaction Networks that Learn.* Carnegie-Mellon University, Department of Computer Science Pittsburgh, PA, ??? (1984)
 - [37] Rumelhart, D.E., McClelland, J.L., Group, P.R., *et al.*: *Parallel Distributed Processing vol. 1.* IEEE New York, ??? (1988)
 - [38] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
 - [39] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**(4), 541–551 (1989)
 - [40] Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14, Microsoft (April 1998)
 - [41] Hinton, G.E.: Training products of experts by minimizing contrastive divergence. *Neural computation* **14**(8), 1771–1800 (2002)
 - [42] Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. *Neural computation* **18**(7), 1527–1554 (2006)
 - [43] Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *science* **313**(5786), 504–507 (2006)
 - [44] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: *Icml* (2010)
 - [45] Salakhutdinov, R., Larochelle, H.: Efficient learning of deep boltzmann machines. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 693–700 (2010). *JMLR Workshop and Conference Proceedings*
 - [46] Bengio, Y., *et al.*: Learning deep architectures for ai. *Foundations and*

- trends® in Machine Learning **2**(1), 1–127 (2009)
- [47] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. *Advances in neural information processing systems* **19** (2006)
 - [48] Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feed-forward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256 (2010). *JMLR Workshop and Conference Proceedings*
 - [49] Le Roux, N., Bengio, Y.: Representational power of restricted boltzmann machines and deep belief networks. *Neural computation* **20**(6), 1631–1649 (2008)
 - [50] Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., *et al.*: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* **29**(6), 82–97 (2012)
 - [51] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25** (2012)
 - [52] Zhang, J., Li, C., Yin, Y., Zhang, J., Grzegorzek, M.: Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artificial Intelligence Review*, 1–58 (2022)
 - [53] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
 - [54] Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* **5**(2), 157–166 (1994)
 - [55] 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)
 - [56] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014)
 - [57] Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules.

Advances in neural information processing systems **30** (2017)

- [58] Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584 (2017)
- [59] Zhang, Z., Cui, P., Zhu, W.: Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering (2020)
- [60] Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering **22**(10), 1345–1359 (2009)
- [61] Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine Learning **109**(2), 373–440 (2020)
- [62] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). Ieee
- [63] Zhang, R., Yin, Y., Deng, W., Li, C., Zhang, J.: Deep learning for finger vein recognition: A brief survey of recent trend. arXiv preprint arXiv:2207.02148 (2022)
- [64] Yin, Y., Liu, L., Sun, X.: Sdumla-hmt: a multimodal biometric database. In: Chinese Conference on Biometric Recognition, pp. 260–268 (2011). Springer
- [65] Asaari, M.S.M., Suandi, S.A., Rosdi, B.A.: Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. Expert Systems with Applications **41**(7), 3367–3382 (2014)
- [66] Kumar, A., Zhou, Y.: Human identification using finger images. IEEE Transactions on image processing **21**(4), 2228–2244 (2011)
- [67] Yang, J., Sun, W., Liu, N., Chen, Y., Wang, Y., Han, S.: A novel multimodal biometrics recognition model based on stacked elm and cca methods. Symmetry **10**(4), 96 (2018)
- [68] Tome, P., Vanoni, M., Marcel, S.: On the vulnerability of finger vein recognition to spoofing. In: IEEE International Conference of the Biometrics Special Interest Group (BIOSIG) (2014). <http://publications.idiap.ch/index.php/publications/show/2910>
- [69] Ton, B.T., Veldhuis, R.N.J.: A high quality finger vascular pattern dataset collected using a custom designed capturing device. In: Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, pp. 1–5 (2013)

- [70] Yang, W., Huang, X., Zhou, F., Liao, Q.: Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion. *Information sciences* **268**, 20–32 (2014)
- [71] Yang, W., Qin, C., Liao, Q.: A database with roi extraction for studying fusion of finger vein and finger dorsal texture. In: *Chinese Conference on Biometric Recognition*, pp. 266–270 (2014). Springer
- [72] Tang, S., Zhou, S., Kang, W., Wu, Q., Deng, F.: Finger vein verification using a siamese cnn. *IET Biometrics* **8**(5), 306–315 (2019)
- [73] Tome, P., Vanoni, M., Marcel, S.: On the vulnerability of finger vein recognition to spoofing. In: *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–10 (2014). IEEE
- [74] Zhang, Z.B., Wu, D.Y., Ma, S.L., Ma, J.: Multiscale feature extraction of finger-vein patterns based on wavelet and local interconnection structure neural network. In: *2005 International Conference on Neural Networks and Brain*, vol. 2, pp. 1081–1084 (2005). IEEE
- [75] Zhang, Z., Ma, S., Han, X.: Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network. In: *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4, pp. 145–148 (2006). IEEE
- [76] Wu, J.-D., Ye, S.-H.: Driver identification using finger-vein patterns with radon transform and neural network. *Expert Systems with Applications* **36**(3), 5793–5799 (2009)
- [77] Kaur, A., Mishra, A.: Finger vein image enhancement using rpl, segmentation with neural networks
- [78] He, C., Li, Z., Chen, L., Peng, J.: Identification of finger vein using neural network recognition research based on pca. In: *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pp. 456–460 (2017). IEEE
- [79] Wu, J.-D., Liu, C.-T., Tsai, Y.-J., Liu, J.-C., Chang, Y.-W.: Development of neural network techniques for finger-vein pattern classification. In: *Second International Conference on Digital Image Processing*, vol. 7546, pp. 108–113 (2010). SPIE
- [80] Wu, J.-D., Liu, C.-T.: Finger-vein pattern identification using principal component analysis and the neural network technique. *Expert Systems with Applications* **38**(5), 5423–5427 (2011)
- [81] Wu, J.-D., Liu, C.-T.: Finger-vein pattern identification using svm and

- neural network technique. *Expert Systems with Applications* **38**(11), 14284–14289 (2011)
- [82] Lovepreet, K., Navjot, K.: Finger vein recognition based on pca feature using artificial neural network. *International Journal on Recent and Innovation Trends in Computing and Communication* **5** (2017)
 - [83] Jang, J.-S.: Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics* **23**(3), 665–685 (1993)
 - [84] Meng, G., Fang, P., Zhang, B.: Finger vein recognition based on convolutional neural network. In: *MATEC Web of Conferences*, vol. 128, p. 04015 (2017). EDP Sciences
 - [85] Li, S., Zhang, H., Yang, J.: Finger vein recognition based on local graph structural coding and cnn. In: *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, vol. 11069, pp. 1007–1014 (2019). SPIE
 - [86] Liu, W., Li, W., Sun, L., Zhang, L., Chen, P.: Finger vein recognition based on deep learning. In: *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 205–210 (2017). IEEE
 - [87] Wang, Y., Chen, T.: Finger vein recognition system based on convolutional neural network and android. In: *Journal of Physics: Conference Series*, vol. 2078, p. 012053 (2021). IOP Publishing
 - [88] Chawla, B., Tyagi, S., Jain, R., Talegaonkar, A., Srivastava, S.: Finger vein recognition using deep learning. In: *Proceedings of International Conference on Artificial Intelligence and Applications*, pp. 69–78 (2021). Springer
 - [89] Fairuz, S., Habaebi, M.H., Elsheikh, E.M.A.: Finger vein identification based on transfer learning of alexnet. In: *2018 7th International Conference on Computer and Communication Engineering (ICCCE)*, pp. 465–469 (2018). IEEE
 - [90] Fairuz, S., Habaebi, M.H., Elsheikh, E.M.A.: Pre-trained based cnn model to identify finger vein. *Bulletin of Electrical Engineering and Informatics* **8**(3), 855–862 (2019)
 - [91] Ou, W.-F., Po, L.-M., Zhou, C., Rehman, Y.A.U., Xian, P.-F., Zhang, Y.-J.: Fusion loss and inter-class data augmentation for deep finger vein feature learning. *Expert Systems with Applications* **171**, 114584 (2021)
 - [92] Huang, B., Dai, Y., Li, R., Tang, D., Li, W.: Finger-vein authentication based on wide line detector and pattern normalization. In: *2010 20th*

- International Conference on Pattern Recognition, pp. 1269–1272 (2010). IEEE
- [93] Huang, Z., Guo, C.: Robust finger vein recognition based on deep cnn with spatial attention and bias field correction. *International Journal on Artificial Intelligence Tools* **30**(01), 2140005 (2021)
 - [94] Hou, B., Yan, R.: Arcvein-arccosine center loss for finger vein verification. *IEEE Transactions on Instrumentation and Measurement* **70**, 1–11 (2021)
 - [95] Maser, B., Uhl, A.: Using cnns to identify the origin of finger vein sample images. In: 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), pp. 1–6 (2021). IEEE
 - [96] Yeh, J., Chan, H.-T., Hsia, C.-H.: Resnext with cutout for finger vein analysis. In: 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 1–2 (2021). IEEE
 - [97] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11531–11539 (2020). <https://doi.org/10.1109/CVPR42600.2020.01155>
 - [98] DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552* (2017)
 - [99] Wang, I.S., Chan, H.-T., Hsia, C.-H.: Finger-vein recognition using a nasnet with a cutout. In: 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 1–2 (2021). IEEE
 - [100] Hong, H.G., Lee, M.B., Park, K.R.: Convolutional neural network-based finger-vein recognition using nir image sensors. *Sensors* **17**(6), 1297 (2017)
 - [101] Huang, H., Liu, S., Zheng, H., Ni, L., Zhang, Y., Li, W.: Deepvein: Novel finger vein verification methods based on deep convolutional neural networks. In: 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1–8 (2017). IEEE
 - [102] Zeng, J., Chen, Y., Qin, C., Wang, F., Gan, J., Zhai, Y., Zhu, B.: A novel method for finger vein recognition. In: Chinese Conference on Biometric Recognition, pp. 46–54 (2019). Springer

- [103] Boucetta, A., Boussaad, L.: Biometric authentication using finger-vein patterns with deep-learning and discriminant correlation analysis. *International Journal of Image and Graphics* **22**(01), 2250013 (2022)
- [104] Prommegger, B., Wimmer, G., Uhl, A.: Rotation tolerant finger vein recognition using cnns. In: 2021 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–5 (2021). IEEE
- [105] Yang, W., Hui, C., Chen, Z., Xue, J.-H., Liao, Q.: Fv-gan: Finger vein representation using generative adversarial networks. *IEEE Transactions on Information Forensics and Security* **14**(9), 2512–2524 (2019)
- [106] Jasmine, S., Trueman, T.E., Narayanasamy, P., Ashok Kumar, J.: Finger vein identification using deep convolutional generative adversarial networks. In: *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, pp. 167–177. Springer, ??? (2022)
- [107] Noh, K.J., Choi, J., Hong, J.S., Park, K.R.: Finger-vein recognition using heterogeneous databases by domain adaption based on a cycle-consistent adversarial network. *Sensors* **21**(2), 524 (2021)
- [108] Hou, B., Yan, R.: Convolutional auto-encoder based deep feature learning for finger-vein verification. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–5 (2018). IEEE
- [109] Hou, B., Yan, R.: Convolutional autoencoder model for finger-vein verification. *IEEE Transactions on Instrumentation and Measurement* **69**(5), 2067–2074 (2019)
- [110] Luo, R., Zhang, K.: Research on finger vein recognition based on improved convolutional neural network. *International Journal of Social Science and Education Research* **3**(4), 107–114 (2020)
- [111] Chen, C., Wu, Z., Li, P., Zhang, J., Wang, Y., Li, H.: A finger vein recognition algorithm using feature block fusion and depth neural network. In: *International Symposium on Computational Intelligence and Intelligent Systems*, pp. 572–583 (2015). Springer
- [112] Chen, C., Wu, Z., Zhang, J., Li, P., Azmat, F.: A finger vein recognition algorithm based on deep learning. *International Journal of Embedded Systems* **9**(3), 220–228 (2017)
- [113] Fang, Z.-M., Lu, Z.-M.: Deep belief network based finger vein recognition using histograms of uniform local binary patterns of curvature gray images. *International Journal of Innovative Computing, Information and Control* **15**(5), 1701–1715 (2019)

- [114] Qin, H., Wang, P.: Finger-vein verification based on lstm recurrent neural networks. *Applied Sciences* **9**(8), 1687 (2019)
- [115] Madhusudhan, M., Udaya Rani, V., Hegde, C.: Finger vein recognition model for biometric authentication using intelligent deep learning. *International Journal of Image and Graphics*, 2240004 (2021)
- [116] Kuzu, R.S., Piciucco, E., Maiorana, E., Campisi, P.: On-the-fly finger-vein-based biometric recognition using deep neural networks. *IEEE Transactions on information Forensics and Security* **15**, 2641–2654 (2020)
- [117] Zhao, D., Ma, H., Yang, Z., Li, J., Tian, W.: Finger vein recognition based on lightweight cnn combining center loss and dynamic regularization. *Infrared Physics & Technology* **105**, 103221 (2020)
- [118] Shen, J., Liu, N., Xu, C., Sun, H., Xiao, Y., Li, D., Zhang, Y.: Finger vein recognition algorithm based on lightweight deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–13 (2021)
- [119] Liu, J., Chen, Z., Zhao, K., Wang, M., Hu, Z., Wei, X., Zhu, Y., Yu, Y., Feng, Z., Kim, H., *et al.*: Finger vein recognition using a shallow convolutional neural network. In: *Chinese Conference on Biometric Recognition*, pp. 195–202 (2021). Springer
- [120] Zhong, Y., Li, J., Chai, T., Prasad, S., Zhang, Z.: Different dimension issues in deep feature space for finger-vein recognition. In: *Chinese Conference on Biometric Recognition*, pp. 295–303 (2021). Springer
- [121] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520 (2018)
- [122] Zhao, J.-Y., Gong, J., Ma, S.-T., Lu, Z.-M., Chu, S.-C., Roddick, J.F.: Finger vein recognition scheme based on convolutional neural network using curvature gray image. *J. Netw. Intell.* **4**(3), 114–123 (2019)
- [123] Zhao, J.-Y., Gong, J., Ma, S.-T., Lu, Z.-M.: Curvature gray feature decomposition based finger vein recognition with an improved convolutional neural network. *International Journal of Innovative Computing, Information and Control* **16**(1), 77–90 (2020)
- [124] Zhang, Y., Liu, Z.: Research on finger vein recognition based on sub-convolutional neural network. In: *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 211–216

- (2020). IEEE
- [125] Hu, H., Kang, W., Lu, Y., Fang, Y., Liu, H., Zhao, J., Deng, F.: Fv-net: learning a finger-vein feature representation based on a cnn. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3489–3494 (2018). IEEE
 - [126] Kamaruddin, N.M., Rosdi, B.A.: A new filter generation method in pcanet for finger vein recognition. *IEEE Access* **7**, 132966–132978 (2019)
 - [127] Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing* **24**(12), 5017–5032 (2015)
 - [128] Gong, Y., Lazebnik, S.: Comparing data-dependent and data-independent embeddings for classification and ranking of internet images. In: *CVPR 2011*, pp. 2633–2640 (2011). IEEE
 - [129] Al-Tamimi, M.S., AL-Khafaji, R.S.: Finger vein recognition based on pca and fusion convolutional neural network. *International Journal of Nonlinear Analysis and Applications* **13**(1), 3667–3681 (2022)
 - [130] Kang, W., Liu, H., Luo, W., Deng, F.: Study of a full-view 3d finger vein verification technique. *IEEE Transactions on Information Forensics and Security* **15**, 1175–1189 (2019)
 - [131] Xu, H., Yang, W., Wu, Q., Kang, W.: Endowing rotation invariance for 3d finger shape and vein verification. *Frontiers of Computer Science* **16**(5), 1–16 (2022)
 - [132] Gabor, D.: Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering* **93**(26), 429–441 (1946)
 - [133] Zhang, Y., Li, W., Zhang, L., Lu, Y.: Adaptive gabor convolutional neural networks for finger-vein recognition. In: 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), pp. 219–222 (2019). IEEE
 - [134] Zhang, Y., Li, W., Zhang, L., Ning, X., Sun, L., Lu, Y.: Adaptive learning gabor filter for finger-vein recognition. *IEEE Access* **7**, 159821–159830 (2019)
 - [135] Muthusamy, D., Rakkimuthu, P.: Trilateral filterative hermitian feature transformed deep perceptive fuzzy neural network for finger vein verification. *Expert Systems with Applications* **196**, 116678 (2022)

- [136] Huang, J., Tu, M., Yang, W., Kang, W.: Joint attention network for finger vein authentication. *IEEE Transactions on Instrumentation and Measurement* **70**, 1–11 (2021)
- [137] Liu, W., Lu, H., Li, Y., Wang, Y., Dang, Y.: An improved finger vein recognition model with a residual attention mechanism. In: *Chinese Conference on Biometric Recognition*, pp. 231–239 (2021). Springer
- [138] Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: *European Conference on Computer Vision*, pp. 483–499 (2016). Springer
- [139] Wang, K., Chen, G., Chu, H.: Finger vein recognition based on multi-receptive field bilinear convolutional neural network. *IEEE Signal Processing Letters* **28**, 1590–1594 (2021)
- [140] Noroz, N.K.B., Saleem, S.A., Kumar, R.K., *et al.*: Finger-vein image enhancement and 2d cnn recognition. *International Journal of Innovations in Science & Technology* **3**, 33–44 (2021)
- [141] Najeeb, S.M.M., Al-Nima, R.R.O., Al-Dabag, M.L.: Reinforced deep learning for verifying finger veins. *International Journal of Online & Biomedical Engineering* **17**(7) (2021)
- [142] Liu, C., Qin, H., Yang, G., Shen, Z., Wang, J.: Ensemble deep learning based single finger-vein recognition. In: *International Conference on Cognitive Systems and Signal Processing*, pp. 261–275 (2021). Springer
- [143] Itqan, K., Syafeeza, A., Gong, F., Mustafa, N., Wong, Y., Ibrahim, M.: User identification system based on finger-vein patterns using convolutional neural network. *ARPN Journal of Engineering and Applied Sciences* **11**(5), 3316–3319 (2016)
- [144] Radzi, S.A., Hani, M.K., Bakhteri, R.: Finger-vein biometric identification using convolutional neural network. *Turkish Journal of Electrical Engineering and Computer Sciences* **24**(3), 1863–1878 (2016)
- [145] Padmavathi, S.S.L.C.U.S.L.: Finger-vein biometric identification using convolution neural networks. *Journal of Engineering Sciences* (2020)
- [146] Simard, P.Y., Steinkraus, D., Platt, J.C., *et al.*: Best practices for convolutional neural networks applied to visual document analysis. In: *Icdar*, vol. 3 (2003). Edinburgh
- [147] Weng, L., Li, X., Wang, W.: Finger vein recognition based on deep convolutional neural networks. In: *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*

- (CISP-BMEI), pp. 266–269 (2020). IEEE
- [148] Sharma, S., Lohchab, S.: Personal authentication using finger vein biometric technology with implementation of transfer learning cnn model. Available at SSRN 3993601 (2021)
 - [149] Shaheed, K., Mao, A., Qureshi, I., Kumar, M., Hussain, S., Ullah, I., Zhang, X.: Ds-cnn: A pre-trained xception model based on depth-wise separable convolutional neural network for finger vein recognition. *Expert Systems with Applications* **191**, 116288 (2022)
 - [150] Xie, C., Kumar, A.: Finger vein identification using convolutional neural network and supervised discrete hashing. *Pattern Recognition Letters* **119**, 148–156 (2019)
 - [151] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)
 - [152] Song, J.M., Kim, W., Park, K.R.: Finger-vein recognition based on deep densenet using composite image. *Ieee Access* **7**, 66845–66863 (2019)
 - [153] Noh, K.J., Choi, J., Hong, J.S., Park, K.R.: Finger-vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images. *IEEE Access* **8**, 96748–96766 (2020)
 - [154] Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856 (2018)
 - [155] Ma, N., Zhang, X., Zheng, H.-T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 116–131 (2018)
 - [156] Zheng, H., Hu, Y., Liu, B., Chen, G., Kot, A.C.: A new efficient finger-vein verification based on lightweight neural network using multiple schemes. In: *International Conference on Artificial Neural Networks*, pp. 748–758 (2020). Springer
 - [157] Zhang, L., Sun, L., Dong, X., Yu, L., Li, W., Ning, X.: An efficient joint bayesian model with soft biometric traits for finger vein recognition. In: *Chinese Conference on Biometric Recognition*, pp. 248–258 (2021). Springer
 - [158] Boucherit, I., Zmirli, M.O., Hentabli, H., Rosdi, B.A.: Finger vein identification using deeply-fused convolutional neural network. *Journal of*

- King Saud University-Computer and Information Sciences (2020)
- [159] Gumusbas, D., Yildirim, T., Kocakulak, M., Acir, N.: Capsule network for finger-vein-based biometric identification. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 437–441 (2019). Ieee
 - [160] Li, J., Fang, P.: Fvgnn: A novel gnn to finger vein recognition from limited training data. In: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 144–148 (2019). IEEE
 - [161] He, J., Shen, L., Yao, Y., Wang, H., Zhao, G., Gu, X., Ding, W.: Finger vein image deblurring using neighbors-based binary-gan (nb-gan). IEEE Transactions on Emerging Topics in Computational Intelligence (2021)
 - [162] Jiang, H., Shen, L., Wang, H., Yao, Y., Zhao, G.: Finger vein image inpainting using neighbor binary-wasserstein generative adversarial networks (nb-wgan). Applied Intelligence, 1–12 (2022)
 - [163] Choi, J., Noh, K.J., Cho, S.W., Nam, S.H., Owais, M., Park, K.R.: Modified conditional generative adversarial network-based optical blur restoration for finger-vein recognition. IEEE Access **8**, 16281–16301 (2020)
 - [164] Choi, J., Hong, J.S., Owais, M., Kim, S.G., Park, K.R.: Restoration of motion blurred image by modified deblurgan for enhancing the accuracies of finger-vein recognition. Sensors **21**(14), 4635 (2021)
 - [165] Hong, J.S., Choi, J., Kim, S.G., Owais, M., Park, K.R.: Inf-gan: Generative adversarial network for illumination normalization of finger-vein images. Mathematics **9**(20), 2613 (2021)
 - [166] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8798–8807 (2018)
 - [167] Li, M., Lin, J., Ding, Y., Liu, Z., Zhu, J.-Y., Han, S.: Gan compression: Efficient architectures for interactive conditional gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5284–5294 (2020)
 - [168] He, B., Shen, L., Wang, H., Yao, Y., Zhao, G.: Finger vein denoising algorithm based on custom sample-texture conditional generative adversarial nets. Neural Processing Letters **53**(6), 4279–4292 (2021)
 - [169] Li, D., Guo, X., Zhang, H., Jia, G., Yang, J.: Finger-vein image inpainting

- based on an encoder-decoder generative network. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp. 87–97 (2018). Springer
- [170] Guo, X.-j., Li, D., Zhang, H.-g., Yang, J.-f.: Image restoration of finger-vein networks based on encoder-decoder model. *Optoelectronics Letters* **15**(6), 463–467 (2019)
 - [171] Bilal, A., Sun, G., Mazhar, S.: Finger-vein recognition using a novel enhancement method with convolutional neural network. *Journal of the Chinese Institute of Engineers* **44**(5), 407–417 (2021)
 - [172] Zhu, C., Yang, Y., Jang, Y.: Research on denoising of finger vein image based on deep convolutional neural network. In: 2019 14th International Conference on Computer Science & Education (ICCSE), pp. 374–378 (2019). IEEE
 - [173] Du, S., Yang, J., Zhang, H., Zhang, B., Su, Z.: Fvsn-net: an end-to-end finger vein image scattering removal network. *Multimedia Tools and Applications* **80**(7), 10705–10722 (2021)
 - [174] Lei, L., Xi, F., Chen, S.: Finger-vein image enhancement based on pulse coupled neural network. *IEEE Access* **7**, 57226–57237 (2019)
 - [175] Jalilian, E., Uhl, A.: Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: The impact of training data. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–8 (2018). IEEE
 - [176] Jalilian, E., Uhl, A.: Enhanced segmentation-cnn based finger-vein recognition by joint training with automatically generated and manual labels. In: 2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA), pp. 1–8 (2019). IEEE
 - [177] Jalilian, E., Uhl, A.: Improved cnn-segmentation-based finger vein recognition using automatically generated and fused training labels. In: *Handbook of Vascular Biometrics*, pp. 201–223. Springer, ??? (2020)
 - [178] Zeng, J., Zhu, B., Huang, Y., Qin, C., Zhu, J., Wang, F., Zhai, Y., Gan, J., Chen, Y., Wang, Y., *et al.*: Real-time segmentation method of lightweight network for finger vein using embedded terminal technique. *IEEE Access* **9**, 303–316 (2020)
 - [179] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1580–1589 (2020)

- [180] Zeng, J., Wang, F., Qin, C., Gan, J., Zhai, Y., Zhu, B.: A novel method for finger vein segmentation. In: International Conference on Intelligent Robotics and Applications, pp. 589–600 (2019). Springer
- [181] Prommegger, B., Söllinger, D., Wimmer, G., Uhl, A.: Cnn based finger region segmentation for finger vein recognition. In: 2022 International Workshop on Biometrics and Forensics (IWBF), pp. 1–6 (2022). IEEE
- [182] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
- [183] Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 603–612 (2019)
- [184] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., *et al.*: Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* **43**(10), 3349–3364 (2020)
- [185] Li, X., Lin, J., Pang, Y., Huang, L., Zhong, L., Li, Z.: Fingertip blood collection point localization research based on infrared finger vein image segmentation. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–12 (2021)
- [186] Nguyen, D.T., Yoon, H.S., Pham, T.D., Park, K.R.: Spoof detection for finger-vein recognition system using nir camera. *Sensors* **17**(10), 2261 (2017)
- [187] Raghavendra, R., Venkatesh, S., Raja, K.B., Busch, C.: Transferable deep convolutional neural network features for fingervein presentation attack detection. In: 2017 5th International Workshop on Biometrics and Forensics (IWBF), pp. 1–5 (2017). IEEE
- [188] Shaheed, K., Mao, A., Qureshi, I., Abbas, Q., Kumar, M., Zhang, X.: Finger-vein presentation attack detection using depthwise separable convolution neural network. *Expert Systems with Applications* **198**, 116786 (2022)
- [189] Qiu, X., Tian, S., Kang, W., Jia, W., Wu, Q.: Finger vein presentation attack detection using convolutional neural networks. In: Chinese Conference on Biometric Recognition, pp. 296–305 (2017). Springer
- [190] Yang, W., Luo, W., Kang, W., Huang, Z., Wu, Q.: Fvras-net: An embedded finger-vein recognition and antispooofing system using a unified

- cnn. *IEEE Transactions on Instrumentation and Measurement* **69**(11), 8690–8701 (2020)
- [191] Liu, Y., Ling, J., Liu, Z., Shen, J., Gao, C.: Finger vein secure biometric template generation based on deep learning. *Soft Computing* **22**(7), 2257–2265 (2018)
 - [192] Shahreza, H.O., Marcel, S.: Towards protecting and enhancing vascular biometric recognition methods via biohashing and deep neural networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **3**(3), 394–404 (2021)
 - [193] Shahreza, H.O., Marcel, S.: Deep auto-encoding and biohashing for secure finger vein recognition. In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2585–2589 (2021). IEEE
 - [194] Ren, H., Sun, L., Guo, J., Han, C., Wu, F.: Finger vein recognition system with template protection based on convolutional neural network. *Knowledge-Based Systems* **227**, 107159 (2021)
 - [195] Yang, W., Wang, S., Hu, J., Zheng, G., Yang, J., Valli, C.: Securing deep learning based edge finger vein biometrics with binary decision diagram. *IEEE Transactions on Industrial Informatics* **15**(7), 4244–4253 (2019)
 - [196] Akers, S.B.: Binary decision diagrams. *IEEE Transactions on computers* **27**(06), 509–516 (1978)
 - [197] Chamara, L., Zhou, H., Huang, G., Vong, C.: Representational learning with extreme learning machine for big data. *IEEE Intelligent Systems* **28**(6), 31–34 (2013)
 - [198] Kim, W., Song, J.M., Park, K.R.: Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (nir) camera sensor. *Sensors* **18**(7), 2296 (2018)
 - [199] Daas, S., Yahi, A., Bakir, T., Sedhane, M., Boughazi, M., Bourennane, E.-B.: Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion. *IET Image Processing* **14**(15), 3859–3868 (2020)
 - [200] El-Rahiem, B.A., El-Samie, F.E.A., Amin, M.: Multimodal biometric authentication based on deep fusion of electrocardiogram (ecg) and finger vein. *Multimedia Systems*, 1–13 (2021)
 - [201] Tyagi, S., Chawla, B., Jain, R., Srivastava, S.: Multimodal biometric

- system using deep learning based on face and finger vein fusion. *Journal of Intelligent & Fuzzy Systems* **42**(2), 943–955 (2022)
- [202] mehdi Cherrat, E., Alaoui, R., Bouzahir, H.: Convolutional neural networks approach for multimodal biometric identification system using the fusion of fingerprint, finger-vein and face images. *PeerJ Computer Science* **6**, 248 (2020)
 - [203] Alay, N., Al-Baity, H.H.: Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors* **20**(19), 5523 (2020)
 - [204] Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
 - [205] Qin, H., El-Yacoubi, M.A.: Finger-vein quality assessment by representation learning from binary images. In: *International Conference on Neural Information Processing*, pp. 421–431 (2015). Springer
 - [206] Qin, H., El-Yacoubi, M.A.: Deep representation for finger-vein image-quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology* **28**(8), 1677–1693 (2017)
 - [207] Zeng, J., Chen, Y., Qin, C.: Finger-vein image quality assessment based on light-cnn. In: *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 768–773 (2018). IEEE
 - [208] Wang, Y., Fang, P.: A finger-vein image quality assessment algorithm combined with improved smote and convolutional neural network. In: *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 1–4 (2020). IEEE
 - [209] Lu, Y., Yoon, S., Xie, S.J., Yang, J., Wang, Z., Park, D.S.: Finger vein recognition using histogram of competitive gabor responses. In: *2014 22nd International Conference on Pattern Recognition*, pp. 1758–1763 (2014). IEEE
 - [210] Zeng, J., Wang, F., Deng, J., Qin, C., Zhai, Y., Gan, J., Piuri, V.: Finger vein verification algorithm based on fully convolutional neural network and conditional random field. *IEEE access* **8**, 65402–65419 (2020)
 - [211] Li, H., Lyu, Y., Duan, G., Chen, C.: Improving finger vein discriminant representation using dynamic margin softmax loss. *Neural Computing and Applications* **34**(5), 3589–3601 (2022)
 - [212] Yang, K., Fang, P., Wu, J.: Deep learning-based region of interest extraction for finger vein images. In: *IOP Conference Series: Materials Science and Engineering*, vol. 782, p. 032056 (2020). IOP Publishing

- [213] Ma, N., Li, Y., Wang, Y., Ma, S., Lu, H.: Research on roi extraction algorithm for finger vein recognition based on capsule neural network. In: International Conference on Frontiers of Electronics, Information and Computation Technologies, pp. 1–5 (2021)
- [214] Yang, H., Fang, P., Hao, Z.: A gan-based method for generating finger vein dataset. In: 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, pp. 1–6 (2020)
- [215] Kauba, C., Kirchgasser, S., Mirjalili, V., Uhl, A., Ross, A.: Inverse biometrics: generating vascular images from binary templates. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **3**(4), 464–478 (2021)
- [216] Piciuccio, E., Kuzu, R.S., Maiorana, E., Campisi, P.: On the cross-finger similarity of vein patterns. In: International Conference on Image Analysis and Processing, pp. 12–20 (2019). Springer
- [217] Wimmer, G., Prommegger, B., Uhl, A.: Finger vein recognition and intra-subject similarity evaluation of finger veins using the cnn triplet loss. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 400–406 (2021). IEEE
- [218] Qin, H., El-Yacoubi, M.A.: Deep representation-based feature extraction and recovering for finger-vein verification. *IEEE Transactions on Information Forensics and Security* **12**(8), 1816–1829 (2017)
- [219] Fang, Y., Wu, Q., Kang, W.: A novel finger vein verification system based on two-stream convolutional network learning. *Neurocomputing* **290**, 100–107 (2018)
- [220] Das, R., Piciuccio, E., Maiorana, E., Campisi, P.: Convolutional neural network for finger-vein-based biometric identification. *IEEE Transactions on Information Forensics and Security* **14**(2), 360–373 (2018)
- [221] Fairuz, S., Habaebi, M.H., Elsheikh, E.M.A., Chebil, A.J.: Convolutional neural network-based finger vein recognition using near infrared images. In: 2018 7th International Conference on Computer and Communication Engineering (ICCCE), pp. 453–458 (2018). IEEE
- [222] Avcı, A., Kocakulak, M., Acır, N.: Convolutional neural network designs for finger-vein-based biometric identification. In: 2019 11th International Conference on Electrical and Electronics Engineering (ELECO), pp. 580–584 (2019). Ieee
- [223] Arican, T.: Optimisation of a patch-based finger vein verification with a convolutional neural network. Master’s thesis, University of Twente

(2019)

- [224] Lu, Y., Xie, S., Wu, S.: Exploring competitive features using deep convolutional neural network for finger vein recognition. *Ieee Access* **7**, 35113–35123 (2019)
- [225] Zhang, J., Lu, Z., Li, M., Wu, H.: Gan-based image augmentation for finger-vein biometric recognition. *IEEE Access* **7**, 183118–183132 (2019)
- [226] Singh, M., Singla, S.K.: Convolutional neural network based deep feature learning for finger-vein identification. In: *Proceedings of the 2019 2nd International Conference on Electronics and Electrical Engineering Technology*, pp. 104–107 (2019)
- [227] Yang, S., Qin, H., Liu, X., Wang, J.: Finger-vein pattern restoration with generative adversarial network. *IEEE Access* **8**, 141080–141089 (2020)
- [228] Weng, L., Li, X., Wang, W.: Finger vein recognition based on deep convolutional neural networks. In: *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 266–269 (2020). IEEE
- [229] Mathew, A., Arul, A., Sivakumari, S.: A review on finger vein recognition using deep learning techniques. (2020)
- [230] Wu, F., Yuan, J., Li, Y., Li, J., Ye, M.: Asa-coronet: Adaptive self-attention network for covid-19 automated diagnosis using chest x-ray images. In: *Workshop on Healthcare AI and COVID-19*, pp. 11–20 (2022). PMLR
- [231] Ba, J., Caruana, R.: Do deep nets really need to be deep? *Advances in neural information processing systems* **27** (2014)
- [232] Gou, J., Yu, B., Maybank, S.J., Tao, D.: Knowledge distillation: A survey. *International Journal of Computer Vision* **129**(6), 1789–1819 (2021)
- [233] Hinton, G., Vinyals, O., Dean, J., et al.: Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* **2**(7) (2015)
- [234] Romero, A., Ballas, N., Kahou, S.E., Chassang, A., Gatta, C., Bengio, Y.: Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550* (2014)
- [235] Miura, N., Nagasaka, A., Miyatake, T.: Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Machine vision and applications* **15**(4), 194–203 (2004)

- [236] Chengbo, Y., Huafeng, Q., Lian, Z.: A research on extracting low quality human finger vein pattern characteristics. In: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, pp. 1876–1879 (2008). IEEE
- [237] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
- [238] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018)
- [239] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., *et al.*: Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877–1901 (2020)
- [240] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., *et al.*: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
- [241] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: *European Conference on Computer Vision*, pp. 213–229 (2020). Springer
- [242] Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., *et al.*: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6881–6890 (2021)