An unconstrained face recognition method based on Siamese networks

Cunli Song\textsuperscript{1,2,a}, Shouyong Ji\textsuperscript{1,b}
\textsuperscript{1} College of Software, Dalian Jiaotong University, Liaoning, Dalian 116052, PR China
\textsuperscript{2} Artificial intelligence key laboratory of Sichuan province, Sichuan, Zigong 643000, PR China
\textsuperscript{a} scunli@163.com
\textsuperscript{b} 1477203992@qq.com

Abstract

It is aimed at the low accuracy and low efficiency of face recognition under unlimited conditions. In this paper, a Siamese neural network model SN-LF (Siamese Network based on LBP and Frequency Feature perception) is designed based on the Local Binary Pattern (LBP) and the Frequency sensing model. Based on Siamese neural networks, the network adopts circular LBP algorithm and frequency feature perception to realize face recognition under unrestricted conditions. The LBP algorithm can eliminate the influence of light on the image and provide directional input to the network model at the same time. Frequency feature sensing divides the image features into low frequency features and high frequency features. The low frequency features are compressed in the Siamese neural network to increase the recognition efficiency of the network. At the same time, information is exchanged with the high frequency features, so that the target noise data can be eliminated while the feature data is retained. In this way, the recognition rate of the network is maintained, and the computing speed of the network is improved. Simulation experiments are carried out on standard face dataset CASIA-Webface and Yale-B, and compared with other network models. The experimental results show that the proposed SN-LF network structure can improve the recognition accuracy of the algorithm, and achieve a good recognition accuracy.

Key words: Artificial intelligence; Face Recognition; Siamese neural network; LBP algorithm; Frequency feature perception

1 Introduction

With the development of artificial intelligence, target recognition technology has also been rapidly developed, especially in pedestrian recognition and vehicle recognition. Due to the improvement of security awareness (Sundararajan K, Woodard D L, 2018; Nguyen K et al, 2018), people's demand for public and personal information security continues to rise. And in terms of public safety, pedestrian identification can better trace and investigate the occurrence of some accidents. Unrestricted face recognition means that there are no restrictions on environmental conditions, and the user's facial images are collected and recognized without the user's consciousness. Unrestricted face detection and recognition are mainly used in real-time surveillance areas of public security, which is of great significance in applications such as capturing suspects (Nanni Loris et al, 2021). At this stage, unconstrained face recognition has the following main problems: (1) In an environment without restrictions, it is easy to produce more noise interference, resulting in a decrease in the accuracy of face recognition; (2) the noise data will cause the algorithm consumes a lot of time in the coring process, affecting the efficiency of recognition; (3) In unconstrained face recognition, due to the user's non-cooperation, part or all of the face may be occluded, leading to missed detections and false detections.

From the classic face recognition algorithm, it can be found that the principal component analysis method (PCA) (Wang qianqian, 2019) reduces the computational complexity by transforming the matrix, but there are external interference factors such as occlusion, which will destroy the real spatial structure when there is interference. As a result, the real subspace structure of the data cannot be obtained. In the case of more noise interference, Tian Lei et al (2018) innovatively proposed adding sparse representation to the image deblurring algorithm in the preprocessing stage. It is used to solve the situation that the current traditional algorithm does not have a good denoising effect on blurred photos under unconstrained conditions, which leads to a decrease in the recognition accuracy. Moreover, traditional denoising algorithms require a large number of relevant reference images to provide templates, which are not suitable for face recognition under unrestricted...
conditions. Tian Lei et al used the prior knowledge of non-local self-similar images, first used sparse representation to defuzzify the image, and then used the local image block obtained from the sparse representation as the input of the data dictionary. Through the use of unsupervised clustering algorithm to generalize the similarity of the image blocks in the image, and finally use the principal component analysis algorithm to build the image data dictionary. The traditional classic algorithm will produce a lot of redundant data while processing noise data. Too much redundant data will interfere with the recognition effect of the algorithm, and too much redundant data will increase the computational cost. With the introduction of Compressed Sensing (Barranca Victor J 2021), the theory proposes that the sampling rate of the signal depends on the sparsity and irrelevance of the signal, rather than the width of the signal. Compressed sensing uses sampling and compression at the output end to decompress the data, so as to avoid a large amount of redundant data at the sampling end and reduce storage and calculation costs (Liu Jun 2015).

Because under unrestricted circumstances, the face of the target person may be occluded. In this case, it may cause the face recognition algorithm to miss the detection during the detection process, and even lead to false detection. Zhang Yu-Kang et al (2021) proposed a cross-modal joint constraint image feature to improve the recognition rate, using a generative confrontation network to separately constrain the visible light image and the infrared light image. Then use the shared decoder generates a potential feature space wherein the infrared light and visible light images. The cross-modal recognition is transformed into single-modal recognition, which is used to solve the modal difference between the pedestrian in the near-infrared and visible light state. Liu Yiming et al (2020) proposed the use of adversarial networks to increase the number of samples, and design a joint function to optimize the generation network, and then add the characteristic information of pedestrian key points to reduce the impact of the background, thereby improving the efficiency of the model. Zhao Zhi-wei and NI Gui-qiang (2019) proposed to recognize the key features of multi-pose faces based on vector machines for the low recognition efficiency, and use filters to extract the key points of the face features to obtain the three-dimensional information of the face features. Normalize the gray image of facial features, and realize face recognition under unconstrained conditions by calculating the distance between the facial features.

Deep learning is developed and innovated on the basis of neural networks. With the development of deep networks, a series of deep learning network models such as convolutional neural networks, Siamese neural networks, and deep belief networks have been proposed. The principle of the deep belief network (Roder Mateus et al 2021) is to use a greedy algorithm to train multi-layer restricted Boltzmann machines (RBM) layer by layer, and the output of the previous layer is used as the input of the next layer to perform unsupervised learning training on the image. Improve the speed and accuracy of model training, and solve the problem of local optimal solutions. Since then, the development of deep neural networks has entered a more rapid development stage. And face recognition has also ushered in rapid development. Scholars such as Liang Shu-fen et al (2014) proposed to improve the deep belief network to process the image, use the LBP algorithm to preprocess the texture features obtained from the image, and then use it as the input of the deep belief network to assist the deep belief network in unsupervised training to improve the efficiency of image recognition, but this method is not efficient in identifying targets that have too much noise data interference, and it is easy to miss the detection. The Siamese Neural Network is a simple and efficient network model framework. The Siamese Network was first proposed to compare the signature verification of American checks. However, the development of the Siamese Neural Network has stalled due to hardware limitations. Until 2010, Nair V and Hinton G E (2010) used the Siamese neural network in the field of face recognition, and combined the convolutional neural network and the Siamese network to form a Siamese neural network. In 2015, Zagoruyko S and Komodakis N (2010) used Siamese neural networks to calculate image similarity. Since then, the Siamese neural network has developed rapidly in target recognition and tracking with its small model structure, a small number of parameter settings and the coexistence of multiple network models. Scholars such as XU Xian-feng et al (2020) proposed to introduce the Inception model into the Siamese network, and add the cyclic learning rate optimization strategy to speed up the training speed. WU Hanzhao (2018) combined Siamese neural network and convolutional neural network, Using the method of local response value normalization makes the sudden change of the eigenvalue with sharp feedback larger than the eigenvalue with small feedback, so that the characteristics of the target are more obvious, and the target can be accurately identified. Shen Yan et al (2018) use Siamese
network as a classifier for image recognition. In order to solve the interference of deformation and rotation of the target in the image, it is proposed to add a spatial transformation network to the Siamese neural network to adaptively change the image and improve the accuracy. However, in order to better classify the Siamese network, it is necessary to manually extract the required reference samples, which consumes more human resources.

In order to solve the problem of low detection accuracy in non-restricted conditions, this paper proposes SN-LF to process non-restricted pedestrian images. This paper uses LBP algorithm for image preprocessing to obtain the texture features of the target, and then uses frequency feature perception to classify the features, divide them into high-frequency features and low-frequency features, and use Octave Convolution for training. The LBP algorithm extracts the target texture feature to ensure the target's texture and structure, and can accurately recognize the picture. At the same time, the frequency feature is used to enhance the receptive field of the pixel space, thereby improving the recognition performance.

2 Unrestricted pedestrian face recognition algorithm based on Siamese network

The algorithm flow chart of this paper is shown in Figure 1. Divide the two branches in the Siamese neural network into the search branch and the model branch, add the LBP algorithm to the search branch of the Siamese network to preprocess the image to be detected, and then use the frequency feature to perform convolution operations on the image. Make the spatial acceptance domain of the image target feature larger, and finally learn the pixel-level face features through the octiva convolutional layer. At the same time, the model branch of the Siamese network extracts the target image from the target library, performs the same operation on the target image, and finally uses the Euclidean distance to judge whether the target in the two images is the same target.

Step 1. Use the LBP algorithm to reduce the image dimension to 32*32 and perform histogram equalization to preprocess the image for noise reduction;

Step 2. Use the training samples as the input of the Siamese neural network, and train the feature data layer by layer;

Step 3. Use high and low frequency convolution on the feature image input in the Siamese neural network to obtain a frequency convolution image, the width and height of the frequency convolution image are 1/2 of the original feature image width and height, which makes the receptive range of high-frequency features in the feature image larger, and reduces the spatial resolution of low-frequency features. It can improve the recognition rate while not reducing the recognition speed;

Step 4. Divide the convolutional layer into 4 groups, corresponding to the width and height of the feature image, so that the image feature data can be exchanged for data and frequency. The 4 groups of convolutional layers are the width in the high frequency—the width in the low frequency, the width in the high frequency—the high in the low frequency, the high in the high frequency—the width in the low frequency, and the high in the high frequency—the high in the low frequency. Make it exchange information between high frequency and low frequency, so that the data of the characteristic image is not lost;

Step 5. After the Siamese neural network training is completed, compare the two outputs in the test sample to determine the Euclidean distance of the feature vectors of the two pictures. If the Euclidean distances of the feature vectors of the two pictures are within the threshold, they are the same target, and if they are greater than the threshold, they are not the same target.

![Fig 1. The flow of the SN-LF algorithm](image-url)
2.1 Image preprocessing based on LBP algorithm

The basic idea of the LBP algorithm is that for any pixel, with this as the center, the gray value of the pixel point is the threshold value, and the gray value of the field pixel is compared with the gray value of the center pixel. If the gray value of the field pixel is less than the gray value of the center pixel, it is marked as 0, and vice versa. Mark it as 1, and finally get a binary code of the center pixel, as shown in Figure 2. Because the LBP algorithm uses the central pixel as the threshold, and the binary code of the surrounding pixels remains unchanged, it is not sensitive to changes in illumination, so the LBP algorithm has gray-level invariance. LBP is to get multiple LBP values by rotating at one point in the field, and the smallest value is selected as the characteristic value, so LBP has the characteristics of rotation invariance. And LBP extracting features is a method of extracting features without parameters, so the LBP algorithm does not need to adjust the parameters for calculation, which reduces the difficulty and complexity of the algorithm in the calculation process, and at the same time can completely extract the main features of the image.

![Figure 2. LBP algorithm conversion](image)

The algorithm formula is:
\[ T = t(s(g_0 - g_c), s(g_1 - g_c), ..., s(g_{8} - g_c)) \]  \hspace{1cm} (1)
\[ s(x) = \begin{cases} 
1 & x \geq 0 \\
0 & x < 0 
\end{cases} \]  \hspace{1cm} (2)
\[ LBP = \sum_{i=0}^{8-1} s(g_i - g_c)2^i \]  \hspace{1cm} (3)

In which, formula (1) is the binary code for calculating the LBP algorithm, \( T \) is LBP binary code, \( t(x_1, x_2, ..., x_n) \) is the coding function, \( g_c \) is the value of the middle pixel, \( g_0, ..., g_8 \) are the pixel values of 8 areas around the center point. In formula (2), \( s(x) \) is the judgment function, Combination formula (1) and formula (2). Compare each pixel value with the center pixel value separately. When the field pixel value is less than the value of the central pixel point is 0, the field pixel value is greater than or equal to is 1.

![Figure 3. LBP for Uniform Pattern](image)

The LBP equivalent mode reduces the types of LBP texture features to \( i(i-1)/2 \), and the algorithm formula is:
\[ f(x,y)_{p,R} = \sum_{p=0}^{r} s(g_p - g_R)2^p LBP_{8}^{p} \]  \hspace{1cm} (4)

In which, \( f(x,y)_{p,R} \) is the texture feature of the image in this circular area, \( x, y \) are the coordinates of the center of the circle, \( g_p \) is \( p \) uniformly distributed pixels on the circular area, \( g_R \) is the pixel at the center of the circle. In order to be able to express the spatial structure characteristics of the face, the face image is divided into \( m(R_1, R_2, R_3, ..., R_m) \) blocks, and the LBP histogram of each block is counted to obtain the spatial histogram characteristics of the face sub-images. Connect them to get a complete spatial histogram of the face image.
\[ R_m = f_m(x, y) \]  
\[ H = [R_m]_m = 0, 1, 2, ..., M, M \in N \]

In which, \( R \) is the LBP histogram of the m-th block of image, and \( H \) represents the LBP histogram of the m-th block of face images connected to obtain a complete face image spatial histogram. Using the LBP operator to obtain the texture features of the image can not only greatly reduce the impact of the illumination, but also express the regional features of the face image, and connect the regional features to form the global feature of the face image. Figure 4 shows the original picture and the effect picture calculated by the LBP algorithm.

![Fig. 4 The left image is the original image, and the right image is the feature image transformed by the LBP algorithm](image)

2.2 Frequency feature perception

Frequency feature extraction (Yuxiang Yang et al. 2020) refers to the process of convolution operation, the image features are divided into high-frequency features and low-frequency features according to different feature maps. High-frequency features are mostly some key points in the image, while low-frequency features are less important feature information than high-frequency features. When performing image processing, noise data is often caused by too many low-frequency features of some images, and too many low-frequency features cause the algorithm to run slower and reduce the accuracy of the operation. So we decompose the feature data of each convolutional layer into high-frequency features and low-frequency features. Then the low-frequency features are compressed and reduced in dimension, and information is exchanged between the high-frequency features and the low-frequency features, so as to reduce the storage and calculation consumption through the convolution operation. The size of the low frequency part is \((0.5h, 0.5w)\), and the length and width are exactly half of the high-frequency part \((h, w)\). Although the low frequency part is compressed, it also effectively expands the receptive field in the original pixel space and improves the recognition performance (Chen Y et al. 2019).

We control the high and low frequency feature segmentation ratio by setting the hyperparameter \( \alpha \). The algorithm formula is:

\[ X \in \mathbb{R}^{c \times h \times w} \]  \hspace{1cm} (7)
\[ X^H \in \mathbb{R}^{(1-\alpha) \times c \times h \times w} \]  \hspace{1cm} (8)
\[ X^L \in \mathbb{R}^{\alpha \times c \times h \times w} \]  \hspace{1cm} (9)

In which, \( X \) is the collection of high-frequency and low-frequency features, \( h \) and \( w \) are the height and width of the feature, \( c \) is the channel number, and \( X^H \) is the high-frequency feature of the image. Capture precise details in the image by threshold \( \alpha \). As shown in formula (8), \( X^L \) is the low-frequency feature of the image. Reduce the spatial dimension by reducing the width and height, and slow down the change of low-frequency characteristics in the spatial dimension. As shown in formula (9). The current frequency feature of the image is that the high-frequency feature and the low-frequency feature have the same height and width, and the receptive field of the low-frequency feature is obviously higher than that of the high-frequency feature. As shown in Figure 5. After the frequency feature is extracted, the width and height of the low-frequency feature are reduced proportionally, which can reduce the receptive field of the low-frequency feature. At the same time, information is exchanged between low-frequency features and high-frequency features, so that the frequency features of the reduced image can save most of the information of the original image.

![Common Feature](image)

**Fig 5 Frequency characteristics of the current image**

In the feature update operation, the high-frequency and low-frequency features will be updated in the corresponding frequency. The feature swap operation will update the high-frequency and low-frequency feature information between different frequencies. Therefore, high-frequency features include not only its information processing, but also mapping from low-frequency to high-frequency, and vice versa. Frequency perception features can obtain a larger acceptance range of low-frequency features while exchanging high- and low-frequency information, and the resulting image features can fully interpret the original image compared with general feature...
extraction, it effectively doubles the acceptance field, obtains more contextual feature information, further improves the learning ability of the Siamese neural network, and improves the recognition performance as shown in Figure 6.

**Fig. 6. Frequency sensing feature extraction**

### 2.3 Siamese neural network

Two classification problems are mainly solved by the Siamese neural network: Data set with few classifications and large amount of data, such as ImageNet, VOC etc.; 2. There are more target types, and the amount of data in the target database is small and the number of classifications is large (or the specific number cannot be determined) (Fu Li-hua et al. 2020). Each type has a small number, such as face recognition, face verification and other tasks. The problem to be solved in this paper is to find a specific target in the crowd, the number of targets is too small and uncertain, and the flow of people in the crowd is uncertain, so the Siamese neural network is used for processing.

There are two identical network structures in the Siamese neural network, use the same weights and parameters for training in two network structures, one as a model for training, and one as a search for training. Enter a picture in each of the two networks, then train the two pictures, and finally compare the distance between the two pictures for image recognition. It is suitable for problems such as target detection under large data sets and multi-target recognition under small data sets. The model structure is shown in Figure 7:

![Figure 7: Structure model of Siamese network](image)

If the two networks have different structures, they are called pseudo Siamese neural networks.

In the Siamese network, metric learning is used to compare the similarity of two pictures (SHI Guoqiang and Zhao Xia 2020). First of all, as shown in Figure 7, input a picture in Network 1, enter the convolutional layer, pooling layer, and fully connected layer through the Siamese neural network in turn, and finally output the feature vector \( f(x_1) \) of the picture. Enter another picture in Network 2, perform the same operation on the picture in Network 2, and finally get its feature vector \( f(x_2) \). In order to be able to judge whether two pictures are the same person, we define a threshold (hyperparameter), If the encoding result distance of the two pictures is less than the threshold, the two pictures are one person, and if the distance is greater than the threshold, the two pictures are not the same person. The algorithm formula is:

\[
d(x_1, x_2) = \| f(x_1) - f(x_2) \|^2_2 \quad (10)
\]

\[
\begin{cases}
d(x_1, x_2) > E_w \quad \text{Both are the same person} \\
  d(x_1, x_2) < E_w \quad \text{The two are not the same person}
\end{cases}
\quad (11)
\]

In which, \( d(x_1, x_2) \) is the distance function, \( f(x_1) \) and \( f(x_2) \) are the feature vectors of the pictures in Network 1 and Network 2, respectively, and \( E_\text{w} \) is the threshold. Combined formula (10) and formula (11). By calculating the distance difference between the feature vectors of the two images, and then comparing the difference with the threshold, if the difference is greater than the threshold, the two images are not the same target, and if the difference is less than the threshold, the two images are the same target.

In order to make the input picture have a good encoding and learn the corresponding parameters autonomously, a triple loss function will be used here. Triple loss function refers to the use of three pictures to calculate a loss function: a fixed image
A, a positive example image P (the same person as the fixed image), and a negative example image N (not the same person as the fixed image). We let the encoding distance d(A,P) between the fixed image A and the positive image P be less than or equal to the encoding distance d(A,N) between the fixed image A and the negative image N. In this way, the pictures of the same person are closer, while the pictures of different people are far away from each other. Triple loss minimizes the distance between the fixed image A and the positive image P. Both have the same identity. At the same time, the distance between the fixed image A and the counter-example image N is maximized (Luo Hao et al. 2019). The model may make the same encoding for different pictures in the learning process, which is equivalent to the distance between the images is 0, so on this basis, we add a hyperparameter α to avoid this situation, so that there is a gap between d(A,N) and d(A,P). There is a formula:

\[ d(A, P) + \alpha \leq d(A, N) \]  

(12)

Combining the distance formula in the above formula:

\[ \| f(A) - f(P) \|^2 + \alpha \leq \| f(A) - f(N) \|^2 \]  

(13)

The formula is further obtained as

\[ \| f(A) - f(P) \|^2 - \| f(A) - f(N) \|^2 + \alpha \leq 0 \]  

(14)

The loss function is:

\[ L(A, P, N) = \max (\| f(A) - f(P) \|^2 - \| f(A) - f(N) \|^2 + \alpha, 0) \]  

(15)

The cost function is the sum of the loss functions in all individuals.

\[ J = \sum_{i=1}^{n} L(A) \]  

(16)

3 Experiments

3.1 Data set and parameter settings

This section uses experiments to verify whether the improved method mentioned above is effective for target recognition under unrestricted conditions. The data set uses the CASIA-WebFace data set. In this paper, 300 different people are randomly selected in CASIA-WebFace, and 15 photos of each person are selected to carry out the experimental research. Some pictures of the data set are shown in Figure 8:

Fig 8. A partial image of the CASIA-Webface dataset

3.2 Experimental environment and parameter settings

The experimental hardware environment is Intel core i3 CPU, Geforce 710m GPU, 2G memory, the operating system is Win10, the deep learning framework is Tensorflow, and the programming language and version are Python 3.6. Determining network layer convolution by comparing the number of layers is layer 5, the highest recognition rate of the network structure; When the number of layers of the same convolution, the convolution of a frequency characteristics perceived layer, the image characteristics obtained network structure of the highest recognition rate of learning, a minimum loss of convergence, Therefore, the overall network structure is determined to be a 5-layer convolutional layer, a 5-layer MaxPool layer, and a 1-layer fully connected layer.

Denote the network structure of Siamese neural network and SN-LF as SN and SN-LF respectively. Determine the learning rate range. Set the initial value of the learning rate to 1e-3 according to experience, in each batch of the training process, the learning rate increases linearly at a growth rate of 1e-4, Record the learning rate and loss value of the Siamese neural network and SN-LF under the CASIA-WebFace data set, and draw the relationship diagram. As shown in Figure 9 and Figure 10:
3.3 Comparative experiment of the algorithm in this paper on the CASIA–WEBFACE dataset

In the CASIA–WEBFACE data set, the classic Siamese neural network and SN-LF are compared. This paper adopts the mean Average Precision (mAP) and the recognition rate to evaluate the performance of the algorithm. Comparing the algorithm in this paper with the classic Siamese neural network algorithm, it is found that the similarity of the algorithm in this paper is about 15% higher than that of the classic Siamese neural network. It shows that the algorithm in this paper can recognize target features better than the classic Siamese neural network in the recognition process. The experimental results are shown in Figure 11.

Table 1 shows the recognition rate of the traditional Siamese neural network and the algorithm SN-LF in this paper under different iteration times. It can be seen from the table that under the same iterative situation, the recognition rate of SN-LF is increased by about 5% compared with the traditional Siamese neural network. When the number of iterations is the same, the recognition efficiency of the traditional Siamese network is not as good as that of SN-LF. As the number of iterations increases, the recognition rate of the traditional Siamese neural network is not as robust as the SN-LF algorithm compared to the recognition rate of SN-LF. It shows that in the image recognition of the traditional Siamese neural network, when the target is occluded, the recognition accuracy and efficiency will
The LBP algorithm in SN-LF can reduce the interference caused by high-frequency noise, and at the same time, the frequency characteristic senses the compression and information exchange of high-frequency and low-frequency characteristics. While reducing noise interference, it does not reduce the recognition accuracy.

From the comparison between the classic Siamese neural network algorithm in Figure 12 and the mAP curve of SN-LF, it is found that the mAP curve of SN-LF and the classic Siamese neural network algorithm is not much different before 1000 iterations.

As the number of iterations increases, the difference between the two becomes larger and larger, reaching the maximum at 3000 iterations, with a maximum difference of 15%. This shows that as the number of iterations increases, the convergence speed of the SN-LF algorithm is faster than that of the classical Siamese neural network. When the number of iterations increases to 5000, the mAP curve reaches about 90%, indicating that the algorithm has strong robustness.

**Fig. 12.** MAP curve of comparison between the classical Siamese neural network algorithm and SN-LF

To explore the comparison of SN-LF and traditional face recognition algorithms in unrestricted face recognition, References in this article(Marsico M D E et al 2012). When the number of iterations of SN-LF is 3000 times, it can be seen from Table 2 that the recognition rate of the algorithm in this paper is compared with that of different algorithms. Compared with other algorithms, they have increased by 51.1%, 41.71%, 39.54%, 25.22%, and 4.86%. In the literature(Lee K C et al 2005), the recognition rate of the Face algorithm is 61%, but in the literature, the algorithm is complicated, and there are many parameters, and manual intervention is complicated. The comparison of mAP between different algorithms is shown in Figure 13. Combined with Table 2, the algorithm in this paper is ahead of the traditional algorithm in accuracy and mAP. From Table 1 and Table 2, we know that in complex samples, the algorithm in this paper can maintain a stable recognition rate. It shows that SN-LF can accurately identify the target under interference, which further proves that the algorithm in this paper is robust to pedestrian recognition under unrestricted conditions and has high anti-interference.

**Table 1.** Comparison of recognition rates between traditional Siamese network and SN-LF under different iterations  

| Algorithm        | 1000  | 2000  | 3000  | 5000  |
|------------------|-------|-------|-------|-------|
| Siamese Network  | 65.89%| 78.34%| 82.16%| 82.45%|
| SN-LF            | 70.23%| 83.25%| 87.12%| 90.23%|

**Table 2** Comparison of recognition rates among different algorithms

| Algorithm    | Recognition rate |
|--------------|------------------|
| PCA          | 36.12%           |
| SVM          | 45.51%           |
| LBP          | 47.75%           |
| FACE         | 61.00%           |
| Siamese Network | 82.36%         |
| SN-LF        | 87.22%           |
under the CIASIA-WebFace data set, Compared with pixel-level input, SN-LF uses texture features as input to help reduce redundant information. The added frequency feature perception improves space utilization. This enables the experiment to accurately recognize face images under unconstrained conditions.

3.4 Comparative experiment of the algorithm in this paper on the Yale-B standard face data set

In order to test the accuracy of the algorithm in the face recognition under light conditions, this paper uses the Yale face database and traditional algorithms for comparison experiments. It can be seen from Table 3 that SN-LF has a better face recognition effect under light conditions. With the increase of iterations, the recognition accuracy improves, and the image recognition accuracy no longer changes greatly after iterating 3000 times. It can be seen from Table 4 that compared with other algorithms in the Yale face database, this algorithm has improved by 17.88%, 9.27%, 7.18%, 4.40%, and 2.20%. It shows that the algorithm in this paper has better recognition efficiency under light conditions, and at the same time shows that the algorithm in this paper can handle noisy data well in the presence of interference. At the same time, combined with Table 3, it is known that in the iterative process, the algorithm in this paper has a faster convergence rate, and has better robustness than the traditional Siamese neural network. It shows that the algorithm in this paper also shows higher performance under restrictive conditions. Table 4 combined with Table 3 shows that, in addition to the faster convergence speed of the algorithm in the iterative process, it has better robustness than the traditional Siamese neural network. It shows that the algorithm in this paper also shows higher performance under restrictive conditions.

Table 3 Comparison of accuracy of different iteration times in Yale-B dataset

| Algorithm          | 1000 iterations | 2000 iterations | 3000 iterations | 5000 iterations |
|--------------------|-----------------|-----------------|-----------------|-----------------|
| Siamese Network    | 75.89%          | 85.44%          | 92.36%          | 92.15%          |
| SN-LF              | 80.23%          | 88.25%          | 94.12%          | 94.23%          |

Table 4 Comparison of accuracy rates of different algorithms in Yale B

| Algorithm        | Recognition rate |
|------------------|-------------------|
| PCA              | 76.62%            |
| SVM              | 85.23%            |
| LBP              | 87.32%            |
| FACE             | 91.10%            |
| Siamese Network  | 92.20%            |
| SN-LF            | 94.50%            |

In order to evaluate the performance difference between each algorithm and the algorithm in this paper, through the comparison of mAP between the algorithms in Figure 14, it is known that the algorithm in this paper also has better performance under the Yale B data set. When the number of iterations is about 1000, the performance of each algorithm is
close, as the number of iterations increases, the performance of each algorithm has differences, and some algorithms have large fluctuations, indicating that the anti-interference ability is reduced under sudden light changes. The algorithm in this paper uses the LBP algorithm’s insensitivity to illumination changes, and can have a better recognition effect in the case of sudden changes in illumination. As the number of iterations increases, the mAP value of SN-LF increases by about 20%, indicating that the algorithm in this paper has good robustness in the learning process. When the number of iterations reaches about 5000 times, the mAP value of SN-LF is increased by about 3%-10% compared with other algorithms, which shows that the algorithm in this paper can extract more target features for learning in the iterative process.

![Graph](image)

**Fig. 14.** Map curve of each algorithm under Yale-B dataset

4 Conclusion

The existing unconstrained face recognition methods mainly start from the key points of the face combined with the direction of the denoising algorithm, which tends to converge too fast or the robustness is not strong, and the algorithm application scene is relatively single and cannot adapt to multiple scenes. This paper proposes an architecture that combines frequency feature perception, LBP and Siamese neural networks. By changing the parameters, the network convergence speed is accelerated and the number of iterations required for the same recognition rate is reduced. The algorithm first uses the LBP algorithm to preprocess the image to obtain the texture feature, which effectively reduces the interference caused by the noise data, and then the texture feature is input into the Siamese neural network. While improving the running speed, it also improves the recognition accuracy of the algorithm and enhances the robustness of the algorithm. Finally, through frequency feature perception, the features are divided into high frequency and low frequency, which effectively expands the spatial acceptance domain and improves the recognition accuracy. Moreover, the proposed network has a simple structure and is suitable for face recognition of small-scale data sets under unconstrained conditions. It has been trained and tested on CASIA-webface and Yale standard face database, and obtained good results, and at the same time, it has high robustness. It is a face recognition algorithm with excellent performance. The next problem to be solved is to reduce the running time of the algorithm, improve the efficiency of the algorithm, and at the same time solve the problem of identifying multiple targets in the image under unrestricted conditions.

**Compliance with ethical standards**

**Ethical approval**

This article does not contain any studies with human participants or animals performed by any of the authors.

**Funding**

This work is supported by open fund projects of artificial intelligence key laboratory of Sichuan province (2020RYJ04); Liaoning provincial natural science foundation (20180551020); Liaoning educational committee program (JDL2019011).

**Conflict of interest**

The authors declare that they have no conflict of interest.

**Informed consent**

Informed consent was obtained from all individual participants included in the study.

**Authorship contributions**

Cuili Song: Conceptualization, Methodology.
Shouyong Ji: Data curation, Writing- Original draft preparation.

**References**

Sundararajan K, Woodard D L (2018) Deep learning for biomet-rics: A survey. ACM Computing Surveys (CSUR) 51(3): 65.
Nguyen K, Fookes C, Sridharan S, et al (2018) Super-resolution for biometrics: A comprehensive survey. Pattern Recognition 78: 23-42.
Nanni Loris, Minchao Giovanni, Brahnam Sheryl, Maguolo Gianluca, Luminia Alessandra (2021) Experiments of Image Classification Using Dissimilarity Spaces Built with Siamese Networks. Sensors (Basel, Switzerland) 21(5).

Wang qianqian (2019) Study of Robust Principal Component Analysis and Its Applications. Dissertation, Xidian University

Tian Lei (2018) Research on feature learning based face recognition under unconstrained scenario. Dissertation, Beijing University of Posts and Telecommunications

Barranca Victor J (2021) Neural network learning of improved compressive sensing sampling and receptive field structure. Neurocomputing 455.

Liu Jun (2015) Research on the key technology of face recognition. Dissertation, Beijing University of Posts and Telecommunications

Zhang Yu-Kang, Tan Lei, Chen Jing-Ying (2021) Cross-Modality Person Re-identification Based on Joint Constraints of Image and Feature. ACTA AUTOMATICA SINICA. https://doi.org/10.16383/j.aas.c200184.

Liu Yi-Min, JIANG Jian-Guo, QI Mei-Bin, LIU Hao, ZHOU Hua-Jie (2020) Video-based Person Re-identification Method Based on GAN and Pose Estimation. ACTA AUTOMATICA SINICA 46(3): 576–584

Zhao Zhi-wei, Ni Gui-qiang (2019) Automatic Recognition Algorithm for Unconstrained Multi-pose Face Key Features under Unqualified Conditions. COMPUTER SCIENCE 46(09):250-253.

Roder Mateus, Passos Leandro Aparecido, de Rosa Gustavo H., de Albuquerque Victor Hugo C., Papa Joao Paulo (2021) Reinforcing learning in deep belief networks through nature-inspired optimization. Applied Soft Computing (prepublish)

Liang Shu-fen, LIU Yin-hua, LI Li-chen (2014) Face recognition under unconstrained based on LBP and deep learning. Journal on Communications 35(06):154-160

Nair V , Hinton G E (2010) Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. International Conference on International Conference on Machine Learning. Omnipress pp:807-814

Zagoruyko S, Komodakis N (2015) Learning to Compare Image Patches via Convolutional Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 4353-4361.

Xu Xian-feng, ZHANG Li, LANG Bin, XIA Zhen (2020) Research on Inception Module Incorporated Siamese Convolutional Neural Networks to Realize Face Recognition. ACTA ELECTRONICA SINICA 48(04):643-647.

Wu Hanzhao (2018) Face tracking using siamese convolutional neural networks. Computer Engineering and Applications 54(14) : 175-179.

Shen Yan, WANG Huan, DAI Yuxing (2018) Deep siamese network-based classifier and its application. Computer Engineering and Applications 54(10) : 19-25.

Wu Xiao-Ting, Feng Xiao-Yi, HUANG An, ZHANG Xue-Yi, DONG Jing, LIU Li (2021) Survey on facial kinship verification. Acta Automatica Sinica 1-25.

Yuxiang Yang, Weiwei Xing, Shunli Zhang, Qi Yu, Xiaoyu Guo, Min Guo (2020) A Learning Frequency-Aware Feature Siamese Network for Real-Time Visual Tracking. Electronics 9(5).

Chen Y, Fan H, Xu B, et al (2019) Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks With Octave Convolution. IEEE/CVF International Conference on Computer Vision (ICCV), pp 3434-3443

Fu Li-hua, ZHAO Yu, SUN Xiao-wei, LU Zhong-shan, WANG Dan, YANG Han-xue (2020) Fast Video Object Segmentation Based on Siamese Networks. ACTA ECEVTORNICA SINICA 48(04):625-630.

Shi Guoqiang, ZHAO Xia (2020) Object tracking algorithm based on jointly-optimized strong-coupled Siamese region proposal network. Journal of Computer Applications 40(10):2822-2830.

Luo Hao, Jiang Wei, Fan Xing, Zhang Si-Peng (2019) A survey on deep learning based person re-identification. Acta Automatica Sinica 45(11): 2032–2049

Yi D., Lei Z., Liao S., et al (2014) Learning face representation from scratch. Computer Science. https://arxiv.org/abs/1411.7923. Accessed 28 Nov 2014

Lee K C, Ho J, Kriegman D J (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence (5) : 684-698.

MARSICO M, NAPPI M, RICCO D (2012) Robust face recognition for uncontrolled pose and illumination changes. IEEE Transactions on Systems, Man and Cybernetic 43(1): 149-163.