Fingerprint Liveness Detection via Learning Multi-Modal Deep Features

Chengsheng Yuan\textsuperscript{1,2}, Mingyu Chen\textsuperscript{1,2}

\textsuperscript{1}School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044
\textsuperscript{2}Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044

Abstract
With the widespread application of fingerprint identification systems, fraudulent attacks based on forged fingerprints have gradually increased, so it is very important to distinguish the authenticity of fingerprints. Fingerprint liveness detection technology was proposed to solve this problem. In order to effectively integrate advantages of existing algorithms, this paper proposes an adaptive feature optimization module filtering distinctive multi-modal features. Firstly, we extract the ROI of fingerprint images and unify them to the same size as subsequent input. Secondly, three convolutional neural networks (CNN)-AlexNet, VGG16 and ResNet, are trained through processed images, whose the last fully connected layer as fingerprint feature. Then genetic algorithm is used to assign different weights to extracted features through these networks, which retain distinctive parts and eliminate invalid parts. Finally, considering that the features are extracted from CNN, optimized features are input to the fully connected layer, and then fake fingerprints are identified by softmax function. Experiments show that when feature dimensions of three networks output are 512, feature optimization module proposed can improve the detection accuracy by an average of 1.0% in the 2011 Livdet database, which finds out the more different parts of features extracted by the multi-modal network, enhancing fingerprint liveness detection performance.

Keywords
Fingerprint Liveness Detection, CNN, Genetic Algorithm, Multi-Modal

1. Introduction
In a highly information-based modern society, people often need to use passwords for identity authentication to obtain access to various accounts. Therefore, the password has become direct target of numerous hackers [1]. News about economic losses of users caused by theft of various passwords is endless, which means this traditional identity authentication scheme has serious security risks. In order to solve this problem, many new methods have been proposed and adopted, such as USB KEY, SMS password, dynamic password, etc. Among them, the authentication method based on biometrics has been favored by people. Compared with other identity authentication methods, this method is simple, fast and reliable, so it has been used widely in all aspects of people’s social life. For instance, lots of business units check attendance of employees through fingerprint, customs in many countries will utilize fingerprints to authenticate immigrants, and smartphones use fingerprints to authenticate their identity or achieve quick payment [2, 3, 4, 5].

However, fingerprint authentication also has certain security risks [6]. For the one hand, human fingerprints are easily stolen and counterfeiters can imitate user fingerprints to achieve illegal authentication; For the other hand, people also can copy fake fingerprints for themselves to deceive the attendance system, as shown in Fig.1. These fake fingerprints can be made of silica gel, gelatin, EcoFlex, Modasil and other materials. In addition, with the development of deep learning technology, generative adversarial network can also generate sufficiently realistic fingerprint images. Forged fingerprint attacks are one of the biggest threats to fingerprint identification systems, which greatly reduce the reliability of the system and put users’ private information at risk. Hence, as an important auxiliary algorithm of finger-
print identification system to identify authentic fingerprints, fingerprint liveness detection has become an academic research hotspot now.

2. Related Work

Nowadays, fingerprint liveness detection (FLD) algorithms are mainly divided into two categories: hardware-based FLD and software-based FLD [7]. The former uses additional professional equipment to identify the authenticity of fingerprint images by measuring skin temperature, conductivity, blood pressure, blood oxygen and other vital signs. Although this method can achieve great detection accuracy, expensive equipment is easy for illegal users to find loopholes because of single identification method. With image processing technology, the latter analyze the difference between real and fake fingerprint images for identification. Compared with the former, the latter is more flexible, save costs, simplify operations and minimize additional equipment [8, 9, 10]. Therefore, this method is also the current research focus of fingerprint liveness detection. Existing software-based algorithms can be divided into three categories: traditional FLD, FLD based on texture features and FLD based on deep learning.

2.1. Traditional FLD

The traditional fingerprint liveness detection algorithm designs an descriptor to extract distinctive features between real and fake fingerprint images through the heuristic algorithm. FLD based on sweat hole is the earliest proposed fingerprint liveness detection algorithm. While recognizing high-resolution fingerprint images, the quality of artificial fingerprint images is often worse than that of the real. On account of rougher surface, detail of fake fingerprints is much weaker than the real. In consequence, Moon [11] et al proposed an idea of FLD based on image quality, which denoises and reconstructs fingerprint images by wavelet. The noise residuals between reconstructed and original images are calculated to authenticate fingerprints. When the finger presses and rotates on the sensor, real fingerprints can produce better elastic deformation than the fake. In consequence, Antonelli [12] et al proposed FLD based on skin elastic deformation for the first time. In the process of imitating fingerprints, sweat holes on the ridge of the finger’s epidermis are difficult to replicate. Accordingly, FLD based on sweat holes is proposed. Manivanan [13] et al used a high-pass filter to extract effective sweat hole features and a correlation filter to locate sweat holes.

2.2. FLD based on texture features

Although fingerprint texture cannot be distinguished by human eyes, it is a common feature in fingerprint images and represented by the gray distribution of center pixels and neighboring pixels. Common texture feature descriptors include Local Binary Pattern (LBP) [14], Binary Statistical Image Feature (BSIF) [15], Local Phase Quantization (LPQ) [16], Histogram of Gradient Direction (HOG) [17], etc. Jhat [18] et al proposed FLD based on gray-level independence to verify and distinguish the liveness of fingerprints. Yuan [19] et al calculated parameters of the co-occurrence matrix as features of the fingerprint image to authenticate fingerprints.

2.3. FLD based on deep learning

According to different classification tasks, deep neural networks can complete complex mapping and feature extraction through self-learning, which is simpler. Nogueira [20] et al introduced Convolutional Neural Network (CNN) technology to fingerprint liveness detection. They designed a random model based on CNN as the feature extractor, took the preprocessed images as the input and obtained the best detection results at the time. While achieving better accuracy, CNN models also have some shortcomings. For example, a fixed-scale input image must be used in the input layer. Although cropping or scaling can solve the scale problem well, they can easily lead to the loss of some key texture information and reduction of image resolution, thereby weakening the generalization performance. To solve this problem, Yuan [21] et al proposed a scale-equalized deep convolutional neural network (DCNNISE) model utilizing the retained subtle texture information to further improve the detection performance of forged fingerprints. Moreover, the confusion matrix was applied to FLD as performance indicator in the performance evaluation for the first time. Zhang [22] et al further found that CNN model used for multi-classification of natural images cannot obtain good accuracy in fingerprint liveness detection, because it ignored the difference between natural and fingerprint images and the shallower network structure cannot mine deep features of fingerprints. Therefore, they proposed a lightweight but powerful network structure Slim-ResCNN. At present, Agarwal S [23] et al found that existing FLD algorithms performed well when test dataset and train dataset sample distribution are the same, but result of cross-sensor fingerprint liveness detection is bad. In order to enhance the generalization, robustness and operability of FLD algorithm, they believed that the learning model need be adaptive to the data and proposed a general EaZy model. This adaptability in the context of cross-sensor datasets embodies significant advantages.
In this paper, the main contributions are summarized as follows:

(1) **Multimodal fingerprint feature extraction** According to different classification tasks, CNNs can complete complex mapping through self-learning and extract high-level features of image. However, due to various depth and architecture of CNNs, the characteristics of real and fake fingerprints extracted by different network models are quite different, making the classifier have stronger performance. In order to integrate the excellent characteristics of multiple CNNs, this paper attempts to use multimodal neural network models to extract various fingerprint features to make the difference between real and fake fingerprints more obvious.

(2) **Genetic optimization module** Full concatenation of multi-dimensional features has big defects. In addition, the feature extraction method based on CNN is similar to the black box operation, which means extracted fingerprint features are not known, making it impossible to determine the optimization direction of features. In consequence, genetic algorithm is innovatively utilized to optimize extracted fingerprint feature, which automatically select the obvious distinguishing part, so as to solve the problem of unknown features. By imitating the genetic processes of life, such as crossover, mutation, and selection, the optimal real and fake fingerprint features are found in the fused feature space. Based on the trained CNN feature extractors, initialized feature populations are put into the fully connected layer to calculate the fitness. Through mutation and crossover operators, excellent performance genetic information from parents is inherited and new genes are generated, promoting feature evolution.

(3) **Broad adaptability** This paper carried out model training and testing on 8 Livedet Datasets (2011, 2013) [24, 25]. In order to improve the stability and generalization of the trained model, some operations were used to expand the training set, including rotation, brightening and flipping. The experimental results show that the accuracy of fingerprint liveness detection on multiple data sets is significantly improved in the highest accuracy of the modal network, which proves the effectiveness and wide adaptability of our model.

### 3. Methodology

#### 3.1. Multimodal Deep Feature Learning

Without expert knowledge, deep neural networks have ability to automatically learn pixel distribution of image. By narrowing the gap between model output and label, model parameters are updated until they converge to a certain limit, thereby completing the complex mapping from two-dimensional images to one-dimensional features. The high-level features extracted by the deep neural network are excellent discriminativness, showing amazing performance in image recognition and classification. In the field of fingerprint liveness detection, CNN models can also achieve good identification results and become a research hotspot in this area, such as AlexNet, VGG16, ResNet, etc. These networks have many differences in the depth level of the model or the width level of the convolutional layer, and the learned fingerprint features will also be multifarious. In order to make full use of the merits of various networks, this paper concatenates features of multiple convolutional networks as general features of model reducing difficulty of forged fingerprint detection.

![A flow diagram of FLD via learning multi-modal deep features](image)

In the field of FLD, some scholars believe that to alleviate the shortage of target samples, transfer learning needs to be applied to the model. [26] They use 1.2 million ImageNet images (source task) to pre-train the convolutional network, and stochastic gradient descent is used to optimize error losses. After training, all trained parameters of the convolutional layer in the source task are transferred to FLD. However, we think that this approach still loses part of the characteristics of the fingerprint image, so the pre-training step should be abandoned. We select three classic CNNs—AlexNet, VGG16 and ResNet, as feature extractors, hoping to get features with different concerns. After training the CNN, freeze network parameters and take out the penultimate fully
connected layer as extracted features. Since the different dimensions of this layer will affect the output, fewer features are difficult to support the classification of true and false fingerprints. In addition, more features can also make model training difficult and key features cannot be extracted. Therefore, we selected multiple dimensions for experimentation, such as 256, 512 and 1024.

3.2. Genetic Algorithm

Because excessive features will cause dimensional disasters, it is obviously not advisable to perform fully connected operations for multiple features [27]. If only some of the features are selected to construct the model, the difficulty of the learning task can be reduced and the interpretability of the model can also be increased. However, how to filter and process these characteristics is a big problem. Traditional feature selection algorithms are proposed by researchers based on the analysis of feature defects. For neural networks, unascertainty of feature extraction process masks the source of the feature. Genetic algorithms can adaptively find better solutions from the feature space, without other selection algorithms. Better features are highlighted after a series of biologically inspired operations, such as crossover, mutation, and selection.

The genetic algorithm mainly includes five steps, as shown in fig.3. Firstly, with value of 0 or 1, N chromosomes are randomly generated as the initial population, which is the same as feature length. 0 means discarding the corresponding location feature value, and 1 means selecting. Secondly, the single-layer fully connected layer as a classifier, the characteristics of the fingerprint image at the corresponding position of the chromosome are separated for train and evaluation. Record the classification accuracy of the test set as fitness. Thirdly, some are randomly generated as the initial population is reinitialized or reversed randomly. For experimentation, such as 256, 512 and 1024.

Algorithm 1 Multimodal feature weight learning

Input: The dataset of fingerprint \(D\); The size of initial population \(N\); The fitness penalty value \(M\); The maximum number of iterations \(T\).

Output: In \(T\) generation, the feature chromosome with the greatest fitness;

1. Population Initializlization: Initialize \(N\) chromosomes with a length of \(S\), the value of which is randomly generated 0 or 1;
2. for \(i = 1, \ldots, T\) do
3. Mutation: According to the evolutionary probability, a gene of length \(L\) in chromosome \(s\)omes is reinitialized or reversed randomly.
4. \(\) Crossover: Two parents are selected from the population for single-point crossover to produce new offspring.
5. Fitness: Extract the fingerprint feature of testset at the corresponding position in the chromosome, whose accuracy is used as the fitness.
6. Selection: The original fitness minus penalty value \(M\) is used as the new fitness. The roulette algorithm generates survival prob -ility value
7. -ability of individual, which means poorly adapted are more likely to be eliminated.
8. end for

The way where the proportion of the fitness of individual chromosomes in the population is as the evolution probability is called the roulette wheel algorithm. Due to the high accuracy of the sub-network in the model, the chromosomes with higher fitness cannot obtain an advantage under the roulette wheel algorithm. So we take the lowest precision in the sub-network as a penalty value, and highlight the gap between chromosomes by introducing a fitness penalty value. Through roulette wheel selection, chromosomes with high fitness are more likely to be retained as parental chromosomes, whose structural information is passed on to the offspring. Fourthly, on the basis of inheriting parental chromosomes, offspring chromosomes will have mutations in a certain length of gene encoding. Some randomly generated or inverted gene codes replace the original part with a certain probability, which means that certain features are reselected. The mutation of gene coding provides the possibility of population evolution. Finally, the roulette wheel selection method is utilized to select parents with greater fitness for crossover from population. Cut off the two parental chromosomes (divided into upper and lower parts) at the same position, and exchange the upper parts to form two brand-new chromosomes. Each chromosome inherits powerful genes from both parents. After the gene mutation and crossover of the \(i^{th}\) generation, many offspring will be produced and
combined with the parental chromosomes to form a new population. The $k$ chromosomes with the highest fitness are selected as the $(t+1)^{th}$ generation population and continue to evolve until the set generation threshold is reached. Choose the best-performing individual as the optimal solution.

4. Experiments

4.1. Dataset

In this paper, LivDet 2011 and 2013 datasets are used to conduct experiments. The entire data set includes 16470 images as the training set and 16439 images as the test set. Six different sensors (Biometrika, CrossMatch, Identix, etc.) are used for image acquisition. The fake fingerprint data set includes 9 different materials (BodyDouble, EcoFlex, gelatin, latex, Silgum, WoodGlue, etc.) to make fake fingerprint images. But the scale of the image is different, so it is very necessary to unify the size of the image. This paper solves the traditional operations that require cropping and zooming through ROI operations. The specific information of LivDet Datasets is shown in the following table:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sensor</th>
<th>train/test</th>
<th>Image size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DigitalPersona(Dig)</td>
<td>2004/2000</td>
<td>355×391</td>
</tr>
<tr>
<td></td>
<td>Italdata(Ita)</td>
<td>2000/2000</td>
<td>640×480</td>
</tr>
<tr>
<td></td>
<td>Sagem(Sag)</td>
<td>2016/2036</td>
<td>352×384</td>
</tr>
<tr>
<td></td>
<td>Italdata(Ita)</td>
<td>2000/2000</td>
<td>640×480</td>
</tr>
<tr>
<td></td>
<td>Crossmatch(Cro)</td>
<td>2250/2250</td>
<td>800×750</td>
</tr>
<tr>
<td></td>
<td>Swipe(Swi)</td>
<td>2200/2153</td>
<td>208×1500</td>
</tr>
</tbody>
</table>

4.2. Performance evaluation criteria

In the field of fingerprint liveness detection, the average classification error (ACE) is a widely accepted evaluation standard. The ACE is defined as the average value of false reject rate (FRR) and false accept rate (FAR), calculated as Eq.(2).

$$ACE = \frac{FRR + FAR}{2},$$  \hspace{1cm} (2)

Based on the original VGG16, AlexNet and ResNet18 models, the final fully connected layer are set to 512 dimensions. Through four different operations (small angle rotation, flip, zoom, and brightness), dataset is enhanced to train each sub-network (15 training epochs with learning rate 0.0002). Then, freeze the network parameters as the feature extractor of our model. Next, the genetic algorithm is implemented in the fingerprint features extracted by the above-mentioned sub-network to find the parts with significant differences. We set the initial population size $N$ to 10, the maximum evolutionary generation $T$ to 20, and the mutation probability of each chromosome to 0.05. According to the roulette algorithm to determine the probability of each chromosome being selected as a parent, all parents have a 50% probability of crossover operations with other parents. In the selection operation, the roulette algorithm is still used to eliminate individuals with lower fitness and maintain the population size $N$. The experimental results show that the accuracy of FLD is about 1% higher than the highest in the sub-network after screening by genetic algorithm. The three sub-networks selected in this paper have the same structure as in the original paper, except that the dimensions of the final fully connected layer are changed. Table 2 shows that the accuracy of the method proposed on the LivDet2011 dataset far exceeds that of other algorithms. Although not every dataset achieves the best accuracy on the LivDet2013 data set, our method is not far from it and the average accuracy also has competitiveness.

5. Conclusion

Due to the difference in the depth and architecture of CNN, the fingerprint features extracted by each network are not the same. In order to combine their advantages, this paper concatenated the fingerprint features extracted by multiple CNNs. In addition, the unknown feature of neural network extraction makes it impossible to find the optimization direction of the connection feature. Regarding the issue above, we introduce a FLD algorithm based on multi-modal features, and uses genetic algorithm to optimize these features. Through genetic algorithm, concatenated features are optimized adaptively and the difference features between real and fake fingerprints are deeply mined. The experimental results show that after the optimization of the genetic algorithm, the accuracy of FLD is improved by about 1% on the basis of the highest accuracy of the sub-network. Compared with other FLD algorithms, our method also has better accuracy and stability.

Acknowledgement

In this paper, the two authors have equally important contributions. This work is supported by the National Key R&D Program of China under
Table 2
The Average Classification Error of different models when datasets are LivDet2011 and LivDet 2013 respectively.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>The Average Classification Error ACE in (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LivDet2011</td>
<td>Ours</td>
<td>1.1 1.8 7.4 3.2 3.3</td>
</tr>
<tr>
<td></td>
<td>VGG16 [31]</td>
<td>2.1 2.6 8.4 4.4 4.3</td>
</tr>
<tr>
<td></td>
<td>AlexNet [31]</td>
<td>3 5.9 18.7 5.9 8.3</td>
</tr>
<tr>
<td></td>
<td>ResNet18 [33]</td>
<td>3.5 4.2 13.3 6 6.7</td>
</tr>
<tr>
<td></td>
<td>CNN [28]</td>
<td>9.2 1.35 12.35 6.1 7.2</td>
</tr>
<tr>
<td></td>
<td>DRN [29]</td>
<td>9.6 1.9 13.5 6.4 7.8</td>
</tr>
<tr>
<td></td>
<td>MBLTP [32]</td>
<td>9.7 7.0 16.0 10.9 10.9</td>
</tr>
<tr>
<td></td>
<td>RWG [34]</td>
<td>5.7 6.2 9.4 3.14 6.11</td>
</tr>
<tr>
<td></td>
<td>LQF [35]</td>
<td>7.4 5.6 11.4 6.7 7.8</td>
</tr>
<tr>
<td>LivDet2013</td>
<td>Ours</td>
<td>2.8 12.6 0.7 4.8 5.2</td>
</tr>
<tr>
<td></td>
<td>VGG16 [31]</td>
<td>3.1 37.4 1.1 6 11.9</td>
</tr>
<tr>
<td></td>
<td>AlexNet [31]</td>
<td>5.6 15 1.3 5.9 6.9</td>
</tr>
<tr>
<td></td>
<td>ResNet18 [33]</td>
<td>5.1 10.8 4.5 6.9 6.8</td>
</tr>
<tr>
<td></td>
<td>CNN [28]</td>
<td>4.35 7 1.4 5.1 4.4</td>
</tr>
<tr>
<td></td>
<td>DBN [30]</td>
<td>1.15 7.91 1.35 6.5 4.2</td>
</tr>
<tr>
<td></td>
<td>RWG [34]</td>
<td>0.6 — 0.85 3.2 1.55</td>
</tr>
<tr>
<td></td>
<td>LQF [35]</td>
<td>5.8 5.2 4.3 14.2 7.4</td>
</tr>
<tr>
<td></td>
<td>DCMBP [36]</td>
<td>0 10.62 0.85 15.65 6.78</td>
</tr>
</tbody>
</table>

References


