Face Detection Algorithm Based on Convolutional Pooling Deep Belief Network

Dandan Wang*, Ming Li and Xiaoxu Li

School of Computer and Communication, Lanzhou University of Technology, Gansu Lanzhou, 730050, China
*Corresponding author

Abstract—**When using a single deep model to solve the problem of face detection, it is easy to have the problem of high false detection rate and low learning efficiency, the mixed model algorithm based on deep learning was proposed to solve these problems of face detection, which is called the CPDBN (Convolutional pooling deep belief network). Experimental results show that the algorithm improves the accuracy of face detection in the face of partial occlusion, and increases the robustness of multi-pose.**

Keywords-face detection; deep learning; CPDBN; partial occlusion

I. INTRODUCTION

The face detection is a critical first step in any face recognition system. In practical applications, the face image is often affected by the surrounding environment, the acquisition of equipment and changes in light, and other factors, resulting in occlusion and complex background in the face detection, then leading to the decline of the face detection performance [1]. In which that the occlusion is one of the main factors in affecting the face detection.

Deep learning can learn deep nonlinear network structure and realize complex function approximation, which has strong ability to learn the essential features of data sets from a small number of samples [2,3]. A typical deep learning structure including deep belief networks (DBN), convolution neural network (CNN), MLP and so on. Ch'Ng S I et al. [4], who proposed using a plurality of DBN to respectively learning face image blocks, the proposed method demonstrating a higher accuracy of recognition, but before the input we need to preprocess the images, likely to cause loss of information, and in the case of occlusion the effect is not substantial. Hurieh Khalajzadeh [5], who proposed a face detection algorithm based on convolution neural network of the hierarchy, greatly improve the detection rate, but this method is more complicated to implement.

Face detection algorithm based on a single depth model, has yielded some results, but there are still some problems. In this paper, by combining the traditional deep learning algorithms (Deep belief networks, DBN [6] and Convolutional neural network, CNN [7]), constructing the convolutional pooling deep belief network model, setting up the number of layers and nodes, using the model to study the position of facing local features and to explore strong relationship between various features. In the deep model of our design, combined the bottom parts to become high-level portion that more complete, depending on the combination of high-level results to judge and predict the bottom shadowed facial feature position.

II. RELATED THEORY

Restricted Boltzmann Machine [8] is a two-layer neural network that consist of hidden and visible layers and it has a graph structure with bidirectional symmetric connection and no self feedback mode. If the RBM has m visual units and n hidden units, h and v are used to represent the visual units and the hidden units, the energy function of RBM can be defined as:

$$E\left(\mathrm{v},h\right)=-\sum_{i=1}^{m}\sum_{j=1}^{n}W_{ij}v_ih_j-\sum_{j=1}^{n}b_jh_j-\sum_{i=1}^{m}c_iv_i \tag{1}$$

Among them, b_j is the bias of the j-th hidden units, c_i is the bias of the i-th visual unit, W_{ij} is the connection weights of the j-th hidden unit and the i-th visual unit. Based on the definition of the energy function, the joint distribution function of the probability of RBM can be defined as the following formula and Z is the normalizes factor.

$$P\left(\mathrm{v},h\right)=\frac{1}{Z}\exp(-E(v,h)) \tag{2}$$

DBN is a multiple RBM stack, which is a probabilistic generation model. Each layer includes a group of binary or real valued units, and it can be extracted the abstraction features of a higher level from the output characteristics of the above layer. Hinton in 2006 proposed a greedy learning algorithm layer by layer training for deep belief networks, simply divided into two steps: first, training used the unsupervised learning layer by layer from the bottom to the top layer; the second is to fine tune the network and top-down for supervised learning.

III. CONVOLUTIONAL POOLING DEEP BELIEF NETWORK MODEL

A. Convolutional Pooling Restricted Boltzmann Machine (CPRBM)

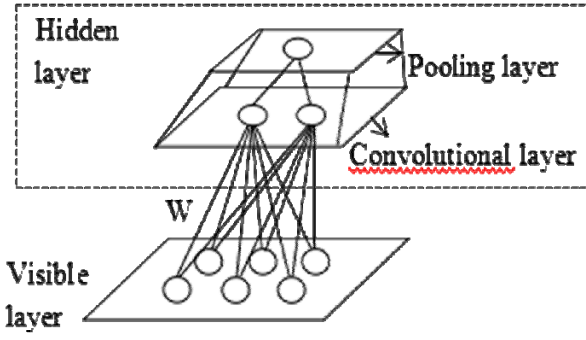Intuitively, CPRBM is similar to RBM. Convolution layer and pool layer are added to the hidden layer of CPRBM.

FIGURE I. THE BASIC STRUCTURE OF CPRBM

We can assumed that the visible layer is composed of a series of binary unit array and convolutional layer and pool layer includes K groups, each of which is composed of a binary unit array and is associated with the filter. In the hidden layer, each hidden layer unit cluster is a feature map, the hidden layer units are separated into different feature maps to represent the different features of the different locations of the visible layers. Among them, the filter weights are shared among the hidden layer groups. Each hidden group has a bias b_k, and all the visible units share an offset value d. Pooling layer reducing the feature representation of convolutional layer by a small integer factor I, that is to say, convolutional layer is divided into multiple n blocks which size is $I \times I$, each block is connected to a binary unit of the pooling layer. We define $\{(i,j) : c_{i,j} \in n\}$. Then we can define the energy function of CPRBM:

$$E(v,c) = -\sum_k \sum_{i,j} \left(c_{i,j}^k \left(\tilde{W} * v \right)_{i,j} + b_k c_{i,j}^k \right) - d \sum_{i,j} v_{i,j} \qquad (3)$$

Among them, $\tilde{W}$ represents the flip array of W in horizontal and vertical directions.

When given a visible layer V, the Gibbs sampling of convolution layer C and the pooling layer need to be further discussed. The K group in the hidden layer can receive signals from the visible layer V:

$$S(c_{ij}^k) \propto (\tilde{W} * v)_{ij} + b_k \qquad (4)$$

Assume that the K group element nodes in the hidden layer share a weight value T, T^m represents the weight matrix connected from the convolutional layer c^m to the pooling layer P^k. Pooling layer P receives a signal from the convolutional layer C as follows:

$$S(p_n^k) \propto \sum_m \left(T^m * c^m \right)_n \qquad (5)$$

Next, independently sampled each block as the input of its

polynomial function, the conditional probability can be written as follows:

$$P(c_{ij}^k = 1 \mid v) = \frac{\exp(S(c_{ij}^k))}{1 + \sum_{(i,j)} \exp(S(c_{ij}^k))} \qquad (6)$$

$$P(p_n^k = 0 \mid v) = \frac{1}{1 + \sum_{(i,j)} \exp(S(c_{ij}^k))} \qquad (7)$$

When given the hidden layer, can be sampled of the visible layer in the same way.

B. Convolutional Pooling Deep Belief Network (CPDBN)

In order to obtain a higher level of representation features, like DBN as the CPRBM stacked into a multi-layer structure.
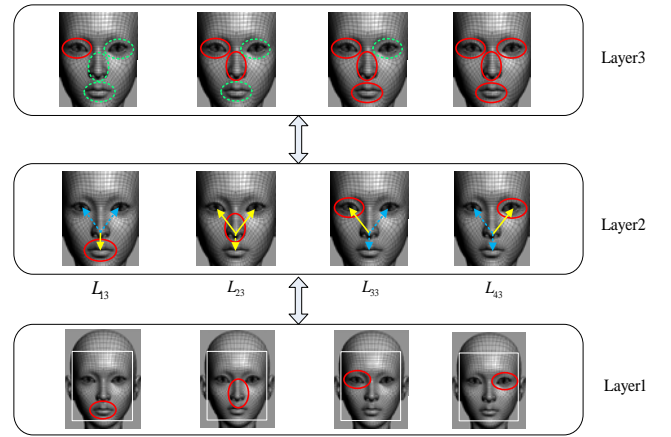

FIGURE II. THE MODEL DETAILS OF CPDBN FOR FACE DETECTION

The deep mixed model proposed in this paper consists of three layers, as shown in figure II. L_{i3} indicates the position of the i-st feature relative to the other 3 features. The first layer detects the position of the face and locates the feature area. The second layer is used to study the relative position L_{i3} of each feature and feature of the face. The third layer of combining the learned facial features and prediction feature position is blocked. The solid circles represent the detected facial feature, a combination of one feature that detected has 4 kinds, a combination of two features has 6 kinds, a combination of three features has 3 kinds, a combination of four features has 1 kind. The dotted line indicates the predicted facial feature position, and the third layer contains the number of features that may be learned.

C. Training Network Model

The model of this paper is to be based on the improvement of DBN, therefore, we use a similar method of DBN to train our model. But the training process of deep model needs a little improvement. The training parameters are divided into two stages:

Stage 1: For l =1 to 2 {Train parameters for layer l and $l+1$ using RBM.}

Stage 2: Fine-tune all the parameters by back propagating error derivatives.

In the first stage, in structure of the multi-layer CPDBN network, all the adjacent two layer structures can be regarded as a RBM, lower hidden layer will be used as the input layer that adjacent thereto a higher hidden layers. the s output feature maps from the $l+1$ layer is:

$$Y_s^{l+1} = sigm(\sum_r T_{rs}\left(Y_r^l * ke_r^l\right) + b_k^l) \qquad (8)$$

After many times of training, the spatial parameters are constantly adjusted, and the spatial parameters $\{W, d, b \mid T, m\}$ of CPDBN can be determined initially, however, it is still necessary to fine tune and optimize the parameters of the CPDBN. In Stage 2, all the parameters are fine tuned by the BP network. The derivation process of the parameters in the hidden layer is followed the method in the [9].

IV. FACE DETECTION BASED CPDBN

In this section, we use F to indicate whether the current window contains a face. If the detection results in the presence of a human face, then F=1, if not, then F=0. As shown in Figure III.
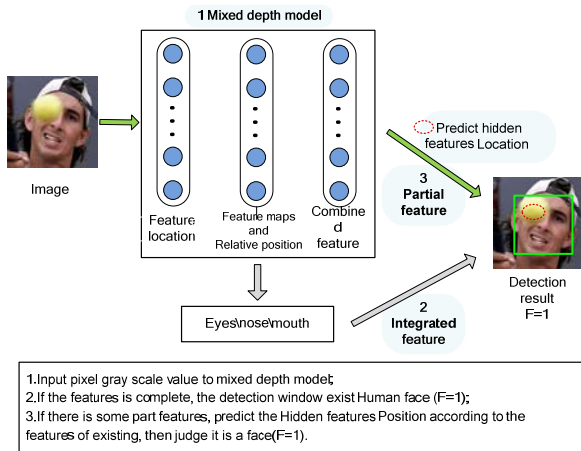

FIGURE III. FACE DETECTION FRAMEWORK BASED ON CPDBN MODEL

Each layer is a mixed model unit, each unit consists of 1 visible layer and 1 convolutional layers and 1 pooling layer. The convolution kernel size of three layer volume were 5 * 5, 1 * 1, 1 * 1. The first layer uses edge filter to learn features from the face image, then we use unlabeled images from the LFW Library to training the other two layers of CPDBN. The first layer is composed of 24 local edge filters, the size of each filter is 10*10px, and the pooling ratio is 3, which is used to extract the edge features of human face organs; in the second layer we set up 40 sets of 10*10px filter, used to study the part of the features of the face and the relative position, such as the eyes,

nose, mouth; the third layer is made up of 24 groups of 14*14px filters, the pooling ratio of the second and third layers is 2, the third layer is used to combine the parts of the facial features from the second layer and also it can predict the hidden feature location.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the performance of the proposed method, the group experiment analyze the performance of the proposed algorithm, the experimental data from LFW database.

A. Performance Analysis Experiment

The purpose of this experiment is to test the performance of this algorithm, the experimental data from the LFW face database that is not marked in the picture. In the experiment, we selected 1500 images as a training set, and 1000 pictures as a test set.

Experimental results: in the test set of 1000 images, including the face number 978, wherein the error detection 21, missed 10, so the False Positive Rate is 2.1%, the Correct Detection Rate was 98.9%. Some of the results are shown in the following figure:

FIGURE IV. THE PARTIAL DETECTION RESULTS OF CPDBN IN LFW DATABASE

In the picture IV, The dotted line circle represents the location of the occluded feature. Because of the deep mixed model we used only had three layer and 22 hidden units, local connections between the hidden layer and the visible layer, and the weights of the model are shared, thus the training and inference time of the model is short, which can meet the requirements of practical application.

B. Comparative Experiment

To further test the effectiveness of this algorithm, the proposed CPDBN and the current advanced depth learning algorithm DBN, CNN experimental comparison at the same time. In the experiment, using the same experimental environment and experimental data, the experimental data from the LFW database contain partial occlusion and multi-pose. We take the average of 10 experiments as the final comparison data. In this paper, we use the ROC curve to

evaluate the algorithm is good or bad. ROC is a curve related to Correct Detection Rate about False Positive Rate, face Correct Detection Rate about Recall Rate. The experimental results are shown below:
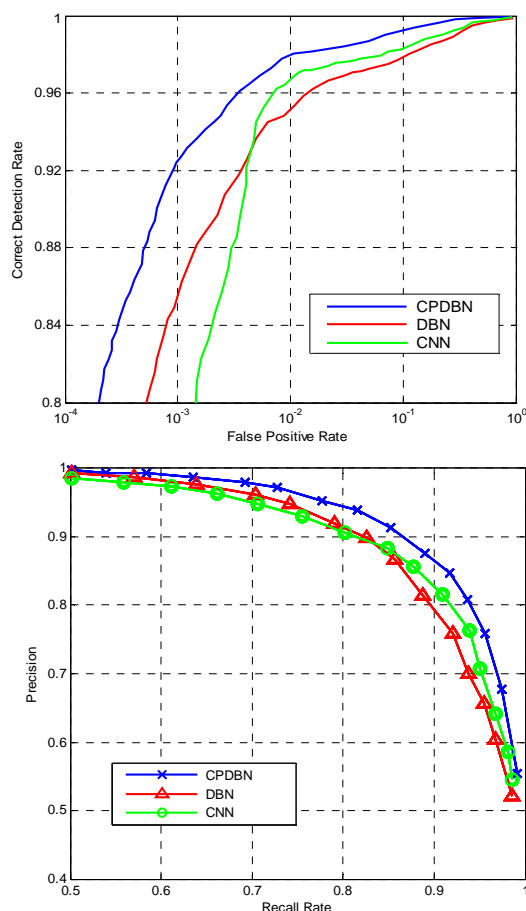

FIGURE V. COMPARISON OF ROC CURVES

The experimental results show that the proposed algorithm is superior to the other two algorithms. As can be seen from the first figure, when the false alarm rate is 0.1%, the accuracy of the algorithm CPDBN is 92.7%, while the accuracy rate of the algorithm DBN is 85.3%, the accuracy of the algorithm CNN is 76.2%; as can be seen from the second figure, when the recall rate is 80%, the accuracy of the algorithm CPDBN in this paper is 94.2%, while the accuracy of the algorithm DBN is 91.6%, and the accuracy of the algorithm CNN is 90.8%.

VI. CONCLUSIONS

The face detection algorithm based on CPDBN has the following features: 1) In order to reduce the loss of information, the algorithm can avoid the complex pre-processing of the image, and input the original image directly. 2) The proposed model CPDBN has a scalable unsupervised learning capability, which can be used to study and predict the feature location of occlusion, therefore, the accuracy of face detection is improved. The experimental results show that the algorithm has better detection effect in partial occlusion and multi-pose.

REFERENCES

[1] H. Hatem, Z. Beiji, and R. Majeed, "Human Facial Features Detection and Tracking in Images and Video", Journal of Computational and Theoretical Nanoscience, vol. 12, no. 11, (2015), pp. 4242-4249(8).

[2] Zheng W, Hu D, Wang J. Fault Localization Analysis Based on Deep Neural Network[J]. Mathematical Problems in Engineering, vol. 2016, (2016), pp. 1-11.

[3] Y. Bengio, and O. Delalleau, "On the Expressive Power of Deep Architectures", Proceedings of the 22nd International Conference on Algorithmic Learning Theory, Espoo, Finland, October 5-7.

[4] S. I. Ch'Ng, K. P. Seng, and L. M. Ang, "Block-based Deep Belief Networks for face recognition", International Journal of Biometrics, vol. 4, no. 2, (2012), pp. 130-143.

[5] H. Khalajzadeh, M. Mansouri, and M. Teshnehlab, "Hierarchical Structure Based Convolutional Neural Network for Face Recognition", International Journal of Computational Intelligence and Applications, vol. 12, no. 3, (2013).

[6] G. E. Hinton, "Deep belief networks", Scholarpedia, vol. 4, no. 5, (2009), pp. 5947.

[7] B. S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face Recognition: A Convolutional Neural Network Approach", IEEE Transactions on Neural Networks, vol. 8, no. 1, (1997), pp. 98-113.

[8] C. X. Zhang, N. N. Ji, and G. W. Wang, "Restricted Boltzmann Machines", Chinese Journal of Engineering Mathematics, vol. 32, no. 2, (2015), pp. 159-173.

[9] G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines", Momentum, vol. 9, no. 1, (2010), pp. 599-619.