Expert Systems With Applications 196 (2022) 116603

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A high compatibility finger vein image quality assessment system based on deep learning

Hengyi Ren [a], Lijuan Sun [a,*], Jian Guo [a], Chong Han [a], Ying Cao [b]

[a] College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[b] College of Computer and Information Engineering, Henan University, Kaifeng 475001, China

ARTICLE INFO

Keywords:
Biometrics
Finger vein
Quality assessment
Automatic labelling
Deep learning
Lightweight network

ABSTRACT

Finger vein feature recognition has been widely studied due to its high security and stability. However, the systematic errors caused by false and missing information in low-quality finger vein images are still concerns. Recently, researchers have proposed a finger vein image quality assessment scheme to eliminate low-quality images. The major drawback of this strategy is the dependence on limited domain knowledge as the strategy cannot effectively filter out low-quality images and is only adapted to a certain recognition method or database. In this paper, we design a new finger vein image quality evaluation method driven by improving the recognition performance of the recognition system and based on the possibility of the false rejection of low-quality images. The method uses statistical knowledge to analyse the matching of different samples between fingers so as to automatically determine the image quality and classify it. Then, we further design a lightweight convolutional neural network to train the high-quality and low-quality images obtained by the above evaluation criteria and mine the common attributes between low-quality images so as to ensure that the prediction system can judge the image quality quickly and accurately in practical use. The proposed scheme is validated on multiple public datasets and compared with current recognition algorithms. The experimental results show that the quality evaluation standard proposed in this paper has good universality and can be combined with different types of recognition methods and acquisition equipment to well screen out the existing low-quality images. Furthermore, the prediction method proposed in this paper has achieved excellent prediction performance on three public datasets, showing that the proposed method can effectively improve the performance of the original recognition system. Taking the local binary pattern (LBP) algorithm as an example, after using the prediction algorithm proposed in this paper to eliminate low-quality images, the recognition system improves the original recognition performance on the three datasets by 12.670% (SDUMLA), 11.923% (MMCBNU_6000) and 9.940% (FV-USM).

1. Introduction

Compared with traditional authentication methods that use specific objects or specific knowledge, biometric recognition methods based on the unique physiological characteristics or behaviour of the human body have attracted widespread attention (Guo, Hsia, Wu, & Chang, 2020). Common biological features include faces, iris, veins and gaits. Among these features, due to their biometrics, internal characteristics, uniqueness and stability, vein features are widely used in various fields, such as convenient payment, smart finance and security prevention and control.

The acquisition of finger vein features requires the employment of near-infrared light with a specific wavelength to illuminate the finger to extract the irregular vascular structures existing in the finger (Chen, Wang, Yang, & He, 2017), usually by using an image as a carrier. Due to the low contrast, uneven illumination and light scattering of skin, the quality of finger vein images varies greatly (Peng, Li, & Niu, 2014), which brings serious challenges to the popularization and use of finger vein recognition systems. In order to overcome the impact of low-quality images on a recognition system, researchers have proposed many effective solutions, which are mainly divided into the following: image enhancement, image quality evaluation and image restoration. According to the application of the recognition system in practical use, one method or a combination of methods are selected. Among the issues, the quality evaluation of finger vein images has been extensively studied because of the following advantages: (1) The process can be regarded as a preprocessing operation and combined with an existing

* Corresponding author.
E-mail addresses: 2018070118@njupt.edu.cn (H. Ren), sunlijuan_nupt@163.com (L. Sun), guoj@njupt.edu.cn (J. Guo), hc@njupt.edu.cn (C. Han), caoyingshiwola@163.com (Y. Cao).

https://doi.org/10.1016/j.eswa.2022.116603
Received 9 April 2021; Received in revised form 15 August 2021; Accepted 20 January 2022
Available online 15 February 2022

recognition system; (2) According to the evaluated image quality, the steps of the recognition system, such as deciding whether to use image enhancement methods based on image quality, can be reasonably adjusted; and (3) Optimizing and improving the recognition performance of the original recognition system.

The quality assessment of biometrics based on images has lagged behind the development of recognition algorithms (Grother & Tabassi, 2007). However, as people have realized that the performance of biometric systems on some abnormal samples is degraded, research on image quality assessment has gradually become a hotspot, and many quality assessment efforts have been proposed (Uhl, Busch, Marcel, & Veldhuis, 2020). The researchers of the National Institute of Standards and Technology (NIST) have performed much work on quality assessment (mainly for fingerprints Tabassi, 2014; Tabassi & Wilson, 2005, irises Lee, Micheals and Phillips, 2009; Lee, Phillips and Micheals, 2009 and other biological characteristics), which has provided effective support for the quality assessment of subsequent finger vein images.

In the past few years, there have been continuous studies on the quality of finger vein images. The existing analyses on the quality of finger vein images can be roughly divided into two categories: analyses of the different attributes of vein images (Huang, Mu, & Xie, 2013; Peng et al., 2014; Xie, Zhou, Yang, Lu, & Pan, 2013; Zhou, Yang, Yang, Yin, & Li, 2015) and statistical analyses of the recognition results (Qin & El-Yacoubi, 2017; Zeng et al., 2019). The methods for analysing the different attributes of a vein image involve analysing the differences in the image according to expert experience to design the judgement standards for evaluating the quality of the vein image, such as the number of vein points, image contrast, and whether there are abnormal points and other attributes. These methods consider human intuition or the inherent attributes of vein images (such as the mean, standard deviation and entropy) as the basis and describe these attributes through manually designed descriptors, such as vein points, image contrast and whether there are abnormal points. Finally, they use a neural network (Esen, Inalli, Sengur, & Esen, 2008a, 2008b; Esen, Ozgen, Esen, & Sengur, 2009a) or support vector machine(SVM) (Esen, Inalli, Sengur, & Esen, 2008c; Esen, Ozgen, Esen, & Sengur, 2009b) for classification. The advantages of this type of method are that it has better descriptiveness and can filter out low-quality images using fixed recognition methods and acquisition equipment. However, there are some shortcomings. First, the designed descriptor is only applicable to the vein dataset considered at the time. When the vein collection equipment changes, the previous descriptor will not be applicable. Second, it is impossible for researchers to investigate all attributes related to image quality in vein images, and images taken by different devices have different sensitivities to different attributes. Finally, the mathematical modelling of some attributes is not easy, which is not conducive to the practical popularization of a system (Esen, Esen, & Ozsolak, 2017; Esen, Inalli, Sengur, & Esen, 2008d).

However, the method of the statistical analysis of recognition results is result-oriented. It is believed that most of the incorrect recognition results are caused by low-quality images. A large number of statistical analyses of the recognition results are used to determine the possible low-quality images. This idea has been applied in the quality evaluation of other biometrics. For example, Tabassi (2014) considered a quality evaluation algorithm to be a black box that converted input samples into output scalars when designing a quality evaluation method of fingerprint images and evaluated it by quantifying the correlation between the image quality and observed matching results. Compared with the method of analysing the attribute difference of vein images, the method of the statistical analysis of recognition results, which was not based on human subjective perception of vein images, has better objectivity and universality. However, there is also a problem in the practical application of this kind of method, that is, how to construct a "black box" between input samples and output scalars. Thanks to the powerful learning ability of convolutional neural network (CNN) when sufficient input data are possessed, some researchers have sought to use CNN to construct a "black box" between input samples and output scalars, which has achieved beneficial results. Among the analyses, previous studies mainly focused on the method of analysing image attributes; however, with the development of deep learning, the method of the statistical analysis of recognition results has received increasing attention.

The quality of finger vein images has been widely studied, and based on the proposed strategies, there are some problems with their designs as follows:

(1) The existing finger vein quality assessment methods are only designed for part of a dataset and a single identification method (Huang et al., 2013; Qin & El-Yacoubi, 2017; Zhou et al., 2015). When finger vein acquisition equipment is replaced or the identification algorithm is updated, the current quality assessment methods are prone to larger errors.

(2) In the actual use of finger vein recognition systems, an important indicator that affects user experience is real-time performance, which requires the quality evaluation system to quickly feedback the image quality within a limited time; however, the existing methods often ignore the time consumed.

Therefore, achieving fast and accurate image quality evaluation without affecting the existing finger vein recognition systems is still a key challenge in designing these systems. In this paper, we propose new vein image quality evaluation criteria to automatically mark high- and low-quality vein images and design a lightweight CNN to predict vein image quality. The main contributions of our work are as follows:

(1) We propose a finger vein image quality evaluation standard based on statistical analysis of the recognition results. Statistical knowledge is used to analyse the matching of different samples among fingers to automatically mark the possible low-quality images. Driven by improving the recognition performance, the evaluation standard can significantly improve the recognition performance of the system while avoiding the possible errors caused by heavy manual labelling of samples and manual operations. We also verify this view in the experimental part.

(2) Although the statistical analysis method can effectively screen out low-quality images, it cannot process a single image. Therefore, it is necessary to mine the common attributes of these low-quality images and use these common attributes as the basis for judging low-quality images. Considering the ability of the CNN to extract image features, we construct a new lightweight CNN to mine hidden different attributes between low-quality images. Based on the deep separable convolution with low time costs, a CNN is constructed to learn and analyse the low-quality images marked in this paper, thereby training an efficient network model that can predict the quality of finger vein images in real time with high accuracy.

(3) The finger vein image quality evaluation criteria proposed in this paper are highly compatible with the quality prediction method. Our method is not limited to a certain recognition algorithm or dataset, but it uses statistical analysis to automatically mark image quality and uses a CNN to mine the common differences between the images. This ensures that the method can be combined with existing finger vein datasets and recognition algorithms.

The remainder of the paper is organized as follows. Section 2 introduces related works on finger vein image quality assessment methods. Section 3 proposes a complete set of finger vein image quality evaluation criteria and an image quality prediction network. The experimental results and detailed analysis are discussed in Section 4, and we conclude the paper in Section 5.

2. Related work

In recent years, an increasing number of researchers have paid attention to biometric identification technology based on finger veins. Many excellent recognition algorithms have been proposed and applied to real life scenarios, but research on the quality evaluation of finger

vein images is still very limited. It is still a challenging task to propose an efficient and reliable finger vein image quality assessment method and combine it perfectly with the existing finger vein recognition system. The following section briefly reviews the latest developments in the evaluation of finger vein image quality.

2.1. Finger vein image quality assessment scheme based on traditional methods

Before the development of CNN, finger vein image processing methods were based on traditional image processing methods. Similar to the research ideas of traditional finger vein recognition algorithms that relied on artificial design features, previous finger vein image quality assessments also used expert experience to design different attributes between different images to determine the quality of an image. Peng et al. (2014) used the gradients, image contrasts and information entropies of finger vein images as difference attributes to calculate the quality score of an image and proposed a score fusion method based on triangle norms to distinguish between finger vein images. Experiments show that this method is helpful for improving the recognition performance of the finger vein system. Zhou et al. (2015) proposed a finger vein image quality evaluation method based on support vector regression (SVR). They first manually annotated the quality scores of finger vein images in the training set, combined it with the five quality features of image contrast, gradient in the spatial domain, Gabor-based features, information capacity, and information entropy to construct the SVR model, and finally used the constructed model to evaluate the quality of the test image. Huang et al. (2013) proposed a rapid evaluation method for finger vein image quality based on the combination of average gradient features of the regional column and texture roughness features to meet the speed and effectiveness requirements of finger vein image evaluation. In the registration phase, the average gradient feature of the regional column is used for one quality judgment. If there is an abnormality, the texture roughness feature is used for the second quality judgment. If the requirements are met, finger registration is carried out. Otherwise, the user is prompted to change the finger to be used for registration. The abovementioned quality assessment schemes are used to analyse the quality of finger vein images by designing different attributes for testing. Although such methods possess good descriptiveness and are easy to understand, it is difficult to find all attributes related to image quality due to the complexity of images and the limitations of artificial design features. Also, the establishment of mathematical models of some attributes is not easy; it requires high labour costs and is not conducive to the popularization and use of finger vein image quality schemes.

2.2. Finger vein image quality assessment scheme based on CNN

With the continuous development of computer hardware, especially GPU computing power, deep learning has made surprising progress in prediction tasks (Altan & Karasu, 2020; Altan, Karasu, & Bekiros, 2019; Altan, Karasu, & Zio, 2021; Karasu, Altan, Bekiros, & Ahmad, 2020). When performing the same task, image processing methods based on CNN continuously break the performance bottlenecks of traditional methods. Inspired by the rapid development of image processing techniques, researchers applied CNN to the prediction of finger vein image quality. Qin and El-Yacoubi (2017) assumed that low-quality images were wrongly rejected in the verification system, and based on this assumption, vein images were automatically marked as either low-quality or high-quality images. At the same time, they also proposed a deep neural network (DNN) for representation learning using very limited knowledge to predict image quality. To further improve the robustness of the DNN, finger vein images were divided into various patches, the quality of each patch from the test image was predicted, and the quality scores from the image patches were jointly input into the probability support vector machine(P-SVM) to improve quality assessment performance. Zeng et al. (2019) proposed a light-CNN-based finger vein image quality assessment method for assessing the impact of low-quality finger vein images on recognition performance. They automatically annotated the finger vein image based on whether the image contained rich and stable vein features and cut the finger vein image into image blocks as the input of the network for training purposes. Finally, the average quality scores of multiple image blocks of an image were used as the final quality score of the image. The above two schemes have been experimentally verified on the public finger vein database. The results show that the reduction of the average error rate is significantly better than the reduction with traditional quality assessment methods. However, there are still some problems, such as only focusing on the prediction accuracy of low-quality images but ignoring the changes in the recognition performance of the system before and after removing the filtered vein images, and a discussion on the compatibility between the quality assessment scheme and the existing recognition system is not present.

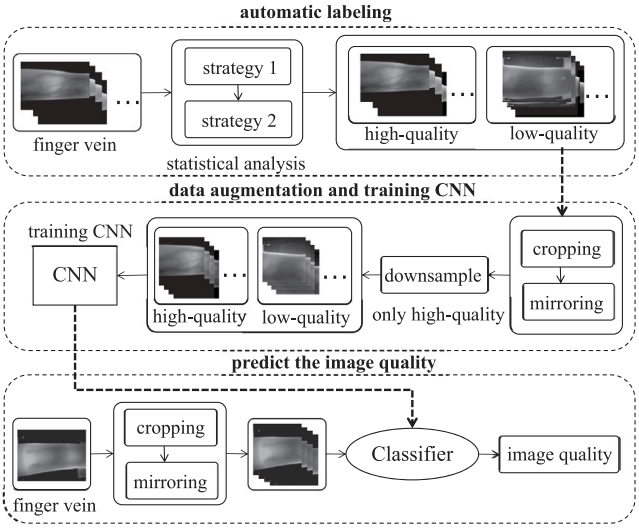

Fig. 1. Finger vein image quality assessment system.

3. Methodology

In this section, we elaborate on the vein image quality evaluation standards and low-quality image prediction methods based on CNN proposed in this paper. As shown in Fig. 1, we design a complete set of image quality assessment standards to detect the quality scores of images and automatically mark low-quality images to avoid heavy manual marking and possible human interference. Then, the screened low-quality images are cut and mirrored to expand the training dataset, and the expanded image blocks are used as the input of the network for feature mining. Finally, the image quality of each image block is predicted, and the image quality of the complete vein image is calculated.

3.1. Detect vein image quality and automatically mark low-quality images

Unlike traditional finger vein image quality assessment schemes that rely on expert experience to design the difference attributes between vein images, we assume that there are a large number of low-quality vein images in images rejected by the verification system. Based on the assumptions above, our proposed finger vein image quality marking method is divided into two parts. First, we analyse the mismatches in the recognition system and determine the abnormal samples that often produce mismatches in the same finger by calculating the average

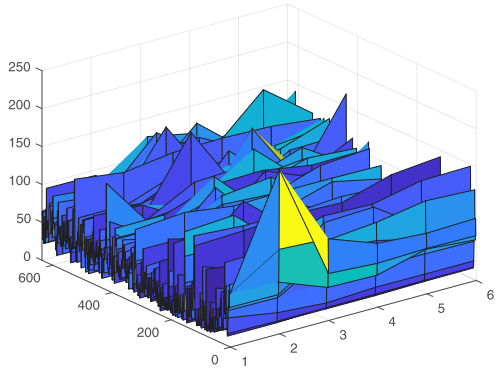

Fig. 2. The average in-class matching similarity calculated by the LBP feature extraction method for different SDUMLA samples.

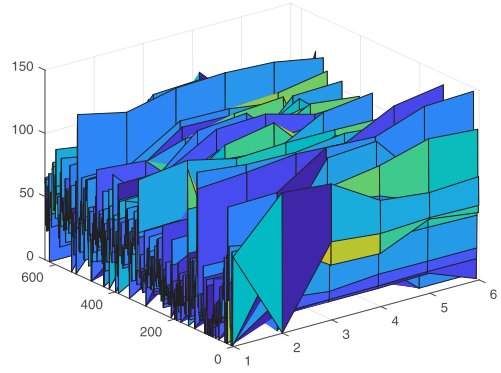

Fig. 3. The average in-class matching similarity of each sample of the SDUMLA dataset after removing low-quality images by using the method in 3.1.1.

intraclass matching similarity of each sample. Then, the possible low-quality images are mined by combining the similarity distribution of the whole dataset with the average intraclass matching similarity between single images.

3.1.1. Analysis of low-quality images present in the same finger

In finger vein recognition systems, errors are often caused by low-quality images. A common problem with such low-quality images is that they are quite different from other images and easily produce large errors in matching. Therefore, these components of different images within a finger must be found first. Our design idea is that in different samples obtained using a single finger, there are few low-quality images, and there is some false or missing information. If low-quality images and normal images are used for similarity comparison, the similarity must be much higher than that calculated using two normal images (please note that we assume that the smaller the similarity is, the greater the probability that the two samples are the same finger). Therefore, we consider calculating the similarity between each sample and other samples in the finger and then judge whether an image is a low-quality image by comparing the mean of the similarity between this sample and other samples. Furthermore, we add a hyperparameter to control the strictness of screening low-quality images. The specific details are as follows: We specify that the j_{th} sample of the i_{th} finger in dataset D is $I_{i,j}$, whose feature is $f_{i,j}$. Please note that our method does not limit the method of feature extraction $f_{i,j}$ and can be obtained by traditional image feature extraction methods (such as local binary pattern (LBP) Ojala, Pietikainen, & Maenpaa, 2002 and histogram of oriented gradients (HOG) Dalal & Triggs, 2005) or by deep learning methods. The extracted features enable the calculation of the similarity between images. After obtaining the characteristics of each sample, we can calculate the similarity $S_{i,j}$ between different samples in the same finger, as shown in formula (1)

$$S_{i,j} = \frac{\sum_{k=1,k\neq j}^{N_i} F(f_{i,j}, f_{i,k})}{N_i - 1} \tag{1}$$

where $S_{i,j}$ represents the average intraclass matching similarity between the j_{th} sample of the i_{th} finger and the remaining samples of the i_{th} finger, N_i is the number of samples of the i_{th} finger, and $F(f_{i,j}, f_{i,k})$ calculates the similarity between features $f_{i,j}$ and $f_{i,k}$, where the method of calculating the similarity is also not limited. Fig. 2 shows the average intraclass matching similarity map calculated after extracting features from different samples of the SDUMLA dataset using the LBP feature extraction method (the similarity calculation method is Euclidean distance). In the figure, the horizontal coordinate axis contains 636 fingers and 6 samples per finger, and the height is the average intraclass matching similarity of each sample. From Fig. 2, we can clearly find that the average intraclass matching similarities of some samples are much higher than those of the remaining samples of the same finger. This is because there is a large difference between this group of samples and other samples of the same finger, which leads to system recognition errors. Next, we must find this group of abnormal samples.

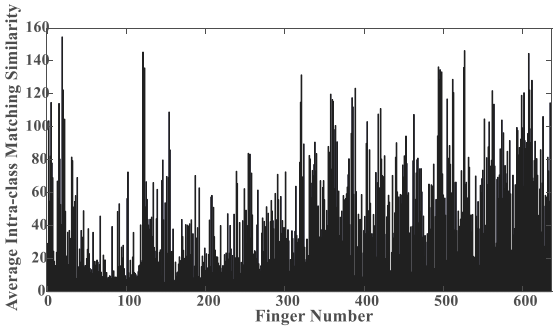

Fig. 4. The average intraclass matching similarity of each finger in the SDMLA dataset after eliminating low-quality images by using the screening mechanism in 3.1.1.

To determine the abnormal samples in different samples of the same finger, we must obtain the overall similarity of all samples in the finger for comparative analysis. We define $P_{i,j}^{avg}$ as the average intraclass matching similarity of samples in the i_{th} finger except for the j_{th} sample, as shown in formula (2):

$$P_{i,j}^{avg} = \frac{\sum_{k=1,k\neq j}^{N_i} S_{i,k}}{N_i - 1} \tag{2}$$

After obtaining the average intraclass matching similarity of a single sample and the average intraclass matching similarity of all fingers except for the j_{th} sample, we can filter out the abnormal samples by comparison. We define the label of abnormal samples as 0 and the label of normal samples as 1. The method is shown in formula (3):

$$label_{i,j} = \begin{cases} 0 & S_{i,j} > P_{i,j}^{avg} * th_1 \\ 1 & otherwise \end{cases} \tag{3}$$

where the threshold th_1 is a hyperparameter that is used to limit the strictness of filtering.

The smaller the th_1 is, the more low-quality images that need to be filtered out. Please note that the smaller the similarity we assume here, the greater the probability that the two samples are from the same finger. If the selected recognition system works based on the property that the greater the similarity, the greater the probability that the two samples are from the same finger, then replace $>$ in formula (3) with $<$.

Through the above screening mechanism, we can easily determine the abnormal samples in each finger. Fig. 3 shows the average intraclass matching similarity of the remaining samples after excluding the filtered low-quality images. Through analysis, it can be found that the amount of abnormal samples in fingers is significantly reduced, and the similarity between different samples of the same finger is also significantly reduced. The change in system performance with different thresholds is verified in the later experimental section.

3.1.2. Analysis of low-quality images present in the entire dataset

After the abnormal samples of each finger are screened out, the performance of the system is significantly improved. However, the above screening mechanism still experiences a problem. When there are multiple low-quality images in one finger, the average intraclass matching similarity of the entire fingers increases, so it is difficult to determine the low-quality images. Fig. 4 shows the average intraclass matching similarity of each finger after eliminating low-quality images through the above screening mechanism. Because some fingers contain multiple low-quality images, the average intraclass matching similarities of these fingers are much higher than those of the remaining fingers. The above mechanism is not able to filter out abnormal samples. Hence, in addition to analysing the internal matching of fingers, we must also consider the similarity distribution of the entire dataset to determine the abnormal images. Our design idea is that after the screening mechanism in 3.1.1, the number of low-quality images will be much lower than the number of normal images. Then, we only need to calculate the average of the similarity of all samples in the entire dataset and determine the remaining low-quality images by comparing the similarity of a single sample with that of the entire dataset. Similar to the strategy in 3.1.1, we also set a hyperparameter here to control the strictness of screening low-quality images.

We stipulate that after the data set D is screened by the method in 3.1.1 for some low-quality images, the remaining samples constitute the data set D_1. For some low-quality images that are removed, the average intraclass matching similarity of each sample must be recalculated. We specify the average intraclass matching similarity of each sample in the data set D_1 as $S'_{i,j}$£, as shown in formula (4):

$$S'_{i,j} = \frac{\sum_{k=1,k\neq j}^{N'_i} F(f_{i,j}, f_{i,k})}{N'_i - 1} \tag{4}$$

where N'_i, represents the number of samples of the i_{th} finger in the data set D_1. Because the similarity distribution of the entire dataset must be considered, we calculate the sample average class matching similarity Q^{avg}_{sum} of the entire dataset, as shown in formula (5):

$$Q^{avg}_{sum} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{N'_i} S'_{i,j}}{\sum_{i=1}^{m} N'_i} \tag{5}$$

where m represents the number of fingers in the dataset D_1. Next, we can compare the similarity of a single sample with the entire dataset to filter out the remaining low-quality images. Similarly, we must set the threshold th_2 to control the strictness of the screening. The specific screening rules are shown in formula (6):

$$label_{i,j} = \begin{cases} 0 & S'_{i,j} > Q^{avg}_{sum} * th_2 \\ 1 & otherwise \end{cases} \tag{6}$$

Similar to formula (3), formula (6) also assumes that the smaller the similarity is, the greater the probability that the two samples will be from the same finger. Fig. 5 shows the average intraclass matching similarity of each sample of the SDUMLA dataset after eliminating low-quality images through the above screening mechanism. By comparing Figs. 2, 4 and 5, it can be found that after two low-quality image filters, the similarity of the intraclass matching of fingers is significantly reduced, which effectively reduces the recognition error of the system. The specific performance analysis is also verified in the experimental section.

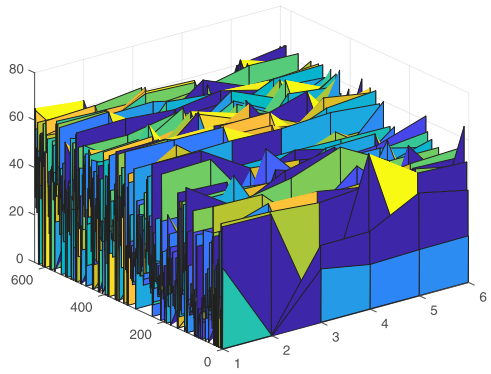

Fig. 5. The average intraclass matching similarity of each sample from the SDUMLA dataset after removing low-quality images by the method in 3.1.2.

3.2. Low-quality image prediction method based on CNN

Through a large number of analyses of recognition results, low-quality images that affect system performance can be effectively screened out. However, these screening mechanisms are based on a large number of recognition results. When the system obtains a single vein image, the above method obviously cannot quickly make a judgment. Therefore, it is necessary to further explore the common attributes in the above low-quality images and use these common attributes to judge the quality of the images quickly and accurately. CNN can effectively mine hidden information in images by learning many training data, and they have achieved success in a variety of visual tasks (Zoph, Vasudevan, Shlens, & Le, 2018). Considering the outstanding ability of CNN in mining hidden image information, we develop a lightweight vein image quality prediction method based on CNN in this paper. The key technologies used in the proposed prediction method are described in detail below.

3.2.1. Key components of CNN

Considering the computing resources and running time of the finger vein recognition system, we construct a lightweight network with goals of low consumption and high speed when constructing the vein image quality prediction network. MobileNet V1 (Howard et al., 2017) proposed a deep separable convolution to replace the traditional convolutional layer, which effectively reduces the operation cost under the premise of ensuring performance. Depth separable convolution decomposes a traditional convolution layer into two parts: a depthwise convolution for spatial filtering and a 1×1 pointwise convolution for feature generation. In the case of identical numbers of weight parameters, compared with the standard convolution operation, the calculation amount can be reduced by several times, thus achieving the purpose of improving the network operation speed.

MobileNet V2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018) introduced a linear bottleneck and inverted residual structure on the basis of depth separable convolution and redesigned depth separable convolution with a residual structure. As shown in Fig. 6, the standard convolution is split into two pointwise convolutions and a depthwise convolution, and linear bottlenecks are used for residual connections, where the pointwise convolutions perform dimensionality addition and reduction operations and the depthwise convolution extracts features.

The SE block (Hu, Shen, & Sun, 2018) is proposed as a substructure embedded in the existing network through learning to obtain the importance of different feature channels, thereby enhancing beneficial features and suppressing invalid or ineffective features for the current task. The core idea is to let the network learn the weights of features according to the loss function, increase the weight of the effective feature map and reduce the weight of the invalid or ineffective map, thus training the model to achieve better results.

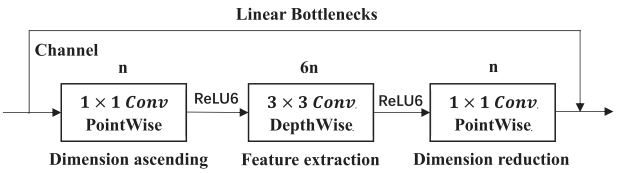

Fig. 6. Depth separable convolution with residual structure.

Table 1
Proposed lightweight network configuration.

Input	Operator	Filter	Stride	Padding	SE	ReLU
$112 \times 112 \times 3$	Conv2d, 3×3	8	2	1	–	√
$56 \times 56 \times 8$	bneck, 3×3	16	2	–	√	√
$28 \times 28 \times 8$	bneck, 3×3	8	2	–	√	√
$14 \times 14 \times 8$	pool, 14×14	–	–	–	–	–
$1 \times 1 \times 8$	Conv2d, 1×1	16	–	–	–	–
$1 \times 1 \times 16$	Conv2d, 1×1	U	–	–	–	–

3.2.2. Complete low-quality image prediction network

In this section, we introduce the proposed lightweight network in detail. The network contains a standard convolution, two depth separable convolutions with residual structures embedded in the SE block, an average pooling layer (Lin, Chen, & Yan, 2013), two fully connected layers and a cross-entropy loss layer. The detailed topology is described as follows:

L0: The input layer of the network possesses an input size defined as $[112 \times 112 \times 3]$. Please note that the normal vein image is a greyscale image, so the image must be copied as a 3-channel RGB image with the pixel value of each channel remaining the same.

L1: The first hidden layer of the network includes a standard convolution, a BatchNorm for normalization and an activation function ReLU. The convolution kernel size of the convolution layer is $[3 \times 3]$, the stride is 2, the padding is 1, and the final output size is $[56 \times 56 \times 8]$.

L2: The second hidden layer of the network includes two deep separable convolutions with residual structures embedded in the SE block, a BatchNorm layer and a ReLU layer. The strides of the two convolutions are set to 2 to control the number of parameters, and the final output size is $[14 \times 14 \times 8]$.

L3: The third hidden layer of the network consists of an average pooling layer. The fully connected layer can reduce the dimensionality of the feature map and input it into the loss function, but it may also cause overfitting. Therefore, we add an average pooling layer to regularize the structure of the entire network to prevent overfitting.

L4: The fourth hidden layer of the network consists of two fully connected layers with an output size of $[1 \times 1 \times U]$, where U represents the number of classifications. We set the image quality classification as a two-classification problem, that is, vein images are divided into high-quality images and low-quality images.

The loss function uses cross-entropy loss. Table 1 shows the detailed configuration information of our proposed lightweight network, including kernel size, stride and padding.

4. Experiments and results

In this section, for objective and accurate evaluation of the effectiveness of the proposed finger vein image quality assessment criteria and the accuracy of the low-quality image prediction methods, we design a series of experiments on three public datasets and compare the proposed criteria with the latest methods. All methods are tested on a computer with an Intel i7 8700k 3.7 GHz CPU, 64 GB RAM and an NVIDIA GTX 1080Ti GPU. We use the deep learning framework PyTorch (Paszke et al., 2019) to train a CNN on a single NVIDIA GTX 1080Ti. The experiment specifies that the basic learning rate of the network is 0.1 and the batch size is 8. The specific experimental contents are as follows. First, in 4.1, we describe the details of the three public datasets used in this paper and provide the download link of the datasets (considering that the download link of the original paper is invalid, we upload our own data to the network to facilitate researchers' access). Second, in 4.2, we discuss the effectiveness of the two strategies (proposed in 3.1.1 and 3.1.2, respectively) in the quality evaluation standard proposed in this paper. Then, we verify the applicability of the quality evaluation standard to different types of recognition methods and acquisition equipment in 4.3 and 4.4 to prove that the quality evaluation standard can be combined with the existing recognition system. Next, we explain the details of dividing and expanding the training data in 4.5, which provides preparation for low-quality image prediction using the prediction network proposed in 3.2. Finally, in 4.7, we discuss the performance changes of the recognition system after using the prediction network to screen low-quality images. A visual assessment and temporal analysis of the methods in this paper are presented in 4.8 to facilitate the reader's understanding of the proposed quality assessment method.

4.1. Dataset

SDUMLA (Yin, Liu, & Sun, 2011). The Machine Learning and Data Mining Laboratory of Shandong University released 3816 finger vein images collected from 106 volunteers in 2011. Each volunteer provided the index finger, middle finger, and ring finger of two hands, producing a total of 636 finger classes, and 6 samples were collected from each finger. The resolution of each image is 320 × 240 pixels without any preprocessing operations, including background areas. (Download Link: https://drive.google.com/drive/folders/18LDp94L9-MuWShCnGuoVG7iYK3GCITVj?usp=sharing;)

MMCBNU_6000 (Lu, Xie, Yoon, Wang, & Park, 2013). In this dataset, researchers collected 6000 finger vein images from 100 volunteers. These 100 volunteers came from 20 countries/regions. Each volunteer provided the index finger, middle finger and ring finger of two hands, generating 600 finger classes with 10 samples per finger. The resolution of each image is 640 × 480 pixels without any preprocessing operations, including the background area. (Download Link: https://drive.google.com/drive/folders/15W_x-c_Hmky1NSokbmry-MpzPJ85NdV?usp=sharing;)

FV-USM (Asaari, Suandi, & Rosdi, 2014). In this dataset, researchers collected finger vein images from 123 volunteers in two different sessions. Each volunteer provided the index finger and middle finger of two hands. There were 6 samples from each finger in a single session, which means a total of 5904 images were collected. In this experiment, we only use images from a single session, i.e., a total of 2952 images from 492 classes. All images are in grey level BMP format with resolutions of 640 × 480 pixels. Although the dataset provides samples of the region of interest (ROI) alone, considering that other datasets contain background regions, we still use samples containing background regions. (Download Link: https://drive.google.com/drive/folders/1oGsp85phS-vnc8p8C840HBZ9QIJHjWn6?usp=sharing;)

4.2. Verification of the effectiveness of two strategies for image quality assessment criteria

The LBP feature extraction method and the SDUMLA dataset are utilized to evaluate the effectiveness of the two strategies for image quality evaluation proposed in this paper. We do not perform any preprocessing operations on the vein images. The LBP algorithm is directly used to extract feature vectors, and then the feature is calculated by the Euclidean distance.

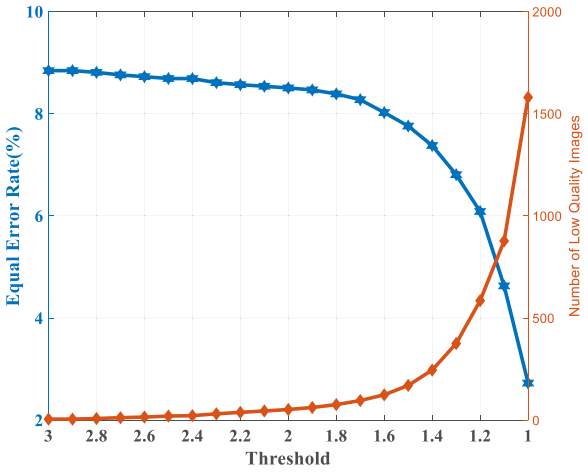

Fig. 7. The improved performance of the system after using strategy 1 to remove low-quality images.

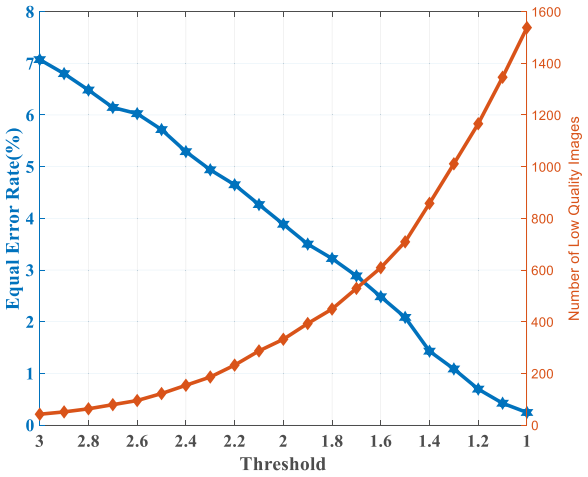

Fig. 8. The improved system performance after using strategy 2 to remove low-quality images.

4.2.1. Strategy 1

First, we verify the change in system identification performance before and after using strategy 1 to filter out low-quality images. Without any low-quality image filtering, we use the LBP algorithm to obtain $EER = 8.883\%$ for the SDUMLA dataset. Next, we set different thresholds th_1 to change the strictness of low-quality image filtering, which means that the smaller the value of th_1, the more low-quality images are filtered. Fig. 7 shows that as th_1 changes, the system screens out increasing amounts of low-quality images, and the identification performance of the system gradually improves. For example, when $th_1 = 1.3$, we remove 375 low-quality images, accounting for 10.417% (375/3600) of the entire dataset, but the identification performance of the system directly improves to 6.806%, an increase of 23.381% from the original performance. Experiments show that the filtering mechanism of strategy 1 can effectively select low-quality images in the dataset and significantly improve the performance of the system.

4.2.2. Strategy 2

Next, we verify the effectiveness of strategy 2 under the same experimental environment. We set the threshold of strategy 1 to 1.5 and verify whether the filtered low-quality images are effective by changing the threshold of strategy 2, th_2. The experimental results are shown in Fig. 8. When the threshold th_2 is gradually reduced from 3 to 1, the system continuously screens out possible low-quality images, and the identification performance of the system is improved from the initial value of 7.758% to 0.423%. When the threshold is set to 1.9, the system removes 393 low-quality images, and the identification performance is also improved to 3.501%, which is a decrease of 54.872% from the original percentage. Through the above two groups of experiments, we find that the image quality evaluation criteria proposed in this paper can effectively filter out the low-quality images of the dataset and significantly improve the identification performance of the system. Furthermore, the above experiments have also verified that the design idea driven by improving the recognition performance of the recognition system and based on the possibility of the false rejection of low-quality images is reliable when we use our design of the finger vein image quality evaluation standard. This occurs because regardless of whether it is strategy 1 or strategy 2 is used, images with low computational similarity, which are often the source of system misrecognition, are screened out as much as possible when screening images.

4.3. Verification of the universality of the image quality assessment criteria for different identification methods

With the development of finger vein biometric identification technology, an increasing number of finger vein feature extraction methods have been proposed (Kang, Lu, Li, & Jia, 2018; Yang et al., 2019; Yang, Yang, Yin, & Xi, 2014). However, most of the current finger vein image quality evaluation schemes only consider a certain identification method and possess poor universality, which greatly hinders the popularization and use of image quality evaluation schemes in actual finger vein identification systems. We can divide the finger vein feature extraction methods into three categories: traditional image processing methods, vein feature processing methods and CNN processing methods. Next, we select two classical algorithms from each types of feature extraction methods to verify the effectiveness of the image quality evaluation criteria proposed in this paper for different types of feature extraction methods. Among them, LBP and HOG are selected as the traditional image processing methods, mean curvature (MC) (Song et al., 2011) and wide line detector (WLD) (Huang, Dai, Li, Tang, & Li, 2010) are selected as the vein feature processing methods, SENet (Hu et al., 2018) and Inception V3 (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016) are selected as the CNN processing methods, and the SDUMLA dataset is selected in all cases.

Fig. 9 shows the changes in the system identification performance based on the six feature extraction methods. Each subfigure contains the identification results without low-quality image filtering and 9 types of identification results after removing low-quality images through different filtering thresholds. By observing the experimental results, we can find that the performance of the recognition system without any low-quality image screening is the worst. The recognition performance of the system gradually improves after continuously increasing the strictness of the screening, that is, continuously screening out images that are not conducive to the recognition results. Regardless of whether the recognition method used is the traditional image processing method or deep learning method, the recognition performance of the system can be improved after screening some low-quality images. All these cases show that the finger vein image quality evaluation criteria proposed in this paper can be applied to different types of feature extraction methods and can significantly improve the identification performance of the system when a small portion of the low-quality images is filtered out. For example, in the WLD algorithm, when the thresholds are set to $th_1 = 0.7$, $th_2 = 0.6$, the system screens 339 low-quality images (accounting for 9.417% of the dataset), and the identification performance improves from 11.838% to 6.834%, a decrease of 42.271%.

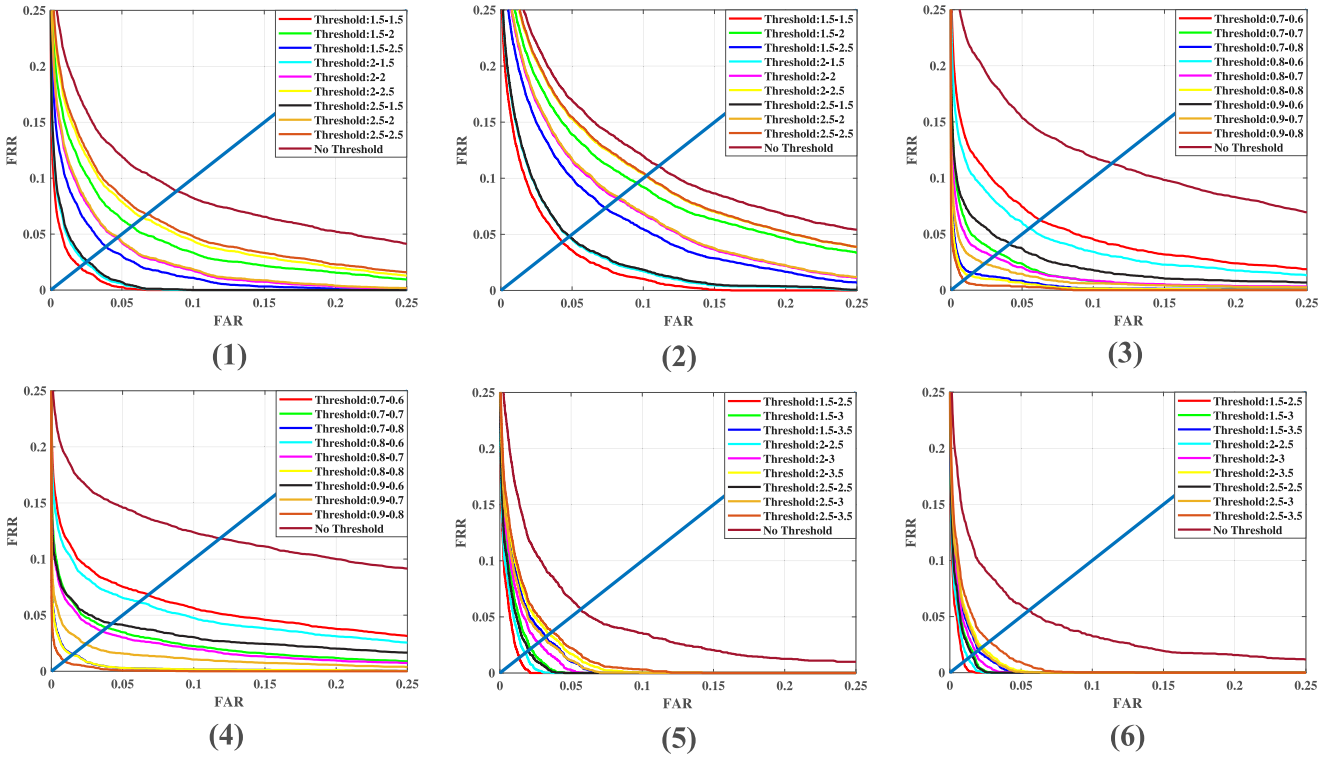

Fig. 9. ROC curves for different datasets using different identification methods. (1) and (2) are LBP and HOG (traditional image processing methods); (3) and (4) are MC and WLD (vein feature processing methods); (3) and (4) are SENet and Inception V3 (CNN processing methods).

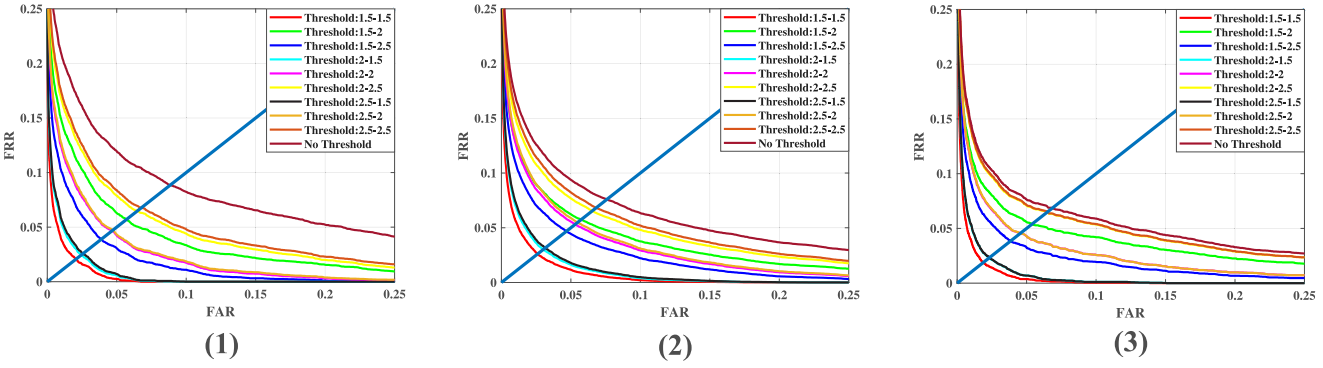

Fig. 10. ROC curve of the LBP algorithm under different datasets. (1) SDUMLA; (2) MMCBNU_6000; (3) FV-USM.

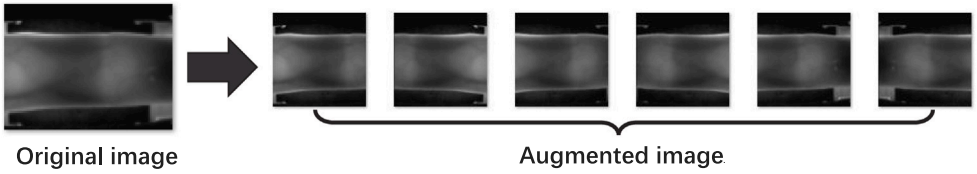

Fig. 11. Schematic diagram of low-quality image data amplification.

Table 2
Number of high-quality and low-quality images after dividing training set and test set for the three public datasets.

Database	Threshold	CIR (%)	Training set	Training set(quality)		Test set	Test set(quality)	
				High	Low		High	Low
SDUMLA	1.5–1.8	3.2225(63.72%)	1908(319–636)	1468	440	1908(1–318)	1729	179
MMCBNU_6000	2.0–1.5	3.0157(60.27%)	3000(301–600)	2584	416	3000(1–300)	2722	278
FV-USM	1.5–1.7	2.8993(57.37%)	1476(247–492)	1343	133	1476(1–246)	1352	124

Table 3
Performance comparison of different methods.

Methods	SDUMLA	MMCBNU_6000	FV-USM
Manual features + SVM (Zhou et al., 2015)	73.72%	64.35%	65.23%
Extended manual features + SVR (Xie et al., 2013)	67.34%	64.76%	–
Radon transform (Qin, Li, Kot, & Qin, 2012)	69.18%	64.57%	64.82%
Hand-crafted features + SVM regression (Yang, Yang, Yin, & Xiao, 2013)	–	–	64.41%
Multi-Scale HOG + light-CNN (Zeng et al., 2019)	74.63%	71.95%	–
Qin(Patch-DNN + P-SVM)(Qin & El-Yacoubi, 2017)	78.15%	72.29%	71.14%
HCGR + Improved SMOTE + CNN (Wang & Fang, 2020)	–	75.79%	–
Our	81.97%	75.27%	73.85%

4.4. Validating the universality of the image quality assessment criteria for different datasets

Due to differences in design principles, hardware selections, working environments and other conditions of finger vein acquisition equipment, collected finger vein images possess great differences (Chen et al., 2017). Assuming that only a certain dataset is considered in the design of finger vein image quality evaluation criteria, the proposed image quality evaluation criteria is useless when the acquisition equipment is replaced. Therefore, when designing the finger vein image quality evaluation criteria, the universality of the criteria for different datasets must be considered; that is, when the identification system is replaced, only a simple operation is required to migrate the original image quality evaluation criteria to the new identification system. In this section, we use the fixed feature extraction method to perform the same experiment on three public datasets to verify the universality of the image quality evaluation criteria proposed in this paper for different datasets.

Fig. 10 shows our experimental results on three public datasets using the LBP method. Similar to Fig. 9, these results include both 1 recognition result without low-quality image screening and 9 recognition results after eliminating low-quality images through different screening thresholds. Fig. 10 summarizes the experimental results and clearly shows that under different datasets, the image quality assessment criteria in this paper can efficiently and accurately screen out low-quality images without additional labour costs. After removing this portion of the low-quality images, the recognition performance of the system is also significantly improved. Through the experiments in Sections 4.3 and 4.4, we find that our proposed finger vein image quality evaluation criteria are not limited by the feature extraction method or the acquisition equipment, and they can be easily combined with the existing finger vein recognition system to improve the performance of the system.

4.5. Production of high-quality and low-quality datasets

The performance of an image processing method based on deep learning depends on the number of training datasets and the degree of discrepancy. An efficient network also requires a large number of discrepant training sets to mine and learn hidden difference attributes (Sun, Shrivastava, Singh, & Gupta, 2017; Wang & Perez, 2017). Therefore, in this section, we elaborate on the details of dataset partitioning and low-quality image amplification. According to the automatic marking method mentioned earlier, we set the thresholds for the SDUMLA dataset to $th_1 = 1.5$ and $th_2 = 1.8$, the thresholds of the MMCBNU_6000 dataset to $th_1 = 2.0$, $th_2 = 1.5$, and the thresholds of the FV-USM dataset to $th_1 = 1.5$, $th_2 = 1.7$. The basis for selecting the above three groups of thresholds is to ensure that the performance improvement of the three groups of datasets is equivalent after screening some low-quality images. That is, after some low-quality images are screened, the recognition performance of the three groups of datasets is improved by approximately 60% compared to the original recognition performance (specifically, the performance on the SDUMLA dataset increased by 63.72%, the performance on the MMCBNU_6000 dataset increased by 60.27%, and the performance on the FV-USM dataset increased by 57.37%). According to the thresholds set above, the divisions of high-quality and low-quality images for the three public datasets are shown in Table 2. The three public datasets are divided equally into two parts, with the first half used as a test set and the second half as a training set. Among them, the training set and test set of the SDUMLA dataset contain 440 and 179 low-quality images, respectively, the training set and test set of the MMCBNU_6000 dataset contain 416 and 278 low-quality images, respectively, and the training set and test set of the FV-USM dataset contain 133 and 124 low-quality images, respectively.

From Table 2, it can be found that the number of low-quality images in each dataset is relatively small compared to the number of high-quality images, which causes trouble for the learning process of the network, so we expand the low-quality images. To ensure the effectiveness of the features extracted by the network from low-quality images, we amplify low-quality images and try our best to ensure that the original attributes of images are maintained. We do not use artificially corrupted high-quality images to act as low-quality images, nor do we use amplification methods that may cause interference to images, because the low-quality features forged by these operations do not necessarily meet the actual use requirements. After analysis, we only use cropping and mirroring to amplify low-quality images. As shown in Fig. 11, we first use cropping to manipulate the image and amplify the original Fig. 3 times (there is much overlap between the cropped block and the block), and then we use mirroring to manipulate the cropped block and amplify the image 2 times more. To facilitate training, we also randomly delete some high-quality images in the training set and downsample them to ensure the consistency of the number of high-quality images and low-quality images in the training set. Ultimately, the training sets for these three public datasets are as follows: the SDUMLA dataset with 5280 images, MMCBNU_6000 dataset with 4992 images, and FV-USM dataset with 1596 images.

4.6. Performance of different methods on test sets

In this section, we evaluate the performance of the proposed finger vein image quality prediction model on three publicly available datasets. Comparisons are made not only with some traditional methods, such as Radon transform (Qin et al., 2012) and SVM (Zhou et al., 2015), but also with some of the finger vein image quality prediction algorithms based on CNN proposed in recent years. In the experiment, we performed image quality prediction on six subblocks of each image. If three subblocks were low-quality images, the image was identified as a low-quality image, and the classification accuracy of high-quality and low-quality images was used as an indicator of prediction performance. The experimental results are shown in 3. The comparison shows that our proposed lightweight vein image quality prediction method based on a CNN has achieved excellent performance on all three public datasets. The method performs only 0.52% worse than Wang and Fang (2020) on the MMCBNU_6000 dataset, but Wang and Fang (2020) did not specify the specific details of the CNN, and the method could only be applied to specific recognition algorithms. At the same time, we also observed an interesting phenomenon where the prediction accuracy of the SDUMLA dataset exceeded 80%, much higher than those of the other two datasets. Through analysis, we find that the size of the

Table 4
Changes in system performance after quality assessment (EER).

Status	SDUMLA	MMCBNU_6000	FV-USM
Original	5.50%	7.93%	7.51%
After quality assessment	4.80%	6.98%	6.76%

training dataset plays a significant role in the prediction accuracy of the network. The SDUMLA dataset contains a relatively high number of low-quality images, the highest of the three datasets at 16.221%, while the MMCBNU_6000 dataset and FV-USM dataset contain only 11.567% and 8.706% low-quality images, respectively. Combined with the automatic labelling image quality method we previously proposed, we observe that the higher the threshold setting requirements are, the more low-quality images that will be screened, the greater the amount of training data, and the more beneficial the method is to predicting the performance of the network. However, the setting of the threshold should be considered in conjunction with the actual application, and different application scenarios will have different considerations for low-quality images. If the performance of the recognition system is poor, the recognition system depends more on the quality prediction network, and then the high-performance prediction network trained by the screened low-quality images can effectively improve the overall performance of the recognition system. If the performance of the recognition system is high, fewer images may lead to low performance of the prediction network, but the recognition system also depends less on the prediction network. Therefore, our quality evaluation standards and prediction network are in line with practical application scenarios.

4.7. Before-and-after comparison of system performance after filtering out low-quality images using prediction algorithms

Finally, we analyse the changes in the performance of the system after filtering out possible low-quality images using prediction algorithms. When the prediction network filters low-quality images, it also mistakes high-quality images for low-quality images. If too many high-quality images are mistaken for low-quality images, it is not only difficult to improve the recognition performance of the finger vein recognition system but also to produce negative optimization of the system. Therefore, we want to verify the change in system performance after removing some vein images that are filtered out by the model. To the best of our knowledge, many finger vein image quality assessment methods have neglected this work. In the experiment, we used the LBP algorithm as a feature extraction method to calculate the recognition performance of three publicly available datasets before and after image quality assessment. Table 4 shows the recognition performance of the three datasets before and after eliminating the predicted images. It can be found that the image quality prediction method in this paper has improved the recognition performance on all three datasets by 12.670%, 11.923% and 9.940%, respectively, relative to the original recognition performance value.

4.8. Visual assessment and time analysis

In this section, we visually evaluate the feature information of high-quality images and low-quality images to obtain a better understanding of the proposed quality evaluation method. Fig. 12 shows our proposed quality evaluation standard used to automatically mark high-quality images and low-quality images (the feature extraction method used is LBP). Specifically, (1), (3), and (5) of Fig. 12 are high-quality images; and (2), (4), and (6) of Fig. 12 are low-quality images. (1) and (2) are the same fingers in the MMCBNU_6000 dataset, (3) and (4) are the same fingers in the SDUMLA dataset, (5) and (6) are the same fingers in the FV-USM dataset. By comparing (1) and (2), we can observe that image (2) is determined to be a low-quality image because of excessive external noise during the acquisition process. Comparing (3) and (4), image (4) is due to finger rotation at a certain angle and serious deformation, resulting in the destruction of vein information. The screening of images (2) and (4) is also in line with our cognition of low-quality images. However, we found an interesting phenomenon when comparing (5) and (6). Compared with the image with external noise (5), the image without external noise (6) is recognized as a low-quality image. In order to explore the reason, we compare the other samples of the finger and find that among the six samples collected for this finger, only image (6) has no external noise. Therefore, taking image (6) as a low-quality image can effectively reduce the error, and this operation is also in line with the actual situation. Because there will always be external noise when collecting finger vein images due to finger placement habits, this external noise can also be understood as a part of their characteristic information. Through the above analysis, we believe that the quality evaluation standard based on the possible false rejection of low-quality images proposed in this paper can handle various types of noise interference, automatically adjust the provisions on low-quality features in combination with the actual situation and effectively screen the low-quality images that are not conducive to the recognition system, which is more in line with the application requirements of real scenes.

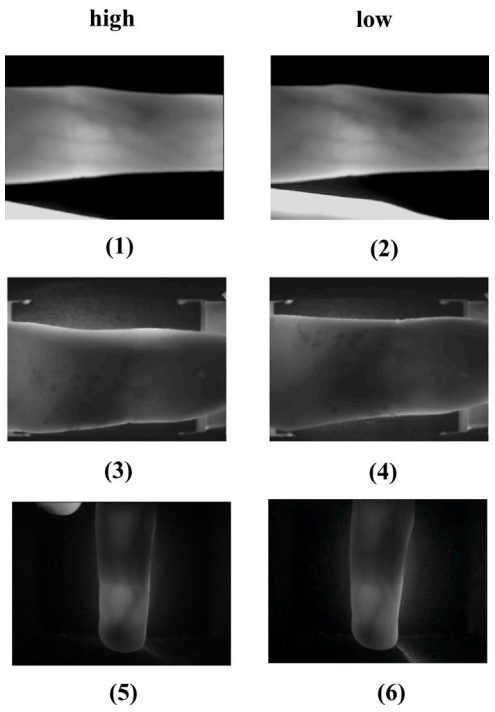

Fig. 12. Visual assessment of quality assessment methods. Among them, (1)(3)(5) are high-quality images, and (2)(4)(6) are low-quality images. (1) (2) is the same finger in the MMCBNU_6000 dataset; (3) (4) is the same finger in the SDUMLA dataset; (5) (6) is the same finger in the FV-USM dataset.

Finally, we consider the impact of the training time and actual running time of the prediction network proposed in this paper on the recognition system. The response time of the vein recognition system is the time required for the user to put his finger into the vein collection device until the system gives the judgement result. The response time is an important indicator to evaluate a vein recognition system. Here, we discuss the time loss of the recognition system due to the extra time generated after incorporating the prediction network proposed in this paper. Fig. 13 shows the training situation of our prediction network on three public datasets. As the number of training sessions increases, the loss value of the network steadily decreases. The three datasets reach the inflection point when the number of training sessions reach 1000, and the network obtains the optimal result. In addition,

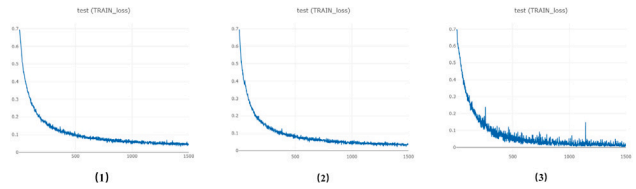

Fig. 13. The training process of the prediction network on three publicly available datasets, the horizontal coordinate is the number of training times and the vertical coordinate is the loss value of the network. Where (1) is the SDUMLA dataset; (2) is the MMCBNU_6000 dataset; and (3) is the FV-USM dataset.

because the number of samples and images of different datasets are different, the time for a single training is also different. Among them, the SDUMLA dataset takes 1699.70 s for 100 training sessions, the MMCBNU_6000 dataset takes 2349.31 s for 100 training sessions, and the FV-SUM dataset takes 1034.10 s for 100 training sessions. That is, in the new application scenario of our prediction network, the highest time consumption is 6.525 h, and the shortest time consumption is only 2.8725 h. Furthermore, we will predict the test set after each training session. If the test set is predicted after multiple control training, the training time of the network will be further controlled.

In terms of predicting the actual time consumption of the network in the operating process, because the lightweight design consideration is adopted in the design of the network, the running time generated by our network is also very low. We counted the running time of 100 finger vein images actually judged by the prediction network. The average time required for three datasets to run a single image is as follows: the SDUMLA dataset requires 9.09 ms, the MMCBNU_6000 dataset requires 9.20 ms, and the FV-USM dataset requires 9.17 ms. We have tested several common recognition systems. Except for the quality evaluation link, the time is approximately 60–80 ms (excluding the time consumed by image acquisition). Our quality evaluation method only produces a less than 15% time increase, which fully meets the requirements for the time overhead of finger vein recognition systems in practical applications.

5. Conclusion and future work

In this paper, we proposed finger vein image quality assessment criteria that can automatically mark low-quality images and develop a complete CNN -based low-quality image prediction method. The prediction method was developed to provide efficient and accurate image quality judgments for existing finger vein identification systems, ensuring that the system does not misidentify some images due to abnormalities. We carried out a series of detailed experiments on three public datasets, which proved that the finger vein image quality assessment criteria in this paper can be perfectly compatible with different recognition methods and acquisition equipment. In addition, the low-quality image prediction method proposed in this paper was able to identify high-quality and low-quality vein images with high accuracy and significantly improve the recognition rate of the system. Finally, our solution is simple and generic. It can be applied to any biometrics that possess similar representation, such as irises, fingerprints, and faces. It is worth noting that the existing quality evaluation of finger vein images can only distinguish high-quality images from low-quality images to a certain extent. It cannot conduct a more detailed classification of low-quality images. For example, fingers may be placed irregularly, there may be foreign objects in the fingers, and the camera may be blurred, which is also a limitation of our work. The cause of low-quality images and reasonable guidance for users will be topics for further study in the next stage of our work.

CRediT authorship contribution statement

Hengyi Ren: Conceptualization, Methodology, Writing – original draft. **Lijuan Sun:** Data curation, Investigation, Supervision. **Jian Guo:** Software. **Chong Han:** Validation. **Ying Cao:** Writing – reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of China under Grant No. 61873131 and 61702284, the Anhui Science and Technology Department Foundation under Grant 1908085MF207, the Postdoctoral Found of Jiangsu Province under Grant No. 2018K009B and the Innovation Project for Postgraduate of Jiangsu Province under Grant No. KYCX19_0915.

References

Altan, A., & Karasu, S. (2020). Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons & Fractals*, *140*, Article 110071.

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons & Fractals*, *126*, 325–336.

Altan, A., Karasu, S., & Zio, E. (2021). A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Applied Soft Computing*, *100*, Article 106996.

Asaari, M. S. M., Suandi, S. A., & Rosdi, B. A. (2014). Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. *Expert Systems with Applications*, *41*(7), 3367–3382.

Chen, L., Wang, J., Yang, S., & He, H. (2017). A finger vein image-based personal identification system with self-adaptive illuminance control. *IEEE Transactions On Instrumentation And Measurement*, *66*(2), 294–304.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Vol. 1* (pp. 886–893). IEEE.

Esen, H., Esen, M., & Ozsolak, O. (2017). Modelling and experimental performance analysis of solar-assisted ground source heat pump system. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(1), 1–17.

Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008a). Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system. *Energy and Buildings*, *40*(6), 1074–1083.

Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008b). Forecasting of a ground-coupled heat pump performance using neural networks with statistical data weighting pre-processing. *International Journal Of Thermal Sciences*, *47*(4), 431–441.

Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008c). Modeling a ground-coupled heat pump system by a support vector machine. *Renewable Energy*, *33*(8), 1814–1823.

Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008d). Performance prediction of a ground-coupled heat pump system using artificial neural networks. *Expert Systems with Applications*, *35*(4), 1940–1948.

Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009a). Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. *Expert Systems with Applications*, *36*(8), 11240–11248.

Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009b). Modelling of a new solar air heater through least-squares support vector machines. *Expert Systems with Applications*, *36*(7), 10673–10682.

Grother, P., & Tabassi, E. (2007). Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 531–543.

Guo, J.-M., Hsia, C.-H., Wu, C.-S., & Chang, L.-Y. (2020). Efficient finger vein technology based on fast binary robust independent elementary feature combined with multi-image quality assessment verification. In *2020 IEEE international conference on consumer electronics-taiwan (ICCE-Taiwan)* (pp. 1–2). IEEE.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, B., Dai, Y., Li, R., Tang, D., & Li, W. (2010). Finger-vein authentication based on wide line detector and pattern normalization. In *2010 20th International conference on pattern recognition* (pp. 1269–1272). IEEE.

Huang, Z., Mu, C., & Xie, J. (2013). Vein image quality evaluation method for feature extraction. *Computer Engineering and Science*, *35*(10), 186–190.

Kang, W., Lu, Y., Li, D., & Jia, W. (2018). From noise to feature: Exploiting intensity distribution as a novel soft biometric trait for finger vein recognition. *IEEE Transactions on Information Forensics and Security*, *14*(4), 858–869.

Karasu, S., Altan, A., Bekiros, S., & Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy*, *212*, Article 118750.

Lee, Y., Micheals, R. J., & Phillips, P. J. (2009). Improvements in video-based automated system for iris recognition (vasir). In *2009 Workshop on motion and video computing (WMVC)* (pp. 1–8). IEEE.

Lee, Y., Phillips, P. J., & Micheals, R. J. (2009). An automated video-based system for iris recognition. In *International conference on biometrics* (pp. 1160–1169). Springer.

Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv: 1312.4400.

Lu, Y., Xie, S. J., Yoon, S., Wang, Z., & Park, D. S. (2013). An available database for the research of finger vein recognition. In *2013 6th International congress on image and signal processing (CISP), Vol. 1* (pp. 410–415). IEEE.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems* (pp. 8024–8035).

Peng, J., Li, Q., & Niu, X. (2014). A novel finger vein image quality evaluation method based on triangular norm. In *2014 Tenth international conference on intelligent information hiding and multimedia signal processing* (pp. 239–242). IEEE.

Qin, H., & El-Yacoubi, M. A. (2017). Deep representation for finger-vein image-quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(8), 1677–1693.

Qin, H., Li, S., Kot, A. C., & Qin, L. (2012). Quality assessment of finger-vein image. In *Proceedings of the 2012 asia pacific signal and information processing association annual summit and conference* (pp. 1–4). IEEE.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510–4520).

Song, W., Kim, T., Kim, H. C., Choi, J. H., Kong, H.-J., & Lee, S.-R. (2011). A finger-vein verification system using mean curvature. *Pattern Recognition Letters*, *32*(11), 1541–1547.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision* (pp. 843–852).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Tabassi, E. (2014). Towards NFIQ II lite: Self-organizing maps for fingerprint image quality assessment.

Tabassi, E., & Wilson, C. L. (2005). A novel approach to fingerprint image quality. In *IEEE International conference on image processing 2005, Vol. 2* (pp. II–37). IEEE.

Uhl, A., Busch, C., Marcel, S., & Veldhuis, R. (2020). *Handbook of vascular biometrics*. Springer Nature.

Wang, Y., & Fang, P. (2020). 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). IEEE.

Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks for Visual Recognition*, 11.

Xie, S. J., Zhou, B., Yang, J., Lu, Y., & Pan, Y. (2013). Novel hierarchical structure based finger vein image quality assessment. In *Chinese conference on biometric recognition* (pp. 266–273). Springer.

Yang, W., Wang, S., Hu, J., Zheng, G., Yang, J., & Valli, C. (2019). Securing deep learning based edge finger vein biometrics with binary decision diagram. *IEEE Transactions on Industrial Informatics*, *15*(7), 4244–4253.

Yang, L., Yang, G., Yin, Y., & Xi, X. (2014). Exploring soft biometric trait with finger vein recognition. *Neurocomputing*, *135*, 218–228.

Yang, L., Yang, G., Yin, Y., & Xiao, R. (2013). Finger vein image quality evaluation using support vector machines. *Optimization and Engineering*, *52*(2), Article 027003.

Yin, Y., Liu, L., & Sun, X. (2011). SDUMLA-HMT: a multimodal biometric database. In *Chinese conference on biometric recognition* (pp. 260–268). Springer.

Zeng, J., Chen, Y., Zhai, Y., Gan, J., Feng, W., & Wang, F. (2019). A novel finger-vein recognition based on quality assessment and multi-scale histogram of oriented gradients feature. *International Journal of Enterprise Information Systems (IJEIS)*, *15*(1), 100–115.

Zhou, L., Yang, G., Yang, L., Yin, Y., & Li, Y. (2015). Finger vein image quality evaluation based on support vector regression. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *8*(8), 211–222.

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697–8710).