Research Article

Yanyu Liu, Liu Zhang*, and Dawei Zhang

Face recognition method based on convolutional neural network and distributed computing

https://doi.org/10.1515/jisys-2024-0121
received March 07, 2024; accepted June 19, 2024

Abstract: Face attribute recognition is also widely used in human–computer interaction, train stations, and other fields because face attributes are rich in information. A convolutional neural network and distributed computing (DC)-based face recognition model is researched to enhance multitask face recognition accuracy and computational capacity. The model is also trained using a multitask learning method for face identity recognition, fatigue state recognition, age recognition, and gender recognition. The model is streamlined by DC to improve the computational efficiency of the model. To further raise the recognition accuracy of each task, the study combines the add feature fusion (FF) principle and concat FF principle to propose an improved face recognition model. The outcomes of the study revealed that the true acceptance rate value of the model reached 0.95, and the face identity recognition accuracy reached 100%. The fatigue state determination accuracy reached 99.12%. The average recognition time of a single photo was 354 ms. The model can quickly recognize face identity and face attribute information in a short time, with short time and high recognition accuracy. It is beneficial in several areas, including intelligent navigation and driving behavior analysis.

Keywords: convolutional neural network, face recognition, multitask learning, feature fusion

1 Introduction

Emotions are people's emotional expression of objective events, and facial expressions are a universal conduction signal for emotions. With the rapid development of traffic, economy, and trade, as well as the continuous improvement of people's living standards, cars, trucks, and other vehicles are increasingly popular and increasing [1]. At the same time, the traffic accident rate is also increasing, in which driver fatigue driving is an important cause of traffic accidents. Therefore, it is particularly important to detect driver fatigue and give corresponding warnings to driver fatigue driving to ensure driver safety [2]. In recent years, with the rapid development of artificial intelligence technology, artificial intelligence has become the core driving force for a new round of industrial reform and has been widely used in speech recognition, text recognition, image recognition, machine manufacturing, and other fields [3]. Multitask face recognition technology not only meets the single task of face identity recognition, but also needs to find more useful face attribute information from a face photo, such as age, gender, and fatigue state. In public transportation, if the driver's basic information set driving state can be identified through real-time face information detection, public transportation safety will be further maintained [4]. Existing methods often use numerous models in parallel to accomplish multiple

* **Corresponding author: Liu Zhang,** School of Electronic Information Engineering, Beihai Vocational College, Beihai, 536000, China, e-mail: zhangliu@bhzyxy.edu.cn
Yanyu Liu: School of Electronic Information Engineering, Beihai Vocational College, Beihai, 536000, China, e-mail: liuyanyu@bhzyxy.edu.cn
Dawei Zhang: School of Electronic Information Engineering, Beihai Vocational College, Beihai, 536000, China, e-mail: zhangdawei@bhzyxy.edu.cn

recognition tasks in order to attain multitask face recognition (FR). Nevertheless, the training period for this strategy is sluggish and expensive [5]. To address this problem, the research utilizes distributed computing (DC) and convolutional neural networks (CNN) to construct a new FR model. The model introduces the idea of multitask learning (MTL), aiming to realize tasks such as face identity recognition and fatigue state recognition. The innovation of the research is to combine the principles of add feature fusion (FF) and concat feature fusion (Concat-FF) to propose an FR model with multilayer FF. The study is broken up into four sections. The first phase is the literature review, which examines and evaluates the present level of both domestic and foreign research in the field of study. The second part is the methodology introduction part, which depicts the specific details of the research design model. The third section, titled "Result Analysis," plans experiments to test the built model's functionality and evaluates the model's performance. The fourth part is the conclusion part, which summarizes the analysis results and model details and gives the outlook.

2 Related works

As the most direct emotional expression of people to objective events, facial expression is a very important conduction signal. FR technology, as the main method of emotion capture, has also been the focus of research by experts and scholars. Ismail et al. proposed a Web-based attendance system for the substitution of answering to attendance in a university classroom that uses a pre-trained model based on open-source deep learning. A dynamic FR process was created by extracting features and storing them in an online database. Experiments showed that the FR accuracy of this method reached 96% [6]. Elaggoune et al. artificially optimized the FR algorithm by applying face hybrid descriptors to feature selection. Face feature extraction was performed by combining the Gabor filter with histogram-oriented gradient and local phase quantization. Experimental results indicated that the FR rate of this method reached more than 90% [7]. Jiao et al. discussed the improvement methods available in the literature in order to improve FR performance. They found that when using the existing loss function in CNN, the training dataset has an overfitting problem, which reduces the FR effect. Therefore, they proposed a new loss function, Dyn-arcFace, and experimentally verified that the model using this loss function obtained a more advanced performance [8]. Fussey et al. addressed the controversy of automated facial recognition in the field of police innovation by presenting several new arguments based on insights from the sociology of policing, surveillance research, and science and technology. The study analyzed and clarified how it is constrained by police discretion [9]. A novel Quaternion Discrete Orthogonal Moment Neural Network model was put up by El Alami et al. to address the issue of insufficiently strong descriptors in the field of color FR. Discrete orthogonal moments are employed to extract pertinent and compact characteristics from an image's quaternion representation. The approach yields more than 96.84% classification accuracy in FR, according to experimental results [10].

CNN is a perceptual model with several layers, which is specialized for solving deep learning problems related to the image domain. Han et al. found that traditional CNNs had serious problems with continuous crack termination and misrecognition of background discrete noise in the semantic segmentation of pavement cracks. Based on this problem, they proposed a jump-level round-trip sampling block structure and implemented Li Russian image segmentation using CNN. The experimental analysis revealed that the model with the sample block structure gives better results in pavement segmentation compared to other models [11]. Pooling layers in CNNs can aid neural networks in learning invariant characteristics and lowering computational complexity, according to research by Nirthika et al. According to the results, class-specific feature pairs in relation to image size scale determine the best pooling strategy [12]. Ilesanmi and Ilesanmi investigated several CNN image denoising methods and evaluated these methods using different datasets. The purpose and guiding principles of CNN approaches, as well as a graphic description of some of the most advanced CNN image-denoising techniques, are outlined in a number of works that were chosen for evaluation and analysis [13]. A CNN-based system for handwritten document recognition was introduced by Ghazal. Following image preprocessing, line, word, and character segmentation was used to divide the input content. When these segmented characters were finally delivered to CNN for recognition, the experimental findings demonstrated that, throughout the training phase, the suggested work approximated the outcomes with an accuracy of 93% [14].

Fang developed a hybrid physical information neural network for partial differential equations. The network borrowed ideas from CNNs and finite volume methods and used local fitting methods to achieve automatic solutions of partial differential equations. Experiments indicated that the network was correct and effective [15].

CNN is one of the most widely used approaches in FR, with extensive application experience in image processing, as demonstrated by the aforementioned literature. However, the current computer performance still has the problems of slow processing speed and low accuracy when processing massive images with a lot of data. For this reason, the research constructs a new FR model based on DC and CNN.

3 Face recognition model based on CNN and DC

The research designs and implements a multitask FR model that is capable of identifying faces, recognizing ages, genders, and detecting fatigue states. The combination of CNN and DC aims to produce more efficient and accurate FR.

3.1 Image data processing based on DC

Traditional facial expression recognition methods are usually aimed at a small number of people in controlled environments such as laboratories, resulting in low data analysis efficiency when faced with real-world scenarios. The research mainly focuses on the recognition of drivers' driving status in public transportation systems, which must involve massive image data. In order to realize efficient and high-precision multitask face recognition, this paper first uses a DC method to process image data to improve the efficiency of facial feature extraction. Then, using CNN, multitask the face information in the image data. Different datasets are required to accomplish different learning tasks, and the study uses the CASIA-Webface dataset as the training dataset for face identity recognition. The dataset includes a total of 484,536 identity photos of 10,576 individuals. The IJB-A dataset was used as the test dataset for the face identity recognition task. IMDB Wiki 500k+ and FG-NET were used as the training dataset and test dataset for the face age recognition task, respectively. The study chose the face tiredness data of Beijing bus drivers as the dataset for fatigue state recognition and used Celeba as the dataset for training and testing of face gender recognition. The related dataset may have missing faces or unbalanced data; for this reason, the study first preprocesses the dataset. The processing flow is shown in Figure 1.

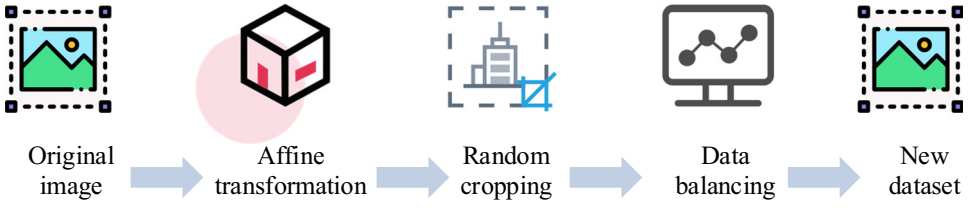

Figure 1: Dataset preprocessing process. Source: Created by the authors.

Different objective functions and training algorithms need to be used for different tasks as a way to achieve the same results as simultaneous training of multiple tasks on multilabeled datasets [16–18]. The forward propagation of FR tasks in multiple FR tasks is shown in the following equation:

$$\text{output}_k = f(W_k X + b_k). \tag{1}$$

In equation (1), f is the activation function, W_k is the convolutional correlation parameter, and the backpropagation can be described as follows:

$$W_{\text{share}} = W_{\text{share}} - \sum_{k=1}^{M} \eta_k \frac{T_k \partial L_k}{\partial W_{\text{share}}}. \tag{2}$$

In equation (2), η_k denotes the learning rate of the kth recognition task, L_k is the loss of the kth recognition task, M is the recognition tasks, T_k is the task weight, and W_{share} is the multitask sharing parameter. In face identity recognition, the study uses the softmax classification function as the objective function for the face identity recognition task. The probability distribution of softmax output is shown as follows:

$$p(y = i|x;\theta) = \frac{e^{\theta_j^T x}}{\sum_{j=1}^{v} e^{\theta_j^T x}}. \tag{3}$$

In equation (3), x is the model output, θ is the convolution kernel and bias term, and y is the output category, n is the category labels, and v is the input layers connected to the softmax layer. The study also uses the softmax-loss as its loss function as follows:

$$\text{Loss}(y, o_y) = -\log(o_y) = -\log\left(\frac{e^{Z_y}}{\sum_{j=1}^{m} e^{Z_j}}\right). \tag{4}$$

In equation (4), y is the real label of the face image, o_y is the probability that the position is y in the probability vector output by softmax, and Z_y and Z_j are the results after convolution operation. After calculating the loss value, backpropagation is performed. For face age recognition, the study designed a unique loss function as follows:

$$\begin{cases} G\_\text{Loss}(y, o_y) = \mu G(y, o_y) + (1 - \mu)\text{Loss}(y, o_y) \\ G(y, o_y) = \left(1 - \exp\left(-\dfrac{(o_y - y)^2}{2\delta^2}\right)\right) \end{cases}. \tag{5}$$

In equation (5), $G(y, o_y)$ is the Gauss loss function, $\text{Loss}(y, o_y)$ is the softmax-loss, μ is the loss weight, which takes the value of 0.5, and δ is the standard deviation of the age distribution of the training dataset. The face age recognition network part outputs a 1 × 100-dimensional softmax probability vector and multiplies the resultant label with the probability of the corresponding position to obtain the final recognition result. This information is displayed as follows:

$$\text{final\_age} = \sum_{j=1}^{K=100} o_j \times j. \tag{6}$$

In equation (6), j is the age labeling category and o_j is the probability value whose position is j in the probability vector output by softmax. Both face gender recognition and fatigue state recognition use the softmax classification function. When the traditional face recognition method is faced with the scene of analyzing multiple face pictures at the same time, it cannot be completed quickly in terms of computational efficiency or processing power. In order to improve the efficiency of research and design of face recognition methods and make them meet the needs of working ability in the case of massive images, a DC method is introduced to improve the computing speed. Hadoop is a collection of sorting, computing, and storing in a distributed information computing system; it can help developers who do not know the parallel operation principle of the case, still can according to their own business logic, set up a cluster system. Therefore, the study introduces Hadoop technology into the field of face recognition. MapReduce first performs a slice Split operation on the uploaded file, but the image itself is indivisible, so the study divides each image into a slice Split. A slice corresponds to a Map task. Image features are extracted in the Map stage and the output is a face feature vector. In Java, files need to use the serialization method in information exchange, need to transfer files in binary format, and then convert the binary data to the original type of file after the transfer is

completed. For this reason, the study uses Hadoop's own development of the sequence framework Writable to serialize data operations. MapReduce processing image process and the data types encapsulated by Hadoop are shown in Figure 2.

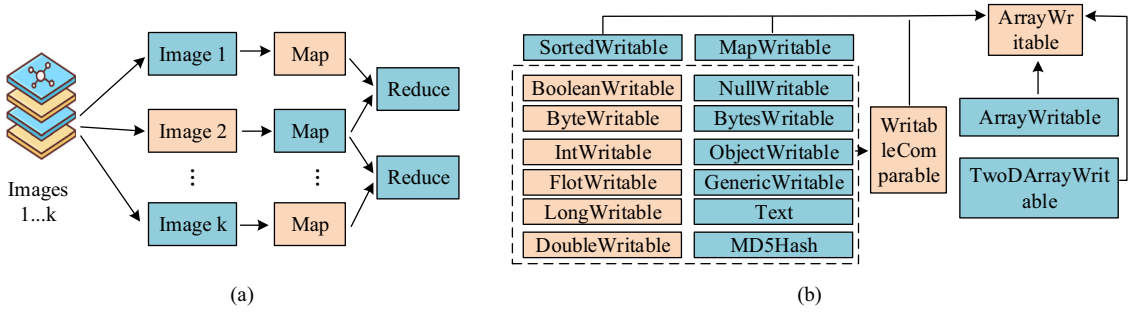

Figure 2: (a) MapReduce image processing flow and (b) Hadoop encapsulated data types. Source: Created by the authors.

However, the traditional Hadoop technology does not take image processing into account and does not provide an image processing interface to the outside world. For this reason, the study designs the image input type before the beginning of the experiment, defines it as ImgWritable inherits the Writable Comparable class, and rewrites the serialization methods write and read File. Use InputFormat to describe the data type of MapReduce jobs. However, there is still no data type for image processing in the InputFormat type provided by MapReduce, but the InputFormat can be customized using MapReduce's extension mechanism. The study inherits the imgInpuFormat class from FileInputFormat and overrides the createRecorder method to modify the Record Reader's generalization to handle image data. Also, rewrite the isSplitable() method to return false; no more slicing of individual image data. Finally, design imgRecorderReader to repackage the key-value pairs. Take the key as the image name and the value as ImgWritable in the actual development; the study found that if the facial feature extraction network is directly set up on a distributed data processing platform for FR, the code execution efficiency is not improved, but rather slower than the running rate when the network is used to extract the facial expression features on a microcontroller. This is due to the fact that the higher the number of small files needed to start, run, and destroy Map tasks more often, which affects the efficiency of Hadoop execution. For this reason, the study uses the Hadoop Image Process Interface (HIPI) image processing tool library to merge facial images into a small number of files for processing. The distributed image data feature extraction method flow is shown in Figure 3.

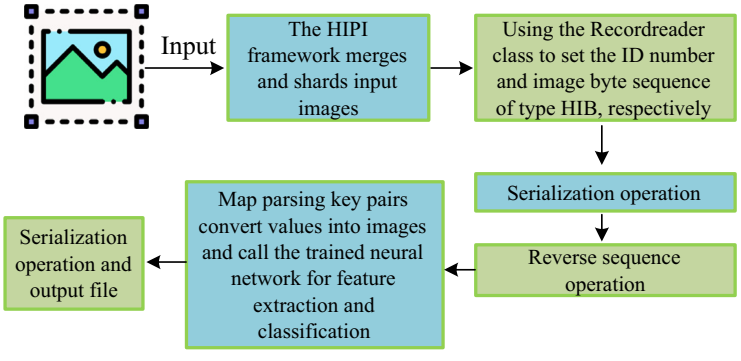

Figure 3: Process flow of feature extraction method for distributed image data. Source: Created by the authors.

3.2 Improved CNN-based face recognition

Before face recognition, the DC method is used to process image data, so as to reduce the time consumed in image feature extraction. After processing the image data based on DC in the previous section, the study uses CNN to perform multitask face recognition on the image data. Traditional multitask face recognition algorithms have low accuracy in each recognition task. Compared with other neural networks, CNN is the main force in the development of current deep neural networks, which can recognize images more accurately than human beings. CNN has a strong image classification ability, and it has a good generalization ability. Therefore, a face recognition algorithm based on deep CNN is proposed [19–21]. There are just nine convolutional layers in the CNN that the study designed overall; the downsampling pooling layer is chosen as the maximum pooling function; the activation function layer is chosen as the maxout activation function; and all training tasks are trained using the softmax function for the pertinent face multiclassification task. The CNN parameter configurations for the four recognition tasks are shown in Table 1.

For single-task CNN training, it is difficult to quickly get the model to focus its attention on the image region that needs attention, which makes it difficult to distinguish between relevant and irrelevant features. However, for multitask CNN training, the recognition tasks with correlation help each other during the training process. For this reason, the study introduces the attention-focusing mechanism to specify the training order according to the complexity of the recognition tasks from high to low, which significantly reduces the training time. The work uses the eavesdropping method to jointly gather the pertinent features of several recognition tasks in an effort to enhance the model's recognition performance. Meanwhile, the study introduces a hidden data addition mechanism to develop the training order according to the amount of training data from the most to the least, which in turn supplements the dataset for the recognition task with less data. Furthermore, a bias method is used to obtain the required generalized features directly from the pooling layer's output following the final convolutional layer. Based on the suitable adversarial mechanism, the rationality of face identity identification and face age recognition tasks coexisting is understood. To improve the accuracy of the FR task further, multilayer FF on top of the traditional CNN is being investigated and the MTL training strategy adjustment. The comparison between the traditional CNN and the CNN network after multilayer FF is shown in Figure 4.

There are two basic principles of CNN multilayer FF, which are the add principle and the concat principle. The basic idea of the add principle is that a combination of feature maps is added, while the number of channels remains unchanged. The basic idea of the concat principle is that combining channel numbers changes the number of channels. Assuming two inputs, after the fusion of the add method, the convolution output at each position is shown in equation (7).

$$Z_{\text{add}} = \sum_{i=1}^{c} (X_i + Y_i) \times W_i. \tag{7}$$

In equation (7), X_i and Y_i are the two channel inputs, W_i is the convolution kernel, and c is the amount of input data for each channel. After fusion based on the concat principle, the convolutional output at each position is shown in equation (8).

$$Z_{\text{concat}} = \sum_{i=1}^{c} X_i \times W + \sum_{i=c+1}^{2c} Y_i \times W. \tag{8}$$

The feature maps corresponding to the add fusion principle share a convolutional kernel, which is equivalent to adding a prior. The concat principle also corresponds to different convolutional kernels for each channel. The study is based on the concat fusion principle to select appropriate CNN features at different levels for fusion. The FF process can be described by equation (9).

$$f_{\text{concat}} = [f_{\text{down}}, f_{\text{middle}}, f_{\text{up}}]. \tag{9}$$

In equation (9), f_{down}, f_{middle}, and f_{up} are the bottom, middle, and top level features in the model, respectively, and [] denotes the successive operations on the quantitative dimension of the vector. The study selected

Table 1: The CNN parameter configurations for the four identification tasks

Parameter		Small batch number	Basic learning rate	Weight attenuation	Momentum	Initialization strategy	Learning rate change strategies
Face identity recognition	Configure the parameter name	Batch_size	Base_lr	Wight_decay	—	—	Lr_policy
	Value	64	0.01	0.001	0.9	Xavier	Inv
Face age recognition	Configure the parameter name	Batch_size	Base_lr	Wight_decay	—	—	Lr_policy
	Value	64	0.003	0.001	0.9	Xavier	Inv
Face gender recognition	Configure the parameter name	Batch_size	Base_lr	Wight_decay	—	—	Lr_policy
	Value	64	0.003	0.001	0.9	Xavier	Inv
Face fatigue recognition	Configure the parameter name	Batch_size	Base_lr	Wight_decay	—	—	Lr_policy
	Value	32	0.001	0.001	0.9	Xavier	Inv

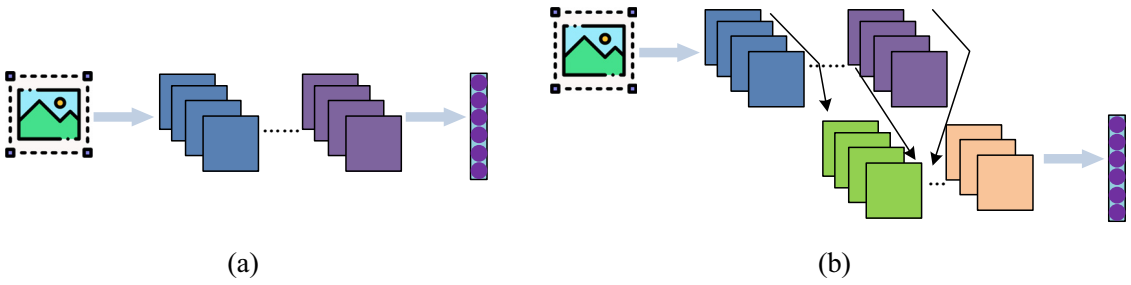

Figure 4: Comparison of (a) traditional CNN and (b) NN networks after multilayer feature fusion. Source: Created by the authors.

three levels of features to be fused and changed only in quantity. Before Concat-FF, the width and height of all concat input features need to be normalized to the same size, and the study uses a convolution operation to do this. After selecting the appropriate level, the study chose to let the multitasking CNN model itself do the feature selection and combination based on the ADD fusion principle. Different levels of CNN features are suitable for different tasks, and it is considered that it is not appropriate to let one feature or multiple features to accomplish a certain task. Therefore, the study designed a convolutional layer with a convolutional kernel size of 1 × 1 after Concat-FF, with the number of convolutional kernels determined by the number of features selected. The fused convolution formula is shown in equation (10).

$$\begin{aligned} f_i &= f_{\text{concat}} \times W_i + b_i \\ &= f_{\text{down}} \times W_i^{\text{down}} + f_{\text{middle}} \times W_i^{\text{middle}} + f_{\text{up}} \times W_i^{\text{up}}. \end{aligned} \tag{10}$$

In equation (10), f_i is the ith output when feature selection is completed, and W_i and b_i denote the convolution parameters and bias of the ith convolutional neuron, respectively. In summary, the study adds an FF network component to the CNN model, as shown in Figure 5.

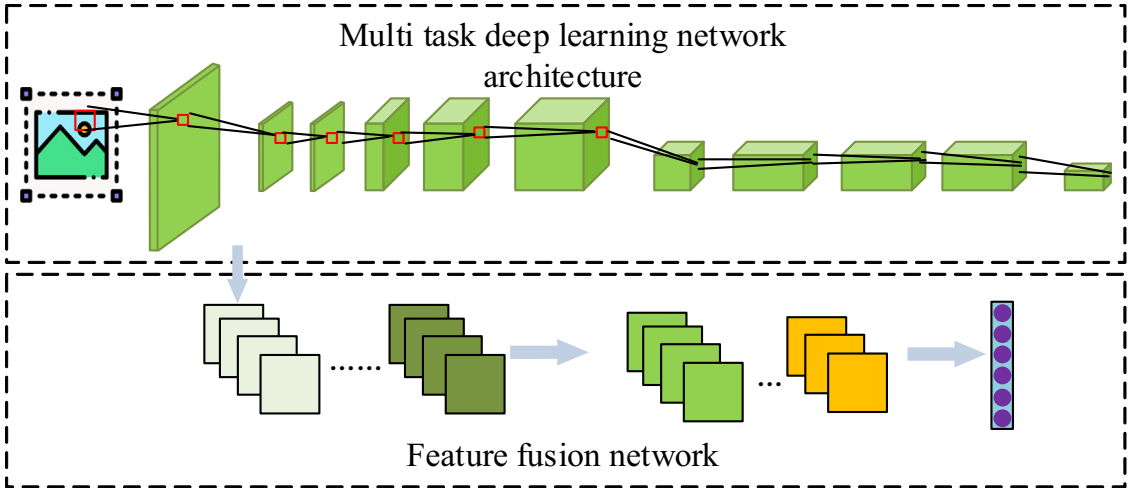

Figure 5: Structure of facial recognition model based on improved CNN. Source: Created by the authors.

The improved CNNFR network structure selects features output from the pooling layer of different convolutional layers, and in order to allow one or more features that are best suited for a particular task to accomplish this task, the study designed a layer of convolutional layer Conv6 with parameter dimensions of 256 × 1 × 1, and let the CNN itself do the feature selection. In order for the features not to lose local information, the study did not use any pooling operation in the fusion network. Furthermore, the study could only achieve

dimensionality reduction sampling by performing convolutional selection on the number of features, hence reducing the computing effort of the FF network.

4 Performance analysis of multitask face recognition model based on CNN and DC

4.1 Experimental environment and parameter setting

The operating system is Microsoft Windows 10 (64-bit), the CPU parameters are Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz(3,600 MHz), the memory is 2,400 MHz, and the graphics card is Radeon (TM) RX550. The development language is Python, and the development tools are PyCharm 2020 and Eclipse. The IJB-A dataset is used as a test dataset for the face identification task. IMDB Wiki 500k+ and FG-NET were used as training and test data sets for face age recognition tasks, respectively. Celeba was used as the dataset for training and testing facial gender recognition, and facial fatigue data of bus drivers in Beijing was selected as the dataset for fatigue state recognition. The indexes involved in the experiment include training loss, recognition error, mean absolute error (MAE), recognition accuracy (acc), true accept rate (TAR), and false accept rate.

4.2 Operational efficiency analysis of the multitask face recognition model

The study trains the model for each of the four distinct recognition tasks in order to evaluate the model's training. The overfitting or underfitting models are also eliminated through training. The loss convergence of the training process for the four different recognition tasks is shown in Figure 6.

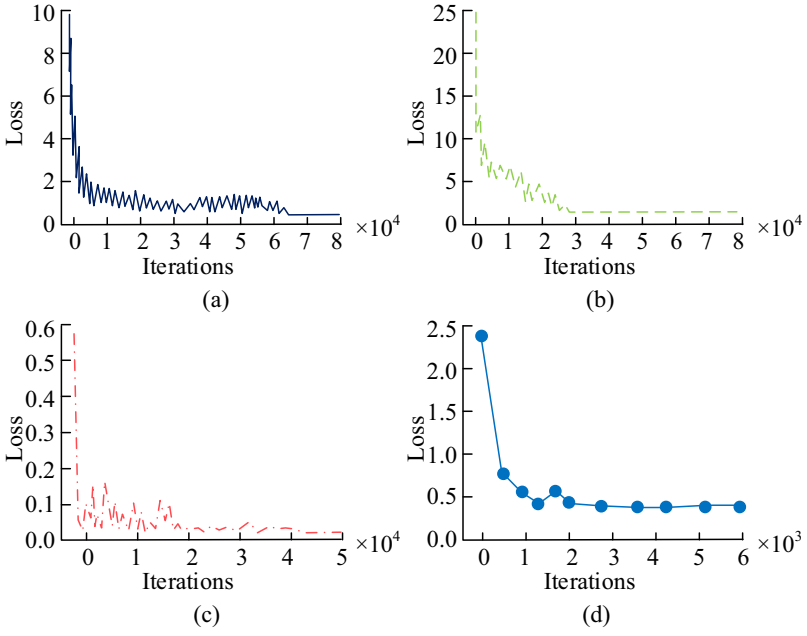

Figure 6: Convergence of loss in the training process of four different recognition tasks. (a) Facial recognition task. (b) Facial age recognition task. (c) Facial gender recognition task. (d) Facial fatigue state recognition. Source: Created by the authors.

In Figure 6(a), the convergence time of the face identity recognition training process is longer, and it starts converging after training up to 67,645 times. In Figure 6(b), face age recognition training to 20,866 times starts converging. In Figure 6(c), face gender recognition convergence occurs after training up to 11,005 times. In Figure 6(d), face fatigue recognition reaches the optimal target loss value after training up to 2,045 times. Comprehensively, Figure 6 shows that all the face attribute recognition tasks converged to the optimal model very quickly, except for the beginning face identity recognition task, which had a longer convergence time to reach the optimal model. The reason why face identification training is the most frequent is that the task is more complex and higher, and the more complex the task is, the more generalized and detailed the extraction of face features, which can provide more information for the subsequent task. Therefore, the training task order is set as face identity recognition, face age recognition, face gender recognition, and face fatigue state recognition.

In order to test the effect of the DC method used in the study on the efficiency of the model operation, the study judges the cluster performance of the model. In the cluster parallelization experiments, the study conducted a stand-alone comparison experiment, an acceleration ratio experiment, and a platform scalability experiment, respectively. The results are shown in Figure 7.

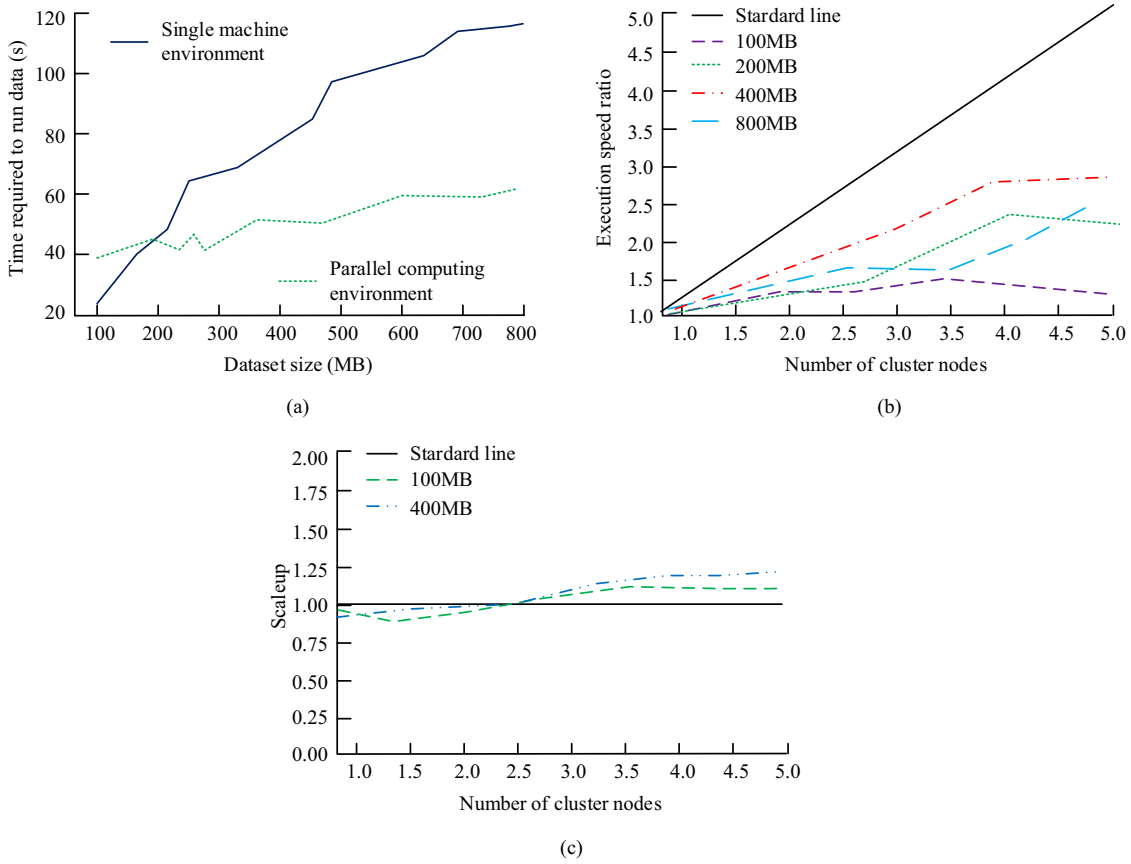

Figure 7: Experimental results of cluster parallelization for DC methods. (a) Comparison between single machine feature extraction and parallelization. (b) Acceleration ratio experiment. (c) Scalability experiment. Source: Created by the authors.

In Figure 7(a), the efficiency of parallelized feature extraction of face images using the Hadoop platform built with clusters is significantly improved, which indicates the feasibility of Hadoop-based FR. In Figure 7(b), the experiment increases the number of cluster nodes sequentially by not changing the data size. All four folds are found to be in a growing trend, which proves that the method is effective. The results of setting four nodes and setting five nodes appear to be almost the same in the 200 MB size of the experimental folds, which is

related to the size of the data block. In Figure 7(c), two sets of comparison curves 400 and 200 MB are used for the experiment, from which it can be seen that the relative running time grows slowly with the increase of the number of nodes and the data size, and the Scaleup value shows an overall increasing trend. It can be seen that 200 MB scaleup data has better scalability on the cluster compared to 400 MB scaleup data. Based on the above content, it can be seen that the research uses the HIPI framework to merge image data, so that several images are merged into a block size for Map processing, which greatly reduces the consumption of system resources. At the same time, DC can effectively improve computing efficiency on different nodes.

4.3 Performance analysis of the multitask face recognition model

The study constructs an FR model based on CNN and DC through the MTL mechanism. In order to test the advantage of using the multitasking mechanism, the study compares the recognition error values of the single-tasking mechanism and the multitasking mechanism model, and the results are shown in Figure 8.

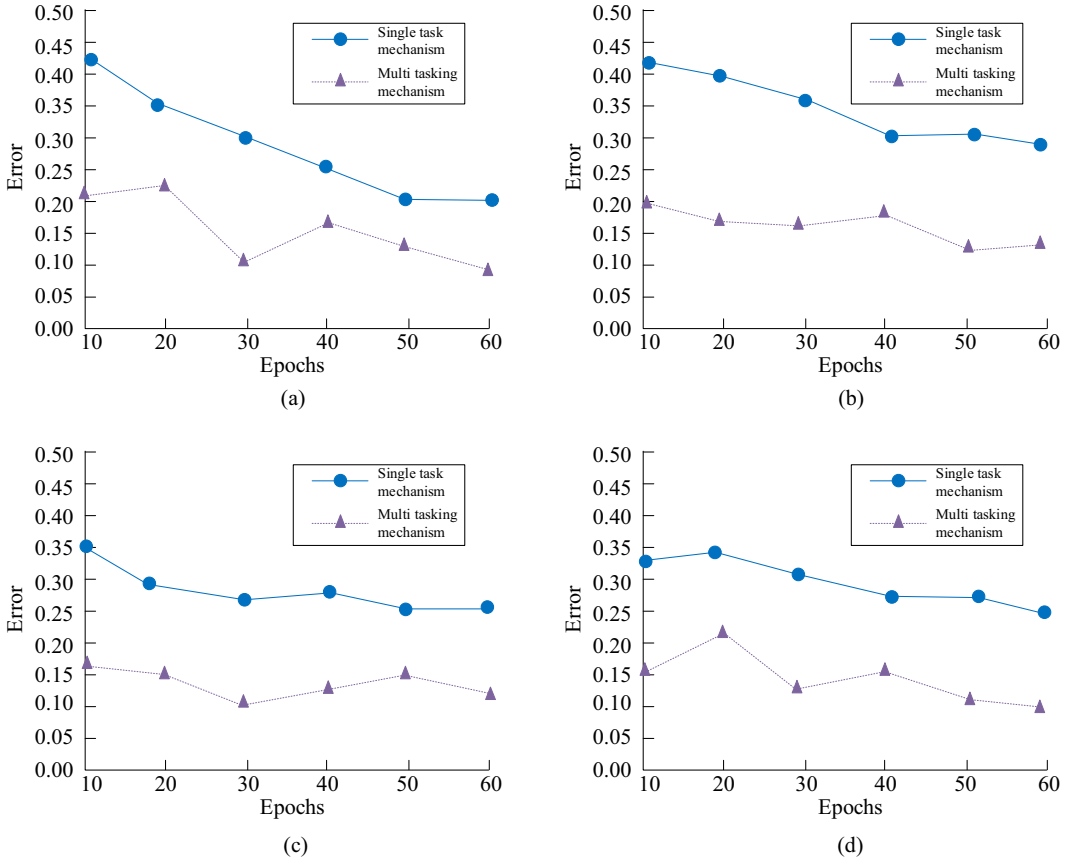

Figure 8: Recognition error values of single-task mechanism and multitask mechanism models. (a) Facial recognition task. (b) Facial age recognition task. (c) Facial gender recognition task. (d) Facial fatigue state recognition. Source: Created by the authors.

In Figure 8(a), the error of the multitasking mechanism fluctuates around 0.2, while the error value of the model with the single-tasking mechanism fluctuates between 0.32. The multitasking mechanism's error profile in Figure 8(b) is noticeably smaller than the single-tasking mechanism's. The multitasking mechanism's error curves vary below the single-tasking mechanism, as seen in Figure 8(c) and (b). This indicates that the multi-tasking mechanism is able to fully exploit the knowledge found in other jobs. In order to test the

reasonableness of the study's improvement approach to the multitask CNN model, the study introduces the MAE and recognition accuracy (acc) to compare the performance of the model before and after the improvement. The recognition results of the model before and after improvement under four recognition tasks are shown in Figure 9.

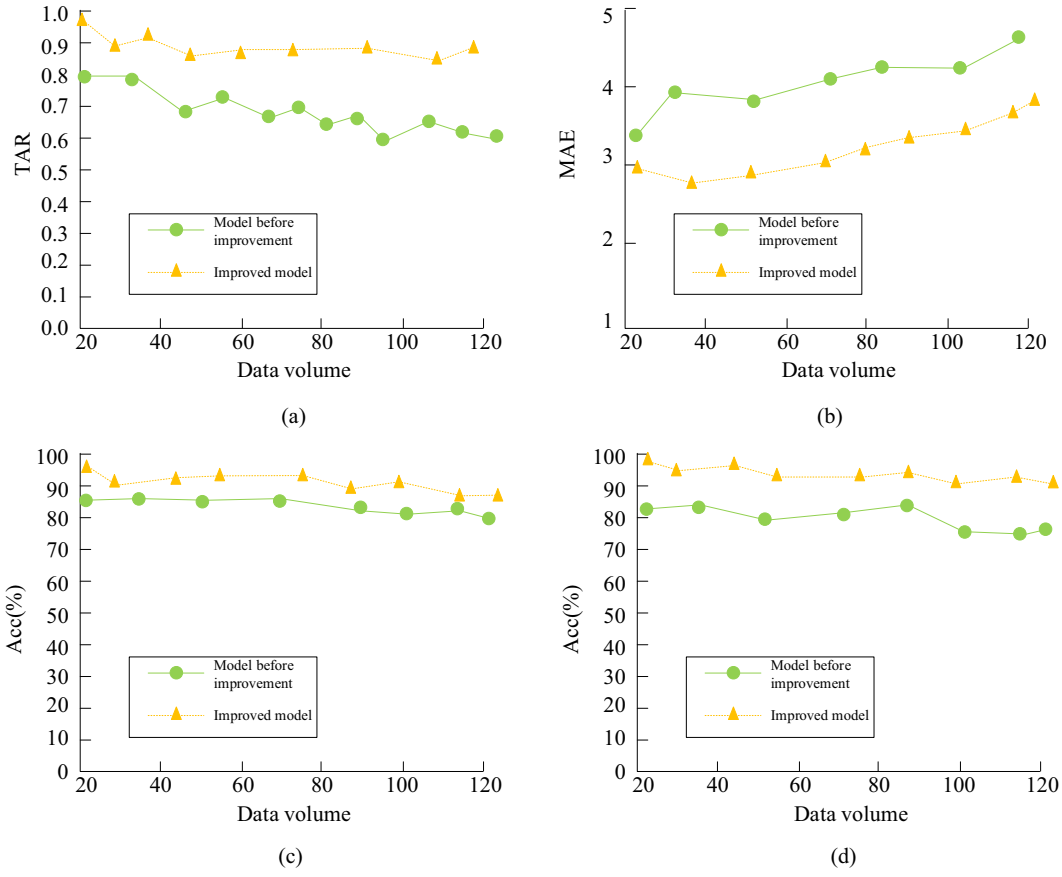

Figure 9: Performance comparison of face recognition models before and after improvement. (a) Facial recognition task. (b) Facial age recognition task. (c) Facial gender recognition task. (d) Facial fatigue state recognition. Source: Created by the authors.

As demonstrated in Figure 9(a), the enhanced model's identity identification accuracy has increased when compared to the unimproved model, reaching 97% recognition accuracy. In Figure 9(b), in the MAE value of age recognition, the improved model is only 3.22, which is much lower than the pre-improved model. The gender recognition rate of the enhanced model is 98.9% in Figure 9(c), 1.4% higher than the pre-improved model. The accuracy of the improved model in identifying the fatigue state is 98.7% in Figure 9(d), an improvement of 0.8% over the pre-enhanced model. As can be summarized in the figure, the performance of the model in each FR task was improved with the addition of the multilayer FF network. To further examine the performance of the FR model (Model 1) designed for the study. Since Model 1 is a multitask FR model, for this reason, the study compares Model 1 with commonly used recognition models in different domains under four recognition tasks. In the face identification task, the study comparison models include a face recognition model based on unconstrained datum (Model 2), visual geometry group-face (Vggface), and unconstrained face recognition based on deep CNN features (Model 3). In the face age recognition task, the study uses the hierarchical model of automatic age estimation considering external factors (Model 4), face age estimation based on label-sensitive learning (Model 5), and face age recognition based on local ordering (Model 6) to compare the models with the study. In the face gender recognition task, the study selected a gender

recognition model based on depth attribute pose alignment (Model 7), a face gender recognition model based on deep learning (Model 8), a face gender recognition model based on face feature representation and residual network (Model 9), and a face attribute classification based on multiple tasks (Model 10). In the driver fatigue state recognition, fatigue state recognition based on eye state recognition (Model 11), fatigue recognition based on support vector machine (Model 12), fatigue recognition based on CNN, and semantic segmentation (Model 13) were selected for comparative analysis. The comparison results are shown in Figure 10.

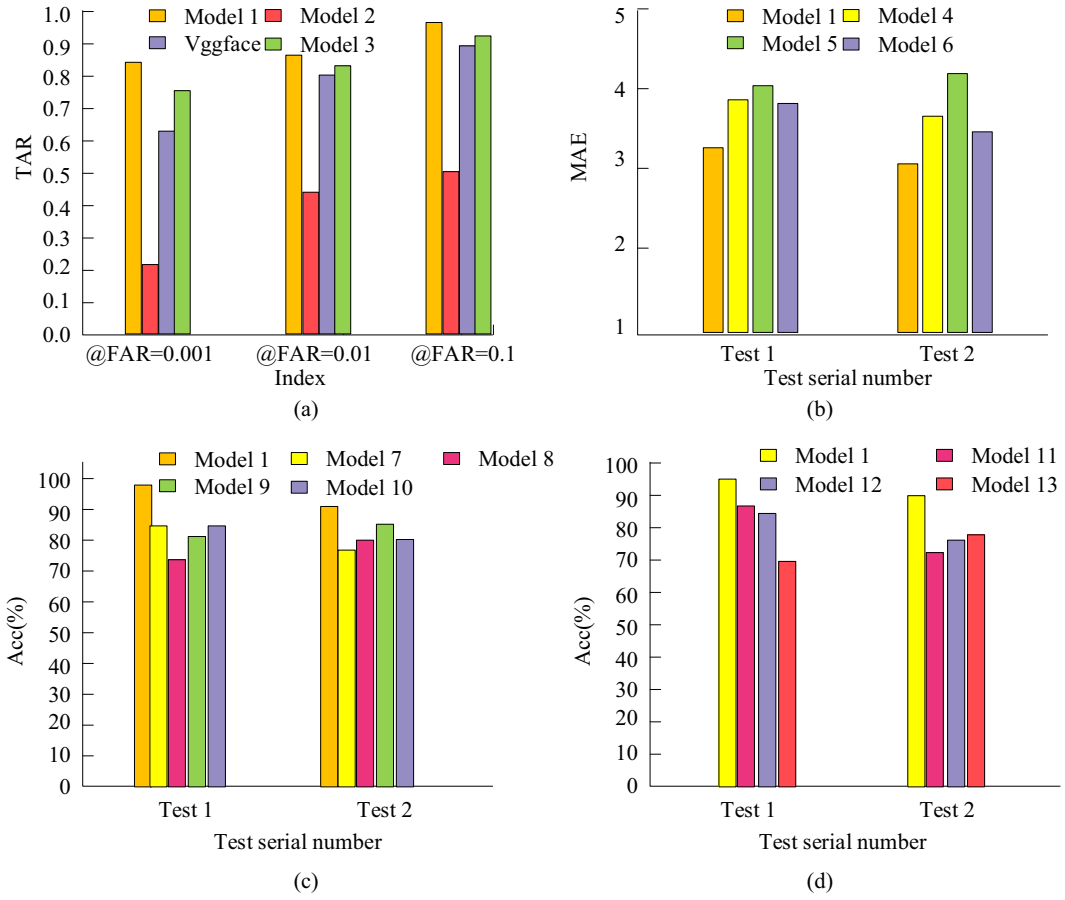

Figure 10: Comparison of recognition effects of different models. (a) Facial recognition task. (b) Facial age recognition task. (c) Facial gender recognition task. (d) Facial fatigue state recognition. Source: Created by the authors.

In the facial identification recognition challenge shown in Figure 10(a), Model 1's TAR value is 0.95, considerably higher than that of the other four models. Model 1 performs noticeably better than the other algorithms in Figure 10(b) for face age recognition, and the MAE value is decreased by more than 1.0. In Figure 10(c) and (d), the recognition accuracy of Model 1 is improved by more than 6% compared with the other algorithms. This is because the proposed multitask face recognition method extracts more generalized features after combining the two FF principles and can further improve the recognition accuracy of the model under the premise of fast model running speed.

4.4 Analysis of the practical application effect of the face recognition model based on DC and CNN

To test the effectiveness of the application of the model designed by the study, the study applies the model to the bus system to recognize the driver's driving status and, at the same time, recognize the driver's identity

and record the bus driver's working condition. After one month of operation, the application effect of the recognition model is evaluated according to the senior leaders of the bus company, the evaluation results are integrated, and the results recognized by more than 70% of the personnel are taken as the final evaluation results. Table 2 displays the test results for each function.

Table 2: Actual application effect data of the model

Performance	Requirement	Test result
Facial detection function	Complete facial detection	Complete
Facial recognition function	Complete facial recognition	Complete
Facial fatigue recognition function	Complete facial fatigue recognition	Complete
Facial recognition accuracy	Over 79%	100.00%
Facial misidentification rate	Below 3%	0.00%
Accuracy of fatigue identification	Over 97%	99.12%
Completion speed of all tasks in a single photo	Within 100 ms	35 ms

In Table 2, in the small dataset of bus driver identification, face identity was recognized a total of 567 times, of which 567 times were successfully recognized, with an accuracy rate of 100%. Fatigue recognition was performed a total of 36,640 times, of which 452 real fatigue photos were taken, and 448 were determined correctly, with a determination accuracy rate of 99.12%. The recognition time of a single photo also reached the actual demand. It can be seen that the FR model constructed by the institute can realize effective multitask FR.

5 Conclusion

Aiming at the problem of how to increase the FR rate and improve the computational efficiency of the model, the study constructs a new FR model based on CNN and DC. The MTL idea is introduced into the training process of the model, and the addFF principle is combined with the Concat-FF principle to improve the recognition accuracy of the model. The experimental analysis revealed that the efficiency of the Hadoop platform built with clusters for parallelized feature extraction of face images was significantly improved, and the four folds were found to be in a growing trend without changing the size of the data and increasing the number of cluster nodes sequentially, which proves that the method is effective. Comparing the performance of the recognition model under the MTL mechanism and the single-task learning mechanism, it was found that the error of the multitasking mechanism fluctuates around 0.2, while the error value of the model with the single-tasking mechanism fluctuates between 0.32, which demonstrates the superiority of MTL. In the face identity recognition task, Model 1 achieved a TAR value of 0.95, which was significantly better than the other models. The accuracy of face identity recognition reached 100%. The fatigue state determination accuracy reached 99.12%. The average recognition time of a single photo was 354 ms. The test results demonstrated that the model performed well in practical applications and could meet practical needs. Therefore, the FR model constructed in the study has high practical value and application prospects. However, this study still has some limitations. There is still room for further expansion of the multilayer FF network designed in the study, and the DenseNet network idea can be utilized to enhance the transfer of features at a later stage.

Funding information: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions: Yanyu Liu: Study design, data collection, statistical analysis, visualization, writing, and revision of the original draft. Liu Zhang: Revised the manuscript, led and supervised this study. The final draft was verified by the authors before submission. Dawei Zhang: Statistical analysis, visualization.

Conflict of interest: The authors declare that they have no conflicts of interest.

Data availability statement: Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] Anwarul S, Dahiya S. Rectified DenseNet169-based automated criminal recognition system for the prediction of crime prone areas using face recognition. J Electron Imaging. 2022;31(4):43055. doi: 10.1117/1.JEI.31.4.043055.

[2] Hu W, Hu H. Orthogonal modality disentanglement and representation alignment network for NIR-VIS face recognition. IEEE Trans Circuits Syst Video Technol. 2021;32(6):3630–43. doi: 10.1109/TCSVT.2021.3105411.

[3] Shiau WL, Liu C, Zhou M, Yuan Y. Insights into customers' psychological mechanism in facial recognition payment in offline contactless services: integrating belief–attitude–intention and TOE–I frameworks. Internet Res. 2023;33(1):344–87. doi: 10.1108/INTR-08-2021-0629.

[4] Lindsay GW. Convolutional neural networks as a model of the visual system: Past, present, and future. J Cognit Neurosci. 2021;33(10):2017–31. doi: 10.1162/jocn_a_01544.

[5] Wang Y, Qiao X, Wang GG. Architecture evolution of convolutional neural network using monarch butterfly optimization. J Ambient Intell Human Comput. 2023;14(9):12257–71. doi: 10.1007/s12652-022-03766-4.

[6] Ismail NA, Chai CW, Samma H, Salam MS, Hasan L, Wahab NHA, et al. Web-based university classroom attendance system based on deep learning face recognition. KSII Trans Internet Inf Syst (TIIS). 2022;16(2):503–23. doi: 10.3837/tiis.2022.02.008.

[7] Elaggoune H, Belahcene M, Bourennane S. Hybrid descriptor and optimized CNN with transfer learning for face recognition. Multimed Tools Appl. 2022;81(7):9403–27. doi: 10.1007/s11042-021-11849-1.

[8] Jiao J, Liu W, Mo Y, Jiao J, Deng Z, Chen X. Dyn-arcface: dynamic additive angular margin loss for deep face recognition. Multimed Tools Appl. 2021;80(17):25741–56. doi: 10.1007/s11042-021-10865-5.

[9] Fussey P, Davies B, Innes M. 'Assisted' facial recognition and the reinvention of suspicion and discretion in digital policing. Br J Criminol. 2021;61(2):325–44. doi: 10.1093/bjc/azaa068.

[10] El Alami A, Berrahou N, Lakhili Z, Lakhili Z, Mesbah A, Berrahou A, et al. Efficient color face recognition based on quaternion discrete orthogonal moments neural networks. Multimed Tools Appl. 2022;81(6):7685–710. doi: 10.1007/s11042-021-11669-3.

[11] Han C, Ma T, Huyan J, Huang X, Zhang Y. CrackW-Net: A novel pavement crack image segmentation convolutional neural network. IEEE Trans Intell Transp Syst. 2021;23(11):22135–44. doi: 10.1109/TITS.2021.3095507.

[12] Nirthika R, Manivannan S, Ramanan A, Wang R. Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study. Neural Comput Appl. 2022;34(7):5321–47. doi: 10.1007/s00521-022-06953-8.

[13] Ilesanmi AE, Ilesanmi TO. Methods for image denoising using convolutional neural network: a review. Complex Intell Syst. 2021;7(5):2179–98. doi: 10.1007/s40747-021-00428-4.

[14] Ghazal TM. Convolutional neural network based intelligent handwritten document recognition. Computers Mater Continua. 2022;70(3):4563–81. doi: 10.32604/cmc.2022.021102.

[15] Fang Z. A high-efficient hybrid physics-informed neural networks based on convolutional neural network. IEEE Trans Neural Netw Learn Syst. 2021;33(10):5514–26. doi: 10.1109/TNNLS.2021.3070878.

[16] Dang VT, Nguyen N, Nguyen HV, Nguyen H, Van Huy L, Tran VT, et al. Consumer attitudes toward facial recognition payment: an examination of antecedents and outcomes. Int J Bank Mark. 2022;40(3):511–35. doi: 10.1108/IJBM-04-2021-0135.

[17] Boo HC, Chua BL. An integrative model of facial recognition check-in technology adoption intention: the perspective of hotel guests in Singapore. Int J Contemp Hosp Manag. 2022;34(11):4052–79. doi: 10.1108/IJCHM-12-2021-1471.

[18] Chen Z, Chen J, Ding G, Huang H. A lightweight CNN-based algorithm and implementation on embedded system for real-time face recognition. Multimed Syst. 2023;29(1):129–38. doi: 10.1007/s00530-022-00973-z.

[19] Thurnhofer-Hemsi K, Domínguez E. A convolutional neural network framework for accurate skin cancer detection. Neural Process Lett. 2021;53(5):3073–93. doi: 10.1007/s11063-020-10364-y.

[20] Nogay HS, Adeli H. Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning. Eur Neurol. 2021;83(6):602–14. doi: doi: 10.1159/000512985.

[21] Preethi P, Mamatha HR. Region-based convolutional neural network for segmenting text in epigraphical images. Artif Intell Appl. 2023;1(2):119–27. doi: 10.47852/bonviewAIA2202293.