Deep Disguised Faces Recognition

Kaipeng Zhang Ya-Liang Chang Winston Hsu
National Taiwan University, Taipei, Taiwan
kpzhang@cmlab.csie.ntu.edu.tw, {b03901014, whsu}@ntu.edu.tw

Abstract

Recently, deep learning based approaches have yielded a significant improvement in face recognition in the wild. However, "disguised face" recognition is still a challenging task that needs to be investigated, and the Disguised Faces in the Wild (DFW) competition is designed for this task. In this paper, we propose a two-stage training approach to utilize the small-scale training data provided by the DFW competition. Specifically, in the first stage, we train Deep Convolutional Neural Networks (DCNNs) for generic face recognition. In the second stage, we use Principal Components Analysis (PCA) based on the DFW training set to find the best transformation matrix for identity representation of disguised faces. We evaluate our model on the DFW testing dataset and it shows better performance over the state-of-the-art generic face recognition methods. It also achieves the best results on the DFW competition - Phase 1.

1. Introduction

Face recognition (FR) is one of the most active areas in the computer vision community. It has been studied for several decades with substantial progress. Recently, researchers utilize Deep Convolutional Neural Networks (DCNNs) for face recognition [14, 7, 13, 12, 11] and achieve nearly 100% accuracy of face recognition in the wild [5].

However, most of the existing research focuses on generic face recognition. Only very limited works deal with disguised face recognition (DFR) where "Disguise" denotes the facial accessories (e.g. glasses, hats and wigs) and makeup. These variances can obfuscate the identity or impersonate someone else's identity. DFR is still a challenging research topic that needs to be investigated.

The Disguised Faces in the Wild (DFW) competition [6] is designed for DFR. It consists of 11,157 images of 1000 individuals. Some examples of the dataset are shown in Fig. 1. There are four types of images in this dataset: normal, validation, disguise and impersonator. Normal and validation face images are non-disguised frontal faces. For a given subject, a disguised face image corresponds to the same identity with an intentional or unintentional disguise. Impersonator face image images are faces that look like the given subject but actually different.

Figure 1. Some example images of the Disguised Faces in the Wild (DFW) dataset. It consists of four kinds of images: normal, validation, disguise and impersonator. Normal and validation faces are general frontal faces. Disguised faces are challenging faces with an intentional or unintentional disguise. Impersonator faces look like the given subject but actually different.

Recent DCNNs-based face recognition methods [14, 7, 13, 12, 11] have achieved very good performance. Still, they require many training images for each given identity while in the DFW training set, there are only a few images for each identity.

In this paper, we propose a two-stage training approach to utilize the provided training data. Specifically, in the first stage, we train Deep Convolutional Neural Networks (DCNNs) for generic face recognition to extract identity features. In the second stage, we use Principal Components Analysis (PCA) on the DFW training set to find the best transformation matrix for identity representation. The second stage can be viewed as an adaptation process from generic face recognition to disguised face recognition.

The main contributions of this paper are summarized as follows:

- We propose a two-stage training approach for DCNNs-based disguised face recognition which can utilize lim-

2160-7516/18 $31.00 © 2018 IEEE
DOI 10.1109/CVPRW.2018.00012

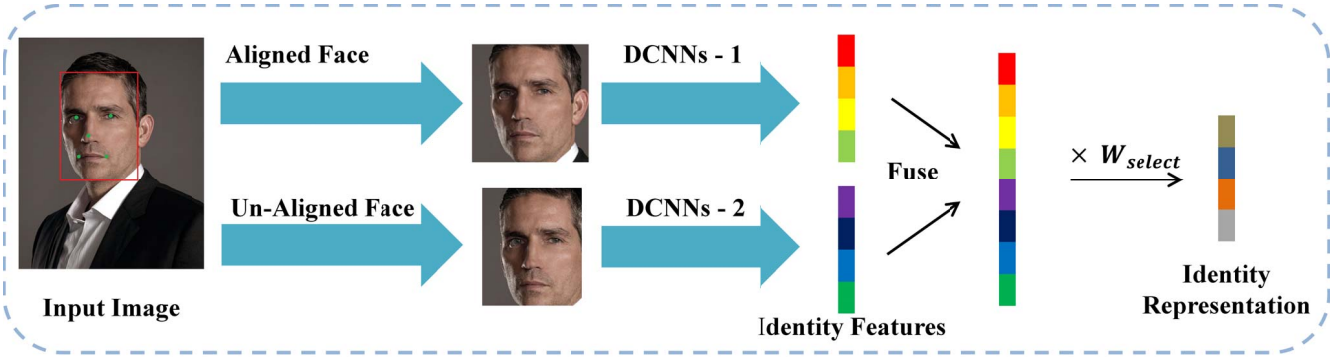

Figure 2. Illustration of our identity representation extraction pipeline. It consists of two stages. In the first stage, we use two DCNNs to extract identity features from aligned and unaligned faces respectively. Then we combine these two identity features. In the second stage, we use learned transformation matrix to transform identity features for disguised face recognition.

ited disguised training data for disguised face adaptation.

- The proposed method achieves superior performance over the state-of-the-art generic face recognition methods on the Disguised Faces in the Wild (DFW) benchmark. It also achieves the best results on the DFW competition - Phase 1.

2. Related Works

Face recognition is a classical problem in computer vision. In terms of testing protocol, it can be evaluated under closed-set or open-set settings [7]. In this paper, we only focus on open-set face recognition which facilities real-world application.

2.1. Generic Face Recognition

Generic face recognition methods are designed for addressing all kinds of face recognition. In recent years we have witnessed the great success of DCNNs-based methods in this task. Researchers proposed different loss functions for open-set face recognition such as contrastive loss [11], triplet loss [9], center loss [14], A-Softmax loss [7] and AM-Softmax loss [13]. However, these modified softmax loss functions and metric learning loss functions cannot perform good results with imbalanced training data and DCNNs-based methods require large-scale training data. Therefore, directly using these methods in the DFW training data will not get good performance.

2.2. Disguised Face Recognition

Disguised face recognition focuses on recognizing the identity of disguised faces and impersonators. There are limited research focus on this topic. [10] proposed a spatial fusion convolutional network to exploit facial part information. [3] proposed a dataset for this tasks. However, these methods and dataset are conducted in controlled scenarios. DFW [6] proposed a dataset collected from uncontrolled scenarios for disguised face recognition, but in the DFW dataset, there are very few images for each subject.

3. Proposed Method

In this section, we will first introduce our overall framework. Then we will explain our two-stage training approach.

3.1. Overall Framework

Our overall identity representation extraction pipeline is shown in Fig. 2. It includes two DCNNs for generic face identity features extraction and a transformation matrix W_{select} for disguised faces adaptation.

For training process, we first train two DCNNs for generic face recognition and then use Principal Components Analysis (PCA) find the transformation matrix for disguised face recognition adaptation. More specifically, in the adaptation training step, we project the identity features, extracted by DCNNs, into the PCA subspace defined by the principal components of greatest sample variance on the DFW training set. Then we pick the number of dimensions based on empirical performance on the DFW training set.

The overall training and testing pipelines are shown in Algorithm 1 and Algorithm 2.

3.2. Generic Face Recognition Training

In the generic face recognition training, we use the 64 layers ResNet-like convolutional neural network introduced in [7] and use AM-Softmax [13] as loss function with the generic face recognition training set. Then we train two DCNNs using un-aligned faces and aligned faces respectively since some alignments are failed. By doing so our method could also fuse pose relevant and pose irrelevant information. Un-aligned faces are cropped from images based

Algorithm 1 Proposed Training Pipeline

Input: Generic face recognition training set Set_G, DFW training set $Set_{D_{train}}$.
Output: DCNNs, W_{select}.

1: Train two DCNNs using Set_G (see Sec. 3.2).
2: Compute the identity features matrix M using above DCNNs.
3: Use Principal Components Analysis (PCA) to get the transformation matrix W for M.
4: **for** $i = 1$ **to** D **do**
5: Compute $M_{transform}$ by projecting M to the PCA subspace using the first i transformation vectors (i.e. principal components) of W.
6: Take $M_{transform}$ as the adapted identity features matrix, evaluate the performance and save the result as $scores_i$.
7: **end for**
8: Find the i in which the $M_{transform}$ achieves the best $scores$. W_{select} is the first i transformation vectors of W.

Algorithm 2 Proposed Testing Pipeline

Input: DFW testing set $Set_{D_{test}}$, DCNNs, W_{select}.
Output: Testing Results

1: Compute the identity features matrix N for $Set_{D_{test}}$ using DCNNs.
2: Compute $N_{transform}$ by projecting N to the PCA subspace using W_{select}.
3: Evaluate the performance.

on provided bounding boxes. Aligned faces are generated using similarity transform based on the five landmarks detected by MTCNN [16]. For a given face, we concatenate the identity features extracted from these two DCNNs.

3.3. Disguised Face Recognition Adaptation

In the disguised face recognition adaptation training, we first use the abovementioned two DCNNs to extract identity features matrix M for the DFW training set. Then we use Principal Components Analysis (PCA) to compute the transformation matrix W for M. Finally, we select the first i transformation vectors (i.e. principal components) of W as W_{select} based upon empirical performance on the DFW training set.

In the testing phase, given a face image, we use the two DCNNs to extract the identity features matrix N. Then, the final identity representation is computed by $N \times W_{select}$.

Methods	GAR@FAR=1%	GAR@FAR=0.1%
Aligned	0.8421	0.6912
Un-Aligned	0.8474	0.7038
Combined	**0.8571**	**0.7131**

Table 1. Evaluation of different DCNNs. These DCNNs are trained using one-stage training.

4. Experiment

4.1. Training Details

For generic face recognition training, we merge several public web-collected face recognition datasets including CASIA-WebFace [15], CelebA [8], MS1M [4] , UMDFACES [1] and VGGFace2 [2] as the generic face recognition training dataset. It roughly goes to 7.6M images of 92,748 unique persons. *We have removed the images or identities overlap between training and testing based on provided identity names.* More details of the removing overlap process can be found in the AM-Softmax paper [13].

For DCNNs training, we use batch size of 640, m (cosine margin constrain) of 0.35 and scale of 30 (norm-scale of features). The learning rate starts from 0.1 and is divided by 10 at the 70K, 90K iterations. The training process is finished at the 100K iterations.

For disguised face recognition adaptation, we select the first 250 transformation vectors to form the subspace projection W_{select}.

4.2. Testing Details

For our two-stage training, we use L2 distance to compute the identity distance. For the one-stage training evaluation, we compute the cosine similarity as identity similarity. For the DFW evaluation, for a given subject, positive pairs are constructed from normal, validation and disguised face images. Negative pairs are constructed from normal and impostor face images as well as cross subject face images. We use provided mask matrix (i.e. pairs) for evaluation. More testing details about the DFW protocol can be found in the DFW paper [6].

4.3. Ablation Experiment

4.3.1 Multiple DCNNs

We use two DCNNs for un-aligned and aligned faces respectively (as shown in Fig. 2). In this experiment, we evaluate the effectiveness of using multiple DCNNs based on one-stage training. From Fig. 3 and Tab. 1, we could see that combining different DCNNs can improve the performance.

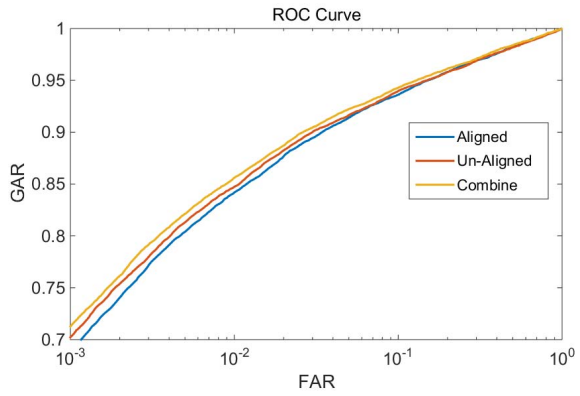

Figure 3. Evaluation of different DCNNs. These DCNNs are trained using one-stage training.

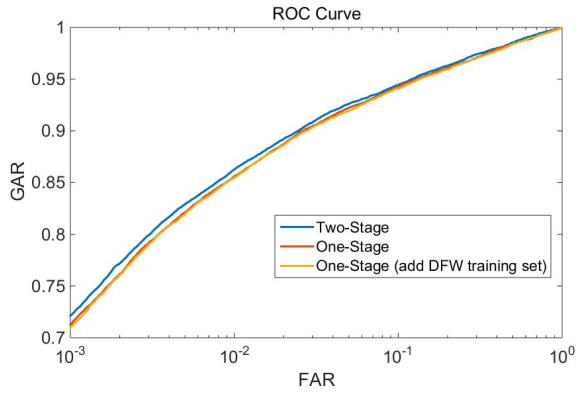

Figure 4. Evaluation of different training approaches.

Methods	GAR@FAR=1%	GAR@FAR=0.1%
One-stage	0.8571	0.7131
Two-stage	**0.8641**	**0.7221**

Table 2. Evaluation of different training approaches.

4.3.2 Two-stage training

We use two-stage training to utilize the small-scale DFW training set. In this experiment, we evaluate the effectiveness of this two-stage training. For ones-stage training, we also try to add the DFW training set into generic face recognition training data. From Fig. 4 and Tab. 2. We observe that two-stage training can improve the performance compared with one-stage training.

4.4. Evaluation on the DFW Testing Data

The Fig. 5 and Tab. 3 are the results of our methods using different testing pairs. 'Provided' denotes using testing pairs provided by organizers. 'Official' denotes the results received from organizers which use different testing pairs. We also test other state-of-the-art methods [13, 7, 14] on the DFW testing set using provided pairs and the results are shown in Fig. 6 and Tab. 4.

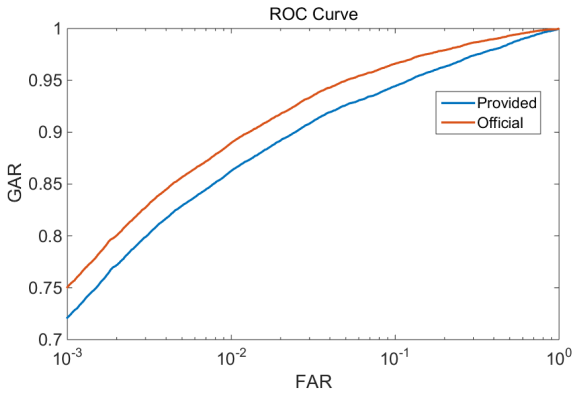

Figure 5. Evaluation results on the DFW testing data using different testing pairs.

	GAR@FAR=1%	GAR@FAR=0.1%
Provided	0.8641	0.7221
Official	0.8904	0.7508

Table 3. Evaluation results on the DFW testing data using different testing pairs.

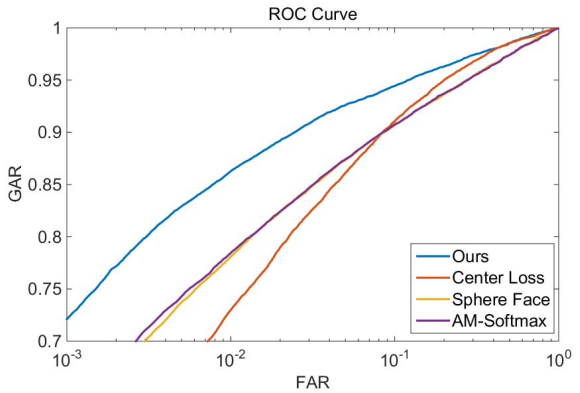

Figure 6. Evaluation results of different methods on the DFW testing data based on provided testing pairs.

Methods	GAR@FAR=1%	GAR@FAR=0.1%
Center Loss [14]	0.7305	0.5136
Sphere Face [7]	0.7814	0.6139
AM-Softmax [13]	0.7863	0.6305
Ours	**0.8641**	**0.7221**

Table 4. Evaluation results of different methods on the DFW testing data based on provided testing pairs.

5. Conclusion

In this paper, we propose a two-stage training approach to utilize the provided small-scale DFW training data for disguised faces adaptation. Specifically, we first train a generic face recognition model and then transform it to disguised face recognition. Our method achieves the best results on the DFW competition - Phase 1.

6. Acknowledgement

This work was supported in part by MediaTek Inc and the Ministry of Science and Technology, Taiwan, under Grant MOST 107-2634-F-002-007. We also benefit from the grants from NVIDIA and the NVIDIA DGX-1 AI Supercomputer.

References

[1] A. Bansal, A. Nanduri, C. Castillo, R. Ranjan, and R. Chellappa. Umdfaces: An annotated face dataset for training deep networks. *arXiv:1611.01484*, 2016.

[2] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. *arXiv*, 2017.

[3] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLOS ONE*, 9(7):1–16, 07 2014.

[4] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. MS-Celeb-1M: A dataset and benchmark for large scale face recognition. In *ECCV*. Springer, 2016.

[5] G. B. Huang and E. Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pages 14–003, 2014.

[6] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa. Disguised Faces in the Wild. Technical report, IIIT Delhi, March 2018.

[7] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. 2017.

[8] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *ICCV*, pages 3730–3738, 2015.

[9] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, pages 815–823, 2015.

[10] A. Singh, D. Patil, G. M. Reddy, and S. Omkar. Disguised face identification (dfi) with facial keypoints using spatial fusion convolutional network. *arXiv*, 2017.

[11] Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. In *CVPR*, pages 2892–2900, 2015.

[12] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, pages 1701–1708, 2014.

[13] F. Wang, W. Liu, H. Liu, and J. Cheng. Additive margin softmax for face verification. *arXiv*, 2018.

[14] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*, pages 499–515, 2016.

[15] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv:1411.7923*, 2014.

[16] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *SPL*, 23(10):1499–1503, 2016.