

A Face Detection Algorithm Based on Deep Learning

Ming Li, Chengyang Yu, Fuzhong Nian and Xiaoxu Li

*School of Computer and Communication
Lanzhou University of Technology
Gansu, Lanzhou, 730050, China
lim3076@163.com, ycy_0323@163.com*

Abstract

To achieve the problem of partial occlusion and multi-pose in the face detection, a face detection algorithm based on deep learning is proposed. Through establishing deep model, the probabilistic correlations of the visibilities are learned in different face local region. Firstly, face local regions are detected by part-based face detector. Secondly, the detection results are feed to deep model, which describes the correlations of different regions. Finally, the face detection is completed according to the correlations. The experimental results show that the proposed algorithm is effective in the case of occlusion and multi-pose.

Keywords: *Face detection, deep learning, deep model, part-based, detection rate, false positive rate, recall rate*

1. Introduction

Face detection is a computer technology that determines the locations and sizes of human faces in digital images, which is a key technology in face information processing. It has been widely applied to pattern recognition, identity authentication, human computer interface and automatic video surveillance, *etc* [1].

At present, the positive face detection has achieved satisfactory results under ideal conditions. However, in the practical situations, this kind of methods cannot be effectively applied to occlusion and multi-pose [2]. Therefore, it is worthwhile to a further investigating for face detection.

To handle the face pose variation problem under different conditions, part-based face detection methods have been proposed [3]. Although the proposed part-based face detection methods have attained good results, their performance will decrease in case of occlusion or large pose variation, since those models only summed the scores of part-based detectors without exploring the correlations among the visibilities of different face parts [4]. If a part of the face is occluded or large pose variation, the score of part-based detector would be very low, which would lead to a relatively low summed score. Therefore, the detection rate is insufficient in this state.

Currently, Deep Learning is a new area of computer vision and machine learning research, which has been successfully introduced to image dimensional reduction [5, 6] and recognition [7, 8]. Deep learning originates from artificial neural networks and consists of multi-layer perceptron (MLP) of multi-hidden layers which is a deep learning structure. By using model architectures composed of multiple non-linear transformations, deep learning finds the high-level features in data. The higher level features are derived from lower level features to form a hierarchical representation [9].

In this paper, the stronger correlations among different face parts are investigated through deep learning (Deep Belief Networks DBN) [9]. The overlap region of different face parts are designed at each layer of DBN. The parts on the higher layer consist of parts on the lower layer. The correlations among the parts on the lower layer are explored

according to the detection results of the parts on the higher layer. The detection result of a part provides valuable information for the estimation on its overlap parts. Binocular and left-eye are overlap parts on different layers. In the detection process, we assume that the part of binocular gets a higher detection score since its visible feature satisfies the corresponding part detector, while the part of left-eye gets a lower score since it doesn't find visible feature to fit the corresponding detector. If the correlations among different parts are modeled by a reasonable way, the detection result of binocular part can be used to infer that the left-eye part is visible. Therefore, the main work is correctly model the correlations of the visibilities in different face parts and combine the detection results of part-based face detectors. Sequentially, the face detection is completed based on correlations and the gross detection result.

2. Related Works

Multi-pose and occlusion are considered as the key problem of face detection. To solve the aforementioned problem, many part-based face detection models have been proposed [10]. For instance, Heisele, *et al.*, [11] have proposed a part-based face detection method, which includes kernel-SVMs used as the part detector and LDA adopted to combine the results of part detector. This method has achieved better performance when using face parts rather than only using the whole face. However, it was not designed to directly deal with partial occlusion faces. Kiyoto Ichikawa, *et al.*, [12] have proposed a face detection method based on partial information, which includes the AdaBoost method used to train the classifiers of partial face and LDA and a decision tree structure adopted to integrate the output of partial classifiers to obtain the final detection results. Its partial information consisted of eyes, noses, and lips. In the case of occluded eyes or mouths, the performance of this method was better. However, this kind of method cannot be effectively applied to the other type of occlusion. In addition, Huang, *et al.*, [13] have proposed a component-based framework for generalized face alignment. Yang, *et al.*, [14] have proposed a face liveness detection method with component dependent descriptor and Zhang, *et al.*, [15] have proposed a face detection method based on local region sparse coding, *etc.*

The existing part-based face detection methods have obtained certain achievements, but they cannot be effectively applied to the case of occlusion [16]. The visibility estimation plays a significant role in dealing with occlusion. The above methods supposed the visibility of every part was independent and adopted hard-threshold to estimate the visibility [17]. The hard-threshold has some disadvantages. Firstly, it can not make a distinction between partial occlusions and full occlusions. Secondly, if a large part is wrongly classified as occlusion, it can not be corrected by the rules of hard-threshold. Thirdly, the rules can be only defined manually, while it can not be learned from the training data. If the rules can be learned from the training data, it would be more robust in the practical situations. Therefore, it is important to propose and develop a method to estimate the visibility.

To solve the above problems, a deep learning model is proposed. It can automatically learn the probabilistic correlations of the visibilities in different face parts. The visibilities can be estimated based on the probabilistic correlations. The proposed deep model has the following advantages. Firstly, the hierarchical structure of the proposed deep model matches well with the multi-layers of our parts model. The visibility of a face part is represented by the hidden unit of our deep model, which is different from DBN in [7, 8], whose hidden units had no concrete semantic meaning. Secondly, it can well model the complex probabilistic correlations of different face parts because of cross-layer with better efficiency on both learning and reasoning. Thirdly, our deep model utilizes the detection result of part face as input, and need not any pose variation and occlusion information. The above discussions focus mainly on the problem of occlusion, but the

problem of multi-pose can be effectively handled. If one part can not be detected due to pose variation, it would be treated as the problem of occlusions to handle.

3. A framework for Face Detection

In this framework, the detection windows is indicated by y , whether it contains a face. The detection scores of face parts are indicated by $s = [s_1, \dots, s_p]^T$, where subscript p indicates the number of face parts. The visibilities of parts are indicated by $h = [h_1, \dots, h_p]^T \in \{0,1\}^p$, where $h_i=1$ indicates the i th part is visible and $h_i=0$ indicates the i th part is invisible. The framework is described in Figure 1.

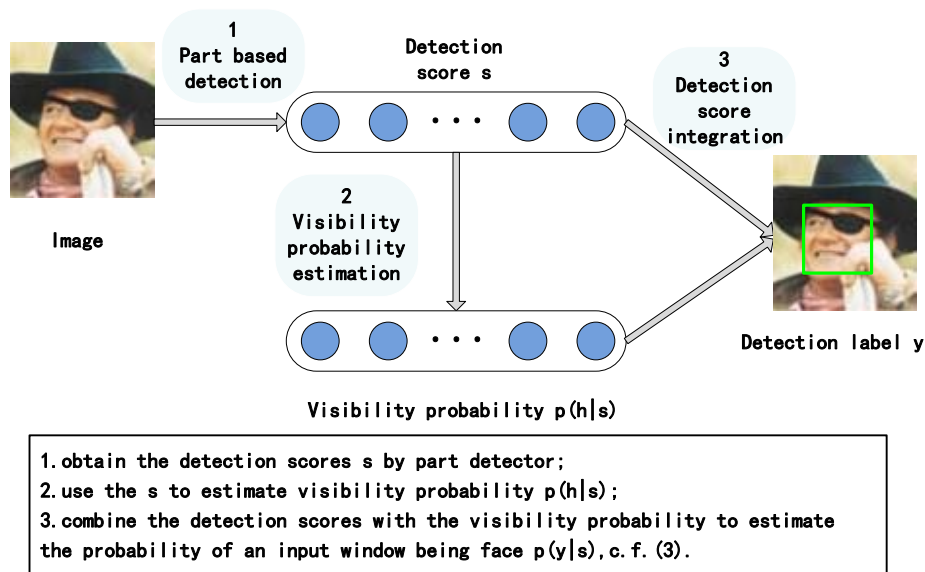


Figure 1. The Proposed Detection Framework

The h is a hidden random vector, because it is not provided in training phase and testing phase. $p(y|s)$ can be achieved by marginalizing out the hidden random variables h .

$$p(y|s) = \sum_h p(y, h|s) = \sum_h p(y|h, s)p(h|s) \quad (1)$$

$p(y|s)$ can be obtained by setting the $p(y|h, s) = e^{\sum_i y h_i s_i}$.

$$p(y|s) = \sum_h e^{\sum_i y s_i h_i} p(h|s) \quad (2)$$

Calculation complexity of the formula (2) is exponential distribution of the dimension of the hidden h . The formula (2) can be calculated through a fast approximation method. The method of calculation is as shown below.

$$p(y|s) \approx e^{\sum_i y s_i \tilde{h}_i} / Z \quad (3)$$

Where $Z = 1 + e^{\sum_i s_i \tilde{h}_i}$ is the partition functions of $\sum_y (e^{\sum_i y s_i \tilde{h}_i} / Z) = 1$ and \tilde{h}_i is obtained from $p(h_i|h, h_i, s)$ or solved by the method of mean-field approximation.

According to $p(h|s)$ of the formula (2), \tilde{h}_i is not the average of the all hidden h , but $\tilde{h} = E[h|s]$. The method of mean-field approximation is also utilized to calculate the posterior probability of DBN and RBM [18, 19, 20]. \tilde{h}_i is visibility term in the formula (3) and has different method to calculate.

The existing part-based face detection methods depended only on a single detection result to estimate the part visibility [17], but neglected the correlations of different parts in this frame. In this paper, a deep model is established to learn the correlations of the visibilities in different face parts.

4. The Deep Model

4.1. The Restricted Boltzmann Machine

In this section, the Restricted Boltzmann Machine (RBM) is investigated. It is the basic modules of our deep model. RBM is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs. It is a type of Markov Random Field and there are not the connections between units within the same layers, *i.e.*, visible-layer to visible-layer and hidden-layer to hidden-layer. However, the connections of the adjacent layers are random binary and symmetric. The binary visible variables of RBM is indicated by the vector $x = [x_1, x_2, \dots, x_N]^T$ and the binary hidden variables of RBM is indicated by $h = [h_1, h_2, \dots, h_N]^T$. A probability distribution is defined on h and x , as follows.

$$p(x, h) \propto e^{[x^T W h + c^T h + b^T x]} \quad (4)$$

Where h constitutes the hidden layers and x constitutes the visible layers. W is the symmetric connection matrix between the hidden layers and the visible layers, but there is not connection for units within the same layers. The model diagram of RBM is described in Figure 2(a). Since the special structure of RBM, the conditional probability distribution is easily calculated.

$$p(x_n = 1|h) = \sigma(w_{n,*} h + b_n)$$

$$p(h_i = 1|x) = \sigma(x^T w_{*,i} + c_i) \quad (5)$$

Where $\sigma(t) = (1 + \exp(-t))^{-1}$ is a logic function. $w_{n,*}$ is the n th row of the matrix W and $w_{*,i}$ is the i th column of the matrix W . The contrastive divergence algorithm (CD) [18] is utilized to learn the parameters W , b and c in the formula (4).

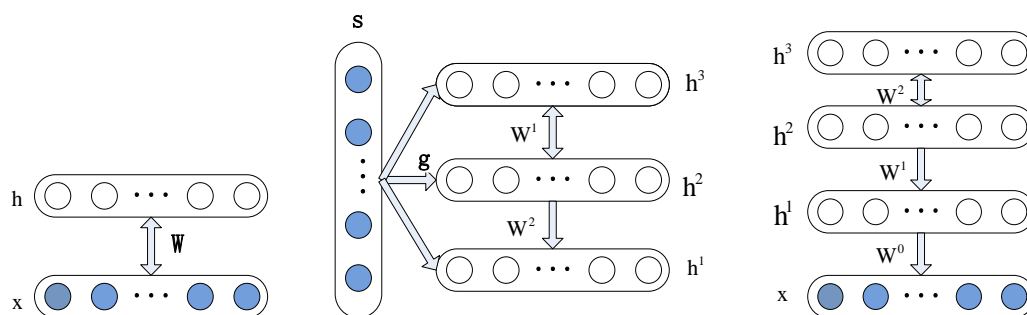


Figure 2. (a) RBM

(b) The Deep Model

(c) Deep Belief Networks

4.2 The Proposed Deep Model of Visibility Estimation

Our Part Detection Model. The part detection model is made up of three layers, which have different detection parts as shown Figure 3. Detection parts of the bottom layer are the smallest, which are the face organs, including eyes, noses and mouths. Detection parts of the higher layer consist of parts at the lower layer. For instance, the first part at the second layer consists of the first part and the third part at the first layer. The parts at highest layer (Layer 3) are the possible states of face. The organs unused border denote that it can be occluded, which can not be detected. Layer 1 and Layer 2 are the face parts, which consist of the face organs. Through combining several parts in the second layer, a possible face state is obtained on the third layer.

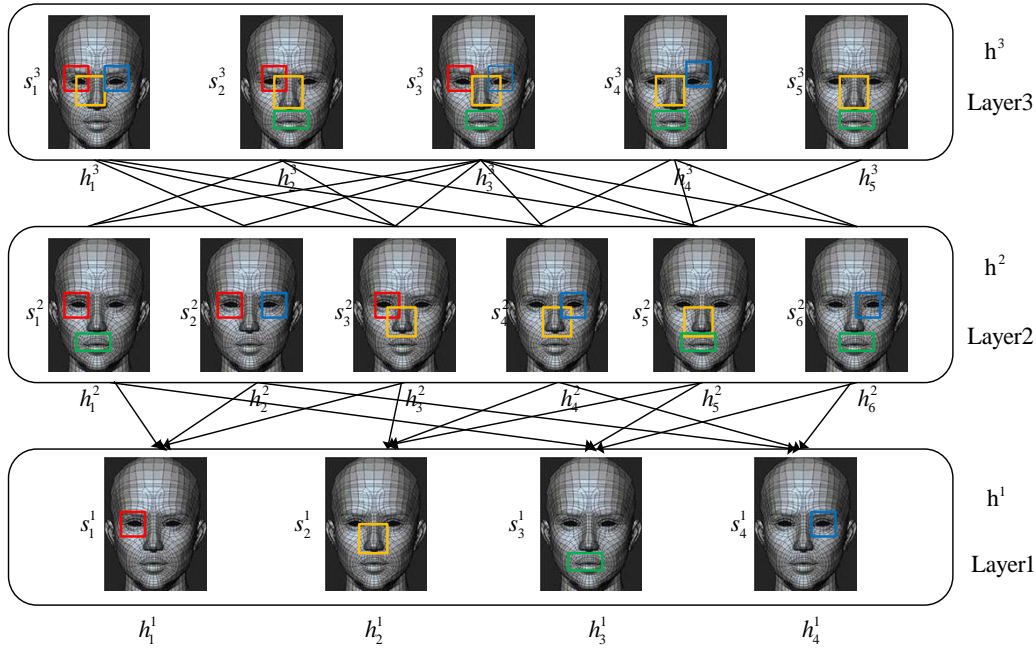


Figure 3. The Proposed Deep Model. h_i^l Indicates the Visibility of Each Part and s_i^l Indicates the Part Detection Score of i^{th} Part on the l^{th} Layer. For Instance, s_1^1 Indicates the Score of the Left-eye Part, whose Visibility is Indicated by h_1^1

Our deep model. The deep model is described in Figure 2(b) and its detail information is given in Figure 3. The visibility of p_i face parts on the l^{th} layer are indicated by $h^l = [h_1^l \dots h_{p_i}^l]^T$. In our model, there are connections for units between the l^{th} layer and $(l + 1)^{th}$ layer and no connections for units within the same layer. Each unit of any layer can be connected with the multiple units of the adjacent layers. Therefore, the visibility of one part on the l^{th} layer is related with the other parts within the same layer by sharing the parts of the $(l + 1)^{th}$ or $(l - 1)^{th}$ layer. Given the detection score s by the part-based detector, the probability distribution of h^1, \dots, h^L is the following.

$$p(h^1, \dots, h^L | s) = \left(\prod_{l=1}^{L-2} p(h^l | h^{l+1}, s) \right) p(h^{L-1}, h^L | s)$$

$$p(h_i^l | h^{l+1}, s) = \sigma(w_{i,*}^l h^{l+1} + g_i^l s_i^l + c_i^l)$$

$$p(h^{L-1}, h^L | s) = e^{[h^{L-1T} w^{L-1} h^L + c^{L-1T} h^{L-1} + c^{LT} h^L + g^{L-1T} s^{L-1} + g^{LT} s^L]} \quad (6)$$

For the model as shown in Figure 3, we get $L = 3$. The g_i^l , c^l and W^l parameters need to be learned. Where the g_i^l is used to balance the weights between the score s_i^l of part face detection and the correlation with other parts, the c^l is the offset term, W^l models the correlations between the h^l and the h^{l+1} , and $w_{i,*}^l$ is the i th row of W^l . Due to the detection s of part-based face is achieved from our model of part-based face detection at the training and testing phases, we assume that s is the input variables observed and $p(s)$ need not to be modeled. The model can be seen as conditional random fields. It is important to note that, to the s , h_j^l and h_i^l are not mutually independent, namely $p(h_i^l, h_j^l | s) \neq p(h_i^l | s) p(h_j^l | s)$, but, to the s and h^{l+1} , they are mutually independent, namely $p(h_i^l, h_j^l | h^{l+1}, s) = p(h_i^l | h^{l+1}, s) p(h_j^l | h^{l+1}, s)$. By this way, we model the correlation of different face parts within the same layer.

The proposed model is a multi-layers deep model, so training is time-consuming. Hinton, *et al.*, [19, 21] have proposed a fast training algorithm to train the deep belief networks (DBN), which has been successfully applied to the field of computer vision. In this paper, a similar training method is used to train our deep model. The DBN is a generation deep structure and there is not concrete meaning for its hidden units. However, our proposed deep model is a conditional model, whose hidden units indicate visibility of a face part. Due to above-mentioned reasons, the training method of DBN can not be directly applied to the proposed deep model. Therefore, the training and reasoning method of DBN in [19] need to be modified.

Our training method consists of two steps to learn the parameters W^l , g^l , c^l .

Step1, Through using RBM method, the parameters of the l th layer and $(l+1)$ th layer is trained layer by layer.

Step2, The all parameters is finely tuned using the method of Back-Propagation.

At step1, the adjacent layers are seen as a Restricted Boltzmann Machine (RBM) to train the parameters layer by layer. The probability distribution of RBM is as follows.

$$\begin{aligned} p(h_i^l | h^{l+1}, s) &= \sigma(w_{i,*}^l h^{l+1} + c_i^l + g_i^l s_i^l), \\ p(h_j^{l+1} | h^l, s) &= \sigma(h_j^{lT} w_{*,j}^l + c_j^{l+1} + g_j^{l+1} s_j^{l+1}) \end{aligned} \quad (7)$$

The parameters of the formula (7) are fast trained by the contrastive divergence algorithm (CD). At step2, we use back-propagation network (BP as shown in Figure 4.) to fine-tune all the parameters.

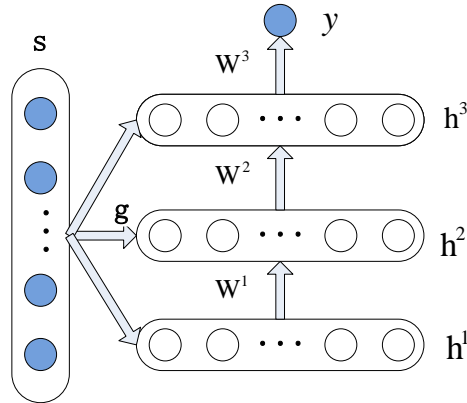


Figure 4. The Back-propagation Network

At reasoning stage, we utilize the part detection score s to infer the y . In this stage, $p(y|s)$ can be obtained using the frame of the formula (3). Meanwhile, we use BP network in Figure 4 to obtain the part visibility probability \tilde{h}_i^{l+1} , as follows.

$$\begin{aligned} \tilde{h}_i^{l+1} &= p(h_i^{l+1}|h^l, s) = p(h_i^{l+1}|h^l, s) \\ &= \sigma(h^T w_{i,j}^l + c_j^{l+1} + g_j^{l+1} s_j^{l+1}) \end{aligned} \quad (8)$$

We define that the visibility correlation parameters matrix W is non-negative, in order to reduce the offset of the training data and adjust the training process. So, the negative correlation among part face visibility, the prior knowledge, is invalid in our training process. In addition, if there is not connection between the unit h_i^l and the unit h_i^{l+1} , the $w_{i,j}^l$ of W^l is set to zero in the formula (6). According to our prior knowledge, the most important marginal information is kept in this paper. There are also other connection methods for the connection among units on the adjacent layers, namely fully-connected model [22], which will increase the complexity of the model, reducing the efficiency and need to be provided more training data.

5. Experiment

In order to verify the performance of the proposed algorithm, two groups experiment are conducted for testing. Group 1, the performance of our algorithm is analyzed on LFW database. Group 2, we conduct a comparison between our algorithm and the traditional classical LBP-Adaboost algorithm. In this experiment, we use the ROC curve to reflect the performance of our classifier. The ROC is a curve of the correct detection rate about false position rate and recall rate. Since there is mutual restricted relationship between correct detection rate and false position rate and recall rate, a better algorithm can guarantee the lower false position rate and higher recall rate in the case of the higher correct detection rate. Therefore, the ROC curve has higher evaluation value for our algorithm.

In this experiment, the Haar is used as the detection feature and the part-based model in [12] is adopted to model the 15 face parts in Figure 3. Our deep model uses the part detection scores as the input.

5.1. The Performance Analysis Experiment

In this experiment, the performance of the proposed algorithm is evaluated. The experimental data is from LFW face database, which is accomplished by the University of Massachusetts in America and can be applied to research the problem of face detection

under non-limitation condition. The LFW is a rich-content face database. It consists of the face images of various angles, postures and occlusions under the different situations.

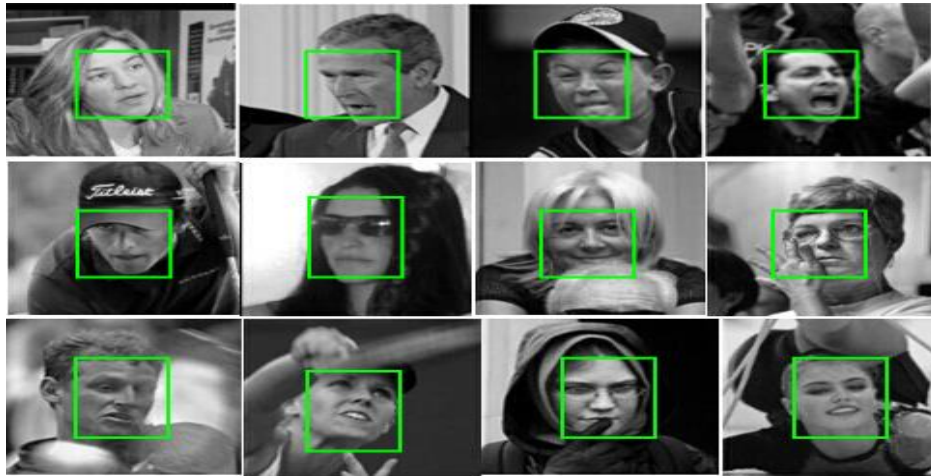
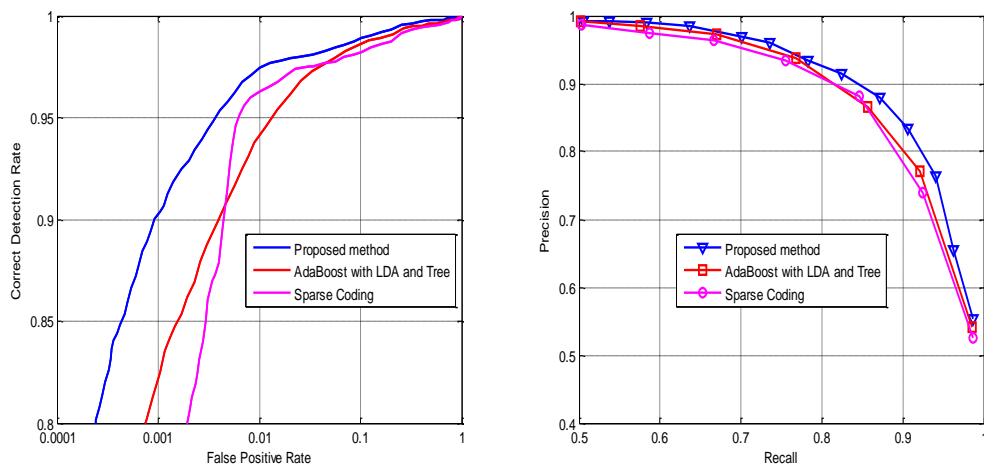


Figure 5. The Part Detection Results on LFW Database

Since we use the Haar as the detection feature, the experimental images need to be performed the pretreatment of gray transformation and histogram equalization. In the experiment, we select 1500 images as the training data and 1000 images as the testing data. The training and testing data consist of the face images of the front, multi-pose and occlusion.

In this experiment, our face detection algorithm detects 987 from 1000 face targets, 13 undetected and correct detection rate is 98.7%. In the detection process, there are 25 false detections. Therefore, false position rate is 2.5%. The partial testing results are as shown in Figure 5.

The proposed algorithm is trained by a fast training method. It is a greedy algorithm that can learn our deep model layer by layer. It is superior to BP algorithm in performance and can effectively avoid the excessive fitting situation because of the too strong expression ability of the network functions [19]. Since the proposed model consists of 3 layers and only 15 hidden units, the consuming time of the training and reasoning is much less than the parts model. So, the proposed model can meet the practical situations.



(a) The Correct Detection and False Position Rate (b) The Correct Detection and Recall Rate

Figure 6. The ROC Curves for Comparison

Table 1. The Correct Detection Rate with False Position Rate of 0.001

Index \ Method	proposed method	AdaBoost With LDA and Tree	Sparse coding
Precision	90.8%	82.2%	70.6%

Table 2. The Correct Detection Rate with Recall Rate of 0.8

Index \ Method	proposed method	AdaBoost With LDA and Tree	Sparse coding
Precision	92.7%	91.3%	90.9%

In addition, we compare our algorithm with the current advanced face detection algorithms of part-based, which consists of AdaBoost with LDA and Tree [12] and sparse coding face detection [15]. In the experiment, the above methods all use the same experimental data containing multi-pose and occlusion, which is from LFW face database.

The experimental result is as shown in Figure 6. From the experimental result, we can see that the performance of our proposed algorithm is superior to others. When false position rate is 0.1%, the correct detection rate of our algorithm is 90.8%, while the correct detection rate of AdaBoost with LDA and Tree is 82.2% and the correct detection rate of sparse coding face detection is 70.6%, as shown in figure 6(a). When the recall rate is 80%, the correct detection rate of our algorithm is 92.7%, while the correct detection rate of AdaBoost with LDA and Tree is 91.3% and the correct detection rate of sparse coding face detection is 90.9%, as shown in Figure 6(b). The experimental results show that the performance of our algorithm is obviously better than the performance of the current advanced part-based face detection algorithms and has more robustness in the case of occlusion and multi-pose.

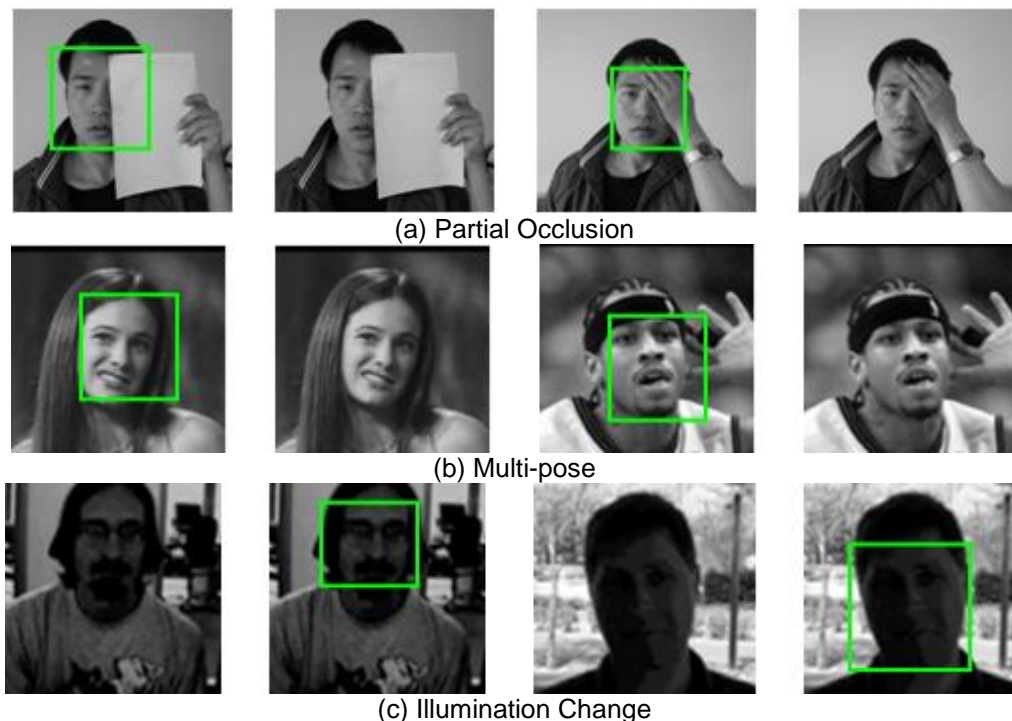


Figure 7. Partial Detection Results

5.2. The Comparative Experiment

In order to further test the performance of our algorithm, we conduct experiment on various natural images. In this experiment, we compare our algorithm with the traditional classic LBP-Adaboost algorithm. The experimental data consist of partial occlusion, multi-pose and illumination change. The experimental results are as shown in Figure 7, where Column 1 and Column 3 is the detection results of our algorithm, Column 2 and Column 4 is the detection results of LBP-Adaboost algorithm. Compared with the LBP-Adaboost algorithm, our algorithm has a better detection performance in the case of partial occlusion and multi-pose, as shown in Figure 7 (a) and (b). The main reason is that although the face targets have great change on the whole, some face local regions have no change. Since our algorithm uses deep model to learn the correlations of the different face local regions, its interference is relatively smaller for face detection. However, LBP-Adaboost uses global features to representation face. When faces have great change, it is difficult to be detected. From Figure 7(c), we can see that our algorithm is inferior to the LBP-Adaboost in the case of poor illumination. The main reason is that since the LBP-Adaboost has illumination invariant at a certain extent, it can well remove the illumination interference.

In order to highlight the superiority of the proposed algorithm, we conduct experiment on the images containing multiple human faces. The experimental images are from the real life, which consist of front, multi-pose, partial occlusion and can reflect the practical situations. A part of the experimental result is as shown in Figure 8, where Figure 8(a) is the detection result of the LBP-Adaboost. There are 10 faces in the image, among them 6 detected correctly, 4 undetected and 1 detected false position. Figure 8(b) is the detection result of our algorithm. There are 10 faces in the image, among them 10 detected correctly, 0 undetected and 0 detected false position. Therefore, from the experimental results, we can see that the proposed algorithm obtains a better performance in the case of occlusion and multi-pose.

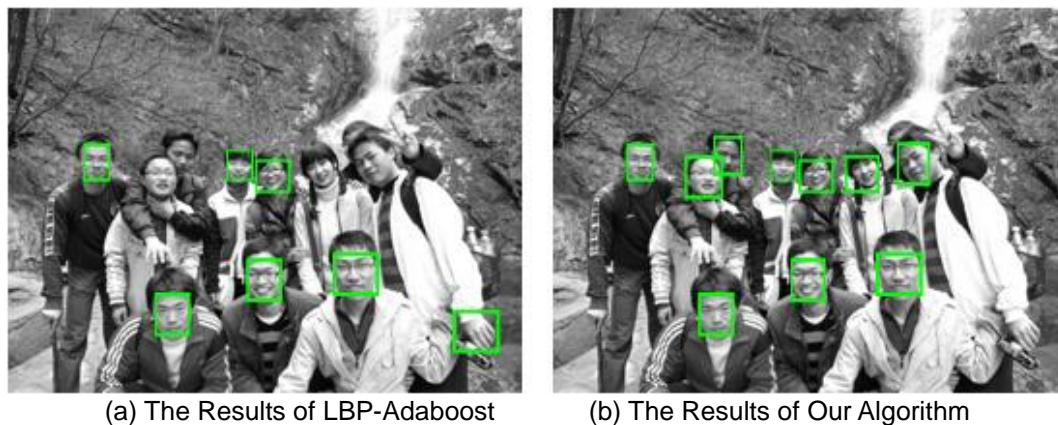


Figure 8. The Comparison of Detection Results on Multiple Human Faces

6. Conclusions

To solve the problem of partial occlusion and multi-pose, a face detection algorithm based on deep learning is proposed. In the algorithm, we combine deep learning theory and part-based model for face detection and construct deep learning model to learn the correlations of the visibilities in different face parts. The experimental results show that our algorithm has more robustness and effectiveness in the case of occlusion and multi-pose. However, there are also some limitations. For example, the detection rate is relatively low under the poor light. The follow-up work of this paper including, firstly, the characteristics of illumination-invariant will be introduced into our algorithm to enhance

the robustness under the light. Secondly, the algorithm will be applied to other objects detection.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No.61263019), the Program for International S&T Cooperation Projects of Gansu province (No.144WCGA166), the Natural Science Foundation of Gansu Province, China (No.1310RJZA034) and the Doctoral Foundation of LUT.

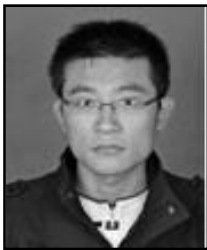
References

- [1] Y. Ban, S. K. Kim and S. Kim, "Face detection based on skin color likelihood", *Pattern Recognition*, vol. 47, no. 4, (2014).
- [2] D. N. Chandrappa, M. Ravishankar and D. R. RameshBabu, "Face detection in color images using skin color model algorithm based on skin color information", *Electronics Computer Technology (ICECT), 2011 3rd International Conference on. IEEE*, (2011) April, pp. 254-258.
- [3] K. M. Lee, "Component-based face detection and verification", *Pattern Recognition Letters*, vol. 29, no. 3, (2008).
- [4] R. Aliradi, N. Bouzera and A. Meziane, "Detection of Facial Components Based on SVM Classification and Invariant Feature", *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE/WIC/ACM International Joint Conferences on IEEE*, (2013) November, pp. 30-36.
- [5] G. E. Hinton, "Learning multiple layers of representation", *Trends in cognitive sciences*, vol. 11, no. 10, (2007).
- [6] Y. Kang, S. Kim and S. Choi, "Deep Learning to Hash with Multiple Representations", *ICDM*, (2012) December, pp. 930-935.
- [7] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification", *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE*, (2012) June, pp. 3642-3649.
- [8] S. Zhong and Y. Liu, "Bilinear deep learning for image classification", *Proceedings of the 19th ACM international conference on Multimedia, ACM*, (2011) November, pp. 343-352.
- [9] G. E. Hinton, "Deep belief networks", *Scholarpedia*, vol. 4, no. 5, (2009).
- [10] K. Ichikawa, T. Mita and O. Hori, "Component-based robust face detection using AdaBoost and decision tree", *Automatic Face and Gesture Recognition, FGR 7th International Conference on. IEEE*, (2006) April, pp. 413-420.
- [11] B. Heisele, T. Serre and M. Pontil, "Component-based face detection", *Computer Vision and Pattern Recognition, CVPR Proceedings of the 2001 IEEE Computer Society Conference on IEEE*, I-657-I-662, (2001).
- [12] K. Ichikawa, T. Mita and O. Hori, "Component-based face detection method for various types of occluded faces", *Communications, Control and Signal Processing, ISCCSP 3rd International Symposium on. IEEE*, (2008) March, pp. 538-543.
- [13] Y. Huang, Q. Liu and D. N. Metaxas, "A component-based framework for generalized face alignment", *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 1, (2011).
- [14] J. Yang, Z. Lei and S. Liao, "Face liveness detection with component dependent descriptor", *Biometrics (ICB), 2013 International Conference on IEEE*, (2013) June 1-6.
- [15] Z. Shu, C. Yong and X. Mei, "Face Detection Based on Local Region Sparse Coding", *Journal of Software*, vol. 24, no. 11, (2013).
- [16] Z. Kovács-Bálint, G. Stefanics and A. Trunk, "Automatic detection of trustworthiness of the face: A visual mismatch negativity study", *Acta biologica Hungarica*, vol. 65, no. 1, (2014).
- [17] C. Zhang, J. Zhang and H. Zhao, "A Part-Based Probabilistic Model for Object Detection with Occlusion", *PloS one*, vol. 9, no. 1, (2014).
- [18] G. E. Hinton, "Training products of experts by minimizing contrastive divergence. *Neural computation*", vol. 14, no. 8, (2002).
- [19] G. E. Hinton, S. Osindero and Y. W. Teh, "A fast learning algorithm for deep belief nets", *Neural computation*, vol. 18, no. 7, (2006).
- [20] Y. Bengio, "Learning deep architectures for AI", *Foundations and trends® in Machine Learning*, vol. 2, no. 1, (2009).
- [21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, vol. 313, no. 5786, (2006).
- [22] M. Bergtholdt, J. Kappes and S. Schmidt, "A study of parts-based object class detection using complete graphs", *International journal of computer vision*, vol. 87, (2010), pp. 1-2.

Authors



Ming Li, was born in Lanzhou, Gansu Province, China in 1959. He received the bachelor degree in mathematics at Xi'an University of Technology in 1982. Now he is a professor in the School of Computer and Communication at Lanzhou University of Technology, Lanzhou, China. His current research interests are intelligent information processing, pattern recognition, signal processing, face analysis and object tracking. In recent years, he has authored about 30 papers in international journals, national issues and international conference proceedings.



Chengyang Yu, was born in Yantai, Shandong Province, China in 1988. He received the bachelor degree in school of information science & technology at Qingdao University of Science & Technology in 2012; he has taken up the subject of Master's Vision Group. His research interests include intelligent information processing and pattern recognition.



Fuzhong Nian, was born in Wuwei, Gansu Province, China in 1974. He received the Ph.D. in electronic information engineering at Dalian University of Technology in 2011. Now he is a Associate professor in the School of Computer and Communication at Lanzhou University of Technology, Lanzhou, China. His current research interests are complex system and complex network, intelligent information processing. In recent years, he has authored about 30 papers in international journals, National issues and international conference proceedings.



Xiaoxu Li, was born in Baicheng, Jilin Province, China in 1983. She received the Ph.D. in Computer Science and Technology at Beijing University of Posts and Telecommunications in 2012. Now she is a Lecturer in the School of Computer and Communication at Lanzhou University of Technology, Lanzhou, China. Her current research interests are machine learning and computer vision. In recent years, she publishes several papers in some top international conferences including ICDM.