Face Recognition Based on CD-RBM and BM-ILM

Weisong Qiao, Binbin Chen, Jiabao Zhao*, Shudong Guo
School of Management and Engineering Nanjing University, Nanjing, China

*Corresponding author: jbzhao@nju.edu.cn

Abstract. In recent years, the application of neural networks in face recognition has improved its accuracy greatly. However, it still suffers from the disappearance of the network gradient and the training cost. In this paper, we propose a new hybrid method which combines the contrast divergence algorithm (CD) accelerated Restricted Boltzmann Machine (RBM) and the Integrated Learning Method (ILM) with the Boosting algorithm (BM). First, we use RBM to establish a minimum energy model of the sample distribution. Then we use CD to speed up the feature extraction of samples in RBM. And then we build a multi-classifier with heterogeneous samples for ILM by BM. Finally, we experience our proposed method on the ORL database and the AR database. The results show that, with a simpler neural network determined by CD-RBM and BM-ILM, the recognition accuracy is better than that without the proposed method.

Keywords: Boosting method, face recognition, IML, RBM, CD method.

1. Introduction

Face recognition is increasingly popularly applied in many fields, such as transportation supervisory systems, mobile-payments, medical identification systems and national security [1]-[2]. The 2D image processing-based face recognition is still the dominant one among the diversity of face recognition methods, which generally includes three main categories [3]: the texture-based (local features) method, the model-based method, and the appearance-based (holistic) method. The appearance-based one is the most popular.

The appearance-based method considers the whole detected face image and represents this whole image from a perspective of a lower-dimensional subspace, which is learned from the training set. There are mainly four types of methods [3]: 1) nonlinear methods, such as ISOMAP LLE, KPCA, 2) linear methods, e.g., LDA, ICA, PCA, etc., 3) sparsity-based methods, and 4) learning-based methods. The first two kinds of methods are hard to control the accuracy of recognition, and the third method is too complex to apply widely. The learning-based method is more suitable for application [3], so how to extract the effective features efficiently is very important.

There are two important parts in these learning methods: 1) image feature extraction [4]-[7] and 2) image classification [8]-[10]. For image features extraction, in 1996, Ojala proposed the LBP (Local Binary Pattern) algorithm, which extracted the image texture information and generated the LBP value by comparing the local pixel value and the central pixel value [4]. In 1999, Lowe [5]-[6] proposed the scale-invariant feature transform algorithm (SIFT), which only selected some key feature points for description, and achieved scale invariance. In 2008, Bay [7] proposed the speed-up robust features
(SURF), including acceleration and robust features, as an improvement to the SIFT, which greatly reduced the running time of the program and increased the robustness of SURF. To the image classification, Osuna [8] proposed a face detection algorithm based on SVM and obtained good results in the recognition of high-definition images in 1997. Dong [9] realized the Bayesian classifier to classify and judge urine sediment images in 2007. Pan [10] used the decision tree method to classify remote sensing images in 2008.

Compared with the traditional classification methods mentioned above, researchers tend to adopt neural networks in face recognition recently [11]-[15]. AlexNet network won the championship in the 2012 ImageNet competition [11]. In 2014, Szegedy designed the GoogleNet network, and proposed the inclusion structure and branch structure [12], which reduced the error rate to 6.7% in the ImageNet data set. The VGG network further improved the recognition accuracy with more network layers [13].

However, although there are different forms of the learning methods, the problems on the disappearance of the network gradient and the rapid increase of the training cost are valuable to be researched. A large part of the research focuses on modifying the depth of the network and adjusting the parameters to improve the recognition accuracy of the network [14]-[17]. For example, [14] proposed the residual network structure, which decreases the network depth, and solves the problem of gradient disappearance effectively. [15] proposed a new solution to the problem of gradient disappearance, which allows maximum transmission of information.

In this paper, we focus on the disappearance of the network gradient and the rapid increase of the training cost in face recognition system. A new learning method is proposed in this paper based on the Restricted Boltzmann Machine (RBM) [16]-[17] and the Integrated Learning Method (ILM). First, we use RBM to establish a minimum energy model of sample distribution. And then, we apply the contrast divergence (CD) algorithm to achieve the feature extraction of the sample in RBM. Finally, the Boosting Method (BM) is used to design a multi-classifier with heterogeneous samples for ILM to improve the accuracy of face recognition, including the integration of the classifiers.

The other in this paper is organized as follows. Section II is the preliminaries for RBM and ILM. Section III is the design of CD-RBM and BM-ILM. Section IV is the results of the method based on CD-RBM and BM-ILM. Section V is the conclusion.

2. Preliminaries for RBM and ILM

2.1. The Restricted Boltzmann Machine (RBM)
RBM is developed based on the Boltzmann machine (BM), which is an intra-layer and inter-layer all-connected model. To reduce the complexity of network calculation, RBM removes the intra-layer connection in the BM network, and only retains the connection between the explicit layer and the hidden layer, as shown in Fig. 1.

![Fig. 1 Comparison of Boltzmann and Restricted Boltzmann](image-url)
The visible layer of RBM is represented by $v$, which node is used to receive the input signal, and the hidden layer is represented by $h$, which can be regarded as the feature extractor of the input signal. Provided that the visible layer has $n_v$ neurons and the hidden layer has $n_h$ neurons, the energy function of the network can be defined as [18]:

$$E(v, h|\theta) = -\sum_{i=1}^{n_v} \sum_{j=1}^{n_h} v_i W_{ij} h_j - \sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j,$$

(1)

Where $i$ represents the $i$th neuron in the visible layer, $j$ is the neuron in the hidden layer, $W_{ij}$ denotes the weight value from visible neuron $i$ to hidden neuron $j$, $a_i$ is the bias of the $i$th neuron in the visible layer, $b_j$ is the bias of the $j$th neuron in the hidden layer. According to (1), we can define the joint probability distribution of the visible layer and the hidden layer:

$$P(v, h|\theta) = \frac{e^{-E(v, h|\theta)}}{Z(\theta)},$$

(2)

$$Z(\theta) = \sum_{v, h} e^{-E(v, h|\theta)},$$

(3)

Where $Z(\theta)$ is called the partition function. On the one hand, the calculation of the normalization factor needs to be performed many times. On the other hand, the calculation will be greater for high-dimensional data, which makes it not directly applicable in practical applications.

However, RBM adopts full connections between layers and no connection within the layer. When we input the signal to the visible layer, the visible layer will determine the state of each neuron in the hidden layer, and there is no connection between the layers. At this time, the hidden layer neuron is conditionally independent, so the probability that the $i$th neuron in the visible layer is active and the probability that the $j$th neuron in the hidden layer is active are respectively denoted as:

$$P(V_i = 1|h) = \sigma(-\sum_i W_{ij}v_j - \sum_i b_i v_i),$$

(4)

$$P(h_j = 1|v) = \sigma(-\sum_j W_{ij}h_j - \sum_j a_j h_j),$$

(5)

Where $\sigma(\cdot)$ is the activation. Different activation functions can be selected according to the different data types of the visible layer. Generally speaking, for binary data, the most frequently used activation function is:

$$\sigma(x) = \frac{1}{1+e^{-x}},$$

(6)

Through a series of derivations, we can get:

$$P(h|v) = \prod_{j=1}^{n_h} P(h_j|v),$$

(7)

$$P(v|h) = \prod_{i=1}^{n_v} P(v_i|h).$$

(8)

Which represents the final activation probability of nodes in the hidden and visible layers of the RBM model, respectively.

The model needs to find a series of suitable parameters to fit the data through training. Since RBM is an energy-based model, it only needs to maximize the log likelihood, which is defined as:
\[ \ln l = \ln \prod_{i=1}^{n} P(v^i) = \sum_{i=1}^{n} \ln P(v^i), \]  

(9)

Where \( v^i \) is the \( i \text{th} \) training sample, and \( n \) is the number of training samples. In order to achieve the goal, the most commonly used method is the gradient ascent method, which is approximated by iteration. The iteration format is defined as \( \theta := \theta + \eta \frac{\partial \ln l}{\partial \theta} \), where \( \eta > 0 \) is the learning rate.

2.2. Integrated Learning Method (ILM)
ILM accomplishes the learning task by building and integrating multiple learners. Unlike traditional machine learning which always tries to learn a hypothesis set from the training set to improve the learning effect, the integrated learning tries to build another hypothesis set and uses these different sets together.

2.3. Boosting Method (BM)
BM can promote a weak learner to a strong learner based on the processing of the training set to construct the intuitive and straightforward type of digits as the result of the ensemble learning, rather than the average one. Different from the typical Bagging algorithm, the core idea of Boosting is to repeatedly apply a primary learner to adjust the training data set, so that the training samples receive more attention in subsequent training. After the iteration of a series of basic learners are produced accordingly, instead of training multiple classifiers through data reconstruction.

In the beginning, all training samples are set to the same weight, each enhancement iteration will generate a new basic learner, and this learner can adapt to the weighted training data set. Meanwhile, the error rate of each iteration is calculated, and then the weight is adjusted, according to the difference between the wrong division and the correct division. The weight of the correctly divided data set will be reduced, and the weight of the wrong one will be increased so that in the following training, the wrong training item will be focused. Repeat this until the number of learners reaches the goal we set, and finally we can get the classifiers combined weights.

3. Design of CD-RBM and BM-ILM
In this part, we design the method based on CD-RBM and BM-ILM.

For CD-RBM, the determination of the number of hidden layer nodes is a crucial step in the design of the RBM network, however, there is currently no relevant theoretical research to give a specific setting method, and it is usually set based on experience. Generally, the setting of the number of hidden layer nodes relates directly to the size of the training data, cannot take it for granted that the more nodes, the better the model. Too many hidden layer nodes which may bring strong learning ability, bring inconsistent extracted attributes and reduced generalization ability as well, thus enlarge the testing errors. However, if the hidden layer nodes number is too small, the learning ability will be insufficient. The following are two empirical formulas for calculating the number of hidden nodes:

\[ k < \sum_{i=1}^{n} C_{i}^{n_{1}}, \]  

(10)

\[ n_{1} = \sqrt{n + m + a}, 1 \leq a \leq 10. \]  

(11)

Where \( k \) is the number of samples, \( n_{1} \) is the number of hidden nodes, \( n \) is the number of the input layer, \( m \) is the number of output layer, and \( a \) is a compensation number. The main procedure can be concluded as: 1) determination of the number \( N \) of the nodes in the hidden layer of RBM, the initial weights \( w_{i,j} \), and the bias; 2) data processing of all hidden nodes for all pixel values; 3) data reconstruction, i.e., data is passed forward and backward multiple times between the visible layer and the hidden layer of the RBM model until the reconstruction error reaches the minimum; 4) usage of the contrast divergence (CD) algorithm to gradient calculation for quick learning.
For BM-ILM, BM is used to upgrade a series of weak learners to strong learners. The core of BM is to repeatedly apply a basic learner to adjust the training data set, so that the training samples can receive more attention in subsequent training. After a certain number of iterations, there are a series of basic learners, instead of training multiple classifiers through data reconstruction. The main procedure of BM-ILM is as follows: 1) Divide the overall data set into two parts: the training set and the testing set, and set initial weights for the training set; 2) Select a quantitative sample from the overall training set according to the weight as the training sample for this time; 3) Input the training samples into the learner for training until the training is completed; 4) Input the overall training set to the trained learner to get the current learner’s predicted label for the overall training set, and compare the predicted label with the correct label of the overall training set. According to the comparison, adjust the overall training set for each sample Weight, get all the training set with new weights. Besides, the weight of the current learner needs to be calculated according to the label comparison and recorded in the corresponding learner weight storage vector; 5) Input all the testing sample into the learner, get the prediction result of the current learner on the testing sample, and store it in the corresponding prediction result matrix; 6) Repeat steps 2 to 5 until the set number of learners are trained or the accuracy rate reaches the desired effect.

4. Results of The Hybrid Method
In this part, we mainly apply the above method based on CD-RBM and BM-ILM in the ORL database and AR database. For example, the ORL data set contains 40 directories, each with 10 images, and each directory represents a different person, some of which are shown in Fig. 2. All images are grayscale images with a width of 92 and a height of 112. Compared with the ORL database, the AR database contains more people images with different genders and ages, and some images are sheltered.

![Fig. 2 A part of face images in ORL database](image)

![Fig. 3 Recognition accuracy with different iteration times](image)
Firstly, find the optimal number of iterations. 300 samples are randomly selected from the ORL database for training and 100 samples for testing in the case of two fixed learning rates: 0.005 and 0.001. The results are shown in Fig. 3, which show that: 1) the accuracy with learning rate 0.005 is always better than that with 0.001; 2) the accuracy with learning rate 0.001 is too small to be selected for long-time iteration; 3) the accuracy with learning rate 0.005 exists an optimal value 95% with the optimal number of iteration 900.

Secondly, determine an appropriate learning rate with three different numbers of iteration: 200, 500, and 800. The results are illustrated in Fig. 4, which show that: 1) when the value of the learning rate changes from 0.003 to 0.01, the impact on the overall recognition rate is not too great; 2) as the number of iterations increases, the impact will continue to decrease; 3) the accuracy becomes pretty when the number of iteration is 800 and above.

Then, in order to verify the method based on CD-RBM and BM-ILM, we divide the data set as follows: 1) the overall training set is 320 images, 8 images per person, in each learner 2) the overall test set is 80 images, 2 images per person, in each learner, and 3) the training sample set in each learner is 280 images. When the number of learners exceeds 1, it is regarded as integrated learning. We mainly consider two kinds of numbers of the learners: 1 and 5. The results are presented in Fig. 5, which show that in most cases, the recognition accuracy after integration is higher than that of the unintegrated case, e.g., when the number of iterations is 900 and 5 learners are integrated, the recognition rate reaches 97%, which verifies that our proposed method is available.

![Fig. 4 Recognition accuracy with different learning rates](image1)

![Fig. 5 Recognition accuracy with different iteration times](image2)
Finally, we also apply our proposed method based on CD-RBM and BM-ILM in the AR database, and the main procedure is similar to the one in the ORL database. Here, for limited pages, we only consider the recognition accuracy with different learners, ranging from 1 to 25, which is enough for the explanation. The results are given in Tab. 1, which show that the highest recognition accuracy reaches 95% with the number of the learner 25.

| Number | 1   | 5   | 10  | 15  | 20  | 25  |
|--------|-----|-----|-----|-----|-----|-----|
| Accuracy | 0.86 | 0.88 | 0.89 | 0.91 | 0.92 | 0.95 |

Further, the data in Tab. 1 is obtained under the condition that the number of hidden nodes in the RBM network is 400, the number of iterations is 1300, and the learning rate is 0.006.

5. Conclusions
In this paper, a new method based on CD-RBM and BM-ILM is proposed for the face recognition problem. First, we use RBM to establish a minimum energy model of sample distribution, which extract the features of the image well. Then, we apply CD to achieve the feature extraction of the sample in RBM, which speeds up the procedure. Finally, the BM is designed to construct a multi-classifier based on ILM with heterogeneous samples to improve the accuracy, including the integration of the classifiers. We also apply this method in the ORL database and AR database. The application procedure contains: the determination of the optimal number of iterations, the appropriate learning rate with different numbers of iteration, and the number of iterations. The results show that, in the ORL database, a single learner can reach the accuracy of 95%, where the hidden node of RBM is 500, the learning rate is 0.005, and the number of iterations is 900. While under ILM, the highest recognition accuracy can reach 97%.

There are still some problems for future work, for example, 1) the selection rules of training data, 2) new algorithms for resolution-robust feature extraction methods, etc.

References
[1] A. Distante and C. Distante. Handbook of Image Processing and Computer Vision, Volume 2: From Image to Pattern. Springer, Cham, Switzerland, 2020.
[2] E. Priya and V. Rajinikanth. Signal and Image Processing Techniques for the Development of Intelligent Healthcare Systems. Springer, Singapore, 2021.
[3] M. Taskiran, N. Kahraman, and C. E. Erdem. Face recognition: Past, present and future (a review). Digital Signal Processing, 2020, 106:102809.
[4] T. Ojala, M. Pietikinen, D. A. Harwood. Comparative study of texture measures with classification based on featured distributions. Pattern Recognition: The Journal of the Pattern Recognition Society, 1996, 29(1): 51-59.
[5] D. G. Lowe. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Sep. 1999.
[6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–110, 2004.
[7] H. Bay, A. Ess, and T. Tuytelaars. Speed-up robust features. Computer Vision and Image Understanding, 2008,110(3): 346-359.
[8] E. Osuna. Training support vector machines: An application to face detection. In Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition,1997.
[9] L. Dong and S. Yuan. Image classification based on Bayesian classifier. Journal of Jilin University, 2007(02): 249-253.
[10] Z. Pan and P. Du. Decision tree classification method and its application in remote sensing image processing. Surveying and Mapping Science, 2008(01): 211-253.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[12] C. Szegedy et al., Going deeper with convolutions. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015: 1-9.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016.

[14] K. He, X. Zhang, S. Ren and J. Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016.

[15] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger. Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017.

[16] C. Zhang et al. Restricted boltzmann machine. Journal of Engineering Mathematics. 2015, 32(2): 159-173.

[17] G. Rebala, A. Ravi, and S. Churiwala. An Introduction to Machine Learning. Springer, Cham, 2020.

[18] J. Yu, J. Gwak, S. Lee and M. Jeon. An incremental learning approach for restricted boltzmann machines. In 2015 International Conference on Control, Automation and Information Sciences (ICCAIS), Changshu, 2015.