Retraction

Retracted: Image Recognition Technology Based on Neural Network

Jianqiu Chen

IEEE Access

10.1109/ACCESS.2020.3014692

<p>Notice of Retraction</p> <p>J. Chen, "Image recognition technology based on neural network," IEEE Access, vol. 8, pp. 157161–157167, 2020, doi: 10.1109/ACCESS.2020.3014692.</p> <p>After careful and considered review by a duly constituted expert committee, this article was retracted owing to irregularities in the peer review process, including acceptance for publication without the minimum number of independent reviews required by IEEE.</p> <p>The authors were contacted about the retraction and did not dispute it.</p>

Image Recognition Technology Based on Neural Network

JIANQIU CHEN [ID]
Computer Science and Engineering, UNSW Sydney, Sydney, NSW 2052, Australia

e-mail: ericchenjianqiu@163.com

ABSTRACT Image recognition is an important part of human-computer interaction. Using deep learning algorithms to recognize and classify image has become a hot issue for scholars from all walks of life. In this paper, the traditional classification algorithm based on convolutional neural network is improved, and the feature information of the key parts of the face is used to integrate the key part features with the global features of the face image to better distinguish similar categories. Therefore, this paper designs a method to locate the key points of the face image, and optimizes the key point positioning method through multiple experiments to facilitate the extraction of the feature information of the key points. For the calculation of classification results, a multi-region test method is used. By calculating multiple regions of the image during the test, the accuracy of image recognition can be improved. The final experimental results show that the model with key point feature information has more advantages in accuracy, and the robustness of the model is improved.

INDEX TERMS Image recognition, artificial neural network, deep learning.

I. INTRODUCTION

In the current image recognition system, the recognition of complex images can only be achieved through different levels of information processing [1], [2]. For a more familiar image, after mastering its main features, you can use this main feature as a recognition unit without paying attention to its details [3], [4]. Image recognition is an important application of artificial intelligence. Its main role is to realize the recognition by compiling computer programs that imitate human brain recognition image activities, process the images and extract effective information from them [5], [6]. Now some of these relatively mature image recognition technologies have been applied to the commercial field [7], [8]. Due to the diversity of problems in the image recognition process, image recognition methods can only adopt different methods for specific problems, so many recognition systems require a lot of research to improve performance breakthroughs on certain specific problems (such as improving recognition Efficiency, reduce the time required for system training, etc.) [9]. The machine learning method is a method that can get a good recognition effect on different recognition problems [10].

Nowadays, image recognition mainly refers to the perception and recognition of objects and environments in the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv [ID].

three-dimensional world, which belongs to the category of advanced computer vision [11]. For example, Li Minqiang, Xu Boyi and Kou Jisong expounded the necessity and feasibility of combining genetic algorithm with neural network, and put forward the idea of using multi-layer feedforward neural network as the problem representation of genetic search. A new method of training the weights of neural network with genetic algorithm is designed. The experimental results show that genetic algorithm can learn the weights of network quickly and get rid of the problem of local poles. Xiating *et al.* [12] combined artificial neural network with genetic algorithm, and proposed an evolutionary neural network method for displacement back analysis. This method is based on the samples obtained from orthogonal experiments. The genetic algorithm is used to search the optimal neural network structure and the best extended predictive learning algorithm is used to train the network [13]. The trained network describes the non-linear relationship between the mechanical parameters of rock mass (soil) and the displacement of rock mass [14]. Kaigui *et al.* [15] presented a combined forecasting model of variable weight coefficients for power system load, i.e. a combined forecasting model based on neural network. The model uses the non-linear relationship between the forecasting results of various methods and the actual load data to establish the corresponding neural network model. The network is a single-output three-layer network.

The input layer is the forecasting value of various forecasting methods, and the output layer is the actual load value. At the same time, the combined forecasting model of fixed weight coefficients based on genetic algorithm is briefly introduced. The annual, monthly and hourly load forecasting of several practical systems shows that the model has high forecasting accuracy. Gaoli and Fangping [16] and others, aiming at the shortcomings of standard BP algorithm, this paper presents several improved BP algorithm based on MATLAB language, expounds the optimization principle, advantages and disadvantages of various BP algorithms, and compares their training speed and memory consumption. It is suggested that the Levenberg-Ma rquardt algorithm should be used first, followed by BFGS algorithm or conjugate gradient method and RPROP algorithm. In Jiaan *et al.* [17] proposed a cellular automata based on neural network and used it to simulate the complex land use system and its evolution. There are many studies on urban simulation using cellular automata in the world, but these models are often limited to simulating the transformation from Non-urban land to urban land. It is much more complex to simulate the dynamic system of multiple land use than to simulate the urban evolution [18]. Many spatial variables and parameters need to be used. It is very difficult to determine the parameters and model structure of the model. The combination of neural network, cellular automata and GIS is used for dynamic simulation of land use, and multi-temporal remote sensing classification image is used to train the neural network [19]. It is very convenient to determine model parameters and model structure and eliminate the drawbacks of conventional simulation methods. Gaofeng *et al.* [20] established a neural network model for wind power prediction based on the factors affecting the output power of wind farms [20], [21]. The effects of measured power data and atmospheric data at different altitudes on the prediction results are analyzed [22]. A prediction model of error band based on neural network is established, and the prediction of error band is realized.

In this paper, the traditional classification algorithm based on convolutional neural network is improved, and the feature information of the key parts of the face is used to integrate the key part features with the global features of the face image to better distinguish similar categories. Therefore, this paper designs a method to locate the key points of the face image, and optimizes the key point positioning method through multiple experiments to facilitate the extraction of the feature information of the key points. This article introduces commonly used data enhancement solutions for missing data. At the same time, starting from the perspective of the model, reduce the complexity of the model. That is to reduce the number of parameters and calculation of the model, while reducing the risk of overfitting, it can speed up the training process and detection efficiency of the model. In addition, for the model structure and training process, through the introduction of FeatureMap channel weighting, the model's effect is optimized and the new model fusion method is used to further improve the effect [23].

II. IMAGE RECOGNITION METHOD BASED ON DEEP LEARNING

A. TEMPLATE MATCHING METHOD

In order to detect the object with known shape in the image, the shape template or window of the object is used to match the image, and the object image is detected under a certain quasi-measurement [24]. The template matching method can detect the lines, curves, patterns and so on in the image.

1) CROSS CORRELATION MATCHING

There are many ways to measure the degree of matching between two images f_1 and f_2 in regions Φ, such as mismatch can be expressed in the form of $\max_\varphi |f_1 - f_2|$; $\iint_\varphi |f_1 - f_2|$; $\iint_\varphi (f_1 - f_2)^2$ [25].

If mismatch is use $\iint_\varphi (f_1 - f_2)^2$, and

$$\iint_\varphi (f_1 - f_2)^2 = \iint_\varphi f_1{}^2 + \iint_\varphi f_2{}^2 + 2 \iint_\varphi f_1 f_2 \quad (1)$$

Obviously, after given $\iint_\varphi f_1{}^2$ and $\iint_\varphi f_2{}^2$, $2 \iint_\varphi f_1 f_2$ it's a measure of matching, the bigger the item is, $\iint_\varphi (f_1 - f_2)^2$ the smaller, the lighter the mismatch degree is, the better the matching degree is [26], [27].

Applying the Canchy-Schwarz inequality, for nonnegative f_1 and f_2, the following conclusions can be drawn:

$$\iint_\varphi f_1 f_2 \leq \sqrt{\iint_\varphi f_1{}^2 \bullet \iint_\varphi f_2{}^2} \quad (2)$$

If and only if $f_2 = cf_1$ the equality sign holds (c is constant). In digital images, integrals are converted to summations, and the result is changed to

$$\sum_i \sum_j f_1(i,j) f_2(i,j) \leq \sqrt{\sum_i \sum_j f_1(i,j) \bullet \sum_i \sum_j f_2(i,j)} \quad (3)$$

Similarly, if and only if $f_2(i,j) = cf_1(i,j)$ the equality sign holds (c is constant) [28].

When the target template f_1 is set and f_2 is the image to be matched, obviously the f_1 should be assumed to be smaller than f_2. Then we will move f_1 at all possible positions in f_2 and calculate $\iint_\varphi f_1 f_2$ to each shift (u, v). According to Cauchy-Schwartz inequality, the following formula holds [20], [30].

$$\iint_\varphi f_1(x,y) f_2(x+u, y+v) dxdy$$
$$\leq \sqrt{\iint_\varphi f_1{}^2(x,y) dxdy \bullet \iint_\varphi (x+u, y+v) dxdy} \quad (4)$$

Because f_1 all of them are equal to 0 outside the region φ, the integral region can be expanded from φ to $(-\infty, +\infty)$,

so that the left part of the upper formula can be changed to

$$\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_1(x, y) f_2(x + u, y + v) dx dy \quad (5)$$

So in upper formula f_1 and f_2 of cross correlation function $cf_1 f_2$.

On the right side of the analytic formula (1), although the $\iint f_1{}^2$ term is a constant, but $\iint f_2{}^2$ is not a constant. It is related to u and v. This is because in practice, the template f_1 is usually fixed and the image f_2 to be matched is moved. Therefore, the content of the image f_2 with the corresponding region of f_1 always varies with u and v. Simple apply $cf_1 f_2$ as a matching measure is not suitable. Generally, the domesticated cross-correlation function is used as a matching measure, that is, the domesticated cross-correlation function is used as a matching measure, namely

$$\frac{C_{f_1 f_2}}{\sqrt{\iint f_2(x + u, y + v) dx dy}} \quad (6)$$

If makes $f_2 = cf_1$ (c is constant) on a displacement (u, v), then (formula 6) has a maximum value $\iint f_1{}^2$. At this time (2) or (3) equation holds, then (1) there is a minimum, that is, the minimum mismatch.

In fact, because of the existence of noise, the above equation (2) will not occur, that is, it is impossible to match completely. Generally, the position of the maximum of formula (6) is chosen as the best matching point.

B. BAYESIAN CLASSIFICATION

This is a kind of classification method based on probability statistics and Bayesian theorem, which belongs to the category of statistics [31]. The specific algorithm flow of Bayesian classification is that the classification problem is expressed in probability form and the related probabilities are known. According to Bayesian theorem, the representative features of the image are extracted, and the posterior probability is calculated to classify the image [32]. Bayesian formula can be expressed as:

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (7)$$

Among them, P (A | B) represents the probability of event A under the premise of event B, which is called the conditional probability of event A under the occurrence of time B. P (AB) represents the probability of events A and B occurring together, P (B) represents the probability of events B occurring.

The disadvantage of Bayesian classification is that it cannot extract the representative features of images well in some cases, which results in poor classification effect. The common Bayesian classification algorithms are Naive Bayesian algorithm and TAN algorithm (tree-enhanced Naive Bayesian algorithm)

C. ARTIFICIAL NEURAL NETWORK METHOD

As we know from the formula, there are

$$D(z) = D(\sum_{j=1}^{n} w_j x_j + b) \quad (8)$$

Expand as follows

$$D(z) = \sum_{j=1}^{n} D(w_j x_j) + D(b) \quad (9)$$

By the properties of variance are expanded as follows

$$D(z) = \sum_{j=1}^{n} E\left\{ \left[w_j x_j - E(w_j x_j) \right]^2 \right\} + D(b) \quad (10)$$

Simplify and organize as follows

$$D(z) = \sum_{j=1}^{n} \left\{ E\left(w_j{}^2\right) E\left(x_j{}^2\right) - 2\left[E(w_j)E(x_j) \right]^2 \right. $$
$$\left. + \left[E(w_j)E(x_j) \right]^2 \right\} + D(b) \quad (11)$$

Collate as

$$D(z) = \sum_{j=1}^{n} \left\{ E\left(w_j{}^2\right) E\left(x_j{}^2\right) - \left[E(w_j)E(x_j) \right]^2 \right\} + D(b) \quad (12)$$

So $E(w_j) = 0, E(x_j) = 0, E(w_j{}^2) = D(w_j), E(x_j{}^2) = D(x_j), D(b) = 1$, So, by substituting formula 12, we can get it.

$$D(z) = nD(w)D(x) \quad (13)$$

D. NEURAL NETWORK TRAINING PROCESS

Since the full-connection layer receives the main parameters of the neural network, the output of some neurons in the first and second full-connection layers is set to 0 by dropout operation with a 50% random probability. By using min-batch method, the parameters in the back-propagation update neurons occur after the average error gradient of 64 samples is calculated each time. Let 0.001 and 0.85 be the attenuation terms of parameters and the potential energy of updating parameters. W parameter updates are as follows:

$$v_{n+1} = 0.85 v_n - 0.001 \alpha A - \alpha (\frac{\partial L}{\partial W}|_{w_n}) D_n \quad (14)$$

$$w_{n+1} = w_n + v_{n+1} \quad (15)$$

Among them, n denotes the number of iterations, V denotes the potential energy term when the parameters are updated, $(\frac{\partial L}{\partial W}|_{w_n}) D_n$ represents the average error gradient of all training samples on the nth min-batch, and α denotes the learning rate of the neural network. In the initial state, A is set to 0.01.

The whole training process lasted 10 days, totally completed 100 rounds of training, the training process curve is shown in Figure 1. From the figure, we can clearly see that in the training process, we adjusted the learning rate four times, corresponding to < 0.01, 0.001, 0.0001, 0.00001>.
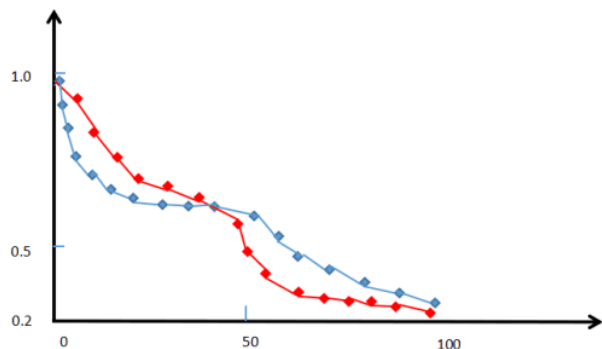
FIGURE 1. Deep neural network training process diagram.

III. EXPERIMENTAL DESIGN AND ALGORITHM IMPROVEMENT

A. DATA SET

Training data set: the same as more researchers, this paper selects CASIA-Web Face as the training sample in the training stage. The database contains 494,414 images of 10,575 people, which are collected from the website IMDB (Internet Movie Database), which contains many face pictures of celebrities. Through the semi-automatic inspection, screening and cleaning of the collected pictures, the accuracy of most data annotations is ensured. Because it contains part of the face in LFW, it is usually necessary to delete the face image in the LFW when conducting the test experiment. Test data set: LFW and YTF are currently the most used test sets. In order to make a better comparison with the current method, this article also selected these two databases for test experiments. Among them, LFW contains 13324 pictures of 5749 objects, and YTF is a picture data set extracted from the video library, containing pictures of 1595 objects. In addition, the traditional multi-pose database is not enough as a test object for deep learning, and in order to better illustrate the advantages of this method in multi-pose face recognition [33], this paper selects a more complex database IJB-A (IARPA Janus Benchmark A, IJB-A) as the test object.

B. IMPROVEMENT OF THE MODEL

Through analysis and comparison of several mainstream network model structures, two most representative networks were selected, ResNet-50 and InceptionV4. Optimization and improvement were made on the "depth" and "width" of the model, respectively, through the numerical addition and the operation of Concat on the channel to increase the feature information contained in the Feature Map. However, these two processing methods are optimized from the Feature Map output from the overall. In the Feature Map, there is a certain difference in the degree of influence of the feature information on different channels on the final prediction result. Therefore, this paper draws on the core part of the SeNet algorithm to weight the feature information on different channels. The weights are also used as part of the model parameters, which are optimized and updated using

backpropagation algorithms. It only needs to add a bypass when performing the convolution operation, which is used to indicate the importance of different Channels. After the convolution operation, the Feature Map of size $H*W*X$ is output, and a weight coefficient of $1*1*C$ is added. The weight coefficient is multiplied by the Feature Map to obtain the weighted Feature Map. Among them, the feature information with a larger prediction result is enhanced, and the weaker information is suppressed, which is more conducive to the screening of feature information, thereby reducing interference information. The increase in calculation is only about 10%, and it is relatively easy to implement.

C. TRAINING PROCESS

In order to further improve the classification effect of the model, taking into account the image of women's clothing, the differences between the categories are mainly reflected in the eyebrows, eyes, mouth and other parts. If the entire picture input is used for classification, it is impossible to highlight the impact of some key point information on the determination of the image type. So in this paper, the Feature Map obtained by the convolution operation of the image of the key part and the Feature Map obtained by the complete image are superimposed on the channel to obtain the fused feature information, and then continue the classification operation. In order to implement this algorithm, it is necessary to locate key points in the face image. Since the number of key points in different face images is different, in order to better measure the positioning effect, the effects of several designed model structures are compared. The calculation formula of NE is as follows.

$$NE = \frac{\sum_k \left\{ \frac{d_k}{S_k} \delta(v_k = 1) \right\}}{\sum_k \{\delta(v_k = 1)\}} \times 100\% \qquad (16)$$

Among them, k is used to represent the number of key points, and its value range depends on the number of key points in the picture. It is used to represent the distance between the position of the key point predicted by the algorithm model and the marked point of the image. In order to avoid the inconsistency of the distance measurement standard due to size, Sk is used as the normalization parameter of the distance and used to indicate whether the key point is visible, making the measurement index more in line with the actual situation.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

A. ACCURACY ANALYSIS

The accuracy of this method is higher than that of CNN method, and much higher than that of SVM method. When there are two hidden layers, the accuracy of this method is basically slightly higher than that of DBN method. With the increase of the number of samples, the correct rate is obviously improved. When there are two hidden layers, the correct rate is higher than that of the single hidden layer.
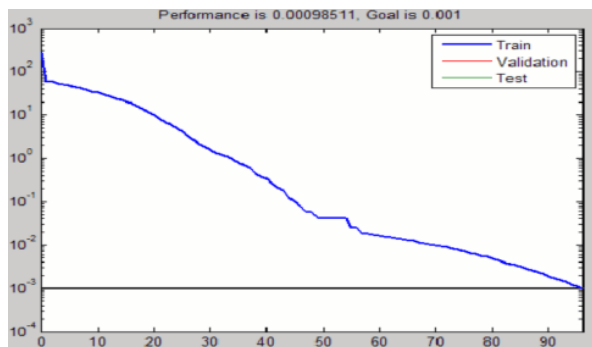
FIGURE 2. Error performance curve and training time diagram of neural network training.

FIGURE 3. Functional images of three different penalty functions.

TABLE 1. Comprehensive recognition rates for different methods.

Method	Recognition rate（%）
LBP-TOP+SVM15	63.77
LBP-TOP+OF+RF16	64.66
MG-LTP+ELM	60.44
CNN+	66.33

Regardless of the training time or the number of training, the recognition effect is ideal when the number of hidden layer nodes is 26. The error performance curve and training time of the neural network training are shown in Figure 2.

B. EFFECT OF SVM PARAMETERS

Generally, the selection of relevant parameters (mainly penalty factor C and kernel function parameter gamma) is very important to obtain the ideal classification prediction accuracy when using SVM for classification prediction. Therefore, there are many methods to optimize the parameters of SVM. Such as genetic algorithm parameter optimization, particle swarm optimization, ant colony algorithm, artificial bee colony algorithm and so on.

(1) The influence of penalty factor C

In the experiment, the number of training samples is 5000, the number of testing samples is 1000, the other parameters remain unchanged, and the gamma γ value is the default value of the toolbox of 0.0033. With different C values, the effects of penalty factor C on the number of support vectors, the accuracy of training samples and the accuracy of testing samples of SVM and RBM-SVM methods were compared [34]. The RBM-SVM method uses a single hidden layer and the number of nodes in the hidden layer is 300. To facilitate comparison, the results are compared. When using two hidden layers and one hidden layer, the same conclusion can be drawn. Fig. 3 is a functional image of three different penalty functions.

(2) The effect of gamma value.

For this experiment, the default gamma value is 0.003. Different gamma values have different effects on the number of support vectors and the accuracy of test samples. Only by choosing the appropriate gamma value can we get better results.

C. EXPERIMENTAL COMPARISON OF AND NON-DEEP LEARNING ALGORITHMS

Recognition rates of different strategies in five types of expressions. When we use a single CNN model to classify facial microexpressions, we adopt dropout strategy and data set expansion strategy to prevent CNN from over-fitting [35].
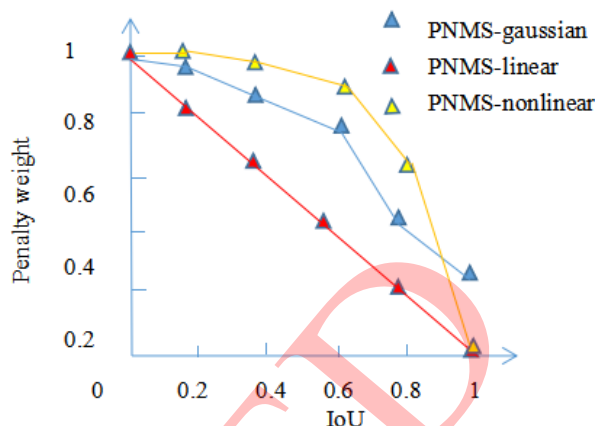
CNN + D denotes the CNN model with dropout strategy, CNN + A denotes the CNN model with data augmentation strategy, that is, the following four transformations are performed on each image: rotation, horizontal translation, vertical translation and horizontal flip. Thus, the data set can be expanded to four times. CNN+AD represents a CNN model with two strategies. The CNN + LSTM representation combines the network model of CNN and LSTM.

We add the strategy CNN, which performs better in image recognition than the STRATEGY-FREE CNN. This may be because there are fewer training pictures and the CNN network is deeper, which makes the STRATEGY-FREE CNN easy to over-fit in the process of training parameters. The performance of CNN + LSTM is better than that of a single CNN model, which shows that LSTM can make full use of the feature information in time domain, so that it can better recognize sequence data. This proves that the model of CNN + LSTM can be used to identify images.

In order to compare the advantages and disadvantages of the traditional machine learning algorithm and the deep learning algorithm, we use the traditional machine learning algorithm to do the same experiment in Casme2 [36]. From Table 1, we can see that the model proposed in this paper has a better performance compared with these traditional machine learning models.

In this experiment, there are three methods of comparative analysis:

(1) Bi-categorized DNN, which trains in-depth neural networks separately for each category

(2) Multi category DNN

(3) Multi-task DNN using Ring training method

FIGURE 4. Comparison of multi-classification dnn, bi-classification dnn and multi-task dnn on training set and test set.

The experimental results are shown in Fig. 4. The results of training data set 1. Show that multi-classification DNN performs worst when the data distribution is not balanced, with an error rate of 48.15%, while multi-task DNN has a error rate of 38.99%. Figure 4 shows the relationship between the predicted probability of one-category in multi-classification DNN and the training data of that category. That is to say, for multi-classification DNNs, when doing classification tasks, they are more inclined to predict the target object as a category with higher frequency in training data. One possible reason is that the category with lower frequency in training data has not been adequately trained. The parameters predicted by these classes in the neural network are all in a relatively small range of values. In multi-classification tasks, some other unrelated classes with high frequency appear in the training data, although they are shared with low frequency classes in the expression of image features. But in the process of classification, the range of parameters corresponding to high frequency category is larger, and the weight obtained by transformation is higher. The performance of multi-classified DNN on dataset 2 is also quite different from that of multi-task DNN. The error rate of multi-classified DNN is 53.27%, while that of dual-classified DNN and multi-task DNN is 45.43% and 40.27%, respectively. The reason for the worst performance of multi-classification DNN on dataset 2 may be the over-fitting problem in the process of training to distinguish two similar classes. This can be seen from the error rate of multi-classification DNN on training dataset 2, which is 6.23%, and 53.27% on test set.

V. CONCLUSION

In this paper, the traditional classification algorithm based on convolutional neural network is improved, and the feature information of the key parts of the face is used to integrate the key part features with the global features of the face image to better distinguish similar categories. Therefore, this paper designs a method to locate the key points of the face image, and optimizes the key point positioning method through multiple experiments to facilitate the extraction of the feature information of the key points.

This article introduces commonly used data enhancement solutions for missing data. At the same time, starting from the perspective of the model, reduce the complexity of the model.

That is to reduce the number of parameters and calculation of the model, while reducing the risk of overfitting, it can speed up the training process and detection efficiency of the model. In addition, for the model structure and training process, through the introduction of FeatureMap channel weighting, the model's effect is optimized and the new model fusion method is used to further improve the effect.

This paper has conducted some research in the field of face image classification algorithms, and put forward some solutions and ideas for improvement, although it has achieved certain effects and improvements. However, there are still many shortcomings, and further work needs to be completed later. For example, whether the structure of the model can be further optimized to better balance model performance and detection efficiency. For the use of key information of the model, more new ways can be tried, combined with traditional image processing methods for better use.

REFERENCES

[1] J. Karnes, R. Robb, P. C. O'Brien, E. H. Lambert, and P. J. Dyck, "Computerized image recognition for morphometry of nerve attribute of shape of sampled transverse sections of myelinated fibers which best estimates their average diameter," *J. Neurolog. Sci.*, vol. 34, no. 1, pp. 43–51, Oct. 1977.

[2] B. Wang, F. C. Zou, and X. W. Liu, "New algorithm to generate the adversarial example of image," *Optik*, vol. 207, Apr. 2020, Art. no. 164477.

[3] A. Khotanzad and Y. H. Hong, "Rotation invariant image recognition using features selected via a systematic method," *Pattern Recognit.*, vol. 23, no. 10, pp. 1089–1101, Jan. 1990.

[4] K. Liu, Y.-Q. Cheng, and J.-Y. Yang, "Algebraic feature extraction for image recognition based on an optimal discriminant criterion," *Pattern Recognit.*, vol. 26, no. 6, pp. 903–911, Jun. 1993.

[5] T. C. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.*, vol. 38, no. 2, pp. 370–381, 1999.

[6] T. Baidyk, E. Kussul, O. Makeyev, A. Caballero, L. Ruiz, G. Carrera, and G. Velasco, "Flat image recognition in the process of microdevice assembly," *Pattern Recognit. Lett.*, vol. 25, no. 1, pp. 107–118, Jan. 2004.

[7] D. Keysers, T. Deselaers, C. Gollan, and H. Ney, "Deformation models for image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1422–1435, Aug. 2007.

[8] B. Crawford, R. Soto, R. Cuesta, and F. Paredes, "Application of the artificial bee colony algorithm for solving the set covering problem," *Sci. World J.*, vol. 1, Apr. 2014, Art. no. 189164.

[9] A. Ebrahimi Zade, A. Sadegheih, and M. M. Lotfi, "A modified NSGA-II solution for a new multi-objective hub maximal covering problem under uncertain shipments," *J. Ind. Eng. Int.*, vol. 10, no. 4, pp. 185–197, Dec. 2014.

[10] M. Elhoseny, "Multi-object detection and tracking (MODT) machine-Learning model for real-time video surveillance systems," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 611–630, Aug. 2019.

[11] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, Jul. 2020.

[12] Y. Bengio, "Learning deep archi tectures I for AI," in *Foundations & Trends in Machine Learning*, vol. 2, no. 1. Hanover, MA, USA: Now, Apr. 2009, pp. 1–127.

[13] S. Pan, R. Hu, S.-F. Fung, G. Long, J. Jiang, and C. Zhang, "Learning graph embedding with adversarial training methods," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2475–2487, Jun. 2020.

[14] S. L. Shuxia, J. Yang, and Z. Cai, "Big data recommendation algorithm based on social relationship and unsupervised learning," *Comput. Appl. Softw.*, vol. 36, no. 5, pp. 310–316 and 327, May 2019.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradi ents I for hluman I detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. l, Jun. 2005, pp. 886–893.

[17] N. Prakash and Y. Singh, "Fuzzy support vector machines for face recognition: A review," *Int. J. Comput. Appl.*, vol. 131, no. 3, pp. 24–26, Dec. 2015.

[18] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.

[19] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 123344.

[20] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, p. 37, 2016.

[21] H. Li, L. Zhang, B. Huang, and X. Zhou, "Sequential three-way decision and granulation for cost-sensitive face recognition," *Knowl.-Based Syst.*, vol. 91, pp. 241–251, Jan. 2016.

[22] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 980–993, Mar. 2015.

[23] R. Wang, Y. Guan, H. Song, X. Li, X. Li, Z. Shi, and X. Song, "A formal model-based design method for robotic systems," *IEEE Syst. J.*, vol. 13, no. 1, pp. 1096–1107, Mar. 2019.

[24] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2041–2056, Oct. 2015.

[25] L. Zhang, W.-D. Zhou, and F.-Z. Li, "Kernel sparse representation-based classifier ensemble for face recognition," *Multimedia Tools Appl.*, vol. 74, no. 1, pp. 123–137, Jan. 2015.

[26] X. Wu, B. Fang, Y. Y. Tang, X. Zeng, and C. Xing, "Reconstructed error and linear representation coefficients restricted by l1-minimization for face recognition under different illumination and occlusion," *Math. Problems Eng.*, vol. 2017, pp. 1–16, May 2017.

[27] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[28] R. Min, N. Kose, and J.-L. Dugelay, "KinectFaceDB: A kinect database for face recognition," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 11, pp. 1534–1548, Nov. 2014.

[29] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 1–42, 2016.

[30] G. Wang, N. Shi, Y. Shu, and D. Liu, "Embedded manifold-based kernel Fisher discriminant analysis for face recognition," *Neural Process. Lett.*, vol. 43, no. 1, pp. 1–16, Feb. 2016.

[31] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[32] L. Zhao, Y. Zhang, B. Yin, Y. Sun, Y. Hu, X. Piao, and Q. Wu, "Fisher discrimination-based $l_{2,1}$-norm sparse representation for face recognition," *Vis. Comput.*, vol. 32, no. 9, pp. 1165–1178, 2016.

[33] Y. Zhang, X. Xiao, L.-X. Yang, Y. Xiang, and S. Zhong, "Secure and efficient outsourcing of PCA-based face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1683–1695, 2020.

[34] B. Zhu, S. Ye, M. Jiang, P. Wang, Z. Wu, R. Xie, J. Chevallier, and Y.-M. Wei, "Achieving the carbon intensity target of China: A least squares support vector machine with mixture kernel function approach," *Appl. Energy*, vols. 233–234, pp. 196–207, Jan. 2019.

[35] B. Cao, J. Zhao, P. Yang, P. Yang, X. Liu, J. Qi, A. Simpson, M. Elhoseny, I. Mehmood, and K. Muhammad, "Multiobjective feature selection for microarray data via distributed parallel algorithms," *Future Gener. Comput. Syst.*, vol. 100, pp. 952–981, Nov. 2019.

[36] L. Wu, C.-H. Chen, and Q. Zhang, "A mobile positioning method based on deep learning techniques," *Electronics*, vol. 8, no. 1, p. 59, Jan. 2019.

JIANQIU CHEN was born in Xingtai, Hebei, China, in 1996. He received the bachelor's degree from Xinjiang University, China. He is currently pursuing the master's degree in computer science and engineering with the University of New South Wales (UNSW Sydney). His research interests include computer vision, feature representation, and pattern recognition.

• • •