International Core Journal of Engineering Volume 7 Issue 8, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

Small Sample Face Recognition Method Under Unnatural Circumstances

Ting Gong, Yaqian Huang, Xiaoqin Liu

Computer Science Department, North China Electric Power University, (Baoding), Baoding, China.

Abstract

For face recognition in the real world can obtain less samples, and when the collection environment is uncontrollable, it will affect the accuracy of face recognition system. This paper proposes a hierarchical model based on meta learning for face recognition system. Firstly, dynamic convolution is used as the feature extractor. The network can greatly improve the ability of extracting the underlying features of images by virtue of the characteristic of generating convolution kernel dynamically according to the input. Then, by using the idea of bi level optimization of meta learning, it can find the appropriate network parameters through one or several steps of gradient descent. The experimental results show that the accuracy of the proposed model is improved by 2% ~ 3% compared with other models.

Keywords

Few-shot Learning; Dynamic Convolution; Meta-learning; Feature Extraction.

1. Introduction

With the development of society, face recognition is playing an increasingly important role in more and more fields, compared to other verification methods, such as fingerprints, iris and other biometric methods. Face recognition has the advantages of security, convenience, and low cost, and has been widely used in various industries such as finance and the Internet. The diversity of scenes and the complexity of tasks have extremely high requirements for the accuracy of the face recognition system. But in most cases, the available face samples are very few, usually only one or a few samples. This conflicts with the idea that deep learning requires a large number of samples to train neural networks, which leads to a sharp decline in recognition accuracy. There are also various noises in face samples, such as occlusion, illumination, posture, expressions, etc., which makes the already difficult recognition process more difficult.

Literature [1] proposed a small sample image classification method under the migration model, but the robustness of the system is poor, and the impact of environmental changes on the face recognition system is not considered. Literature [2] proposes a compensation method in the training set to deal with the problem of small sample face recognition, which better solves the problem of poor robustness encountered in the migration process of small sample data, but it still does not consider To the impact of the environment on the accuracy of face recognition. Literature [3] et al. proposed a method for face recognition under unnatural circumstances, but did not consider the impact of small samples on its accuracy. The decrease in the number of training samples led to a sharp drop in accuracy.

Based on meta-learning [4] and the above-mentioned enlightenment on face recognition processing methods under unnatural situations, this paper proposes a small sample face recognition method under unnatural situations. At present, based on meta-learning target detection, the image classification method [5-11] has been very successful. Literature [12] proposed a two-layer optimization idea. During training, it automatically finds the appropriate model parameters, that is, learns to learn so

ISSN: 2414-1895
DOI: 10.6919/ICJE.202108_7(8).0080

that it can be obtained after one or several gradient descents on a new task. Very good recognition performance. Secondly, considering the special situation of the real world, literature [13] proposed several occluded face recognition processing methods. Experiments show that a neural network-based method for processing occluded face recognition has good performance. After experiments on commonly used face recognition data sets, it is verified that the method in this paper has high robustness for small sample face recognition problems with noise interference, and has improved accuracy compared with the previously proposed face recognition algorithm.

2. Related work

2.1 Meta-learning

There have been many studies on small sample learning before this. Literature [11] proposed a two-layer optimization idea to optimize our network, so that it automatically finds the appropriate network parameters, so that subsequent tasks can obtain better performance network parameters through one or several steps of gradient descent. However, because the double-layer optimization needs to calculate the second step degree, the calculation amount is too large, therefore, the literature [10] proposed a method of approximating the first step degree to the second step degree, but the experimental results found that some information was lost. Subsequently, Nichol [12] et al. proposed the Reptile method on the basis of the predecessors, which reduces the amount of calculation while retaining the image information.

2.2 Face recognition method with occlusion

For the occluded face image, the low-rank matrix recovery (LRMR) model is usually used at present. Its core idea is: if we regard an image as a matrix, then the fewer the number of its bases, the fewer the number of linearly independent vectors corresponding to the bases, and the smaller the rank of the matrix. Some rows or columns of a low-rank matrix can be expressed linearly with other rows or columns, which means that this matrix contains more redundant information. Using this redundant information, we can restore the image information, thereby effectively removing the disturbing noise information, and can also restore the erroneous image information. This model based on Low-Rank Matrix Approximation (LRMA) is the Low-Rank Matrix Recovery Model (LRMR). At present, LRMR mainly has three types of modes: robust PCA (RPCA), matrix completion (MC), and low-rank representation (LRP). In this paper, we adopt the first mode. Low-rank matrix restoration.

RPCA wants to decompose matrix D into the sum of two matrices: D=A+E, where A is the low-rank matrix part, and E is the noise-polluted sparse matrix part. Figure 1 shows its graphical description. Then the restoration of the matrix at this time can be described by the following optimization problem:

$$\min_{A,E} rank(A) + \lambda\|E\|_0 \quad s.t. \quad D = A + E \tag{1}$$

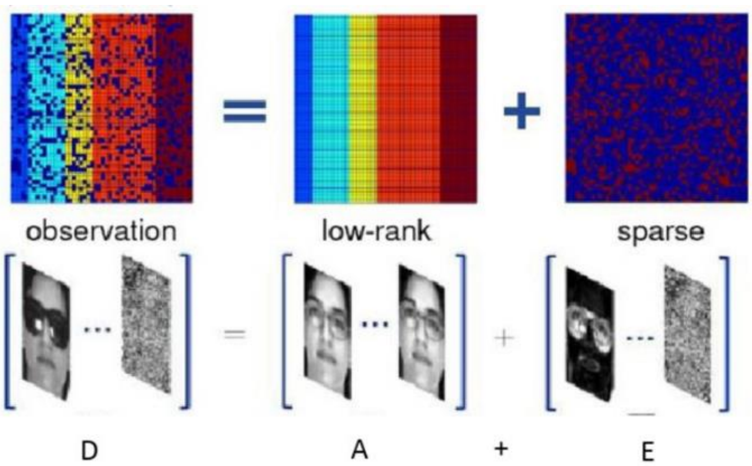

Fig. 1 Low rank decomposition process of matrix

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

Where is the observation matrix, and the objective function is the rank of the matrix and the zero norm of the noise matrix, which represents the weight of the noise. (1) The rank function and 0 norm in the formula are both non-convex, which becomes an NP-hard problem, so it needs to be relaxed to optimize. From the matrix filling theory, we know that both the rank and norm of the matrix can be relaxed convexly, so the formula (1) can be relaxed into the following convex optimization problem:

$$\min_{A,E}\|A\|_* + \lambda\|E\|_{1,1} \quad s.t. \quad D = A + E \tag{2}$$

Solving equation (2) is Robust Principal Component Analysis (RPCA). In this paper, we choose to adopt the augmented Lagrangian multiplier algorithm, namely the Alternating direction methods (ADM), also known as the inexact Lagrangian multiplier method (Inexact ALM, IALM).

For the optimization of formula (2), first construct an augmented Lagrangian function:

$$L(A,E,Y,u) = \|A\|_* + \lambda\|E\|_{1,1} + < Y, D - A - E > + \frac{u}{2}\|D - A - E\|_F^2 \tag{3}$$

When $Y = Y_k, u = u_k$, use the alternate method to solve the block optimization problem:

$$\min_{A,E} L(A,E,Y_k,u_k) \tag{4}$$

Use Exact ALM (EALM) to iterate the matrices alternately and until the termination condition is met. If, then

$$A_{k+1}^{j+1} = arg\min_A L\left(A, E_{k+1}^j, Y_k, u_k\right) = arg\min_A \|A\|_* + \frac{u_k}{2}\left\|A - \left(D - E_{k+1}^j + \frac{Y_k}{u_k}\right)\right\|_F^2$$

$$= D_{\frac{1}{u_k}}(D - E_{k+1}^j + \frac{Y_k}{u_k}) \tag{5}$$

Update the matrix E according to (5):

$$E_{k+1}^{j+1} = arg\min_E L\left(A_{k+1}^{j+1}, E, Y_k, u_k\right) = arg\min_E \lambda\|E\|_{1,1} + \frac{u_k}{2}\left\|E - \left(D - A_{k+1}^{j+1} + \frac{Y_k}{u_k}\right)\right\|_F^2$$

$$= S_{\frac{\lambda}{u_k}}(D - A_{k+1}^{j+1} + \frac{Y_k}{u_k}) \tag{6}$$

Remember A_{k+1}^* and E_{k+1}^* are the exact values of A_{k+1}^{j+1} and E_{k+1}^{j+1}respectively, then the update formula of the matrix Y is:

$$Y_{k+1} = Y_k + u_k(D - A_{k+1}^* - E_{k+1}^*) \tag{7}$$

The parameters u_k are updated to:

$$u_{k+1} = \begin{cases} \rho u_k \frac{u_k\|E_{k+1}^* - E_k^*\|}{\|D\|_F} < \varepsilon \\ u_k otherwise \end{cases} \tag{8}$$

Among them, $\rho > 1$ and $\rho \in R$,$\varepsilon > 0$.

The above method requires multiple updates and singular value decomposition in the inner loop. For this reason, the researchers proposed the inexact Lagrangian multiplier method (Inecact ALM, IALM), and the update formula becomes the following form:

$$A_{k+1} = arg\min_A L(A, E_{k+1}, Y_k, u_k) = D_{\frac{1}{u_k}}(D - E_{k+1} + \frac{Y_k}{u_k}) \tag{9}$$

$$E_{k+1} = arg\min_E L(A_{k+1}, E, Y_k, u_k) = S_{\frac{\lambda}{u_k}}(D - A_{k+1} + \frac{Y_k}{u_k}) \tag{10}$$

2.3 Dynamic convolution

The idea of dynamic convolution is mainly based on trying to find a balance between network performance and computational load. In the traditional sense, the method of improving network performance is mainly by constructing a wider or deeper network model, but this usually leads to more calculations, so it is not friendly to the pursuit of high-efficiency networks.

International Core Journal of Engineering Volume 7 Issue 8, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

The dynamic convolution proposed in [13] does not increase the depth and width of the network. It improves the expressive ability of the network model by fusing multiple convolution kernels. The convolution kernel obtained by the dynamic convolution generator is related to the input, that is, different image data have different convolutions. The dynamic perceptron and dynamic convolution will be described in detail below.

First, define the traditional perceptron as $y = g(\omega^T x + b)$, where ω, b, g represents the weight, bias and activation function respectively; then, define the dynamic perceptron as follows:

$$y = g\left(\tilde{\omega}^T x + \tilde{b}\right), \tilde{\omega} = \sum_{k=1}^{K} \pi_k(x)\tilde{\omega}_k$$

$$\tilde{b} = \sum_{k=1}^{K} \pi_k(x)\tilde{b}_k \;\; s.t. 0 \le \pi_k(x) \le 1, \sum_{k=1}^{K} \pi_k(x) = 1 \tag{11}$$

Among them, π_k represents the attention weight, which is not fixed, it will vary with the input. Therefore, dynamic convolution has a stronger feature expression ability than traditional static convolution.

Compared with the static perceptron, the dynamic perceptron has two additional calculations: (1) attention weight calculation; (2) dynamic weight fusion. Even so, the calculation of these two parts is completely negligible compared to the calculation of the perceptron:

$$O(\tilde{\omega}^T x + \tilde{b}) >> O(\sum \pi_k \tilde{\omega}_k) + O(\sum \pi_k \tilde{b}_k) + O(\pi(x)) \tag{12}$$

Dynamic convolution is similar to a dynamic perceptron, and both have a core. Referring to the classic design in convolutional neural networks, literature[13] connects BatchNorm and ReLU after dynamic convolution. Figure 2 shows the process of dynamic convolution.

Literature [13] uses a lightweight squeeze and excitation to extract the weight of attention, which is different from SENet in that the latter is to give attention to the channel, while the former is regarded as the convolution kernel to give attention to the mechan ism; Since the convolution kernel is usually relatively small, the process of kernel fusion does not consume more computing resour- ces, so it is very efficient. Table 1 shows the comparison of the calculation amount of dynamic convolution and static convolution. From the data in the table, it can be seen that the increase in calculation amount can be completely ignored.

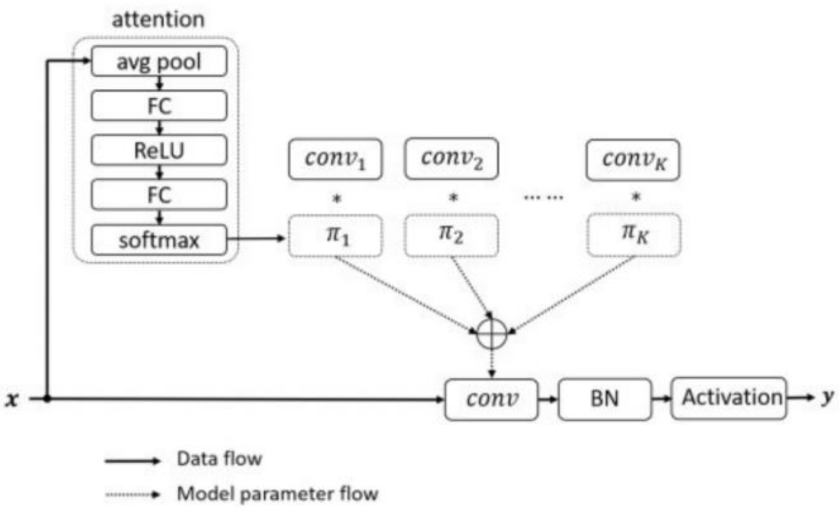

Fig. 2 Dynamic Convolution

In view of the many advantages of dynamic convolution, it can be easily embedded in the convolution of the existing network model to replace the traditional different types of static convolution. Therefore, this article intends to adopt dynamic convolution to improve the traditional network structure.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

Table 1. The comparison of the computation of static convolution and dynamic convolution

	V2 ×1.0	V2 ×0.5	V3-large	V3-small
static	300M	97M	219M	66M
K=2	309.5M	100.5M	224.9M	67.8M
K=4	312.9M	101.4M	227.3M	68.5M
K=6	316.3M	102.3M	229.8M	69.3M

3. Network structure

The traditional face recognition model faces the problem of lack of face data and large noise in some data. This type of face recognition algorithm will cause the appearance of data imbalance, which will affect the accuracy of face recognition. In order to solve the above proposal For the problem, this paper proposes a face recognition model with occlusion in the case of small samples based on meta-learning.

First, perform low-rank matrix restoration image processing on the pictures input into the network, and perform preliminary processing on the occluded images. Encode it according to the input data, pass the encoded data to the balance generator for statistical processing, generate balance variables to deal with the imbalance problem, the latter generates balance variables α, β through the input to solve the problem of category imbalance and category selection imbalance, The two are used as the input of the network, and the idea of meta-learning is used to optimize and update the network parameters. Subsequently, the more popular ResNet network in recent years is used as the underlying feature extractor, and on this basis, the dynamic convolution layer is used to replace the convolution layer in the ResNet network, which greatly improves the ability of the network to extract features, and then borrows meta-learning The two-layer optimization idea of Optimized the network parameters. Finally, softmax is used as the classifier of the final result. The network structure is shown in Figure 3.

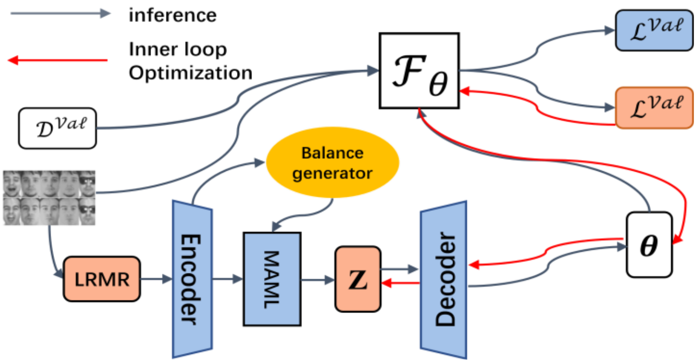

Fig. 3 The structure of network

4. Data processing

(1) In the real world, there are great imbalances in various samples. In order to simulate the imbalance of data, this article performs special processing on the data set used to simulate the imbalance of data. Assuming that the total number of categories of data is N, the data of category M in category N is processed so that the amount of data in category M differs from the amount of data in category N-M by more than 50%.

(2) Perform preprocessing operations on the image data, inclu ding removing invalid and error images, and performing low-rank matrix image restoration operations on the images.

(3) In the traditional convolutional neural network, the convo- lution kernel is fixed. In this paper, dynamic convolution is used instead of traditional convolution. Formula (14) is used to calculate and output feature maps.

$$y_j^t = \sum_j w_{ij}^t * x_i^t \tag{14}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

x_i^t is the ith input feature map of the sample t, and w_{ij}^t is the ij input convolution kernel of the sample t, y_j^t as the jth output feature map sample. In the backward propagation, the dynamic convolution is the same as the traditional convolution operation. The dynamic convolutional layer calculation x_j^t gradient loss function L is as follows:

$$\frac{\partial L}{\partial x_i^t} = \sum_j (\frac{\partial L}{\partial y_j^t}) \underline{*} (w_{ij}^t) \tag{15}$$

Where $\underline{*}$ represents zero convolution padding. The gradient of the gradient loss function L with respect to w_{ij}^t is calculated as follows:

$$\frac{\partial L}{\partial k_i^t} = \sum_j (\frac{\partial L}{\partial y_j^t})^* x_i^t \tag{16}$$

Among them, $\hat{x}_i^t$ is the transposition of x_i^t. Compared with the traditional convolutional layer, w_{ij}^t is not a parameter or layer, as a function of the input x. The improved residual network [14] model is used as the meta-learner of MAML to extract features from the data. First, after improvement, a new MAML learner is used to learn ϕ. As the initial parameters of the verification set, formula (17) is used to update the underlying parameters of meta-learning. At this time, the network parameters are not updated, and formula (18) is used to calculate the network parameters. For the loss value, SGD is used for subsequent gradient descent. When the data training of a batch is completed, the isomorphic initialization parameters can be obtained according to formula (19), and our parameters can be updated.

$$\phi \leftarrow \phi - \eta \nabla_\phi L(\phi) \tag{17}$$

$$L(\theta) = \sum_{n=1}^{N} l^n (\hat{\theta}^n) \tag{18}$$

$$\hat{\theta} = \phi - \varepsilon \nabla_\phi l(\phi) \tag{19}$$

(4) The weight matrix $h \in R^{nm}$; $b \in R^m$, corresponding to the convolution kernel in the convolution layer; assuming the ReLU activation function is f, then the output at the first position is:

$$T_i = f(h^* a_{1:l+n-1} + b) \tag{20}$$

Downsampling using max pooling:

$$\tilde{T_i} = max\{T_i\} \tag{21}$$

After pooling:

$$T = [\tilde{T}_1, \tilde{T}_2, \ldots\ldots, \tilde{T}_{j-n+1}] \tag{22}$$

(5) The output is predicted by the softmax classifier. Forecast probability:

$$y = soft\,max(W_V V + b_V) \tag{23}$$

Choose the output with the highest probability as the predicted category.

5. Experimental results and analysis

5.1 Datasets

This paper uses Yale, Cropped AR, ORL face recognition image data sets, and compares the model proposed in this paper with the traditional face recognition model to verify the effectiveness of the method in this paper.

5.1.1 Yale face recognition dataset

The Yale dataset was created by the Center for Computational Vision and Control of Yale University. It contains 165 pictures of 15 volunteers, including changes in lighting, expressions, and poses. The Yale face database collects 10 samples of volunteers. Compared with the samples collected by each object in the ORL face database Yale database, the samples contain more obvious changes in lighting, expression, posture, and occlusion.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

5.1.2 Cropped AR face recognition dataset

The Cropped AR face database, which is necessary for face recognition, includes 1680 images, a total of 120 people, 14 images per person, corresponding to faces under different expressions and lighting conditions. It is a highly recognized database. Contains two data formats, so the model and accuracy of the model can be evaluated very well.

Table 2. Partitioning of datasets

Type	Types	Training set	Validati- on set	Test set
Count	120	100*10	100*4	20*14

5.1.3 ORL face dataset

The ORL face data set contains a total of 400 images of 40 different people, each with ten face images. All are grayscale images with a size of 92×112. It was created by Olivetti Research Laboratory in Cambridge, England from April 1992 to April 1994.

5.2 Experimental environment

This experimental environment uses the Windows 10 Enterprise Edition operating system, 64GB of memory, the processor is Intel Core i5-7200U CPU@2.50 GHz* 4, and the Python language is used under the Tensorflow framework.

5.3 Experimental details

In the process of training the model, the pictures are normalized uniformly, and the data set is divided according to the ratio of 7:3. The processing of the picture is shown in the network structure diagram, the learning rate adopts the dynamic preselected annealing learning rate, and the Adam optimizer is used. The parameters of the convolutional network refer to literature [15], and finally after constant debugging, it is determined that the size of the convolution kernel is 4*64, and the size of the picture is also 64*64. The Batchsize size is set to 128, the training batch period is 20, the results are output every 100 rounds, and the best results are saved. If the training result has not improved after 10 epochs, stop training. The curves of the accuracy of the three data sets with epochs are shown in Figure 4.

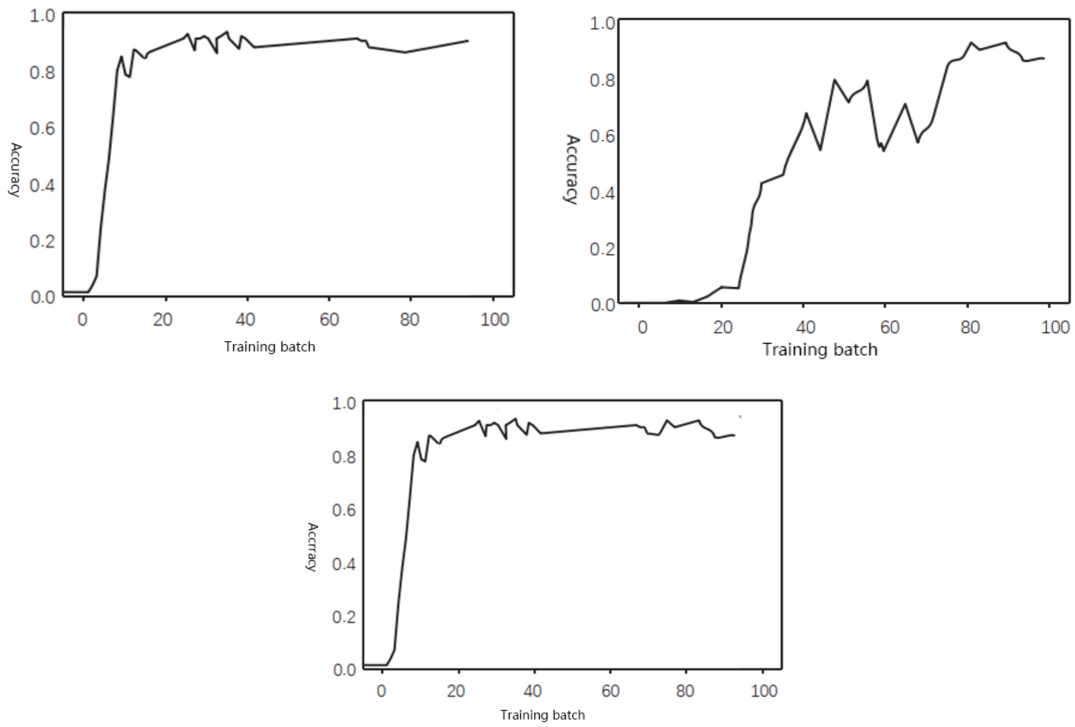

Fig. 4 Curve of recognition accuracy with training batch on three data sets

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

At the same time, other network models are selected on the three data sets for comparison experiments with the models proposed in this paper. The accuracy rates of the different models are shown in Tables 3 to 5.

Table 3. Accuracy results of YALE experiment (unit:%)

Algorithm	25* 25	50* 50	75* 75	100* 100
SRC	85.68	85.68	85.68	84.87
CRC	85.00	85.00	85.00	85.00
CSSVDL	92.13	92.76	92.76	92.76
OURS	93.94	94.82	95.23	95.23

Table 4. Accuracy results of AR experiment(unit:%)

Algorithm	50 Dim	100Dim	150Dim	200Dim
SRC	16.50	29.70	26.30	25.95
CRC	29.25	44.25	52.25	59.75
CSSVDL	66.00	78.83	79.67	83.85
OURS	67.25	82.17	86.96	89.40

Table 5. Accuracy results of ORL experiment(unit:%)

Algorithm	50 Dim	100 Dim	150 Dim	200 Dim	250 Dim
SRC	54.90	63.92	61.18	42.74	21.18
CRC	45.88	60.78	61.96	59.61	59.21
CSSVDL	50.63	62.53	66.82	70.73	70.57
OURS	59.71	66.42	70.88	75.31	77.26

5.4 Model comparison experiment analysis

In order to ensure that the model evaluation effect is similar to the real situation, the training set, test set, and verification set are used to evaluate the model. In order to verify the effectiveness of our model, we compare the model in this article with models such as SRC, CRC, and CSSVDL. The accuracy of the training set on the three data sets varies with the number of iterations, and the results are shown in Figure 8. Table 3 to Table 5 show the recognition rate results of each model on the three data sets.

It can be seen from Figure 8 that as the number of iterations increases, the accuracy of our model training continues to increase. It can be seen from the figure that in the 80th to 100th rounds, our method converges quickly, and then it is in a stable and fluctuating state. According to the early stop setting of the training, if the results of 20 rounds do not improve, stop training. In summary, the learning ability and generalization ability of our proposed model are relatively excellent.

The data in Table 3~Table 5 show that for input images of different dimensions, the accuracy of the model proposed in this article on the three data sets is higher than that of other comparison models, and the higher the input dimension, the image that the model can extract The richer the features, the higher the final recognition accuracy, which is about 3 to 7 percentage points higher than other models, confirming the effectiveness of the model. The data in the table shows that compared with the previous model, the model proposed in this article has relatively improved the accuracy of each input dimension. In summary, our proposed model has advantages over the previous models.

6. Conclusion

This paper proposes a small sample face recognition model under unnatural circumstances. First, use dynamic convolution instead of traditional convolutional network as the underlying feature extractor,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0080

which greatly improves the network's feature extraction ability of data. At the same time, our method fully considers the situation of face recognition with noise under unnatural circumstances. The obtained data is processed, and finally, the idea of double-layer optimization of meta-learning is used to enable the network to automatically find suitable network parameters, so that the model can achieve good performance through one or several steps of gradient descent on a new task. To a certain extent, the impact of low data volume on model accuracy is reduced. Experiments show that our model has a certain degree of improvement in accuracy compared to the traditional deep learning-based face recognition system.

References

[1] ARQUITECTURA E Y, INTRODUCCI T I, He Xiaoxia, et al. Small sample face recognition algorithm under transfer learning model[J/OL].Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2015, 53(9): 1689–1699. http://publications.lib.chalmers.se/records/fulltext/245180/245180.pdf%0Ahttps://hdl.handle.net/20.500.12380/245180%0Ahttp://dx.doi.org/10.1016/j.jsames.2011.03.003%0Ahttps://doi.org/10.1016/j.gr.2017.08.001%0Ahttp://dx.doi.org/10.1016/j.precamres.2014.12. DOI:10.16772/j.cnki.1673-1409.2019.07.017.

[2] MA X, ZHUANG W, FENG J. Loose Sparse Representation Based Undersampled Face Recognition with Auxiliary Dictionaries [J]. Moshi Shibie Yu Rengong Zhineng/ Pattern Recognition and Artificial Intelligence, 2016, 29(5): 439–446. DOI:10.16451/j.cnki.issn1003-6059.201605007.

[3] Dong Yanhua, Zhang Shumei, Zhao Junli. Overview of face recognition methods with occlusion[J]. 2020.

[4] VANSCHOREN J. Meta-Learning: A Survey[J/OL]. 2018: 1–29. . http://arxiv.org/abs/1810.03548.

[5] HILLIARD N, PHILLIPS L, HOWLAND S, et al. Few-Shot Learning with Metric-Agnostic Conditional Embeddings[J]. ArXiv, 2018.

[6] XU J, TON J-F, KIM H, et al. MetaFun: Meta-Learning with Iterative Functional Updates[J/OL]. 2019. http://arxiv.org/abs/1912.02738.

[7] CELLA L, LAZARIC A, PONTIL M. Meta-Learning with Stochastic Linear Bandits[J/OL]. 2020: 1–22. http://arxiv.org/abs/2005.08531.

[8] CHEN W Y, WANG Y C F, LIU Y C, et al. A Closer Look at Few-Shot Classification[J]. 7th International Conference on Learning Representations, ICLR 2019, 2019(2018): 1–17.

[9] HOSPEDALES T, ANTONIOU A, MICAELLI P, et al. Meta-Learning in Neural Networks: A Survey [J/OL]. 2020: 1–23. http://arxiv.org/abs/2004.05439.

[10] NICHOL A, ACHIAM J, SCHULMAN J. On First-Order Meta-Learning Algorithms[J/OL]. 2018: 1–15. http://arxiv.org/abs/1803.02999.

[11] FINN C, ABBEEL P, LEVINE S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks [J]. 34th International Conference on Machine Learning, ICML 2017, 2017, 3: 1856–1868.

[12] NICHOL A, SCHULMAN J. Reptile: A Scalable Metalearning Algorithm[J]. [no date]: 1–11.

[13] CHEN Y, DAI X, LIU M, et al. Dynamic Convolution: Attention over Convolution Kernels[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020: 11027–11036. DOI:10.1109/CVPR42600.2020.01104.

[14] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 2016-Decem: 770–778. DOI:10.1109/CVPR.2016.90.

[15] KIM Y. Convolutional Neural Networks for Sentence Classification[J]. EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014: 1746–1751. DOI:10.3115/v1/d14-1181.