

RESEARCH

Open Access



# Deep representation for partially occluded face verification

Lei Yang<sup>1,2</sup>, Jie Ma<sup>3</sup>, Jian Lian<sup>4,6\*</sup>, Yan Zhang<sup>5\*</sup> and Houquan Liu<sup>1\*</sup>

## Abstract

By using deep learning-based strategy, the performance of face recognition tasks has been significantly enhanced. However, the verification and discrimination of the faces with occlusions still remain a challenge to most of the state-of-the-art approaches. Bearing this in mind, we propose a novel convolutional neural network which was designed specifically for the verification between the occluded and non-occluded faces for the same identity. It could learn both the shared and unique features based on a multiple network convolutional neural network architecture. The newly presented joint loss function and the corresponding alternating minimization approach were integrated to implement the training and testing of the presented convolutional neural network. Experimental results on the publicly available datasets (LFW 99.73%, YTF 97.30%, CACD 99.12%) show that the proposed deep representation approach outperforms the state-of-the-art face verification techniques.

**Keywords:** Face verification, Machine vision, Convolutional neural network, Loss function)

## 1 Introduction

Face recognition has become the primary biometric technique used for personal authentication and identification in various fields, including finance [1, 2], public security [3, 4], and education [5, 6]. A plethora of computer vision-based approaches have been proposed from the early 1990s and could derive the low-dimensional representation under specific priors on the features in the facial images. However, these techniques would result in limited performance while the presented assumptions might not be compatible with the practical scenarios.

Recently, deep learning especially the convolutional neural network (CNN) has been widely accepted as the state-of-the-art approach for face verification and face identification [7, 8]. CNNs have shown excellent performance in various face recognition tasks, e.g., Rajeev et al. [9] and Schroff et al. [10] presented that their proposed method achieved the accuracy of 99.78% and 99.63% on the facial dataset of Labeled Faces in the Wild (LFW) [11], respectively. However, it remains difficult for them

to obtain satisfactory accuracy on faces varying in pose, illumination, and occlusion, among which facial occlusion has always been considered as an extremely challenging mission.

On the one hand, data imbalance in the prevailing facial datasets should be one possible explanation for this phenomenon. Despite most of the face recognition training datasets contain large amount of identities, they still suffer from the deficiency of difficult facial images with partial occlusions such as sunglasses, hats, and hairs. An intuitive solution to this problem is that more occluded facial images should be included into the training process of the CNN framework.

On the other hand, the loss function could also significantly affect the training of CNN-based face verification systems and lead to poor performance while it might be biased to the data distribution. For instance, softmax loss, which was not specifically designed for complicated samples, would neglect the faces with occlusion by enlarging the conditional probability of the entire samples. To address this issue, numerous loss functions and different constraints on the traditional loss functions have been presented [9, 12–14].

Bearing the abovementioned analysis in mind, we propose a novel CNN architecture trained by manually collected 6,178 facial image pairs from 560 different identities

\*Correspondence: [lianjianlianjian@163.com](mailto:lianjianlianjian@163.com); [120650354@qq.com](mailto:120650354@qq.com); [hqliu@cumt.edu.cn](mailto:hqliu@cumt.edu.cn)

<sup>4</sup>Department of Electrical Engineering Information Technology at Shandong University of Science and Technology, Jinan 250031, China

<sup>6</sup>Shandong Normal University, Jinan 250014, China

Full list of author information is available at the end of the article

with occlusion. Our proposed CNN adopts the typical multi-path framework, and its convolutional layers are implemented by introducing the maxout operator previously introduced by [15]. After the initialization, the layers including the convolutional layers in the CNN architecture are then fine-tuned with the collected facial images. The output layer is divided into two separate feature vectors that contain the shared identity information between one face and its occluded counterpart and the unique information from each input image, respectively. Meanwhile, one newly proposed mutual information constraint and the introduced maxout operator are integrated to reduce the dimensionality of the parameters and eliminate the possibility of overfitting that might appear in small dataset. By using the alternate minimization approach, the objective function for the proposed CNN model could be iteratively optimized for the heterogeneous images both in the training and testing procedures.

To evaluate the performance of the proposed method, we conducted comparison experiments on several publicly available facial benchmarks between state-of-the-art approaches and ours. Experimental results demonstrate that our mutual information constrained CNN framework learns occlusion-invariant representation and outperforms the state-of-the-art face verification techniques.

Generally, our work offers three contributions as follows.

- A novel deep CNN architecture or so-called deep representation is proposed to extract the shared information between the input pair of facial images that were manually collected and could be optimized through alternating minimization.
- We propose a novel loss function. It can both maximize the intra-identity distances and minimize the inter-identity similarity of the features extracted. Through combining the mutual information loss and the softmax loss, the proposed method can produce the highly discriminative features that would contribute to enhance the accuracy for face verification.
- Experiments on the public available datasets by our approach outperforms the state-of-the-art techniques with an impressive superiority.

The rest of this paper is organized as follows. Firstly, we briefly reviewed the related work on face verification methods in Section 2. Then, in Section 3, we describe both the materials that we used and the details of the proposed approach. Section 4 presents the experimental results and the discussion. Finally, we draw the conclusion and presents our future work in Section 5.

## 2 Related work

In general, the previously proposed deep face recognition approaches differ from each other in at least one of the following aspects: the network architecture and the loss function.

### 2.1 Types of network architecture

Most of the previously proposed CNNs have been exploited to implement the face recognition applications. According to the input channels used in the network architectures, these CNN networks could be roughly categorized into two types as follows.

#### 2.1.1 Single network

Alexnet [16] was presented in 2013 and has shown its great performance in different machine vision systems. It contains five convolutional layers combined with rectified linear unit (ReLU), dropout operator, and three fully connected layers.

As a very deep CNN, VGGNet [17] was proposed for large-scale image recognition in 2014. It has 16–19 layers, which significantly contributes to the enhancement of the image classification accuracy.

In 2015, GoogleNet [18] with 22 network layers could integrate the information from multi-resolution images by concatenating all of the features maps.

In 2016, ResNet [19] was presented. Instead of learning the underlying mapping directly, it learns a residual mapping from the input layer.

#### 2.1.2 Multiple networks

According to the input images, the multiple networks could be roughly divided into multi-view, multi-patch, and multi-task.

In 2016, [20] and [21] proposed to address the variations from view and pose with multiple network CNNs.

== Table 1 ==

Both [22] and [23] proposed the multiple networks to handle with the different input face patches from the same image in each iteration.

Meanwhile, several multi-task networks were proposed to implement various tasks in one architecture, e.g., [24].

Generally, we summarize the state-of-the-art CNN-based approaches with the corresponding database and their performance as shown in Table 1.

**Table 1** Face verification performances of state-of-the-art CNN-based techniques

Method	Dataset	Accuracy (%)
DDML [49]	YTF	82.30
DeepFace-single [50]	YTF	91.40
DeepID2 [51]	YTF	93.20
FaceNet [52]	YTF	95.12
VGG-Face [53]	YTF	97.30

## 2.2 Loss function

Previously, a great deal of loss functions was proposed for face recognition. Despite it has inner limitations, the softmax loss is the broadly adopted loss function [25]. And many modifications have been added to enhance the performance of original softmax loss. For instance, [9] proposed a L2-constrained softmax loss to restrict the extracted features to lie on a hypersphere with a fixed radius. In [26], a margin loss was proposed to combine with the softmax loss, which could encourage the inter-class separability and the intra-class compactness together.

Triplet loss is one of the most typical Euclidean distance-based loss functions that embeds the faces into Euclidean space. In [27] and [28], with the input triple images including the anchor image, the positive image, and the negative image, the proposed methods could maximize the distance between the anchor image and the positive image and minimize the distance between the anchor image and the negative image.

Recently, the angular loss and its different modifications were proposed. Instead of the employment of Euclidean space, the angular loss functions could realize the separation of the output features with angular distance.

## 3 Methodology

Benefiting from the success of CNN in recent years, face verification has obtained significant progress. In this section, we propose the CNN-based approach and the novel loss function. Since the structure of CNN has been presented in a great deal of studies [29], we focused on the network architecture of the proposed CNN.

### 3.1 Network architecture

To address the difficulty of occluded facial verification, we propose a novel CNN-based image classification approach. The proposed CNN architecture correlates with the CNN presented in [29], e.g., they both are multiple networks. However, the size, the number of their layers, and the loss functions are different from each other. Furthermore, the proposed architecture was mainly designed to extract the shared information between the occluded part and the non-occluded part of the face images while the CNN in [29] was used to highlight the subtle difference between the different bananas' ripening stages.

The proposed CNN architecture and the corresponding parameters are firstly trained on the publicly available facial dataset LFW [11] and VGG Face [30]. Then, the initialized model is fine-tuned with the manually collected 6178 images including the non-occluded faces and the corresponding occluded images. As shown in Fig. 1, for each input facial image, 8 convolutional layers with

corresponding maxout operators and 3 fully connected layer were utilized.

Two parameter-sharing channels (as shown in Fig. 1) are exploited to process the non-occluded and occluded faces. And the output feature layer is exploited to obtain both the shared features from the input face pair and the unique feature from each input single image. The details for each CNN channel are listed as follows.

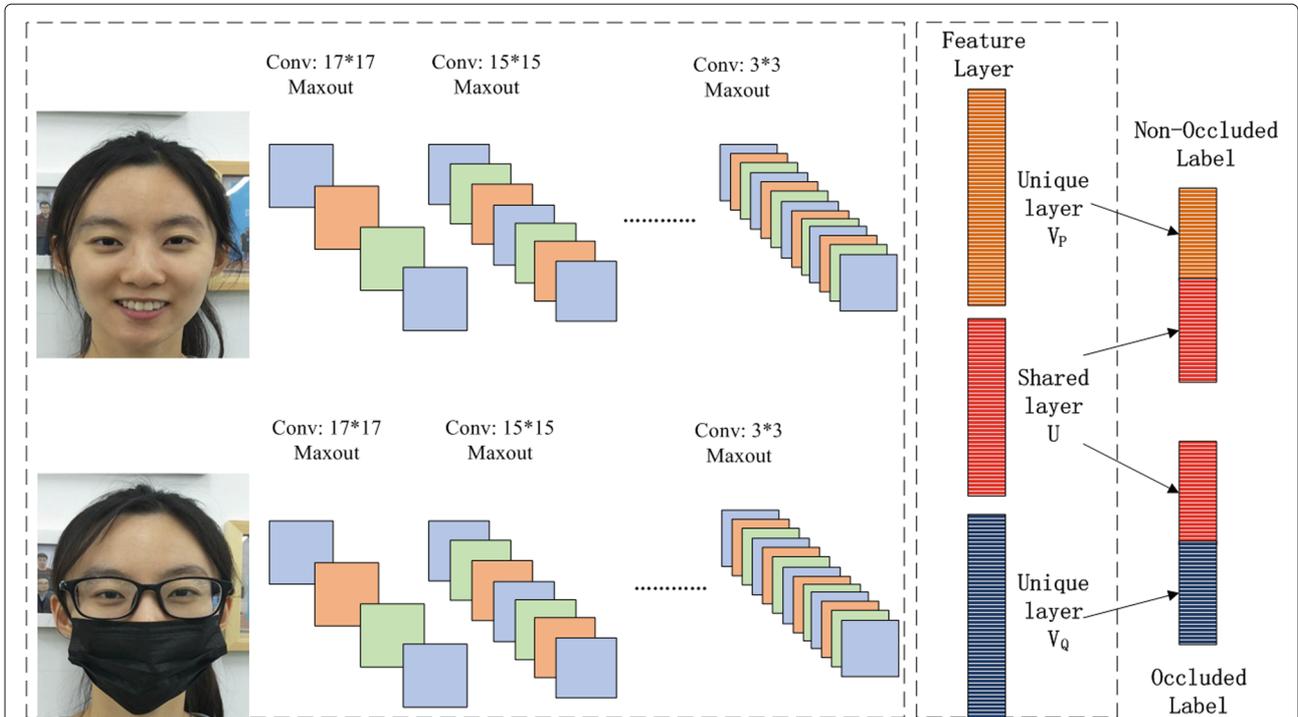
- Convolutional layer 1. There are 48 kernels (size  $17 \times 17$ , stride of 2) in the first convolutional layer, which is combined with one maxout operator and one max pooling layer.
- Convolutional layer 2. There are 96 kernels (size  $15 \times 15$ , stride of 2) in the second layer, which is combined with one maxout operator and one max pooling layer.
- Convolutional layer 3. There are 128 kernels (size  $13 \times 13$ , stride of 2) in the third layer, which is combined with one maxout operator and one max pooling layer.
- Convolutional layer 4. There are 128 kernels (size  $11 \times 11$ , stride of 2) in the fourth layer, which is combined with one maxout operator.
- Convolutional layer 5. There are 128 kernels (size  $9 \times 9$ , stride of 2) in the fifth layer, which is combined with one maxout operator.
- Convolutional layer 6. There are 128 kernels (size  $7 \times 7$ , stride of 2) in the sixth layer, which is combined with one maxout operator.
- Convolutional layer 7. There are 384 kernels (size  $5 \times 5$ , stride of 2) in the seventh layer, which is combined with one maxout operator.
- Convolutional layer 8. There are 384 kernels (size  $3 \times 3$ , stride of 2) in the eighth layer, which is combined with one maxout operator.
- Fully connected layer 1. 512 neurons combined with ReLU.
- Fully connected layer 2. 512 neurons combined with ReLU.
- Fully connected layer 3. 512 neurons combined with ReLU.

### 3.2 Mutual information regularized softmax loss function

Firstly, the widely used softmax loss function is formulated as follows.

$$L_s = \sum_{i=1}^m \log \frac{e^{W_{y_i}^T X_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T X_i + b_j}} \quad (1)$$

where  $X_i \in \mathcal{R}^d$  denotes the feature of the  $i$ th input image that belongs to the  $y_i$ th class.  $W_j \in \mathcal{R}^d$  is the  $j$ th column of the weights  $W \in \mathcal{R}^{d \times n}$  in the last fully connected layer,



**Fig. 1** Our proposed CNN architecture. The maxout operator is used to extract invariant features and avoid the possibility of overfitting. Both the features from the occluded face and its corresponding non-occluded counterpart can be extracted from the shared layer and compared in cosine distance

and  $b \in \mathbb{R}^n$  is the bias. And  $m$  and  $n$  denote the batch size and the number of identities, respectively.

Let  $I_P$  and  $I_Q$  denote the occluded face and the facial image without occlusion, respectively. The general feature extraction process is defined in following formulation.

$$X_i = \text{Conv}(I_i, \theta_i) (i \in \{P, Q\}) \tag{2}$$

where Conv denotes the feature extraction function with the proposed CNN,  $X_i$  is the corresponding output feature, and  $\theta$  denotes the feature maps in the CNN architecture that need to learn in the training phase. One fundamental prior used in the proposed method is that there should be shared component between the non-occluded image and its occluded counterpart. Accordingly, we introduce three different matrices ( $U$ ,  $V_P$ , and  $V_Q$ ) to represent the shared information of the features and the unique feature, which can be formulated as follows.

$$F_i = \begin{bmatrix} F_{\text{share}} \\ F_{\text{unique}} \end{bmatrix} = \begin{bmatrix} UX_i \\ V_i X_i \end{bmatrix} \quad (i \in \{P, Q\}) \tag{3}$$

where  $UX_i$  denotes the shared feature, and  $VX_i$  denotes the unique feature. Due to the characteristics of mutual information in the shared feature and unique feature, we impose the mutual information as a regularization term

on the commonly used softmax loss function, which can be formulated as follows.

$$\mathcal{L}(F, c, \theta, U, V) = \sum_{i \in \{P, Q\}} \text{softmax}(F_i, c, \theta, U, V_i) \tag{4}$$

*s.t.*  $MI(U, V_i) = 0 \quad (i \in \{P, Q\})$

where  $c$  denotes the class of the identity, and  $MI(\cdot)$  [31] denotes the function to compute the mutual information of the input pair of matrix.

### 3.3 Optimization

Then, the final objective function of the proposed CNN model could be expressed as follows according to the lagrange multiplier.

$$\mathcal{L}(F, c, \theta, U, V) = \sum_{i \in \{U, V\}} \text{softmax}(F_i, c, \theta, U, V_i) + \lambda_i \sum_{i \in \{U, V\}} MI(U, V_i) \tag{5}$$

where  $\lambda_i$  denotes the Lagrange multiplier for  $x_i$ . By using the alternating minimization algorithm and the back-propagation mechanism, the  $\theta$ ,  $U$ , and  $V_i$  can be iteratively optimized. The gradients of  $U$  and  $V_i$  can be expressed as:

$$\frac{\partial \mathcal{L}}{\partial U} = \sum_{i \in \{P, Q\}} \frac{\partial \text{softmax}(F_i, c, \theta_i, U, V_i)}{\partial U} + \sum_{i \in \{P, Q\}} \frac{\partial \text{MI}(U, V_i)}{\partial U} \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial V_i} = \sum_{i \in \{P, Q\}} \frac{\partial \text{softmax}(F_i, c, \theta_i, U, V_i)}{\partial V_i} + \sum_{i \in \{P, Q\}} \frac{\partial \text{MI}(U, V_i)}{\partial V_i} \tag{7}$$

Thus, the  $\theta$ ,  $U$ , and  $V_i$  should be updated with alternating minimization with a learning rate  $\gamma$  and expressed as follows.

$$\theta^{(t+1)} = \theta^{(t)} - \gamma \frac{\partial \mathcal{L}}{\partial \theta^{(t)}} \tag{8}$$

$$U^{(t+1)} = U^{(t)} - \gamma \frac{\partial \mathcal{L}}{\partial U^{(t)}} \tag{9}$$

$$V_i^{(t+1)} = V_i^{(t)} - \gamma \frac{\partial \mathcal{L}}{\partial V_i^{(t)}} \tag{10}$$

And the initial value of the  $\theta$  could be obtained from the trained CNN model;  $U$  and  $V_i$  are initialized with random feature maps.

#### 4 Results and discussion

Extensive experiments were conducted to assess the performance of our proposed face verification method on several publicly available face recognition benchmarks, including LFW [11], YouTube Faces (YTF) [32], and Cross-Age Celebrity Dataset (CACD) [33]. Both the experimental results and the analysis are demonstrated in this section.

##### 4.1 Dataset and pre-processing

The dataset employed to train the CNN model would significantly influence the performance of the corresponding tasks [34]. Therefore, a variety of face recognition datasets have been presented. LFW [11] was released in 2007 and contains 13,233 facial images from 5749 different identities. As the most popular benchmark used to evaluate the performance of the deep learning techniques under unconstrained conditions, its accuracy has achieved to nearly 100% [9]. However, the faces in LFW are mainly frontal without severe pose or illumination, while there are not enough difficult instances. VGG-Face [30] and VGG-Face2 [35] includes 2.6M and 3.32M from 2622 and 9131 identities, respectively. Unlike LFW, these two datasets are not publicly available, and they both contain the faces with pose-related variations. MS-Celeb-1M [36] is the largest publicly available face recognition dataset. There are 10M facial images from 100,000 famous celebrities with some annotation noise. MegaFace

[37] contains 4.7M faces from 672,057 unique individuals. Meanwhile, it also provides two subsets of the images that could be used to verify the pose and age variations. Although IJB-A [38] includes only 25,809 faces of 500 different subjects, it has been widely considered as a difficult face recognition database because it was designed for joint face detection and recognition tasks since it contains both images and videos of faces with pose variations.

Besides the LFW [11] dataset, we collected 6178 facial images including both non-occluded and occluded ones (samples shown in Fig. 2) to train the proposed CNN architecture. There are about 6 images captured for each identity. To increase the diversity of the input images and decrease the possibility of potential overfitting in the small-scale facial images, we enlarge the original dataset with data augmentation methods including translations (varying from 10 pixel to 100 pixels with a gap of 10 pixels) and vertical and horizontal reflections. After the procedure of data augmentation, the images are then resized into  $256 \times 256$ .

##### 4.2 Training and evaluating

We manually labeled the samples into different categories corresponding to the identities. Fifty percent of the images are taken as training dataset, 30% percent are chosen into the evaluation dataset, and the other 20% percent are used in the testing process. In the training process, the proposed framework is refined with the back-propagation mechanism which originally calculates the minimization of the squared difference between the classification ground truth and the corresponding output prediction. The training is performed on GPU of high performance and implemented in TensorFlow [39]; the learning rate started from 0.01; it takes  $10^5$  iterations in total. For each iteration, it takes only about 0.5 s. Finally, the similarity score is calculated with the cosine distance of two output features, and the threshold comparison is used for the face verification in the testing process.

##### 4.3 Experiments on the LFW

To evaluate the performance of the proposed face verification technique, we conducted the comparison experiments between the state-of-the-art methods (including DeepID3 [7],  $L_2$  Softmax [9] (3.7M), FaceNet [10] (200M), VGG-Face [30] (2.6M), Baidu [40] (1.3M), Deep Face [41] (4M), Range Loss [42] (1.5M), Deep Visage [43] (4.48M)) and ours on three publicly available datasets. Following the protocol “unrestricted with labeled outside data,” we firstly performed the experiments on 6000 pairs of facial images in LFW, and the experimental results are shown in Table 2.



**Fig. 2** Image samples in the manually collected database

As shown in Table 2, the proposed approach outperforms the state-of-the-art face verification methods, while the number of images used in the training set to train our model is less than most of the other techniques like FaceNet, which exploits 200 millions of images in its training process.

#### 4.4 Experiments on the YTF

To assess the performance of the proposed face verification technique, we conducted the comparison experiments between the state-of-the-art methods (including

DeepID3 [7],  $L_2$  Softmax [9] (3.7M), FaceNet [10] (200M), VGG-Face [30] (2.6M), Baidu [40] (1.3M), Deep Face [41] (4M), Range Loss [42] (1.5M), Deep Visage [43] (4.48M) and ours on three publicly available datasets. Following the protocol “unrestricted with labeled outside data,” we firstly performed the experiments on 5000 pairs of facial frames in YTF, and the experimental results are shown in Table 3.

As shown in Table 3, the proposed approach outperforms most of the state-of-the-art face verification methods except the VGG-Face [30].

**Table 2** Face verification performances on LFW

Methods	Images	Single loss	Accuracy (%)	Time (sec)
DeepID3	–	No	99.47	1.3
$L_2$ Softmax	3.7M	Yes	99.60	1.5
FaceNet	200M	Yes	99.63	2.6
VGG-Face	2.6M	No	98.95	1.9
Baidu	1.3M	No	99.13	1.8
Deep face	4M	No	97.35	1.2
Range loss	1.5M	No	99.52	3.8
Deep visage	4.48M	No	99.62	1.8
Our method	2.62M	No	99.73	1.2

The entries in the “Images” column represent the number used to train the face verification methods. The “Time” column represents the execution time for single input image

#### 4.5 Experiments on the CACD

To assess the performance of the proposed face verification technique, we conducted the comparison experiments between the state-of-the-art methods (high-dimensional LBP [44], hidden factor analysis [45], LF-CNN [46], center loss [47], and marginal loss [12]) and ours on three publicly available datasets. Following the protocol “unrestricted with labeled outside data,” we firstly performed the experiments on 4000 pairs of facial frames with different types of occlusion in CACD, and the experimental results are shown in Table 4. Notably, there are only several methods have reported their performance on CACD.

As shown in Table 4, the proposed approach achieve superior performance over the state-of-the-art methods.

#### 4.6 Experiments on the datasets with different $\lambda$

Furthermore, to evaluate the influence of different value of  $\lambda$  in Eq. (5), we carried out experiments with our method on the abovementioned datasets, and the experiments are shown in Fig. 3.

**Table 3** Face verification performances on YTF

Methods	Images	Single loss	Accuracy (%)	Time (sec)
DeepID3	–	No	93.20	1.4
$L_2$ Softmax	3.7M	Yes	95.54	1.4
FaceNet	200M	Yes	95.12	2.2
VGG-Face	2.6M	No	97.30	2.1
Baidu	1.3M	No	–	1.5
Deep face	4M	No	91.4	1.3
Range loss	1.5M	No	93.70	2.7
Deep visage	4.48M	No	96.25	1.9
Our method	2.62M	No	96.61	1.1

The “Time” column represents the execution time for single input image

**Table 4** Face verification performances on CACD

Methods	Accuracy (%)	Time (sec)
High-Dimensional LBP	81.60	2.4
Hidden factor analysis	84.40	2.6
LF-CNN	98.50	2.1
Center loss	97.48	1.7
Marginal loss	98.95	1.5
Our method	99.12	1.5

The “Time” column represents the execution time for single input image

As shown in Fig. 3, while the value of  $\lambda$  is greater than 0.40, the accuracy of the proposed method would start to degenerate. It demonstrates that the value of  $\lambda$  should be set to around 0.40.

#### 4.7 Analysis

From the experimental results on LFW, YTF, and CACD, we can observe the availability and reliability of the proposed loss function. By integrating the shared features from the input image pairs to the proposed CNN model (the non-occluded face and occluded face) and unique information from each input facial image, the presented loss function could provide the constraint from a pair of input facial images on the original softmax loss and have shown its performance in the face verification tasks. Meanwhile, similar to human being’s visual system, our proposed approach can extract the global and local features of the images of faces. The global features combined with local features form a layout for each identity.

The proposed approach have proven to improve the face verification accuracy by integrating the softmax loss and the mutual information loss while the  $\lambda_i$  in Eq. (5) is used to implement the trade-off between them. As shown in Fig. 3, the optimal value of  $\lambda$  should be set to 0.35–0.45. To note, the introduction of the new loss function contributes substantially to the image classification by combining the complementary information from both the non-occluded image and occluded image.

## 5 Conclusion

To implement the partially occluded face verification, we propose a deep learning strategy-based two-channel CNN architecture and a newly presented loss function. In the proposed CNN architecture, two parameter-sharing CNN channels are exploited to respectively process a pair of face images: the non-occluded facial image and occluded facial image. At the end of the network, both the shared feature and the unique feature could be obtained in a feature layer. The mutual information regularized softmax loss is iteratively optimized through the alternating minimization algorithm. To evaluate the performance of

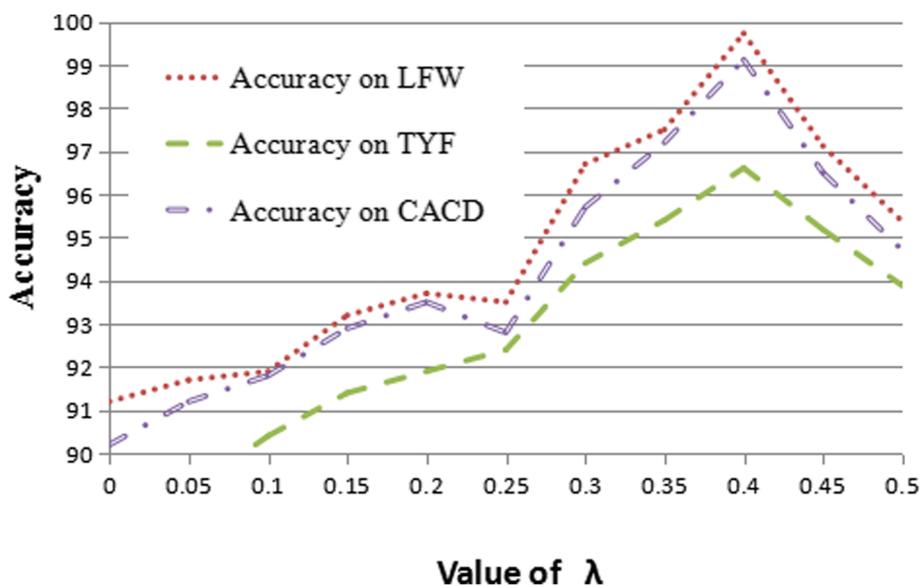


Fig. 3 Performance of the proposed method with different  $\lambda$

the proposed approach, we conducted comparison experiments between the state-of-the-art methods and ours on several publicly available face image datasets. Experimental results show that the proposed approach outperforms the state-of-the-art methods in accuracy.

This paper offers several contributions. First of all, a novel deep CNN is proposed to implement the face verification task. Secondly, this is probably the first attempt to introduce the novel loss function in the CNN architecture. Meanwhile, it is also an early application of the shared information between the non-occluded image and occluded image into the same CNN model. Thirdly, our approach performs with superiority to the state-of-the-art face verification techniques.

In our future works, we will continue to implement more applications [31, 48] of the presented CNN architecture. For instance, we would use more practical images, e.g., the blurry images, and evaluate the accuracy of the proposed CNN on these images. To achieve this objective, we will continue to collect more face images and create a publicly available dataset.

**Abbreviations**

AUC: Area under curve; CACD: Cross-age celebrity dataset; CNN: Convolutional neural network; LFW: Labeled Faces in the Wild; ReLU: Rectified linear unit; ROC: Receiver operating characteristics; YTF: Youtube Faces

**Acknowledgements**

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

**Funding**

There is no funding received by this project.

**Availability of data and materials**

We would like to very much share our image dataset with the public upon getting a permission from the hospital where the dataset was acquired. We will try our best to do it because we think it can facilitate the related field's growth and help on advertising our approach.

**Authors' contributions**

All authors take part in the discussion of the work described in this paper. The contributions of the proposed work are mainly in the following aspects: A novel deep CNN architecture is proposed to extract the shared information between the input pair of facial images that were manually collected and could be optimized through alternating minimization. We propose a novel loss function (named after mutual information loss). It can both maximize the intra-identity distances and minimize the inter-identity similarity of the extracted features. By combining the mutual information loss and the softmax loss, our proposed method could produce the highly discriminative features that would contribute to accurate face verification. Experiments on the public datasets our approach outperforms the state-of-the-art techniques with an impressive superiority. All authors read and approved the final manuscript.

**Authors' information**

1. Lei Yang. He is now a doctor candidate in China University of Mining and Technology. His research interests include graphics, mage processing, and data mining.
2. Jie Ma. She is a lecturer in Jiangsu Normal University. Her main research interests include graphics, image processing, and data mining.
3. Jian Lian. He is now an instructor in Shandong University of Science and Technology. His interest includes machine learning and image processing.
4. Yan Zhang. She is a doctor candidate in Shandong University of Science and Technology. Her research interests include machine learning, machine vision, and image analysis.
5. Houquan Liu. He is a Professor in China University of Mining and Technology. Main researches are graphic, image processing, and virtual reality.

**Competing interests**

All authors declare that they have no competing interests. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221008, China. <sup>2</sup>Xuhai College China University of Mining and Technology, Xuzhou 221008, China. <sup>3</sup>School of Education Intelligent Technology, Jiangsu Normal University, Xuzhou 221116, China. <sup>4</sup>Department of Electrical Engineering Information Technology at Shandong University of Science and Technology, Jinan 250031, China. <sup>5</sup>College of Mining and Safety Engineering, Qingdao 266590, China. <sup>6</sup>Shandong Normal University, Jinan 250014, China.

Received: 26 August 2018 Accepted: 20 November 2018

Published online: 13 December 2018

## References

1. D. S. Bartoo, *Financial services innovation: opportunities for transformation through facial recognition and digital wallet patents*. (Dissertations & Theses - Gradworks, Ann Arbor, 2013)
2. M. Ketcham, N. Fagfæ, *The algorithm for financial transactions on smartphones using two-factor authentication based on passwords and face recognition*. *International Symposium on Natural Language Processing*. (Springer, Cham, New York, 2016), pp. 223–231
3. J. Xiao, Research on application of face recognition in area of public security. *Comput. Sci.* **43(11A)**, 127–132 (2016)
4. D. Chawla, M. C. Trivedi. A comparative study on face detection techniques for security surveillance (Springer, Singapore, 2018), pp. 531–541
5. Q. Zhao, M. Ye, *The application and implementation of face recognition in authentication system for distance education*. *International Conference on NETWORKING and Digital Society*. (IEEE, Los Alamitos, 2010), pp. 487–489
6. D. Yang, A. Alsadoon, P. W. C. Prasad, et al., An emotion recognition model based on facial recognition in virtual learning environment. *Procedia Comput. Sci.* **125**, 2–10 (2018)
7. Y. Sun, D. Liang, X. Wang, et al., DeepID3: face recognition with very deep neural networks. *Comput. Sci.* **abs/1502.00873** (2015)
8. M. A. Hasnat, J. Bohn, J. Milgram, et al., von Mises-Fisher mixture model-based deep learning: application to face verification. arXiv: 1706.04264 (2017)
9. R. Ranjan, C. D. Castillo, R. Chellappa, L2-constrained softmax loss for discriminative face verification. arXiv preprint arXiv: 1703.09507 (2017)
10. F. Schroff, D. Kalenichenko, J. Philbin, et al., FaceNet: a unified embedding for face recognition and clustering[J]. *Computer vision and pattern recognition*, 815–823 (2015)
11. L. J. Karam, T. Zhu, Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments[J]. *Proceedings of SPIE - The International Society for Optical Engineering*. 9394:93940B-93940B-10 (2015)
12. J. Deng, Y. Zhou, S. Zafeiriou, *Marginal loss for deep face recognition*. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. (IEEE Computer Society, Los Alamitos, 2017), pp. 2006–2014
13. W. Liu, Y. Wen, Z. Yu, et al., Large-margin softmax loss for convolutional neural networks. *International Conference on International Conference on Machine Learning*, 507–516 (2016)
14. J. Deng, J. Guo, S. Zafeiriou, ArcFace: additive angular margin loss for deep face recognition. arXiv: 1801.07698 (2018)
15. I. J. Goodfellow, D. Warde-Farley, M. Mirza, et al., *Maxout networks*. (ICML, Atlanta, 2013), pp. 1319–1327
16. A. Krizhevsky, I. Sutskever, Hinton G.E., in *NIPS*. Imagenet classification with deep convolutional neural networks (MIT Press, Cambridge, 2012), pp. 1097–1105
17. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 1–9 (2014)
18. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al., Going deeper with convolutions. *Cvpr*, 1–9 (2015)
19. K. He, X. Zhang, S. Ren, J. Sun, in *CVPR*. Deep residual learning for image recognition (IEEE, Los Alamitos, 2016), pp. 770–778
20. M. Kan, S. Shan, X. Chen, in *CVPR*. Multi-view deep network for cross-view classification (IEEE, Los Alamitos, 2016), pp. 4847–4855
21. I. Masi, S. Rawls, G. Medioni, P. Natarajan, in *CVPR*. Pose-aware face recognition in the wild (IEEE, Los Alamitos, 2016), pp. 4838–4846
22. Y. Sun, X. Wang, Tang X., in *ICCV*. Hybrid deep learning for face verification (IEEE, Los Alamitos, 2013), pp. 1489–1496
23. Y. Sun, X. Wang, Tang X., in *CVPR*. Deep learning face representation from predicting 10,000 classes (IEEE, Los Alamitos, 2014), pp. 1891–1898
24. R. Ranjan, S. Sankaranarayanan, C. D. Castillo, Chellappa R., in *FG 2017*. An all-in-one convolutional neural network for face analysis (IEEE, Los Alamitos, 2017), pp. 17–24
25. X. Qi, L. Zhang, Face recognition via centralized coordinate learning. arXiv preprint arXiv: 1801.05678 (2018)
26. W. Liu, Y. Wen, Z. Yu, M. Yang, in *ICML*. Large-margin softmax loss for convolutional neural networks (ACM, New York, 2016), pp. 507–516
27. S. Sankaranarayanan, A. Alavi, C. D. Castillo, R. Chellappa, in *BTAS*. Triplet probabilistic embedding for face verification and clustering (IEEE, Los Alamitos, 2016), pp. 1–8
28. S. Sankaranarayanan, A. Alavi, R. Chellappa, Triplet similarity embedding for face verification. arXiv preprint arXiv: 1602.03418 (2016)
29. Y. Zhang, J. Lian, M. Fan, et al., Deep indicator for fine-grained classification of bananas ripening stages. *Eurasip. J. Image Video Process.* **2018(1)**, 46 (2018)
30. O. M. Parkhi, A. Vedaldi, A. Zisserman. Deep face recognition (British Machine Vision Association, Durham, 2015), pp. 41.1–41.12
31. J. Lian, Y. Zheng, W. Jiao, et al., Deblurring sequential ocular images from multi-spectral imaging (MSI) via mutual information[J]. *Med. Biol. Eng. Comput.* **56(6)**, 1107–1113 (2018)
32. L. Wolf, T. Hassner, Maoz I., in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. Face recognition in unconstrained videos with matched background similarity (IEEE, Los Alamitos, 2011), pp. 529–534
33. B.-C. Chen, C.-S. Chen, W. H. Hsu, in *European Conference on Computer Vision*. Cross-age reference coding for age-invariant face recognition and retrieval (Springer, New York, 2014), pp. 768–783
34. E. Zhou, Z. Cao, Q. Yin, Naive-deep face recognition: touching the limit of LFW benchmark or not? *Comput. Sci.* **abs/1501.04690** (2015)
35. Q. Cao, L. Shen, W. Xie, et al., VGGFace2: a dataset for recognising faces across pose and age, 67–74 (2017)
36. Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, in *ECCV*. Ms-celeb-1m: a dataset and benchmark for large-scale face recognition (Springer, New York, 2016), pp. 87–102
37. I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, E. Brossard, *The megaface benchmark: 1 million faces for recognition at scale*. (IEEE, Los Alamitos, 2016), pp. 4873–4882
38. B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, A. K. Jain, in *CVPR*. Pushing the frontiers of unconstrained face detection and recognition: larpa janus benchmark a (IEEE, Los Alamitos, 2015), pp. 1931–1939
39. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv: 1603.04467 (2016)
40. J. Liu, Y. Deng, T. Bai, et al., Targeting ultimate accuracy: face recognition via deep embedding. arXiv: 1603.04467 (2015)
41. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Deepface: closing the gap to human-level performance in face verification (IEEE, Los Alamitos, 2014), pp. 1701–1708
42. X. Zhang, Z. Fang, Y. Wen, et al. Range loss for deep face recognition with long-tail (IEEE, Los Alamitos, 2016)
43. A. Hasnat, J. Bohne, J. Milgram, et al., *DeepVisage: making face recognition simple yet with powerful generalization skills*. *IEEE International Conference on Computer Vision Workshop*. (IEEE Computer Society, Los Alamitos, 2017), pp. 1682–1691
44. D. Chen, Cao X., F. Wen, J. Sun, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification (IEEE, Los Alamitos, 2013), pp. 3025–3032
45. D. Gong, Z. Li, D. Lin, J. Liu, X. Tang, in *Proceedings of the IEEE International Conference on Computer Vision*. Hidden factor analysis for age invariant face recognition (IEEE, Los Alamitos, 2013), pp. 2872–2879

46. Y. Wen, Z. Li, Y. Qiao, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Latent factor guided convolutional neural networks for age-invariant face recognition (IEEE, Los Alamitos, 2016), pp. 4893–4901
47. Y. Wen, K. Zhang, Z. Li, Y. Qiao, in *European Conference on Computer Vision*. A discriminative feature learning approach for deep face recognition (Springer, New York, 2016), pp. 499–515
48. X. Ren, Y. Zheng, Y. Zhao, et al., Drusen segmentation from retinal images via supervised feature learning. *IEEE Access*. **PP**(99), 1–1 (2017)
49. J. Hu, J. Lu, Y. P. Tan, Discriminative deep metric learning for face verification in the wild. *Comput. Vis. Pattern Recognit. IEEE*, 1875–1882 (2014)
50. Y. Taigman, M. Yang, M. Ranzato, et al., *DeepFace: closing the gap to human-level performance in face verification*. *IEEE Conference on Computer Vision and Pattern Recognition*. (IEEE Computer Society, Los Alamitos, 2014), pp. 1701–1708
51. Y. Sun, X. Wang, X. Tang, et al., *Deeply learned face representations are sparse, selective, and robust*. *Comput. Vis. Pattern Recognit.* (Los Alamitos, 2015), pp. 2892–2900
52. F. Schroff, D. Kalenichenko, J. Philbin, et al., FaceNet: a unified embedding for face recognition and clustering. *Comput. Vis. Pattern Recognit.*, 815–823 (2015)
53. O. M. Parkhi, A. Vedaldi, A. Zisserman, et al., *Deep face recognition*. (British Machine Vision Conference, 2015), pp. 41.1–41.12

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---