Face Verification with Disguise Variations via Deep Disguise Recognizer

Naman Kohli[1], Daksha Yadav[1], Afzel Noore[2]
[1]West Virginia University, [2]Texas A&M University-Kingsville
[1]{nakohli, dayadav}@mix.wvu.edu, [2]afzel.noore@tamuk.edu

Abstract

The performance of current automatic face recognition algorithms is hindered by different covariates such as facial aging, disguises, and pose variations. Specifically, disguises are employed for intentional or unintentional modifications in the facial appearance for hiding one's own identity or impersonating someone else's identity. In this paper, we utilize deep learning based transfer learning approach for face verification with disguise variations. We employ Residual Inception network framework with center loss for learning inherent face representations. The training for the Inception-ResNet model is performed using a large-scale face database which is followed by inductive transfer learning to mitigate the impact of facial disguises. To evaluate the performance of the proposed Deep Disguise Recognizer (DDR) framework, Disguised Faces in the Wild and IIIT-Delhi Disguise Version 1 face databases are used. Experimental evaluation reveals that for the two databases, the proposed DDR framework yields 90.36% and 66.9% face verification accuracy at the false accept rate of 10%.

1. Introduction

Automated face recognition is a complex and critical task which has received significant attention from researchers over the past couple of decades. Numerous approaches ranging from Eigenfaces [1] and Principal Component Analysis (PCA) [2] to Convolutional Neural Networks [3] have been proposed for designing accurate face recognition systems. However, due to the widespread deployment of face recognition applications, especially for security, it is vital to note that various challenges/covariates such as facial aging [4, 5], plastic surgery [6, 7], and facial disguise [8] hinder the performance.

Facial disguise involves intentional or unintentional alterations in the facial features with the aim of identity impersonation or obfuscation [9]. Figure 1 demonstrates an example of drastic variations in the appearance of an individual by utilizing different facial disguises. Facial disguise is a key area of concern for face recognition researchers as Dhamecha et al. [8] demonstrated that commercial face recognition systems exhibit poor performance while matching face images with disguises. For applications in border crossing and airport security, it is crucial to develop better face recognition algorithms that are able to mitigate the impact of facial disguises.

Figure 1: Illustrating variations in facial appearances of the same individual due to the usage of different disguises.

In the literature, several techniques have been developed for matching faces with disguises. Ramanathan et al. [4] utilized PCA along with Mahalanobis Cosine distance to match disguised faces. Singh et al. [10] used 2D log-polar Gabor features to accurately verify faces with disguises. Several other approaches involving PCA [11] and various texture descriptors [8, 12] have also been proposed. The state-of-the-art technique for verifying visible spectrum faces with disguises [9] comprises classifying face patches as biometric or non-biometric followed by matching of biometric patches using local binary patterns. Table 1 summarizes the different algorithms published in the literature for disguised face verification.

Recently, deep learning based algorithms have been successfully utilized for encoding feature representations in various image classification tasks including face recogni-

Table 1: Summary of published algorithms for detecting disguises in faces and face recognition with disguise variations.

Year	Authors	Face Disguise Classification	Face Recognition with Disguise Variations	Database Used
2002	Martinez et al. [13]	No	Yes	AR
2004	Ramanathan et al. [4]	Yes	Yes	National Geographic, AR
2005	Kim et al. [11]	No	Yes	AR, FERET
2009	Singh et al. [10]	No	Yes	AR, Private DB
	Wright et al. [14]	No	Yes	AR, Yale
2010	Yang and Zhang [15]	No	Yes	AR, Yale
2011	Min et al. [12]	Yes	Yes	AR
2013	Dhamecha et al. [8]	Yes	Yes	I^2BVSD
2014	Dhamecha et al. [9]	Yes	Yes	ID V1
2017	Singh et al. [16]	Yes	No	Simple and Complex Face Disguise DBs
2018	**Proposed**	No	Yes	DFW, ID V1

tion [3, 17, 18]. Singh et al. [16] employed deep learning features for disguised face classification. It should be noted that this paper did not address the problem of verifying disguised faces. However, there is a distinct lack of any study in the literature which utilizes deep learning based features for addressing this research challenge. In this paper, we propose a novel deep learning based framework for enhancing the performance of automatic face recognition with disguised faces. The key contributions of this paper are:

- Proposing a transfer learning framework, Deep Disguise Recognizer, which employs Inception-Net based features for verifying faces with disguise variations.
- Utilizing center loss for training which embeds features of the same class as close as possible. Thus, an individual having images consisting of disguise variations are clustered together.
- Evaluating the performance of the proposed framework on two disguise databases to demonstrate its state-of-the-art accuracy on the challenging databases.

2. Proposed Deep Disguise Recognizer Framework

Figure 2 showcases the proposed **D**eep **D**isguise **R**ecognizer (DDR) framework for verifying faces with disguise variations. The proposed framework utilizes Inception Residual Network [19] for learning facial features. The details of the proposed DDR framework are described below.

2.1. Inception-ResNet model

He et al. [20] introduced residual connections in deep networks which use additive merging of signals to learn optimally. Experimental evaluations showed their efficacy in different deep learning architectures and ability to train faster [19]. Inception-ResNet models have demonstrated superior performance in face recognition with triplet loss [18], automatic speech recognition [21], and even fingerprint minutia extraction [22].

As shown in Figure 2, Inception-ResNet network consists of the following six blocks followed by an average pooling layer alongside a bottleneck layer:

- Stem: Consists of three convolutional layers, followed by a max pooling layer, 1×1 convolutional layer and two convolutional layers
- Inception-ResNet-A: Consists of several 3×3 convolution layers (shown in Figure 3)
- Reduction-A: Uses 1×1 convolutions for dimensionality reduction
- Inception-ResNet-B: Branch consists of 7×1 and 1×7 convolutions

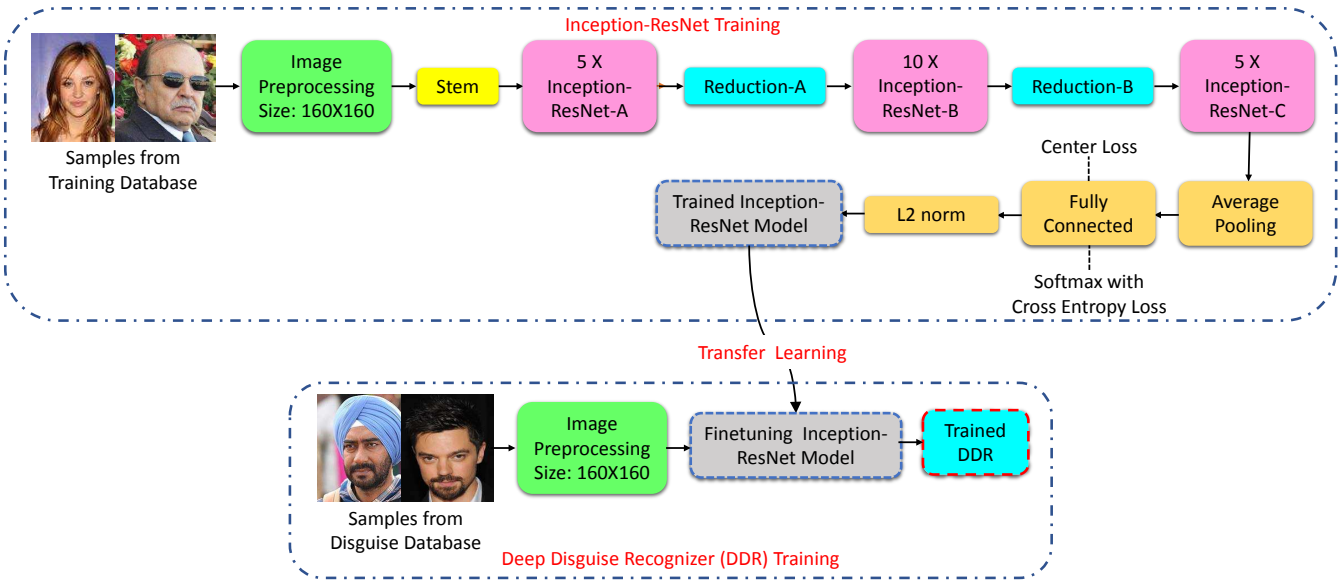


Figure 2: Flowchart of the proposed Deep Disguise Recognizer (DDR) showcasing the Inception-ResNet training and Deep Disguise Recognizer training steps.

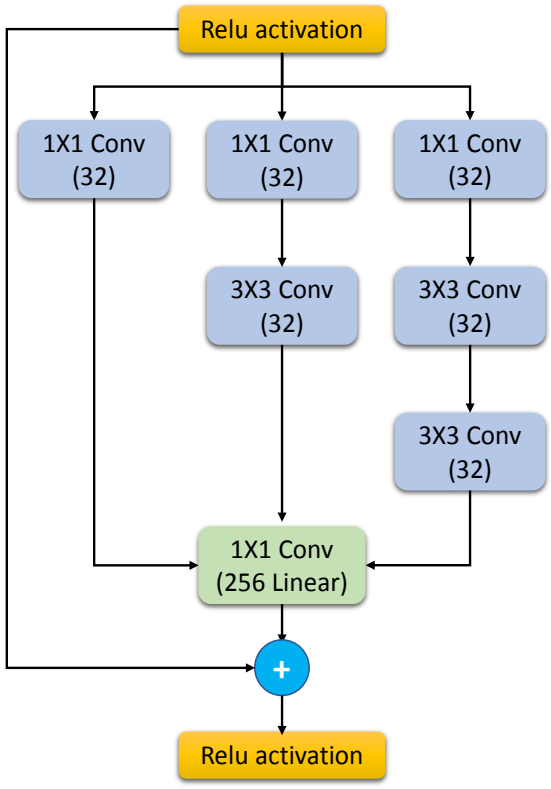


Figure 3: Diagrammatic representation of Inception-ResNet-A module of Inception-ResNet network.

- Reduction-B: Uses 1×1 convolutions for dimensionality reduction
- Inception-ResNet-C: Branch consists of 3×1 and 1×3 convolutions

Each of the Inception-ResNet blocks is repeatedly used to learn features deeply. As illustrated in Inception-ResNet training block of Figure 2, Inception-ResNet-A module is used 5 times, while Inception-ResNet-B and Inception-ResNet-C modules are used 10 and 5 times respectively. Softmax loss is utilized to predict the identity of the input face image.

2.2. Center Loss

Center loss (L_c) learns an embedding where the features of every class are closer to their corresponding class center [23]. It is defined as

$$L_c = \frac{1}{2} \sum_{i=1}^{m} \| x_i - c_{y_i} \|_2^2 \quad (1)$$

where x_i represents a sample feature vector and c_{y_i} denotes the y_ith class center of the deep features. In every iteration, centers are computed by averaging the features of corresponding classes in a mini-batch. The total loss of the model in practice is a sum of the cross-entropy softmax loss and weighted center loss.

2.3. Proposed Deep Disguise Recognizer

Torrey and Shavlik [24] suggest transfer learning is highly beneficial when the initial skill demonstrated on the

source model is high and can be leveraged for the target task. Thus, in the proposed Deep Disguise Recognizer (DDR) framework, a transfer learning approach is utilized for learning disguise invariant facial features.

For this, an Inception-ResNet model is trained using center loss on large public face databases for face recognition. The model embeds different centers for every individual identity using center loss. This trained Inception-ResNet model is able to accurately learn face representation of an input face image. Next, the trained model is used for inductive transfer learning for disguise invariant face verification. The idea here is to minimize the distance between the various disguises of an individual to their center identity feature vector. The last layer of the pre-trained source model is removed and the model is re-trained by adding a new softmax layer based on the total number of identities in the training partition of the disguise database.

Once the DDR framework is finetuned using the training set of the disguised faces database, the similarity score between a pair of input face images is computed as follows. First, each input image is preprocessed and then passed to the trained DDR framework which extracts the feature representations for the images (F_1 and F_2) from the fully connected layer. Next, the similarity between the two feature vectors (F_1 and F_2) is calculated using cosine-similarity score:

$$cos(F_1, F_2) = \frac{F_1 \cdot F_2}{||F_1|| \cdot ||F_2||} \quad (2)$$

3. Databases

For experimental evaluation, the following two databases are selected which contain face disguise variations:

- Disguised Faces in the Wild (DFW) [25]: The recently released DFW dataset consists of images from 1000 subjects and with a total of 11,155 face images. The images in the database have been collected from the Internet and hence, contain real-world variations as shown in Figure 4. The folder of each subject in the database contains face images of the subject with no disguise, with a disguise as well as some images of a similar appearing impersonator. The face image with disguise variations increase the intra-class variations while the impersonator images decrease the inter-class variations, hence, increasing the challenging nature of the problem.

- IIIT-Delhi Disguise Version 1 Face (ID V1) [9, 8]: Dhamecha et al. [9, 8] introduced the ID V1 database consists of 6,81 visible spectrum images of 75 subjects. The number of images per subject varies from 6 to 10. The database contains examples of various disguises such as mustache, sunglasses, hat, and masks. Sample images from ID V1 are shown in Figure 5.

Figure 4: Sample images of subjects from the DFW database [25] along with face images with disguise variations and impersonators.

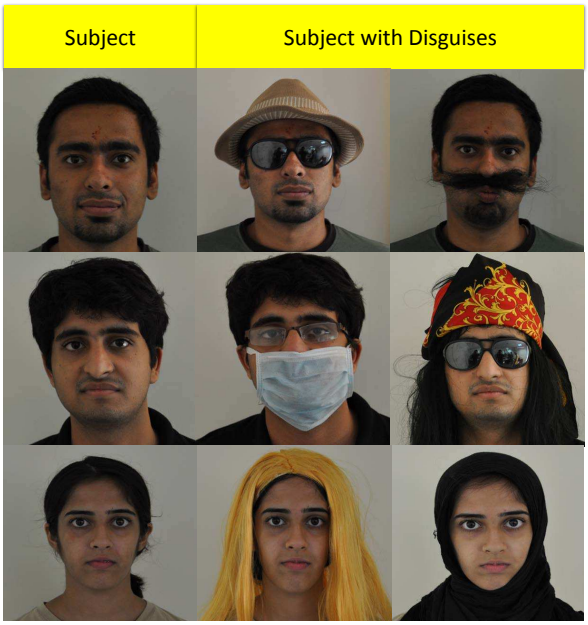


Figure 5: Samples from ID V1 database [9] illustrating the diverse variations in facial disguises.

4. Experimental Evaluation

In this section, we present the experimental protocols followed for training the proposed DDR framework on the two above described disguised face databases.

4.1. Training Inception-ResNet models

Two different large-scale face databases are used to train the Inception-ResNet model for face recognition:

- MS-Celeb [26]: Consists of over 10 million images comprising of 100,000 identities. The proposed DDR framework where Inception-ResNet is trained using MS-Celeb is termed as **DDR-MSCeleb**.
- VGG2 [27]: Consists of over 3 million images corresponding to 9,131 identities and has both pose and age variations. The proposed DDR framework where Inception-ResNet is trained using VGG2 database is referred to as **DDR-VGG2**.

In the proposed DDR framework, firstly, the images are aligned using multi-task cascaded convolutional neural network [28] and are resized to 160×160 size. The models are trained using RMS-Prop optimizers for 90 epochs and the best model based on validation data is saved. The dimensionality of the extracted face representations is 128.

4.2. Transfer Learning for Deep Disguise Recognizer

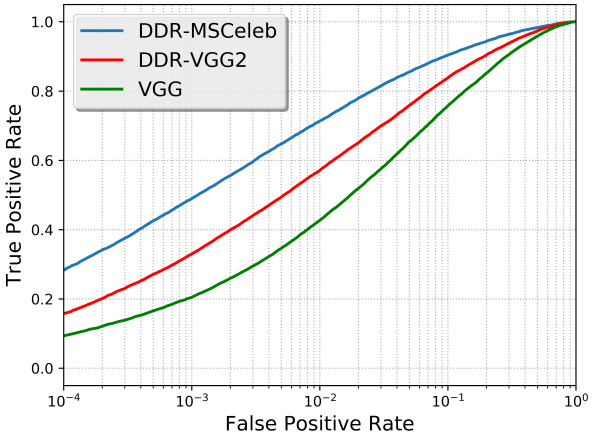
For DFW database, images from 400 subjects comprise the training set and 600 subjects comprise the testing set. For ID V1 database, images from 35 subjects form the training set while the images from the remaining 40 subjects are used for testing purposes. The training data from these databases are separately used for employing the transfer learning of the proposed deep disguise recognizer. We utilize the previously available coordinates to extract face images for the DFW Database. The DDR-MSCeleb and DDR-VGG2 are retrained using the training partitions of the disguise face databases. The training partition of DFW consists of 3,386 images but only the genuine face pairs are utilized for finetuning the Inception-ResNet model.

5. Experimental Results

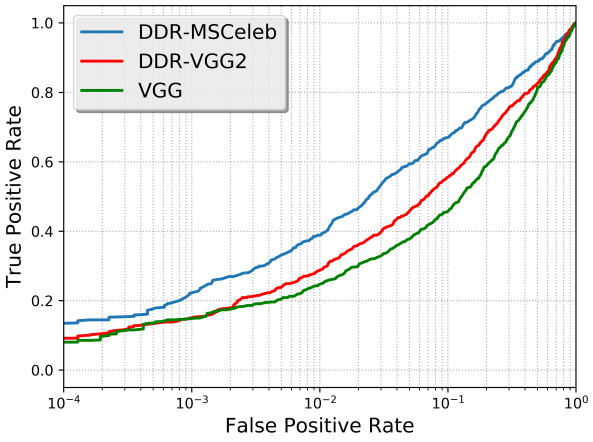
In the next section, we present the experimental results for face verification on the two disguise face databases: DFW and IDV1. The Receiver Operating Characteristic (ROC) curves of the two databases are shown in Figure 6.

5.1. Results on DFW Database

The testing partition of the DFW database contains 7771 face images, forming more than 18 million face pairs. The performance of the proposed DDR framework on DFW database is shown in Figure 6a and Table 2. For comparison purposes, performance of pre-trained VGG [3] network on the DFW database is also shown.



(a) Disguised Faces in the Wild (DFW) Database [25]



(b) IIIT-Delhi Disguise Version 1 Face (ID V1) Database [9]

Figure 6: ROC curves illustrating the face verification performance of the proposed Deep Disguise Recognizer (DDR) framework on disguised faces databases.

The best performing DDR framework, i.e. DDR-MSCeleb, achieves 49.08%, 71.43%, and 90.36% face verification accuracy when false accept rate = 0.1%, 1%, and 10% respectively. It is observed that the proposed DDR-MSCeleb consistently outperforms the proposed DDR-VGG2 framework for different false positive rate values. It exceeds the verification accuracy of DDR-VGG2 by 16.24%, 14.37%, and 6.59% when false accept rate = 0.1%, 1%, and 10% respectively.

The superior performance of the proposed DDR-

Table 2: Genuine Accept Rate (%) for the face verification experiment on the DFW database [25].

Algorithm	Verification Accuracy @ FAR		
	0.1%	**1%**	**10%**
VGG [3]	20.51	42.50	75.54
DDR-VGG2	32.84	57.06	83.77
DDR-MSCeleb	49.08	71.43	90.36

Table 3: Genuine Accept Rate (%) for the face verification experiment on the ID V1 Database [9]. * represents results reported from [9].

Algorithm	Verification Accuracy @ FAR		
	0.1%	**1%**	**10%**
SRC [14]*	5.6	15.5	37.7
Anavrta [9]*	7.4	16.6	38.1
VGG [3]	14.8	24.7	45.7
DDR-VGG2	15.2	28.7	55.6
DDR-MSCeleb	22.4	38.9	66.9

MSCeleb framework as compared to the proposed DDR-VGG2 can be attributed to a larger number of training face images as well as identities in MSCeleb database. The higher number of images with diverse real-world variations allows the DDR-MSCeleb to learn disguise invariant face representations. Some sample image pairs from the DFW database and their predicted similarity scores by the proposed DDR-MSCeleb framework are shown in Figure 7.

5.2. Results on ID V1 Database

The testing partition of the ID V1 database contains 31,800 face pairs. The face verification performance of the proposed DDR framework on IIIT-Delhi Disguise Version 1 (ID V1) face database is shown in Figure 6b and Table 3. The performance of state-of-the-art, Anavrta [9], VGG network [3], and Sparse Representation Classifier (SRC) [14] is also reported.

It is observed that the proposed DDR-MSCeleb framework for verifying faces with disguise variations outperforms the other approaches. It outperforms the current state-of-the-art, Anavrta [9], by 15%, 22.3%, and 28.8% for false accept rate = 0.1%, 1%, and 10% respectively. Also, both the versions of the proposed DDR framework (DDR-MSCeleb and DDR-VGG2) demonstrate higher face verification than the popular VGG [3] model. This demonstrates the superior efficacy of the proposed DDR framework for recognizing disguised faces.

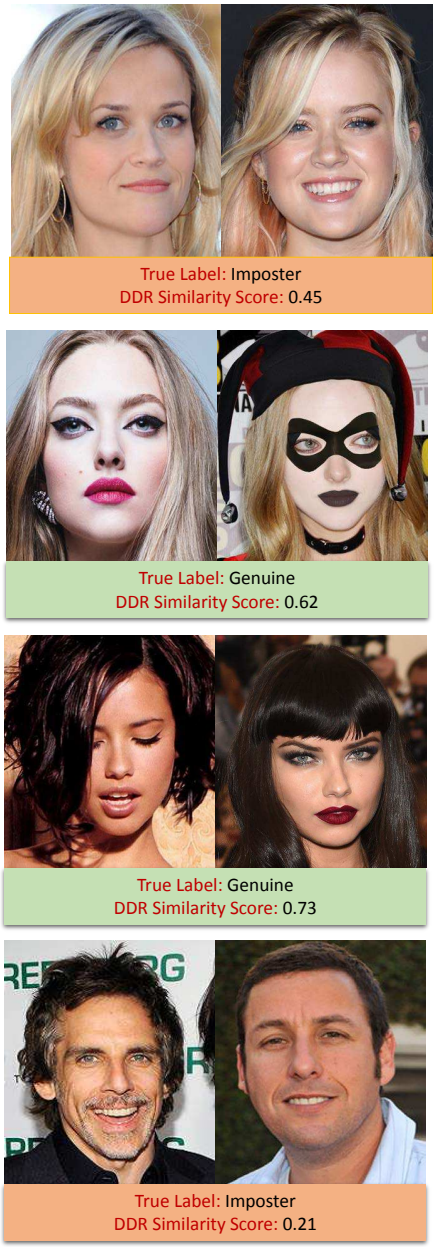
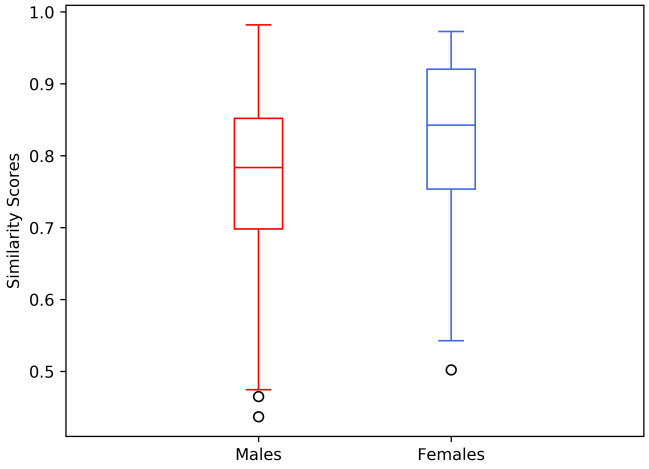


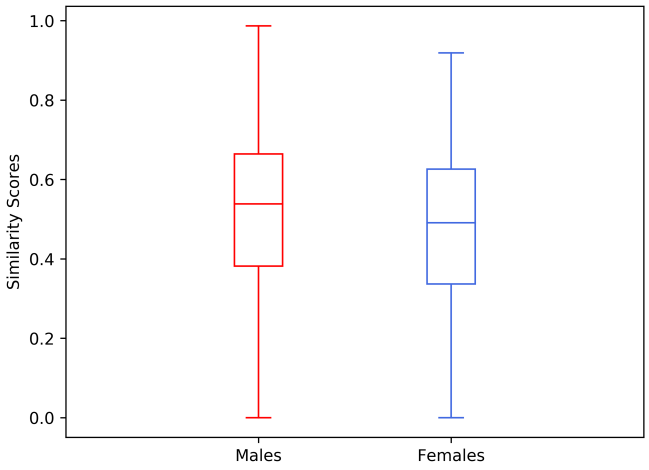
Figure 7: Sample face pairs from the DFW database [25] and the corresponding similarity score predicted by the proposed DDR-MSCeleb framework.

5.3. Gender-based Analysis

Next, the genuine scores produced by DDR-MSCeleb are analyzed based on the gender of the subject. The box plot shown in Figure 8a displays the similarity scores for male and female subjects in the ID V1 database. It is seen that the median genuine similarity score of female subjects is higher as compared to male subjects on this database. One of the reasons for this could be the limited number of



(a) IIIT-Delhi Disguise Version 1 Face (ID V1) Database [9]



(b) Disguised Faces in the Wild (DFW) Database [25]

Figure 8: Boxplots depicting genuine similarity score distribution by the proposed DDR-MSCeleb framework based on the genders. Median genuine similarity score by DDR-MSCeleb for male subjects is higher as compared to female subjects on the DFW database.

samples for female subjects as compared to male subjects. On the larger database, i.e. DFW, an opposite result is obtained as observed in Figure 8b. It is observed that the median genuine similarity score of male subjects is higher than female subjects. This might suggest that the proposed DDR framework may perform better in verifying male disguised faces as compared to female disguised faces.

6. Conclusion

In this paper, we proposed a novel deep learning based framework for recognizing faces with disguise variations. The proposed Deep Disguise Recognizer (DDR) involves a two-step training process: (1) training deep Inception-ResNet network using a large-scale face database for learning face representations and (2) transferring the trained Inception-ResNet model using disguised faces to encode representation which mitigates the effect of facial disguises. Experimental evaluation of the proposed DDR framework trained using MSCeleb face database reveals its superior performance on two disguised face databases: Disguised Faces in the Wild database and IIIT-Delhi Disguise Version 1 Face database. We also analyzed the performance of the proposed DDR framework based on the gender of the subject and observe that the DDR-MSCeleb framework currently shows better results for genuine male subjects as compared to genuine female subjects.

Acknowledgment

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 and Titan X Pascal GPUs used for this research.

References

[1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591. 1

[2] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004. 1

[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, vol. 1, no. 3, 2015, p. 6. 1, 2, 5, 6

[4] N. Ramanathan, R. Chellappa, and A. R. Chowdhury, "Facial similarity across age, disguise, illumination and pose," in *IEEE International Conference on Image Processing*, vol. 3, 2004, pp. 1999–2002. 1, 2

[5] B. Chen, C. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *European Conference on Computer Vision*, 2014, pp. 768–783. 1

[6] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan, "Plastic surgery: A new dimension to face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 441–448, 2010. 1

[7] N. Kohli, D. Yadav, and A. Noore, "Multiple projective dictionary learning to detect plastic surgery for face verification," *IEEE Access*, vol. 3, pp. 2572–2580, 2015. 1

[8] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *IEEE International Conference on Biometrics*, 2013, pp. 1–8. 1, 2, 4

[9] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, "Recognizing disguised faces: Human and machine evaluation," *PLOS ONE*, vol. 9, no. 7, 2014. 1, 2, 4, 5, 6, 7

[10] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," *Image and Vision Computing*, vol. 27, no. 3, pp. 245–257, 2009. 1, 2

[11] J. Kim, Y. Sung, S. M. Yoon, and B. G. Park, "A new video surveillance system employing occluded face detection," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2005, pp. 65–68. 1, 2

[12] R. Min, A. Hadid, and J.-L. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, 2011, pp. 442–447. 1, 2

[13] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002. 2

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009. 2, 6

[15] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *European conference on computer vision*, 2010, pp. 448–461. 2

[16] A. Singh, D. Patil, G. M. Reddy, and S. Omkar, "Disguised face identification (DFI) with facial keypoints using spatial fusion convolutional network," *arXiv preprint arXiv:1708.09317*, 2017. 2

[17] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015. 2

[18] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823. 2

[19] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning." in *AAAI*, vol. 4, 2017, p. 12. 2

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. 2

[21] C. Zhang and K. Koishida, "End-to-end text-independent speaker verification with flexibility in utterance duration," in *IEEE Automatic Speech Recognition and Understanding Workshop*, 2017, pp. 584–590. 2

[22] D.-L. Nguyen, K. Cao, and A. K. Jain, "Robust minutiae extractor: Integrating deep networks and fingerprint domain knowledge," in *IEEE International Conference on Biometrics*, 2018. 2

[23] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*, 2016, pp. 499–515. 3

[24] E. S. Olivas, *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*. IGI Global, 2009. 3

[25] V. Kushwaha, M. Singh, R. Singh, and M. Vatsa, "Disguised faces in the wild," IIIT-Delhi, Tech. Rep., 2018. 4, 5, 6, 7

[26] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*, 2016, pp. 87–102. 5

[27] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *IEEE Conference on Automatic Face and Gesture Recognition*, 2018. 5

[28] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct 2016. 5