Research Article

Deblurring Method of Face Recognition AI Technology Based on Deep Learning

Weilong Li ⓘ,[1] Jie Li,[2] and Junhui Zhou[3]

[1]*College of Educational Information Technology, Hunan Vocational College for Nationalities, Yueyang 414000, Hunan, China*
[2]*Organization and Personnel Department, Yueyang Vocational and Technical College, Yueyang 414000, Hunan, China*
[3]*Information Construction and Management Center, Hunan Vocational College for Nationalities, Yueyang 414000,
 Hunan, China*

Correspondence should be addressed to Weilong Li; liweilong1127@163.com

As a common method of deep learning, a convolutional neural network (CNN) shows excellent performance in face recognition. The features extracted by traditional face recognition methods are greatly influenced by subjective factors and are time-consuming and laborious. In addition, these images are susceptible to illumination, expression, occlusion, posture, and other factors, which bring a lot of interference to the computer face recognition and increase recognition difficulty. Deep learning is the most important technical means in the field of computer vision. The participation of this technology reduces manual participation and can identify the identity of visitors from multiple aspects. This study, based on the introduction at all levels and on the fundamental principle of the colloidal neural network, combines the basic model and the common exploitation methods of aspects to make a model of a combination of multiple aspects. Then, an improved CNN-based multifeature fusion face recognition model is proposed, and the effectiveness of the model in face feature extraction is verified by experiments. With the experimental results, the identification rate for the ORL and Yale data sets is 98.2% and 98.8%, respectively. Accordingly, an online face detection system and recognition system based on the combination of element models are designed. The system can obtain dynamic facial recognition and meet the recognition rate of the design requirements. The system is training four detection models and online recognition, and the test results show that the noise component model has the highest recognition rate, and the recognition rate has improved by 13% compared with the baseline capacity, further verifying that a model of a combination of features can achieve more effectively.

1. Introduction

1.1. Research Background. Today's society has a higher and higher demand for identification technology, especially in the fields of finance and criminal investigation. Traditional identification methods such as certificates, keys, and passwords cannot satisfy the needs of today's society.

Although the traditional identification method is very mature, it can be combined with other verification technologies for corresponding protection (such as mobile phone verification code). However, identification technology relies on the additional identification of individuals, which is more prone to forgery, theft, and other incidents than the features of the body itself (such as the face). The threat of instability and misidentification can also cause

more trouble. Biometric identification technology mainly relies on fingerprint, face, voice, and other unique physiological characteristics of individuals to carry out biological identification, including face recognition, fingerprint recognition, voice recognition, and iris recognition. It is an interdisciplinary subject involving pattern recognition, digital image processing, machine learning, psychology, and other disciplines, providing a theoretical source for face recognition models.

1.2. Research Status at Home and Abroad. Many scholars have made deep research in deep learning, face recognition, and other aspects. Vanitha CN believed that face recognition technology with the ability to detect, recognize, and examine

faces is now being used everywhere. Artificial intelligence plays an important role in recognizing faces; the AI face recognition system captures an image of a person from any recorded image and stores it in a database for analysis of the captured image. However, this method requires a huge amount of data and lacks suitable face recognition tools [1]. Litjens et al. first verified the effectiveness of stacked autoencoders through other classifications. Secondly, a classification method based on spatial dominant information is mentioned. Then, a new deep learning framework is proposed to integrate features to achieve the highest classification accuracy. Experimental results on widely used hyperspectral data show that the classifier based on a deep learning framework can provide an effective classifier. Because of the complexity of the research process, the results are not very accurate [2]. Wang and Chen introduced deep learning into supervised speech recognition, which greatly accelerated the process and improved the separation performance. The introduction of deep learning can greatly promote the learning of algorithms and provide space for progress in identity recognition. However, the technology still has applications in face recognition [3].

1.3. Main Innovation. The main innovations of this study are as follows: firstly, the deep learning method is introduced briefly and the basic principle of CNN is elaborated in detail. Then, the structure of deep CNN is designed using Caffe's first-class deep learning framework. Then, an improved multifeature fusion face recognition model based on CNN is proposed, and the effectiveness of this model in face feature extraction is verified by experiments. Finally, an online face detection and recognition system is designed and implemented. The system has trained the two algorithm models involved in this study and achieved good results. Therefore, a multilayer neural network face recognition feature fusion model structure based on convolution is proposed. Feature fusion is the basic model for feature extraction and fusion of each layer of convolution pooling features. The feature extraction methods used in PCA, LDA, and LPP are based on the basic features of each layer of convolutional models and then tested on the face dataset. Finally, the test results and analysis are given.

2. Proposed Method

Some common biological characteristics are often used for identification. Figure 1 shows some commonly used biometric features. Biometric identification technology mainly refers to a technology for identity authentication through human biometrics. The core of so-called biometrics is how to obtain these biometrics, convert them into digital information, store them in a computer, and use reliable matching algorithms to complete the process of verifying and identifying individuals. Compared with traditional biometric methods, biometric identification technology on the one hand has the advantages of robustness (not affected by the external environment), uniqueness (different for different subjects), and universality (each subject has). On the other
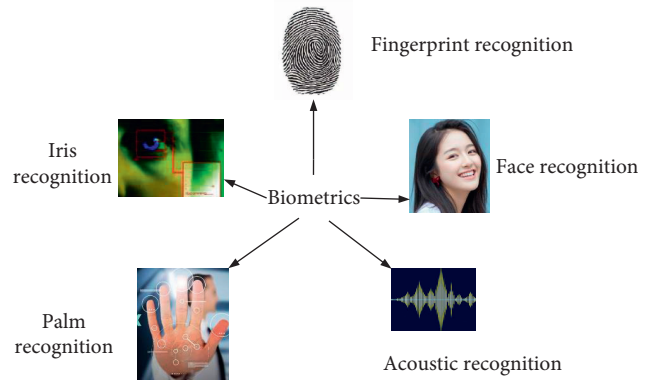

Figure 1: Commonly used biological characteristics.

hand, it means that what can be used to identify the identity of the subject can be the physiological characteristics and behavioral characteristics of the subject to be identified, which are relatively easy to obtain [4–6].

The initial research on face recognition began in 1888, and the research on face recognition has been developing continuously over the past 100 years. In the past 30 years, face recognition research has been a hot stage [7, 8]. The knowledge-based representation method mainly obtains the feature data that are helpful for face classification according to the shape description of the face organs and the distance characteristics between them. The feature components usually include the Euclidean distance, curvature, and angle between the feature points. On the one hand, face recognition research promotes continuous application in real life. On the other hand, the power of face recognition research comes from more and more extensive practical applications. At first, the face recognition method about the global features of the face has been widely concerned, and then, people found that many interference factors of the face image have a great impact on the global features, so the face recognition method based on the local features is gradually rising. In recent years, deep learning has achieved surprising results in face recognition and has been used in more fields and is popular with more and more people [9–11].

As shown in Figure 2, the face recognition system can be divided into detection and recognition of two parts. Firstly, face color, contour, texture, and structure of face are detected, and templates are extracted from existing stored face database. Then, a template matching strategy is adopted to match the collected face images with templates extracted from the template library [12–14].

However, the disadvantage is the loss of information when the characteristics of input data are expressed at a high level with the continuous increase in network level. CNN is another common model of deep learning [15]. The following will focus on the working principle of each part of CNN.

2.1. Face Recognition Method Based on Deep Learning. For a long period of time, the technology of face recognition has stagnated, and with Hinton et al. in 2006, the deep belief network opened the prelude to the application of deep
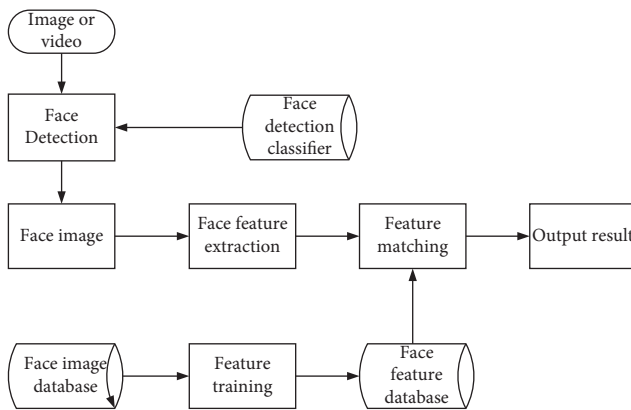
Figure 2: Face recognition system composition.

learning in face recognition, which finally solved the problem. Over the years, multilayer neural networks are prone to local optimization problems. This enables the rapid development of deep networks. At present, deep learning application scenarios involve many aspects, including pattern recognition, target segmentation, information retrieval, natural language processing, and other fields.

The face recognition method based on deep learning can obtain super learning ability after training, which is very suitable for solving various complex nonlinear problems. The deep learning network structure is very complex, with multiple hidden layers and a large number of neurons. In general, network training requires the use of unsupervised learning algorithms to train the network layer by layer. However, due to the lack of sufficient prior knowledge, it is difficult for humans to label categories, or the cost of manually labeling categories is too high. Naturally, we want computers to do the work for us, or at least to help. Solving various problems in pattern recognition based on training samples of unknown categories (not labeled) is called unsupervised learning. After the network training is completed, supervised learning is used to learn fine-tuning of the network. After the fine-tuning is over, the network can obtain strong classification and recognition capabilities. With the continuous improvement of deep learning theory, the development of face recognition based on deep learning is very rapid. Chai Ruimin proposed a deep belief network multichannel face recognition algorithm. The algorithm combines the Gabor wavelet transform and restricted Boltzmann machine to extract facial features, which further improves the accuracy of face recognition. The Gabor wavelet is sensitive to the edge of the image and can provide good direction selection and scale selection characteristics, and it is insensitive to the illumination transition and can provide good adaptability to the illumination transition. Domestic scholar Wang Haibo adopted a face recognition system based on deep beliefs. In the process of preprocessing, PCA was used to complete the dimensionality reduction in the face, which achieved good experimental results, but the number of network model layers used was still small, and n is small. Sabzevari et al. first used the local constraint model (CLM)

to generate the basic shape of facial expressions and then used a three-layer DBN model to determine facial expressions. Combined with the local coordinate coding theory and sparse constraints, the adjacent shape vector set based on the manifold structure replaces the nonrigid deformation set in the point distribution model to realize the fusion of the local tangent space arrangement method in the manifold learning and the point distribution model. A constrained local model of manifold embedding is developed. The face image obtained by this method is more dimensional than the original image by preprocessing. Huang GB et al. used the convolutional deep belief network (CDBN) to verify the high-resolution face image. The experiment showed that optimizing the convolution kernel parameters can not only help to obtain better feature representation, but also improve the system. Robustness: Liu et al. proposed an improved deep belief network (BDBN), which uses the BDBN model to realize facial recognition, which can capture effective facial features well. The deep belief network is a probabilistic generative model. Compared with the traditional neural network of the discriminant model, the generative model is to establish a joint distribution between observation data and labels. The evaluation is done, and the discriminant model only evaluates the latter.

In summary, traditional face recognition algorithms still have many shortcomings, and deep learning, an emerging branch of machine learning algorithms, has shown its strong advantages in many fields, especially in the field of face recognition. Achievement: on the other hand, a lot of research has found that if you directly use the deep learning framework to learn the features of the face, the network will automatically learn the redundant information of the face, and the features extracted through such learning and training are not representative. In most cases, the local and gradient information of the face cannot be well represented. Therefore, to make the best use of the characteristics of the deep learning framework, this study first extracts the local features and gradient features in the face image and then merges these two features into a hybrid feature by means of vector connection. Such a network can obtain more effective information through layer-by-layer learning. It can solve the abovementioned problems caused by directly using deep learning training and then performing face recognition to a certain extent.

2.2. Convolution Layer. The input of the convolutional layer in the CNN also comes from the output of the upper layer, in which different convolutional cores are used for the winding machine operation. Each convolutional layer in the convolutional neural network is composed of several convolutional units, and the parameters of each convolutional unit are optimized by the backpropagation algorithm. The purpose of the convolution operation is to extract different features of the input. The first convolution layer may only extract some low-level features such as edges, lines, and corners. More layers of the network can iteratively extract more complex features from the low-level features.

Characteristics: these convolution checks are repeated over and over again across all the receptive fields of the entire input image, so that after the convolution operation you get different feature maps of the input data. The parameters in the convolution kernel, including the weight matrix A and bias term B, need to be learned through the convolution layer [16, 17]. The detailed calculation formula of convolution is as follows:

$$y_m = f \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} x_{m+i,n+j} a_{ij} + b, \quad 0 \leq n < M, 0 \leq n < N. \quad (1)$$

In fact, the convolutional layer is a linear correlation calculation method, where x is a two-dimensional input vector with a size of (M,N), and f represents an activation function [18, 19].

2.3. Pooling Layer.

Computational load in CNN is a problem that needs to be considered. Generally, the characteristic dimension of the input image will not decrease too much after the calculation of the convolutional layer, so the computational load will be extremely large. In addition, the learning speed of the network model will be affected and the learning accuracy will be greatly reduced under the large computational load. The pooling layer is also called the lower sampling layer, and its main function is to solve the problem of large amount of computation.

The pooling layer is a nonlinear subsampling method, which is usually the layer to scale down the feature map of the network model. Feature map: by convolving a certain range of feature maps, the pattern composed of multiple features can be extracted into one feature, and the next feature map can be obtained. After that, the feature map is convolved, and the features are combined to obtain a more complex feature map. In the CNN, the input of the pooling layer comes from the output of the previous layer, and the pooling layer will subsample all the feature graphs input to itself, so the input feature graphs will decrease significantly after passing the pooling layer. Assuming that the pooling layer adopts the method of uniform sampling and the sampling size is 1×2, the pooling formula is as follows:

$$y_{mn} = \frac{1}{s_1 s_2} \sum_{i=0}^{s_1-1} \sum_{j-1}^{s_2-1} x_{m^* s_1 + i, n^* s_2 + j}. \quad (2)$$

The above formula is the uniform sampling formula. x and y, respectively, represent the input two-dimensional vector and the output value after the pooling layer. The feature images of each layer input into the network are first divided into regions of specified size, and then, the average value of all the values in different regions is taken.

Weight sharing is another important feature of CNNs. In neural networks, the number of parameters can be reduced by assigning power values. Assuming the resolution of the output image is 1000/1000 and the number of hidden neurons is 1 meter, as you can see, the weight distribution reduces the number of parameters by one order.

In the CNN, the mapping of the input image features formed by the convolution kernel in the convolutional layer and the input activation function after the convolution operation with the input image is called the feature graph. The feature graph group composed by superposition of the feature graph generated by different convolutional kernels represents a variety of features of the input image.

After the convolution, you have to add a bias term b. As the input of nonlinear activation function, the calculation formula of neuron output is shown in the equation, where b_j^l represents the additive bias of the jth feature graph of the current layer. During initialization, b_j^l is generally set to 0 and f is the activation function.

$$y_j^l = f\left(\sum_{i=1}^{n_j^{i-1}} a_{i,j} \otimes x_i^{l-1} + b_j^l \right), \quad j = 1, 2, \ldots, m. \quad (3)$$

The pooling layer is also known as the lower sampling layer. The pooling layer compresses the data, keeping the number of feature graphs unchanged, but making the feature graphs smaller, thus reducing the storage space. At the same time, you can also keep the translation and scaling of the network model invariant. There are two common pooling methods, maximum and average, which are widely used and simple in algorithm. The difference between the max pooling operation and the mean pooling operation is that it is necessary to record which pixel has the largest value during the pooling operation, that is, max id. This variable is used to record the location of the maximum value, because it is used in back propagation.

2.4. The Connection Layer.

CNNs typically have one or more fully connected layers before the output layer and after the stacked convolutional layers and pooling layers. Each neuron in the full-connection layer is connected with all neurons in the previous layer, but there is no connection between neurons in the same layer. A neuron function is very similar to a human neuron; in other words, it has some input and then gives an output. Mathematically, a neuron in machine learning is a placeholder for a mathematical function whose only job is to take a function on an input and give an output. The full-connection layer increases the nonlinear mapping ability of the neural network, limits the network size, and acts as a classifier. Its mathematical expression is as follows:

$$\phi_j^l = f\left(\sum_{i=1}^{n} x_i^{(l-1)} \cdot a_{ji}^{(l)} + b_j^{(l)} \right), \quad (4)$$

where l represents the current number of layers.

2.5. Hidden Layer Neural Network.

The calculation formula of the neural network structure of a hidden layer is as follows:

$$d_1^{(2)} = f\left(a_{11}^{(1)} x_1 + a_{12}^{(1)} x_2 + a_{13}^{(1)} x_3 + b_1^{(1)}\right),$$
$$d_2^{(2)} = f\left(a_{21}^{(1)} x_1 + a_{22}^{(1)} x_2 + a_{23}^{(1)} x_3 + b_2^{(1)}\right),$$
$$d_3^{(2)} = f\left(a_{31}^{(1)} x_1 + a_{32}^{(1)} x_2 + a_{33}^{(1)} x_3 + b_3^{(1)}\right),$$
$$h_{a,b}(x) = d_1^{(3)} = f\left(a_{11}^{(2)} d_1^{(2)} + a_{12}^{(2)} d_2^{(2)} + a_{13}^{(2)} d_3^{(2)} + b_3^{(2)}\right). \tag{5}$$

The simplified expression of the formula makes it clearer. Here, we use symbols to represent the result of adding the input weights of unit I in layer l and the offset items, where the calculation formula of unit I in layer 2 can be written as follows:

$$h_i^2 = \sum_{j=1}^{n} a_{ij}^{(l)} x_j + b_i^{(1)}, d_i^{(2)} = f\left(z_i^{(2)}\right), \tag{6}$$

$$f\left(\left[h_1, h_2, h_3\right]\right) = \left[f(h_1), f(h_2), f(h_3)\right]. \tag{7}$$

The parameters can be matricized, and the extended expression of the excitation function can be expressed by a vector, which is equation (7). The package of equation contains the detailed calculation steps from the input to the output of the neural network with a single hidden layer, which is called the forward propagation algorithm of the neural network. The core of forward propagation algorithm is to calculate the corresponding intermediate excitation value for each layer. The simplified expression is as follows: assume that the excitation value of the floor l is given as $h^{(l)}$, and then, the calculation of the activation value of the floor $l + 1$ is $h^{(l+1)}$, which is given as follows:

$$h^{(l+1)} = a^{(l)} d^{(l)} + b^{(l)},$$
$$d^{(l+1)} = f\left(z^{(l+1)}\right). \tag{8}$$

Figure 3 shows a simple neural network with hidden layers and output units. n_i is used to represent the number of layers contained in the neural network, L_1 represents the input layer, and L_{ni} represents the output layer. The forward propagation process of the multi-hidden layer neural network is shown in Figure 3, and the key point of the network calculation lies in the calculation of the excitation value of each layer.

2.6. The Activation Function. Sigmoid and tanh are two activation functions commonly used in neural networks, which make neural networks to have nonlinear mapping ability. A large value in the central region means that these two functions have a strong enhancement effect on the signal in the central region, but relatively small enhancement effect on the signals on both sides, so a good mapping effect can be achieved in the feature space. The expressions of sigmoid function and tanh function are as follows:

$$f(h) = \text{sigmoid} = \frac{1}{1 + \exp(-h)},$$
$$f(h) = \tanh(h) = \frac{e^h - e^{-h}}{e^h + e^{-h}}. \tag{9}$$

As shown in Figure 4, it can be concluded that the central zone of sigmoid function and tanh function and the active state of neurons in the human brain are similar, and the lateral region of both functions is similar to the inhibitory state of neurons in human brain. Therefore, in the process of traditional neural network learning, the important features of data to be learned are concentrated in the central region; relatively insignificant features tend to be gradually diluted.

The mathematical model of ReLU activation function is shown in the figure. The introduction of the activation function is to increase the nonlinearity of the Q model of the neural network. Without the activation function, each layer is equivalent to matrix multiplication. The output of each layer is a linear function of the input of the upper layer. No matter how many layers of the neural network, the output is a linear combination of the input; that is, the most primitive perceptron adds an activation function and introduces nonlinear factors to the neuron Q. The neural network can arbitrarily approximate any nonlinear function, so that the neural network can be applied to many nonlinear models. The formula is as follows:

$$f(x) = \max(0, x). \tag{10}$$

ReLU activation function will selectively respond to the input signal, and part of the signal will be shielded, as shown in Figure 5. After repeated tests, researchers obtained the effectiveness of ReLU function in learning sparse signal representation and improving the accuracy of learning. ReLU and tanh activation functions were applied to carry out comparative experimental tests in the ALEX network, and the experimental results show that ReLU greatly reduces the training cycle of the network. A simple CNN model with a total of four layers is built on the CIFAR 10 data set to carry out experimental tests, and it is found that the ReLU function has great advantages in improving the network learning speed and improving the recognition rate. Therefore, the ReLU activation function is generally selected in the neural network.

3. Multilayer Feature Fusion Face Recognition Experiment of CNN Experiments

3.1. Experimental Data. In this study, the ORL and Yale data sets are selected as the training sets of the subsequent deep neural network. The basic introduction of the two data sets is as follows:

(1) ORL Face Data Set. The ORL face data set was compiled by AT&T Laboratories at the University of Cambridge, UK. The face database contains a total of 40 people, each of whom has 10 grayscale images of faces. All face photographs in the data set are positive and have a uniform black background. To collect their face features more comprehensively, each photograph has certain angle changes and expression changes.

(2) Yale Face Data Set. Yale, also collected by AT&T Labs, took 165 grayscale images of faces, including 11 created for each of 15 people in different poses, angles, and lighting conditions.
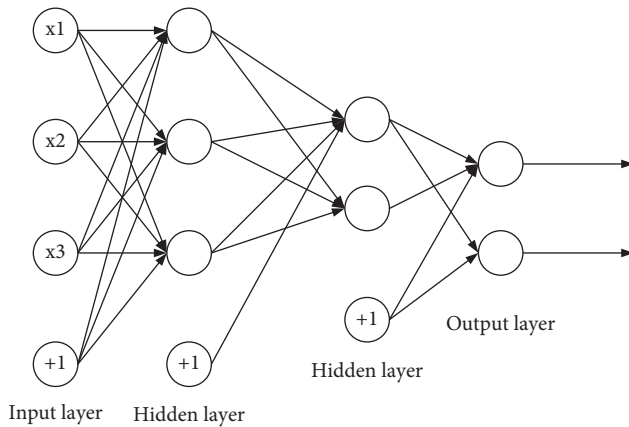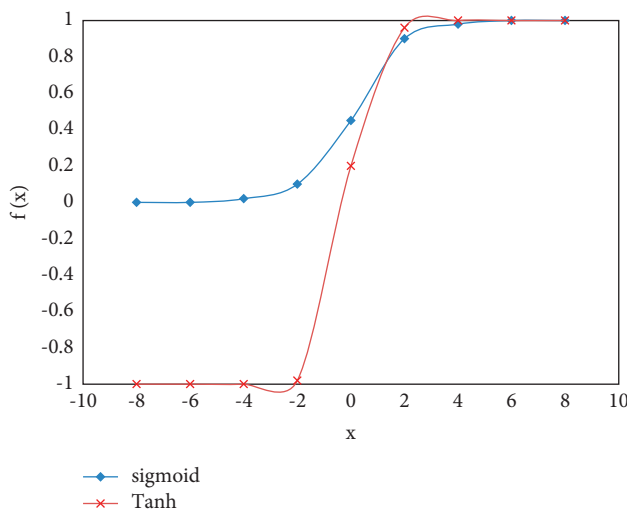
Figure 3: Multi-hidden neural network structure.

Figure 4: Sigmoid and tanh function model.

3.2. The Experimental Steps. The selected face data set is used for testing. In the experiment, two activation functions were selected to test the influence on the model, namely ReLU and sigmoid. The fused face feature *T* obtained from the fusion feature model was sent into the classifier, and the recognition rate was estimated by the support classifiers and the recognition rate was compared and analyzed with that of the basic model. At the same time, it pays attention to the execution efficiency of each method while comparing the accuracy of different methods.

Finally, each of the convolutional pooling features is extracted and fused to get the final face features, and the face features are input into SVM and KNN classifier to carry out face recognition test, to build a huge facial feature database, for the next extraction, utilization, and comparison.

4. Comparative Analysis Results of Recognition Rate

4.1. ORL Data Set Experimental Results and Analysis. In the experiment, 6 images of each experimenter were selected

from the ORL data set using the principle of randomness for training, and the remaining 4 images of each person were used for testing, so there were 240 training pictures and 160 test pictures in total. The method was repeated 10 times to obtain the final identification accuracy and detection consumption time. In Table 1, the recognition rate and time efficiency are obtained under the basic model, and in Table 2, they are obtained under the SVM classifier.

First of all, it can be seen from Table 1 that the method of the base model has a great influence on the time efficiency of the algorithm in the selection of activation function. When sigmoid function is the activation function of the base model, the efficiency of the algorithm is lower than that of ReLU activation function. The time efficiency of the feature fusion method is shown in Table 2, and the time consumed by the algorithm is higher. A represents CNN (ReLU) + PCA, B represents CNN (sigmoid) + PCA, C represents CNN (ReLU) + PCA + LDA, and D represents CNN (sigmoid) + PCA + LDA, E represents CNN(ReLU) + PCA + LPP, and finally F denotes CNN (sigmoid) + PCA + LPP.

Tables 1 and 2 are represented by CNN activation function, and it is effective to use four RLU functions in sigmoid function gradient.

To further verify the advantages of feature fusion model in feature extraction, KNN classifier is used to estimate the recognition rate and efficiency of the algorithm. In Table 3, the classification effect of A and B, C and D, and E and F on KNN is better than that of the method using the basic model, and the recognition rate of C and D reaches 97.6%. However, the algorithm takes more time than the basic model.

Different dimensions are set in the dimension of output function of basic model (CNN) method (the activation function is RLU) to classify and identify them. Feature fusion models CNN + PCA + LDA and CNN + PCA + LPP are used to extract features with different feature dimensions by reducing the dimension input into SVM classifier. Therefore, Figure 5 can be obtained.

From the trend analysis in Figure 6, the recognition rate of different characteristic variable dimensions is 13.0%, indicating that the eigen even function model responds quickly to the change in characteristic variable dimensions.

4.2. Yale Data Set Experimental Results and Analysis. In the experiment, seven photographs are extracted from the selected data set, and the remaining four photographs are trained and tested. There are 105 training images and 60 test images. Table 4 shows the use of the basic model in the light color data set.

In terms of activation function, choosing ReLU as the network structure has a better effect. In the basic model, the method recognition rate is 88.9% when using ReLU function, which is also higher than that using sigmoid function. Moreover, in terms of time efficiency, using ReLU activation function is also higher than sigmoid function. This is because LDA is a supervised learning method, which can effectively use the label information of samples. LPP algorithm and PCA algorithm are not
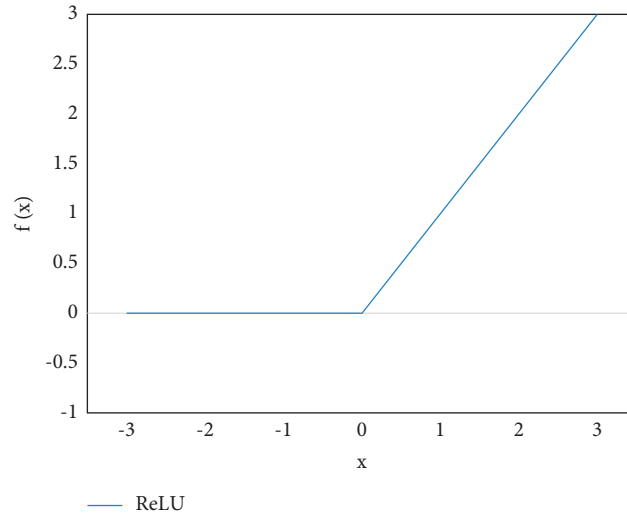
FIGURE 5: ReLU function model.

TABLE 1: Recognition rate and time efficiency under the basic model.

Algorithm model	Recognition rate (%)	Time efficiency
Base model CNN (ReLU)	87.6	120 min 38 s
Base model CNN (sigmoid)	85.9	125 min 6 s

TABLE 2: Recognition rate and time efficiency.

Classifier	Algorithm model	Recognition rate (%)	Time efficiency
	A	96.3	129 min 21 s
	B	90.3	×
	C	97.8	136 min 47 s
SVM	D	92.7	×
	E	90.2	137 min 3 s
	F	91.1	×

TABLE 3: Recognition rate and time efficiency of different algorithm models under KNN classifier.

Classifier	Algorithm model	Recognition rate (%)	Time efficiency
	CNN + PCA	96.1	139 min 21 s
KNN	CNN + PCA + LDA	97.6	147 min 28 s
	CNN + PCA + LPP	95.8	157 min 3 s

traditional learning algorithms. The time of feature model function fusion is also higher than that of other supervised learning algorithms. Experimental results on the Yale data set also show that the feature fusion model mentioned is more effective in face feature extraction, and the corresponding efficiency is reduced. Similar to the experiment on the ORL data set, the recognition results of different feature fusion model algorithms under KNN classifier are shown in Table 5.

Under the KNN classifier, the basic model only selects the ReLU function. As shown in Table 6, the recognition rate of the method of C also reaches the highest 96.8%. The time efficiency of the algorithm is similar to that of the ORL data set, while the efficiency of the feature fusion model is lower than that of the basic model. In conclusion, the feature fusion model algorithm proposed in this study is more effective than the face features extracted by the basic model method, but the execution efficiency of the algorithm is somewhat reduced.
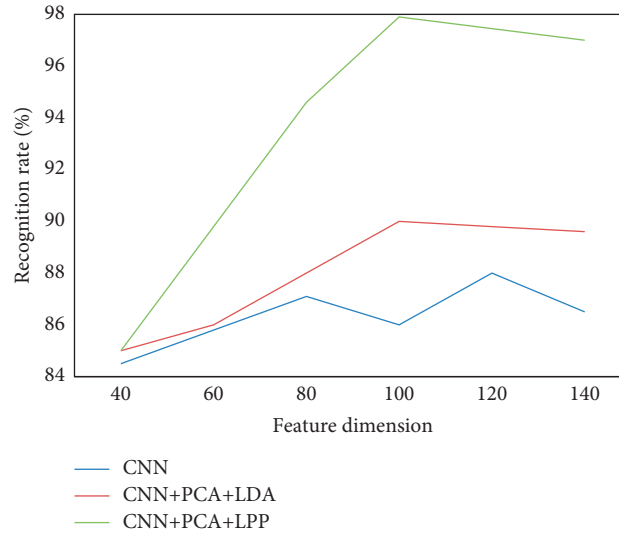
Figure 6: Comparison of discrimination rates in different feature dimensions.

Table 4: Recognition rate and time efficiency under the basic model.

Algorithm model	Recognition rate (%)	Time efficiency
Base model CNN (ReLU)	88.9	65 min 38 s
Base model CNN (sigmoid)	85.6	67 min 6 s

Table 5: Recognition rate and time efficiency of different algorithm models under SVM classifier.

Classifier	Algorithm model	Recognition rate (%)	Time efficiency
	A	96.3	68 min
	B	94.9	×
	C	98.6	60 min 47 s
SVM	D	96.2	×
	E	91.8	69 min 3 s
	F	89.3	×

Table 6: Recognition rate and time efficiency of different algorithm models under KNN classifier.

Classifier	Algorithm model	Recognition rate (%)	Time efficiency
	CNN + PCA	94.2	76 min 12 s
KNN	CNN + PCA + LDA	96.8	84 min 28 s
	CNN + PCA + LPP	91.2	80 min 6 s

5. Conclusion

As one of the deep learning methods, CNN can effectively combine the two tasks of face image feature extraction and classification and recognition and has an excellent performance in image classification. Compared with the basic model, it is improved by 10.4% and 9.6%, respectively. However, the disadvantage brought by the improvement of the recognition rate is the reduction in the algorithm execution efficiency, and the online detection and recognition system designed based on this model meets the design requirements, but there are still some shortcomings, which can be further studied in the following aspects: (1) the feature fusion model in this study can adopt more machine learning methods on the feature extraction of each convolutional pooling feature map instead of being limited to PCA, LDA, LPP, etc.; the efficiency of the algorithm is the next step to optimize, that is to improve the model recognition rate without losing the efficiency of the algorithm or even improve the execution efficiency. (2) In the process of online detection and recognition, there are interference factors such as complex face background and different quality of captured images, so it is necessary to explore appropriate machine learning methods to reduce the impact of these factors. In addition, the design of face online detection and recognition system can be combined with specific practical life applications.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] C. N. Vanitha, S. Suwathika, A. Mathura, and P. Saranyadevi, "Facial recognition processing using uniform pattern histogram with AI in multimedia applications," *Solid State Technology*, vol. 64, no. 2, pp. 788–798, 2021.

[2] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.

[3] D. L. Wang and J. Chen, "Supervised speech separation based on deep learning: an overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702–1726, 2017.

[4] A. K. Singh, A. Anand, and Z. Lv, "A survey on healthcare data: a security perspective," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 2, 2021.

[5] K. Jai Raman, A. Azman, W. Kit Wen et al., "Driver face recognition: anti-theft system," *Indian Journal of Science and Technology*, vol. 11, no. 20, pp. 1–6, 2018.

[6] T. Ibrayev, U. Myrzakhan, O. Krestinskaya, A. Irmanova, and A. P. James, "On-chip face recognition system design with memristive Hierarchical Temporal Memory," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 3, pp. 1393–1402, 2018.

[7] Y. Zhang, X. Xiao, L. X. Yang, Y. Xiang, and S. Zhong, "Secure and Efficient Outsourcing of PCA-Based Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 15, 2019.

[8] R. J. Bennetts, E. Murray, T. Boyce, and S. Bate, "Prevalence of face recognition deficits in middle childhood," *Quarterly Journal of Experimental Psychology*, vol. 70, no. 2, pp. 234–258, 2017.

[9] W. Deng, J. Hu, Z. Wu, and J. Guo, "Lighting-aware face frontalization for unconstrained face recognition," *Pattern Recognition*, vol. 68, no. C, pp. 260–271, 2017.

[10] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[11] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep Learning Models for Real-Time Human Activity Recognition with Smartphones," *Mobile Networks and Applications*, vol. 25, no. 1, pp. 1–13, 2019.

[12] K. Guo, S. Wu, and Y. Xu, "Face recognition using both visible light image and near-infrared image and a deep network," *Caai Transactions on Intelligence Technology*, vol. 2, no. 1, pp. 39–47, 2017.

[13] J. Lee, "Integration of digital twin and deep learning in cyber-physical systems: towards Smart Manufacturing," *IET Collaborative Intelligent Manufacturing*, vol. 38, no. 8, pp. 901–910, 2020.

[14] H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," *Journal of Lightwave Technology*, vol. 30, no. 5, pp. 713–726, 2021.

[15] L. June-Goo, J. Sanghoon, C. Young-Won et al., "Deep learning in medical imaging: general overview," *Korean Journal of Radiology*, vol. 18, no. 4, pp. 570–584, 2017.

[16] D. Ravi, C. Wong, F. Deligianni et al., "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.

[17] K. Quevedo, M. Harms, M. Sauder et al., "The neurobiology of self face recognition among depressed adolescents," *Journal of Affective Disorders*, vol. 229, pp. 22–31, 2018.

[18] Y. Xu, Z. Li, B. Zhang, J. Yang, and J. You, "Sample diversity, representation effectiveness and robust dictionary learning for face recognition," *Information Sciences*, vol. 375, no. C, pp. 171–182, 2017.

[19] B. Rossion, "Humans are visual experts at unfamiliar face recognition," *Trends in Cognitive Sciences*, vol. 22, no. 6, pp. 471-472, 2018.