Intelligent Data Analysis 28 (2024) 377–392
DOI 10.3233/IDA-227309
IOS Press

Robust partial face recognition using multi-label attributes

Gaoli Sang[a,*], Dan Zeng[b], Chao Yan[c,f], Raymond Veldhuis[d,e] and Luuk Spreeuwers[d]
[a]*College of Information Science and Engineering, Jiaxing University, Jiaxing, Zhejiang, China*
[b]*Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, Guangdong, China*
[c]*College of Computer Science, Sichuan University, Chengdu, Sichuan, China*
[d]*Chair of Data Management and Biometrics (DMB), Faculty of Electrical Engineering, Computer Science and Mathematics (EEMCS), University of Twente, Enschede, Netherlands*
[e]*Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Gjϕvik, Norway*
[f]*Artificial Intelligence Laboratory, Sichuan Eface Technology Co., LTD., Chengdu, Sichuan, China*

Abstract. Partial face recognition (PFR) is challenging as the appearance of the face changes significantly with occlusion. In particular, these occlusions can be due to any item and may appear in any position that seriously hinders the extraction of discriminative features. Existing methods deal with PFR either by training a deep model with existing face databases containing limited occlusion types or by extracting un-occluded features directly from face regions without occlusions. Limited training data (i.e., occlusion type and diversity) can not cover the real-occlusion situations, and thus training-based methods can not learn occlusion robust discriminative features. The performance of occlusion region-based method is bounded by occlusion detection. Different from limited training data and occlusion region-based methods, we propose to use multi-label attributes for Partial Face Recognition (Attr4PFR). A novel data augmentation is proposed to solve limited training data and generate occlusion attributes. Apart from occlusion attributes, we also include soft biometric attributes and semantic attributes to explore more rich attributes to combat the loss caused by occlusions. To train our Attr4PFR, we propose an implicit attributes loss combined with a softmax loss to enforce Attr4PFR to learn discriminative features. As multi-label attributes are our auxiliary signal in the training phase, we do not need them in the inference. Extensive experiments on public benchmark AR and IJB-C databases show our method is 3% and 2.3% improvement compared to the state-of-the-art.

Keywords: Partial face recognition, multi-label attributes, discriminative feature learning

1. Introduction

In the past few years, face recognition has received significant attention and has become a successful application of pattern recognition and machine learning. Often, in unconstrained real-world face recognition applications, faces are captured without user collaboration, and it can happen that only parts of the face are visible. As illustrated in Fig. 1, faces can be obscured by glasses, scarves, veils, smartphones, shadows, hands, or self-occlusion due to poses. Partial face recognition (PFR), recognizing partial faces in unconstrained environments, remains a challenging and still a largely unsolved problem [1,2,3].

*Corresponding author: Gaoli Sang, College of Information Science and Engineering, Jiaxing University, Jiaxing, Zhejiang, China. E-mail: glsang@zjxu.edu.cn.

1088-467X/$35.00 © 2024 – IOS Press. All rights reserved.

Fig. 1. Partial face images from unconstrained environments (these pictures are collected from the internet).

According to whether detection of the occluded area is required, existing PFR methods can be roughly divided into two categories. The first category consists of occlusion region-based methods, which mainly use an effective representation of the unoccluded area [4,5,6,7,8,9]. They first detect the occluded and un-occluded regions of the input face image, and then extract features from un-occluded regions. These methods have achieved impressive results especially on the faces with limited occlusions. However, their performance largely depends on the accuracy of occlusion detection, which itself is a challenging task.

The methods in the second category, called training-based methods, directly learn robust features from training data [10,11,12,13,14,15]. Methods in this category extract user-defined texture features from the cropped face images, and then use the extracted features to train a face recognition classifier. Compared with the occlusion region-based methods, these training-based methods do not need to explicitly detect the occluded and un-occluded regions, but directly extract features from the whole face image. A major limitation of training-based methods is that they rely heavily on training data. Training data can affect training-based methods in two aspects: first, most databases do not contain sufficiently many occluded faces; second, face images in most databases usually contain a finite number of types of occlusions. An example is the AR [16] database. This is one of the widely used databases in face recognition under occlusion. The face images in AR contain occlusions by glasses and scarves. However, in real-world applications, such as face recognition for video surveillance, arbitrary occlusions may occur. In other words, the available training data only provides a limited representation of all possible cases of occlusions. The large-scale IJB-C database [17] contains a subset of 18 labeled occlusion types. Although it contains more occlusions than earlier databases, a drawback is the nonuniform distribution of types of occlusion, and with sometimes only one or two examples of a type. This unbalance can affect the performance of such training-based methods.

In this study, our work focuses on the second category of methods. We focus on improving the PFR accuracy by mitigating the limitations of the training data. Specifically, in order to cover more cases of occlusion, we augment un-occluded face images with various pre-defined types of occlusions. In the training phase, for each face image, we not only provide the identity but also multi-label attributes, including occlusion attributes, soft biometric attributes and semantic attributes. The main contributions of this paper include:

– We propose a novel data augmentation method that ensures that training data is equipped with faces with diverse occlusions. In addition to obtaining occluded faces without cost, we can identify the occluded area in the augmentation without applying occlusion detection. We also propose to use the aforementioned occluded area as an occlusion attribute to help train our Attr4PFR. In this way, such occlusion attributes work as soft spatial attention mechanisms for Attr4PFR, leading to an occlusion-free representation.

– We proposed Attr4PFR to extract discriminative features. We also introduce soft biometric attributes (i.e. gender, race, age) and semantic attributes (i.e. mole, scar, freckle) to form multi-label attributes. Together with identity (i.e. supervised by softmax loss), we further improve discriminative feature learning by implicit attributes loss. The proposed Attr4PFR is easy to extend to include more meaningful attributes for better performance.
– The proposed Attr4PFR can deal with arbitrary occlusions (e.g., pose variations) for PFR and yields consistent and significant improvements on the most popular AR database and the most challenging IJB-C database. Specifically, Attr4PFR is 3% improvement on AR database and 2.3% improvement on the IJB-C database.

The remainder of this paper is organized as follows. Section 2 provides a review of the related work. Section 3 details our proposed method. Section 4 describes our experimental results. The conclusions are presented in Section 5.

2. Related work

In this section, we review some existing PFR methods that are most relevant to our proposed method. The review will be organized into two main sections: face recognition with attributes and deep learning based PFR.

2.1. Face recognition with attributes

Facial attributes, also considered as soft biometric attributes, refer to a set of biological characteristics of the human face, providing a wide variety of identifying information like gender, age, race, hairstyle, accessories, etc. [18,19,20,21]. Over the past years, people attempted to use such soft biometric attributes to improve face recognition performance. An early study combining face attributes for this purpose was reported in 2009, the work [22] proposed to train attribute classifiers at first; then fused these attributes with other features for the face recognition task. However, in the context of deep learning, attribute-assisted face recognition does not gain too much attention. One related work, reported in [23], exploits CNN-based face attributes features for authentication, the facial attributes are used to train a deep CNN architecture. Later, the work [24] reformulated the fusion of features for face recognition and features of facial attributes as a gated two-stream neural network. More recently, the work [25] proposed to utilize face attributes to improve a CNN-based RGB-D face recognition feature learning task.

Unlike soft biometric attributes that every individual owns, there are some other semantic attributes such as scar, mole, freckle that play an important role in face recognition and can be used to improve recognition accuracy by complementing conventional face recognition. Especially, under occlusion and pose variation, these semantic attributes provide alternative features for face comparison. However, only a limited number of studies have been done so far on exploiting semantic attributes for face recognition. The work [26] detects potential moles in the facial region by using a multi-scale template matching algorithm. Similar work of detecting moles by template-based matching for face identification is presented by [27]. The work [28] presented their approach on prominent mole detection by applying a homomorphic filter on the contrast-enhanced image. These aforementioned schemes of mole detection do not consider other types of facial marks. Recently, the work in [29] proposed to use facial marks combined with a deep convolutional neural network by the weighted score sum approach.

Based on the assumption that face attributes could provide independent features from a feature representation learning perspective, we develop a robust PFR using multi-label attributes. In our method,

except occlusion attributes, soft biometric attributes, as well as semantic attributes together form our multi-label attributes are used to supervise PFR feature learning in the implicit way, thus they are not required during the testing stage.

2.2. Deep learning based PFR

For deep learning based methods, training data plays a key role in the performance, and in general more training data leads to better performance. In particular, a well-balanced sample distribution positively affects the performance of the algorithm. Due to the randomness of the occlusion position and occlusion content, both reconstructed and augmented methods for PFR have great challenges.

Methods attempting to reconstruct the occluded regions have gained popularity in solving PFR problems [30,31]. For example, the method in [30] used a novel mapping-autoencoder for occlusion detection and an iterative stacked denoising autoencoder for image reconstruction. Another related method was proposed in [31]. A stacked sparse denoising autoencoder with two channels was proposed to discard noise activations in the encoder network and achieve better image reconstructions. Recently, the work [32] proposed to use LSTM autoencoders with two channels to reconstruct faces in the wild. The drawback of these methods is that the reconstruction can only be achieved for face images of the same kind as the training images and does not generalize well to other kinds of facial images. Our proposed Attr4PFR can do PFR with arbitrary occlusions because we propose a novel data augmentation to solve limited training data and generate occlusion attributes.

Several other methods of training with augmented occluded faces have been proposed for improving the performance of PFR. For example, the work in [33] augments occluded faces with various hairstyles and glasses to enable the CNN model to be robust to various hairstyles and glasses. The method indeed relieves the data deficiency problem and results in improved performance. However, its use is limited to handling sunglasses and hair in recognition. Instead of using synthetic occluded faces directly, the work [34] first identifies the importance of face regions in an occlusion sensitivity experiment and then trains a CNN with identified face regions covered to reduce the model's reliance on these regions. Specifically, they propose to augment the training set with face images with occlusions located in high effect regions (the central part of the face) more frequently than in low effect regions (the outer parts of the face). This forces the model to learn more discriminative features from the outer parts of the face, resulting in less degradation when the central part is occluded. However, pre-defined occlusions may cause performance degradation when dealing with face images with an occlusion that is not of the same size. More recently, aiming at dealing with arbitrary sizes of partial faces, the work [35] proposed a method combining a fully convolutional network with sparse representation classification. Indeed, fully convolutional networks can be applied to extract spatial feature maps of arbitrary input image size. In particular, a sliding window of the same size as the input feature maps is used to decompose the gallery feature maps into several sub-feature maps.

Unlike existing occluded face augmentation methods [30,31,33,35], we augment occlusion faces with common and random occlusion masks in small cells coupled with occlusion attributes. The distinctiveness of our method lies in two aspects: i) First, our proposed method ensures that clean faces are trained with greater occlusion diversity; at the same time, it enables the identification of occluded areas in the augmentation without necessitating applying occlusion detection. ii) Second, the occluded areas, functioning as occlusion attributes, also assist in training our Attr4PFR to extract discriminative features. In this way, we can address various occlusion positions encountered in real-world scenarios and guidethe learning process, leading to Attr4PFR an occlusion-free method.

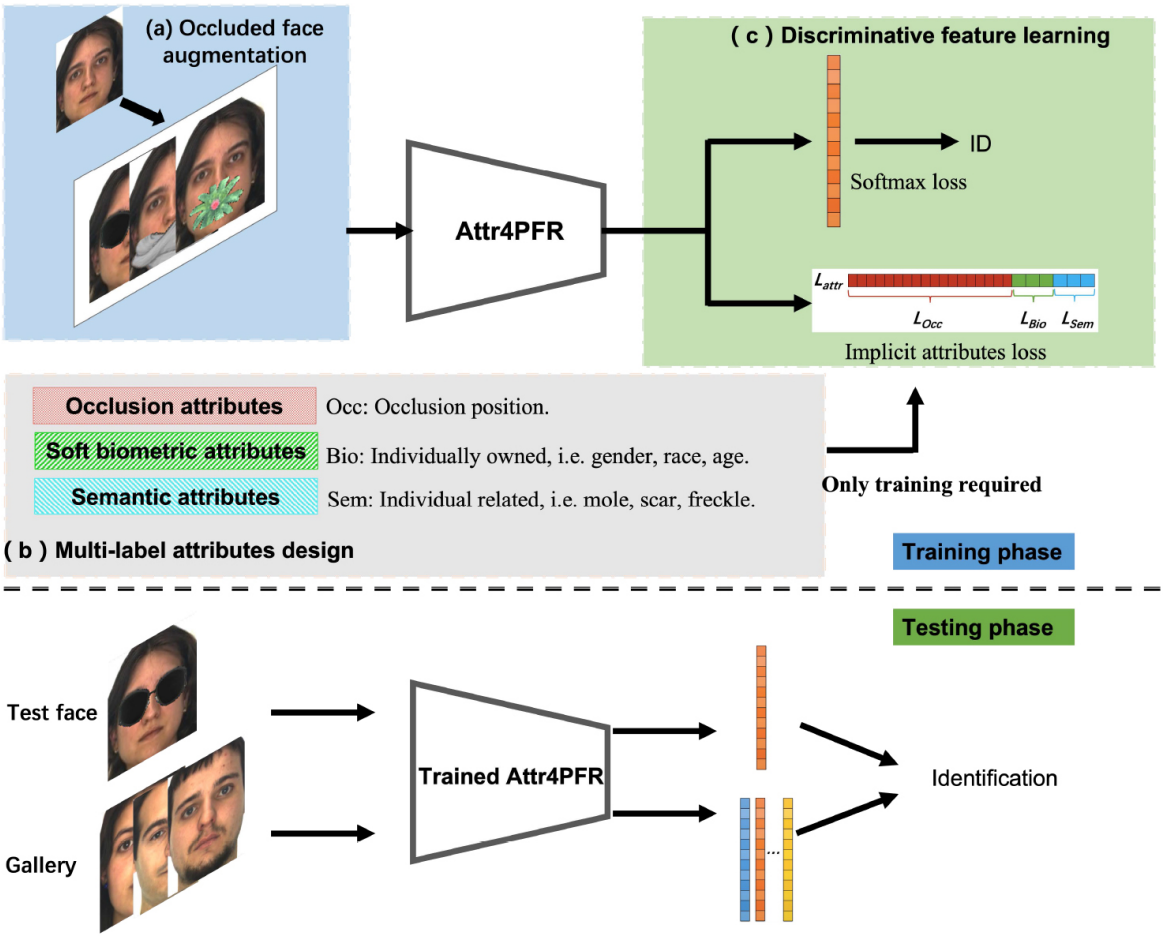


Fig. 2. The framework of the proposed method. In the training phase, there are three modules. a. Occluded face augmentation can generate common and random occlusion faces. b. Multi-label attributes contains occlusion attributes, soft biometric attributes and semantic attributes. c. Discriminative feature learning contains softmax loss and implicit attributes loss. As multi-labels attributes are our auxiliary signal in the training phase, we do not need them in the inference.

3. The proposed Attr4PFR

3.1. Overview

We propose a robust PFR method to deal with aforementioned issues: limited training data, occlusion detection, and discriminative feature learning. Figure 2 shows the framework of Attr4PFR which contains three parts: (a) occluded face augmentation, (b) multi-label attributes design, and (c) discriminative feature learning. At the training phase, occluded face images are fed into the Attr4PFR network. The output features of the Attr4PFR are fed into a fully connected layer and followed by the softmax loss layer. Furthermore, the output features are also followed with the implicit attributes loss layer together with corresponding multi-label attributes vector for further discriminative feature optimization in an implicit way. Figure 3 gives the detailed architecture of Attr4PFR. At recognition time, only the output of the first fully connected layer is used for biometric comparison, and the attribute-aware layer is not needed. In the remainder of this section, we introduce our proposed method in detail.

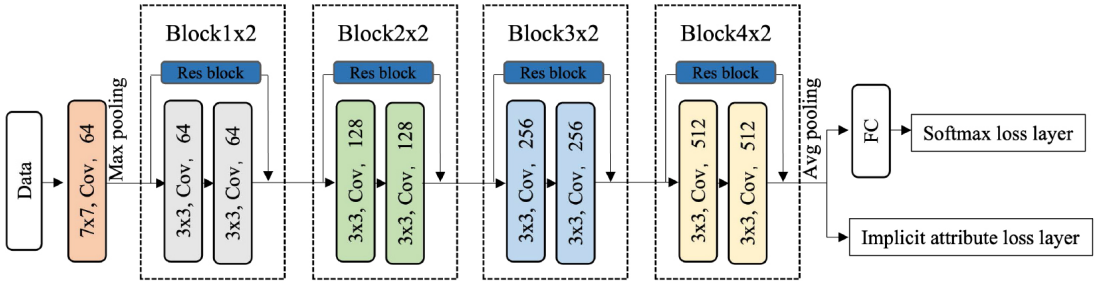


Fig. 3. The architecture of Attr4PFR contains a convolutional layer, a max-pooling layer, 4 × 2 residual blocks, an average pooling layer, a fully connected layer, a softmax loss layer, and a implicit attributes loss layer. The residual block is composed of two convolutional layers and a residual connection.

(a) Occlusion position.

(b) Occlusion types including common occlusion and random occlusions.

Fig. 4. An illustration of occlusion attributes as multi-label attributes.

3.2. *Occluded face augmentation*

As mentioned before, most face databases do not present enough occlusion cases and samples for a CNN to be trained well. We propose a novel data augmentation method for generating real-occluded face images in a strategic manner.

In real-world applications, occlusion may occur in various parts of the face. In order to represent various occlusions more accurately and flexibly, we divide the faces into small cells firstly. As shown in Fig. 4a, faces are evenly divided into 18 cells. So the occlusion attributes can be represented by a binary vector of length 18, with each element representing a cell. An element is set to 1 if a cell is occluded, otherwise it is set to 0.

In our proposed method, face images are randomly covered by a set of pre-designed occlusion masks (see Fig. 4b) before training. All the occlusion positions can be regarded as random occlusions due to factors, such as glasses, scarves, veils, masks or occluded by poses, hands, objects. For these occlusions, we propose occluded face augmentation method as below.

Two types of occlusions are considered: i) common occlusions such as glasses, scarves, veils, masks, the occlusion position are usually fixed; ii) random occlusions caused by other uncontrolled factors,

such as hands, objects, the occlusion positions are usually random and not fixed. For these two types of occlusions, we propose to use two types of occlusion masks to simulate real-world occlusions: common occlusion masks (the first 6 occlusion masks in Fig. 4b) and random occlusion masks (the last 7 occlusion masks in Fig. 4b). Then faces are covered with these occlusion masks with fixed or random locations. Our proposed method for occluded face augmentation that is easy to extend to more common and random occlusion masks, but it is a trade-off between method performance and efficiency.

It should be noted that when applying these pre-defined occlusion masks, the occlusion attributes of each augmented face are labeled meanwhile. For common occlusion, we utilize objects such as glass and scarf with sizes of 224×36 and 224×72, respectively. For random occlusion, it may occur in diverse regions of the face, attributable to theoretically any object of varying sizes. In this paper, we only enumerate a few examples of random occlusions with different sizes for illustration purposes. However, this concept can readily be expanded to incorporate a wide array of occluding objects with different sizes.

The deficiency of using pre-defined masks to obtain occluded faces has two causes: i) it is difficult to cover all kinds of occlusion situations, especially any object that can appear in a face; ii) lack of guidance during learning. In this paper, we propose to deal with this problem in two ways: i) dividing a face into smaller areas (18 small cells) and annotating occlusion situations by occlusion attributes (see 3.3); ii) extracting the un-occluded face regions with occlusion attributes advised (see 3.4). In such a way, the various positions of occlusions in the real world can be covered and guided for learning, which can make up for the shortcomings of the previous artificial masks.

3.3. Multi-label attributes

In this paper, we consider occlusion attributes, soft biometric attributes, and semantic attributes as our multi-label attributes. Our method is flexible to extend our multi-label attributes with more rich attributes if any.

Occlusion attributes

We consider face occlusion cases as an attribute called occlusion attribute. As a result, each occluded face image will associate with a definite occlusion attribute. As we defined before, the occlusion attribute vector contains the occlusion label of 18 cells, where 1 means the cell is occluded, and 0 means the cell is un-occluded. To be precise, we define the occlusion attribute vector by $Occ \in R^{H_1}$, where H_1 represents the number of cells, and each element labels the occlusion of one cell. The occlusion attributes vector as we defined them can deal with all kinds of occlusion cases.

Soft biometric attributes

In order to make the features more discriminative, we introduce several soft biometric attributes to improve the training process. Soft biometric attributes for the face refer to some individually owned traits that can be extracted from face images, such as gender, race, age, and so on. Our method is flexible to include more other soft biometric attributes for higher recognition accuracy. In this paper, three soft biometric attributes gender, race and age are involved in our multi-label attributes vector. Denoted by $Bio \in R^{H_2}$, H_2 represents the number of semantic attributes. For gender, 0 means male, 1 means female. For race or ethnicity, Asian, African and Caucasian is considered, represented by 0, 1, 2 respectively. For the age, three stages such as childhood, youth and old are considered, represented by 0, 1, 2 respectively. More soft biometric attributes can be easily obtained by some soft biometric attributes detector methods [19,36,37]. We propose a framework that is easy to extend to more soft biometric attributes, but it is a trade-off between method performance and efficiency.

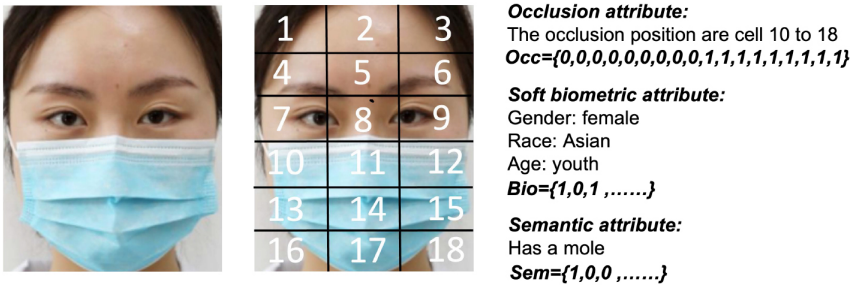


Fig. 5. An illustration of multi-label attributes including occlusion attributes, soft biometric attributes and semantic attributes.

Semantic attributes

We assume that semantic attributes can help to improve the extraction of discriminative features. For example, there are many criminal cases that use some original characteristics (such as having a mole) of the face on the surveillance image to narrow down the search and finally determine the identity of the criminal. The presence of a mole can make a notable difference between the faces of two individuals. Semantic attributes for the face refer to individually related attributes that can be summarized up on a human face semantically, such as mole, scar and freckle and so on. Therefore, we introduce semantic attributes into our multi-label attributes vector. Denoted by $Sem \in R^{H_3}$, where H_3 represents the number of semantic attributes. For simplicity, only mole, scar and freckle are considered. For example, 1 means a face has a mole (scar/freckle), 0 means a face has no mole (scar/freckle). As before, it is easy to extend to more semantic attributes, but it is a trade-off.

To this end, our multi-label attributes vector A constitute occlusion attributes *Occ*, soft biometric attributes *Bio* and semantic attributes *Sem*. Figure 5 shows some examples and their multi-label attributes.

3.4. Discriminative feature learning

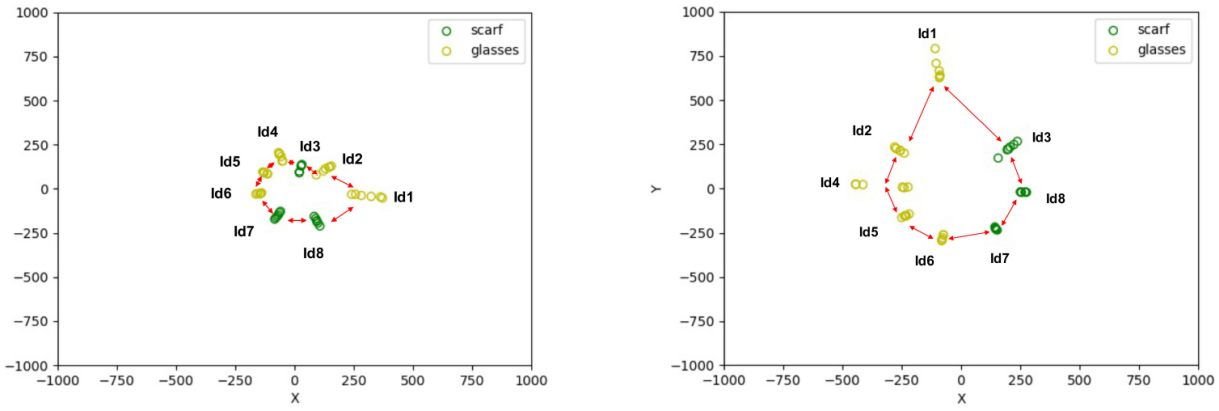
Discriminative features are learned from occluded face images by our PFRNet in a supervised manner. As shown in Fig. 2, we use softmax loss and implicit attributes loss to supervise Attr4PFR jointly in an implicit way. Unlike existing multi-task methods, which directly learn the prediction of two tasks, our Attr4PFR focuses on learning the association between discriminative features and attributes. During the training phase, multi-label attributes act as an auxiliary constraint for learning discriminative features. However, during testing, predicting multi-label attributes is not necessary, as we have already learned the mapping matrix of discriminative features from the multi-label attributes.

Using softmax as the loss function, the network pays attention to which identity has the highest probability, but not to the degree of distinction between the identity. While implicit attributes loss is optimized from the attribute level, and tries to perform attribute-related probability based on the extracting features. Particularly, utilizing implicit attributes loss directly mapping discriminative feature from related multi-label attributes. In other words, the loss function combining softmax and implicit attributes loss will determine the maximum probability that the sample belongs to a certain category under similar attribute conditions.

The formulations of softmax and implicit attributes loss functions will be presented separately.

Suppose training set is notated as $\{x_i\}_{i=1}^N$ with $x_i \in R^{m \times n}$, and the corresponding identity labels donated as $\{y_i\}_{i=1}^N$ with $y_i \in 1, 2, \ldots, C$. The softmax loss function L_s is formulated as:

$$L_s = -\sum_{i=1}^{N} \log \left(\frac{\exp(w_{y_i}^T f(x_i) + b_{y_i})}{\sum_{j=1}^{C} \exp(w_j^T f(x_i) + b_j)} \right) \tag{1}$$



(a) Softmax loss. (b) Softmax loss + occlusion attritube loss.

Fig. 6. The distribution of learned features under (a) softmax and (b) the joint supervision of softmax and occlusion attribute losses. There are eight identities with two different occlusions. The points with different colors denote features with different occlusions.

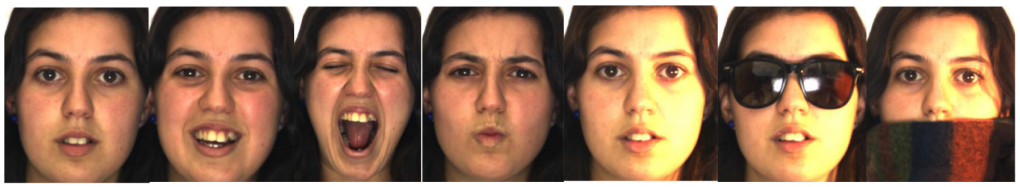
where $f(\cdot)$ is the learned feature mapping by training Attr4PFR, K is the dimension of deep feature $f(x_i)$. $W = [w_1, \ldots, w_C] \in R^{K\times C}$ and $b = [b_1, \ldots, b_C] \in R^{1\times C}$ are the weights and biases in the last fully connected layer.

Multi-label attributes here are used as an auxiliary supervision signal for discriminative feature learning, which can be combined with softmax loss. As shown in Fig. 2, implicit attributes loss contain three parts, occlusion attributes loss L_{Occ}, soft biometric attributes loss L_{Bio} and semantic attributes loss L_{Sem}. For a face image x_i, the multi-label attributes are composed of occlusion attributes (18-dimensional vector), soft biometric attributes (3-dimensional vector) and semantic attributes (3-dimensional vector). It is easy to merge the multi-label attributes to one vector $A_i \in R^{H_4} = \{Occ_i, Bio_i, Sem_i\}$. Then the implicit attributes loss can be formulated as:

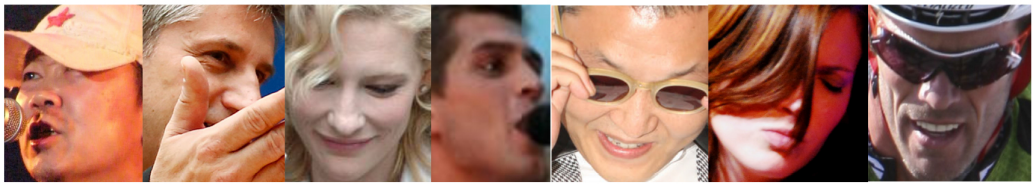
$$L_{attr} = \frac{1}{2}\sum \|((f_i - G(A_i))\|_2^2, \tag{2}$$

where $G \in R^{K\times H_4}$ is a parameter matrix to be trained of the implicit attributes loss layer, K and H_4 are the dimensions of the deep facial feature and the multi-label attributes vector, respectively. With global linear mapping $G(\cdot)$, features of different attributes remain separated, and features of the same attributes remain clustered. That is, the implicit attributes loss can improve the learned feature mapping by utilizing these attributes.

To show the effectiveness of our occlusion attribute-aware loss, we present a toy example on a small face set. This dataset is selected from the AR face database that will be introduced in Section 4.2, and contains only eight identities of the same gender and ethnicity but with different occlusions. Five of the identities are occluded by glasses, and three by scarf. We use ResNet [38] with 10 layers to train two models, one with softmax only, the other with both softmax loss and the occlusion attribute loss. Then we use t-SNE [39] for the visualization of high-dimensional features of two models, respectively. According to t-SNE, we reduce the output dimension of the penultimate fully connected layer to two dimensions and visualization them in Fig. 6a and b. We can see that the adjacent distance between identities in Fig. 6a is much larger than those in Fig. 6b. This indicates that the two-dimensional features of each identity become more discriminative through the regularization using occlusion attribute loss. We can also observe that features for identities of the same occlusion in Fig. 6b are closer to each other than in Fig. 6a, verifying



(a) Some example face images of a subject in the AR face database.



(b) Some occluded face images in the IJB-C face database.

Fig. 7. Example face images from the AR and IJB-C databases.

the occlusion property of the attribute loss. This also revels that the problem of distribution shift caused by adding occlusion can be alleviated when occlusion attribute loss is applied.

Finally, the whole discriminative features learning objective function can be formulated as:

$$L = L_s + \lambda L_{attr}, \tag{3}$$

where λ is a user-defined hyper-parameter to balance the two loss terms.

During the testing phase, the features are extracted from the first fully connected layer of trained Attr4PFR. Then cosine similarity is used to compute the similarity between the query feature and all the gallery features, which can be formulated as follows:

$$S(i,j) = \frac{\sum_{n=1}^{K} f^i \times f^j}{\sqrt{\sum_{n=1}^{K}(f^i)^2}\sqrt{\sum_{n=1}^{K}(f^j)^2}}, \tag{4}$$

where f^i and f^j are the features of query and gallery face image, respectively; K is the feature dimension.

4. Experiments and results

4.1. databases and experimental setup

To evaluate the performance of the proposed method, we compare our proposed method with several state-of-the-art methods both on the in-the-lab database (AR [16]) and on the in-the-wild databases (IJB-C database [17]).

The AR face database contains 4,000 face images of 126 different subjects with different facial expressions, illumination conditions and occlusions. Each subject is captured in two sessions and each session has 13 face images for each subject, where 3 of them wearing sunglasses and 3 wearing scarves, 3 of them are taken under various illumination conditions, 4 of them have different expressions. Figure 7a shows some examples from AR database. All the faces are captured in the laboratory environment.

The IJB-C database extends IJB-A, IJB-B, with more emphasis on occlusion, diversity of subject occupation and geographic origin. IJB-C includes real-world unconstrained faces from 3,531 subjects with full poses, illumination and occlusions variations. Many unconstrained faces were fully occluded by

any objects in any position. It is the most challenging occlusion face database. Figure 7b shows some examples from IJB-C database.

The models we used is based on the ResNet architecture [38] with our modifications of the loss function. Before training and testing, all the face images are detected by using an automated face detector and normalized to improve the recognition performance. Typically, five facial landmarks (the eye and mouth corners and nose tip) are localized by MTCNN [40]. Employing these facial landmarks for face alignment enables us to obtain occlusion attributes when generating occluded faces through our occluded face augmentation method. The MTCNN can be replaced by a more advanced facial landmarking method such as Retinaface [41]. For each training sample, we generate common and random occlusion situations, that are exactly total 13 generated samples with multi-label attributes are used for training. We train the models using stochastic gradient descent with the combination of softmax and multi-label attributes loss function. The batch size is 128. The learning rate begins at 0.1, and is divided by ten after 40K and 60K iterations, respectively.

During testing, the first fully connected layer is taken as the learned feature representation of the input face image. Then the similarity between two features is computed using their cosine distance.

4.2. Experiment on AR face database

In the first experiment, we evaluate the robustness of our method in dealing with real disguise on AR face database. 80 individuals is randomly selected as the training set, and the rest 46 subjects are divided into the test (30 individuals) and the validation set (16 individuals). For the training set, eight frontal face images (four images of Session 1 and 2) of each subject without occlusion are used to training. While, for the testing set, two kinds of sub-test sets are constructed: i) Six images with sunglasses from both Sessions, and ii) Six images with scarves from both Sessions. We compare our method (denoted by Attr4PFR) with some representative partial and occlusion face recognition approaches iincluding occlusion region-based method (denoted by DDL) in [6] and training based method (denoted by RDLRR) in [10], and other state-of-the-art methods including the method proposed (denoted by NNMR) in [42] and in [43] (denoted by IDI).

Because the AR database does not provide standard protocols for use, the experimental settings are slightly different. In order to make a better comparison, the comparison methods can be classified into 3 categories according to the experimental settings. 1) Face images for training and testing belong to the same individual, without overlapping. We use "S-TR-XOccs" to abbreviate the situation when the training set contains X occluded images per person. 2) Face images for training and testing belong to the same individual and without overlapping. We use "S-TR-Xps" to abbreviate the situation when the training set contains X un-occluded images per person. It is reasonable that the setting "S-TR-Xps" is more difficult than setting 'S-TR-XOccs'. 3) The subject for training and testing come from different individual without overlapping. We use 'D-TR-Xps' to abbreviate the situation when the training set contains X un-occluded images per person. It is obvious that 'D-TR-Xps' is the most difficult setting of the three experimental settings, there is no overlapping between training and testing.

Table 1 depicts a detailed comparison of the proposed method (Attr4PFR) with the other four representative algorithms. It is evident that even with more strict experimental settings, our proposed Attr4PFR achieves a recognition rate of 97.78% and 97.22% for sunglass and scarf occlusion, respectively, which is superior to all counterpart methods. In comparison with the DDRC and NNMR method for the scenarios of sunglass and scarf occlusion, the recognition rates for our proposed method of the two scenarios have little difference, which demonstrates that our method is less sensitive to the type of occlusions in the

Table 1
Recognition results for sunglasses and scarf occlusion on AR face database

Method	Experiment settings	Sunglass	Scarf
NNMR [42]	S-TR-8ps	96.9%	73.3%
DDL [6]	S-TR-8ps	79.92%	75.38%
DDRC [44]	S-TR-8ps	90%	99%
RDLRR [10]	S-TR-3Occs	91.7%	90%
Attr4PFR	D-TR-8ps	**97.78%**	**97.22%**

Table 2
Recognition results on IJB-C database with occlusions

Method	Occlusion types	Rank-1	Rank-5
DR-GAN [47]	Common + random occlusions	70.8%	82.8%
IFR [48]	Common + random occlusions	90.3%	93.2%
Attr4PFR	Common occlusions	92.55%	**95.36%**
	Random occlusions	**92.65%**	95.24%
	Common + random occlusions	92.6%	95.3%

training set. Recently, several mask-based face recognition methods (PDSN [45] and CAMFR [4]) have emerged due to the epidemic of COVID-19. They mainly training on a large-scale database and good at recognizing regular mask faces, but not for all random occlusions. Although, PDSN achieved the recognition of 98.20% and 98.33% for sunglass and scarf and CAMFR achieved the average recognition of 98.4%. Note that our proposed method only trained with part of the AR face database and is an ease to extend framework that can be further improved by adding more masks and effective attributes.

4.3. Experiment on IJB-C face database

In the second experiment, we evaluate the robustness of our method in dealing with more natural real disguise on IJB-C face database.

We train our modified ResNet [38] with CASIA-WebFace [46]. CASIA-WebFace database is a large scale public database containing 10,575 subjects and 494,414 images. We tested our proposed method on a standard protocol which is the 1: N mixed identification. There are about 1,800 templates for the gallery and about 20,000 templates for the probe. Since many probe sets include only a single image without occlusion. In this experiment, we select images with at least one occluded area as the test set. For better comparison of the importance of common and random occlusion masks, we divided the whole test set into two sub-set: common occlusion test set and random occlusion test set according that faces are occluded by common (occluded by glasses, scarves, or masks) and random(occluded by microphones, hands, poses or other random objects) occlusion.

Table 2 lists the Rank-1 and Rank-5 recognition results of our proposed method and other two state-of-the-art methods under the same protocol. Although our training data from CASIA-WebFace and our trained model is not fine-tuned on any gallery of IJB-C, our model still achieves the recognition accuracy of 92.6% on IJB-C occlusion subset, which proves the effectiveness of the proposed method again. Remarkably, the accuracy achieved on the common and the random set have small difference, which proves the effectiveness of our occlusion augmentation masks.

4.4. Ablation study

The effect of parameter λ

We conduct exploratory experiment to investigate the effect of λ used in two loss function combination.

Table 3
Rank-1 recognition results comparison of different λ on AR database with different λ

λ	0	0.05	0.15	0.25	0.35	0.45
Rank-1	94.17%	95.64%	**97.78%**	96.5%	95.22%	94.94%

Table 4
Rank-1 recognition results on AR database with different occlusion augmentation

Method	Training data	Label	Loss	Rank-1
Model A	Original clean data(baseline)	ID	Softmax	91.94%
Modle B	Common occlusion	ID	Softmax	93.33%
Modle C	Random occlusion	ID	Softmax	92.22%
Model D	Common + random occlusion	ID	Softmax	**94.17%**

Table 5
Rank-1 recognition results on AR database with different label and loss function

Method	Label	Loss	Rank-1
Model A	ID(baseline)	L_s (baseline)	94.17%
Model C	ID + occlusion attributes	$L_s + L_{Occ}$	96.39%
Model D	ID + soft biometric attributes	$L_s + L_{Bio}$	94.72%
Model E	ID + semantic attributes	$L_s + L_{Sem}$	95.56%
Model F	ID + multi-label attributes	$L_s + L_{attr}$	**97.78%**

By varying λ from 0 to 0.45, we evaluate our proposed method on the AR face database. The probe set contains AR faces with sunglasses and scarf occlusions and the gallery set contains 8 clean face for every subject. The rank-1 identification accuracy is given in Table 3. As λ being increased, the rank-1 recognition rate first rises up and then moves down as λ approaching 0.45. The best recognition rate is achieved at $\lambda = 0.15$.

The effect of occlusion augmentation

To further explore the importance of occlusion augmentation, we performed additional experiments with results in Table 4. First, by comparing the 'common' and 'random' occlusion masks, we see that 'common' occlusion masks noticeably increase performance. We speculate that it's due to the probe images are common occlusion as sunglass and scarf. Then by comparing the 'common' and 'common + random' occlusion masks, 'common + random' occlusion masks can improve performance somehow. Obviously, the importance of occlusion augmentation depends on the occlusion of the test data. If there are more common occlusions, then more common masks can be used for better performance; otherwise, more random occlusion masks is useful to improve the performance.

The effect of different attributes loss

To investigate the effectiveness of multi-label attributes and implicit attributes loss, we performed additional experiments with results in Table 5. As can be seen, compared with the baseline model A, model C, D, E and F with different attributes have different degrees of improvement in performance. It can also be seen that the best result is achieved by plugging softmax and implicit attributes loss. Specifically, among occlusion attributes, soft biometric attributes and semantic attributes, occlusion attributes have the greatest improvement, followed by semantic attributes, and soft biometric attributes have the smallest improvement. The reason for this could be that the probe set we tested, which consists of AR faces with

sunglasses and scarf occlusions. Our proposed Attr4PFR can precisely handle occlusion problems. On the other hand, soft biometric attributes and semantic attributes have relatively less variation in this database, thus their performance is less.

5. Conclusion

In this paper, we propose a multi-label Attributes for Partial Face Recognition (Attr4PFR) to explicitly extract discriminative features. We propose a novel data augmentation method to cover diverse occlusion and generate occlusion attributes. Explore soft biometric attributes (i.e. gender, race, age) and semantic attributes (i.e. mole, scar, freckle) as multi-label attributes for better performance. In addition to apply classic softmax loss, we propose implicit attributes loss to supervise discriminative feature learning. Extensive qualitative and quantitative experiments show that our Attr4PFR method on the most popular AR database and the most challenge IJB-C database achieves state-of-the-art recognition accuracy. Specifically, our method is 3% improvement on AR database and 2.3% improvement on the IJB-C databaset. In addition, ablation study has demonstrated that the effectiveness of each module.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 62206123), "Hundred Youth" Program of Jiaxing University (No. CD70621004), Scientific Research Foundation of Zhejiang Provincial Education Department (No. Y202249424) and Chengdu Major Science and Technology Innovation Project (No.2019-YF08-00264-GX).

References

[1] L. Zhang, B. Verma, D. Tjondronegoro and V. Chandran, Facial expression analysis under partial occlusion: A survey, *ACM Computing Surveys (CSUR)* **51**(2) (2018), 1–49.
[2] D. Zeng, R.N.J. Veldhuis and L.J. Spreeuwers, A survey of face recognition techniques under occlusion, *IET Biometrics* **10**(6) (2021), 451–470. doi: 10.1049/BME2.12029.
[3] A. Maafiri and K. Chougdali, Robust face recognition based on a new Kernel-PCA using RRQR factorization, Intelligent data analysis, 2021, 25.
[4] Y. Li, K. Guo, Y. Lu and L. Liu, Cropping and attention based approach for masked face recognition, *Applied Intelligence* **51**(5) (2021), 3012–3025.
[5] U. Jayaraman, P. Gupta, S. Gupta, G. Arora and K. Tiwari, Recent development in face recognition, *Neurocomputing* **408** (2020), 231–245.
[6] T. Zhang, Z. Yang, Y. Xu, B. Yang and W. Jia, Discriminative Dictionary Learning with Local Constraints for Face Recognition with Occlusion, in: *International Conference on Cloud Computing and Security*, Springer, 2018, pp. 733–744.
[7] D.L. Sánchez, J.M. Corchado and A.G. Arrieta, A CBR system for efficient face recognition under partial occlusion, in: *International Conference on Case-Based Reasoning*, Springer, 2017, pp. 170–184.
[8] H. Yang, X.-J. Yu, H.-S. Liet al., MSML: Enhancing Occlusion-Robustness by Multi-Scale Segmentation-Based Mask Learning for Face Recognition, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, AAAI Press, 2022, pp. 1381–1388. doi: 10.1609/aaai.v36i02.161.
[9] Y. Zhang, X. Wang, M.S. Shakeel, H. Wan and W. Kang, Learning upper patch attention using dual-branch training strategy for masked face recognition, *Pattern Recognition* **126** (2022).
[10] G. Gao, J. Yang, X.-Y. Jing, F. Shen, W. Yang and D. Yue, Learning robust and discriminative low-rank representations for face recognition with occlusion, *Pattern Recognition* **66** (2017), 129–143.
[11] N. McLaughlin, J. Ming and D. Crookes, Largest matching areas for illumination and occlusion robust face recognition, *IEEE Transactions on Cybernetics* **47**(3) (2016), 796–808.

[12] M.M. Alrjebi, N. Pathirage, W. Liu and L. Li, Face recognition against occlusions via colour fusion using 2D-MCF model and SRC, *Pattern Recognition Letters* **95** (2017), 14–21.
[13] Y. Duan, J. Lu, J. Feng and J. Zhou, Topology preserving structural matching for automatic partial face recognition, *IEEE Transactions on Information Forensics and Security* **13**(7) (2018), 1823–1837.
[14] B. Huang, Z. Wang, K. Jiang, Q. Zou, X. Tian, T. Lu and Z. Han, Joint Segmentation and Identification Feature Learning for Occlusion Face Recognition, IEEE Transactions on Neural Networks and Learning Systems, 2022.
[15] Z.X.S.H.Z.X.L.Z. Zhao Weisong, Consistent Sub-Decision Network for Low-Quality Masked Face Recognition, Ieee Signal Processing Letters, 2022, 1147–1151.
[16] A. Singh, D. Patil, M. Reddy and S. Omkar, The AR face database, 1998, 24.
[17] B. Maze, J. Adams, J.A. Duncan, N. Kalka, T. Miller, C. Otto, A.K. Jain, W.T. Niggel, J. Anderson, J. Cheney et al., Iarpa janus benchmark-c: Face dataset and protocol, in: *2018 International Conference on Biometrics (ICB)*, IEEE, 2018, pp. 158–165.
[18] X. Li and H. Zhang, Adapting geometric attributes for expression-invariant 3D face recognition, in: *IEEE International Conference on Shape Modeling and Applications 2007 (SMI'07)*, IEEE, 2007, pp. 21–32.
[19] Z. Liu, P. Luo, X. Wang and X. Tang, Deep learning face attributes in the wild, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.
[20] Z. Zhang, P. Luo, C.C. Loy and X. Tang, Learning deep representation for face alignment with auxiliary attributes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(5) (2015), 918–930.
[21] F. Taherkhani, N.M. Nasrabadi and J. Dawson, A deep face identification network enhanced by facial attributes prediction, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 553–560.
[22] N. Kumar, A.C. Berg, P.N. Belhumeur and S.K. Nayar, Attribute and Simile Classifiers for Face Verification, in: *Computer Vision, 2009 IEEE 12th International Conference on*, 2009.
[23] P. Samangouei and R. Chellappa, Convolutional neural networks for attribute-based active authentication on mobile devices, in: *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 2016, pp. 1–8.
[24] G. Hu, Y. Hua, Y. Yuan, Z. Zhang, Z. Lu, S.S. Mukherjee, T.M. Hospedales, N.M. Robertson and Y. Yang, Attribute-enhanced face recognition with neural tensor fusion networks, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3744–3753.
[25] L. Jiang, J. Zhang and B. Deng, Robust RGB-D face recognition using attribute-aware loss, IEEE transactions on pattern analysis and machine intelligence, 2019.
[26] J. Pierrard and T. Vetter, Skin Detail Analysis for Face Recognition, in: *IEEE Conference on Computer Vision & Pattern Recognition*, 2007.
[27] V.K.R. Ramesha K K B Raja and L.M. Patnaik, Template based mole detection for face recognition, *International Journal of Computer Theory and Engineering* **2**(5) (2010), 797–804.
[28] P.S.D.B.B.K.D. Usha Rani Gogoi Mrinal Kanti Bhowmik, Facial mole detection: An approach towards face identification, *Procedia Computer Science* **46** (2015), 1546–1553.
[29] S. Riaz, U. Park and P. Natarajan, Improving face verification using facial marks and deep CNN: IARPA Janus benchmark-A, *Image and Vision Computing* **104** (2020), 104020.
[30] Y. Zhang, R. Liu, S. Zhang and M. Zhu, Occlusion-robust face recognition using iterative stacked denoising autoencoder, in: *International Conference on Neural Information Processing*, Springer, 2013, pp. 352–359.
[31] L. Cheng, J. Wang, Y. Gong and Q. Hou, Robust deep auto-encoder for occluded face recognition, in: *Proceedings of the 23rd ACM International Conference on Multimedia*, 2015, pp. 1099–1102.
[32] F. Zhao, J. Feng, J. Zhao, W. Yang and S. Yan, Robust lstm-autoencoders for face de-occlusion in the wild, *IEEE Transactions on Image Processing* **27**(2) (2017), 778–790.
[33] J.J. Lv, X.H. Shao, J.S. Huang, X.D. Zhou and X. Zhou, Data augmentation for face recognition, *Neurocomputing* **230**(MAR.22) (2016), 184–196.
[34] D.S. Trigueros, L. Meng and M. Hartnett, Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss, *Image and Vision Computing* **79** (2018), 99–108.
[35] L. He, H. Li, Q. Zhang and Z. Sun, Dynamic feature learning for partial face recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7054–7063.
[36] S.A. Aly and B. Yanikoglu, Multi-label networks for face attributes classification, in: *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, 2018, pp. 1–6.
[37] S. Yang, F. Nian, Y. Wang and T. Li, Real-time face attributes recognition via HPGC: Horizontal pyramid global convolution, *Journal of Real-Time Image Processing* **17**(6) (2020), 1829–1840.
[38] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[39] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research* **9**(86) (2008), 2579–2605.

[40] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Processing Letters* **23**(10) (2016), 1499–1503.
[41] X. Zhang, S. Zhang, C. Liu, X. Shen et al., Retinaface: Single-stage dense face localisation in the wild, arXiv preprint arXiv:1905.00641, 2020.
[42] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang and Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(1) (2017), 156–171.
[43] S. Ge, C. Li, S. Zhao and D. Zeng, Occluded face recognition in the wild by identity-diversity inpainting, *IEEE Transactions on Circuits and Systems for Video Technology* **30**(10) (2020), 3387–3397.
[44] F. Cen and G. Wang, Dictionary representation of deep features for occlusion-robust face recognition, *IEEE Access* **7** (2019), 26595–26605.
[45] L. Song, D. Gong, Z. Li, C. Liu and W. Liu, Occlusion Robust Face Recognition Based on Mask Learning with PairwiseDifferential Siamese Network, IEEE, 2019.
[46] D. Yi, Z. Lei, S. Liao and S.Z. Li, Learning Face Representation from Scratch, arXiv: Computer Vision and Pattern Recognition, 2014.
[47] Luan, Quoc, Tran, Xi, Yin, Xiaoming and Liu, Representation Learning by Rotating Your Faces, IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018.
[48] B. Yin, L. Tran, H. Li, X. Shen and X. Liu, Towards interpretable face recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9348–9357.