Recognizing Disguised Faces in the Wild

Maneet Singh, *Student Member, IEEE,* Richa Singh, *Senior Member, IEEE,* Mayank Vatsa, *Senior Member, IEEE,* Nalini Ratha, *Fellow, IEEE,* and Rama Chellappa, *Fellow, IEEE*

Abstract—Research in face recognition has seen tremendous growth over the past couple of decades. Beginning from algorithms capable of performing recognition in constrained environments, existing face recognition systems achieve very high accuracies on large-scale unconstrained face datasets. While upcoming algorithms continue to achieve improved performance, many of them are susceptible to reduced performance under disguise variations, one of the most challenging covariate of face recognition. In this paper, the Disguised Faces in the Wild (DFW) dataset is presented which contains over 11,000 images of 1,000 identities with variations across different types of disguise accessories. The dataset is collected from the Internet, resulting in unconstrained face images similar to real world settings. This is a unique dataset that contains *impersonator* and genuine *obfuscated* face images for each subject. The DFW dataset has been analyzed in terms of three levels of difficulty: (i) easy, (ii) medium, and (iii) hard, in order to showcase the challenging nature of the problem. The dataset was released as part of the First International Workshop and Competition on Disguised Faces in the Wild at the International Conference on Computer Vision and Pattern Recognition, 2018. This paper presents the DFW dataset in detail, including the evaluation protocols, baseline results, performance analysis of the submissions received as part of the competition, and three levels of difficulties of the DFW challenge dataset.

Index Terms—Face Recognition, Disguise in the Wild, Impersonation, Obfuscation, Face Verification.

1 Introduction

Extensive research in the domain of face recognition has resulted in the development of algorithms achieving state-of-the-art performance on large-scale unconstrained datasets [1], [2], [3], [4]. However, it has often been observed that most of these systems are susceptible to digital and physical adversaries [5], [6], [7], [8], [9], [10]. Digital adversaries refer to manipulations performed on the image being provided to the recognition system, with the intent of *fooling* the system. It has been shown that traditional systems based on hand crafted features [5] degrade gracefully with digital attacks while deep learning systems deteriorate rapidly. Recently, the issue of digital attacks has garnered attention, with perturbation techniques such as Universal Adversarial Perturbation [11] and DeepFool [12] demonstrating devastating adversarial performance on different algorithms. On the other hand, physical adversaries refer to the variations brought to the individual before capturing the input data for the recognition system. In case of face recognition, this can be observed due to variations caused by different spoofing techniques or disguises. While the area of spoof detection and mitigation is being well explored [7], [13], research in the domain of disguised face recognition is yet to receive dedicated attention, despite its significant impact on both traditional and deep learning systems [14], [15].

Disguised face recognition encompasses handling both *intentional* and *unintentional* disguises. Intentional disguise refers to the scenario where a person attempts to hide his/her identity or *impersonate* another person's identity, in order to fool a recognition system into obtaining unauthorized access. This often results in utilizing external disguise accessories such as wigs, beard, hats, mustache, and heavy makeup, leading to obfuscation of the face region. This renders low inter-class variations between different subjects, thereby making the problem challenging in nature. Unintentional disguises cover a range of images wherein the face is obfuscated by means of an accessory such as glasses, hats, and masks. It can also be due to aging, resulting in an increase or decrease of facial hair such as beard or mustache, and variations in the skin texture. Unintentional disguises create challenges for the face recognition system by increasing the intra-class variations for a given subject. The combination of both intentional and unintentional disguises render the problem of disguised face recognition an arduous task. Fig. 1 presents sample images of intentional and unintentional disguises, along with non-disguised enrolled face images. The authentication system faces

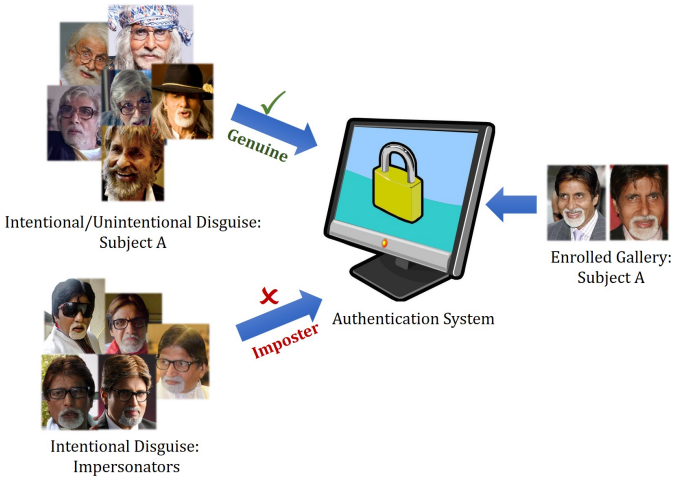


Fig. 1: Authentication systems often face the challenge of matching disguised face images with non-disguised enrolled images.

- *M. Singh, R. Singh, and M. Vatsa are with IIIT-Delhi, India, 110020 (e-mail: maneets@iiitd.ac.in, rsingh@iiitd.ac.in, mayank@iiitd.ac.in).*
- *N. Ratha is with IBM TJ Watson Research Center, New York, USA (e-mail: ratha@us.ibm.com).*
- *R. Chellappa is with Department of Electrical and Computer Engineering and UMAICS, University of Maryland, College Park, MD, 20742 (e-mail: rama@umiacs.umd.edu).*
- *DFW dataset link: http://iab-rubric.org/resources/dfw.html*

Manuscript received October 30, 2018.

TABLE 1: Summary of disguise face datasets in literature.

Name	Controlled Disguise	Number of		Availability of Impersonators	Publicly Available
		Images	Subjects		
AR Dataset (1998) [16]	Yes	3,200	126	No	Yes
National Geographic Dataset (2004) [17]	Yes	46	1	No	No
Synthetic Disguise Dataset (2009) [18]	Yes	4,000	100	No	No
Curtin Faces Dataset (2011) [19]	Yes	5,000	52	No	Yes
IIITD I^2BVSD Dataset (2014) [20]	Yes	1,362	75	No	Yes
Disguised and Makeup Faces Dataset (2016) [21]	No	2,460	410	No	Yes
Spectral Disguise Face Dataset (2018) [22]	Yes	6,480	54	No	Yes
DFW Dataset (2018)	**No**	**11,157**	**1,000**	**Yes**	**Yes**

the challenge of verifying an image containing unconstrained disguise variations against a frontal non-disguised face image.

This paper presents the Disguised Faces in the Wild (DFW) dataset[1], containing 11,157 face images of 1,000 identities. Almost the entire dataset is collected from the Internet resulting in an unconstrained set of images. One of the key highlights of the dataset is the availability of (i) *normal*, (ii) *validation*, (iii) *disguised*, and (iv) *impersonator* images for a given subject. This is a unique dataset containing multiple types of in-the-wild images for a subject in order to evaluate different aspects of disguised face recognition, along with three pre-defined evaluation protocols. Here, for a given subject, disguised face images are images of the same subject with disguise accessories, while impersonators correspond to images of different subjects. It is our assertion that the availability of a large-scale dataset, containing images captured in unconstrained settings across multiple devices, pose, illumination, and disguise accessories would help in encouraging research in this direction. The dataset was released as part of the DFW challenge, in the Disguised Faces in the Wild Workshop at International Conference on Computer Vision and Pattern Recognition (CVPR), 2018. We present the DFW dataset, along with the findings across the three evaluation protocols. Performance of participants in the DFW challenge along with the baseline results, and analysis of three difficulty levels has also been provided. The organization of this paper is as follows: Section 2 presents the motivation of the DFW workshop and challenge, followed by a detailed description of the DFW dataset in Section 3. Section 4 elaborates upon the DFW challenge, its submissions, and performance across the three protocols. Section 5 presents the DFW dataset's three degree of difficulties in terms of *easy*, *medium*, and *hard*.

2 Motivation

Table 1 presents the characteristics of existing disguise face datasets, along with the DFW dataset. One of the initial datasets containing disguise variations is the AR dataset [16]. It was released in 1998 and contains a total of 3,200 face images having some images containing controlled disguise variations. This was followed by the release of different datasets having variations across disguise accessories and dataset size. Most of the datasets are moderately sized having controlled disguise variations. Other than disguised face datasets, a lot of recent research in face recognition has focused on large-scale datasets captured in unconstrained environments [23], [24], [25], [26], [27].

1. Shorter version of this paper was presented at the CVPR Workshop on DFW, 2018 [14] which summarized the phase-I results of the competition. This manuscript presents the final results of the DFW 2018 competition, along with additional analysis and observations on the DFW dataset.

TABLE 2: Statistics of the DFW dataset.

Characteristic	Count
Subjects	1,000
Images	11,157
Normal Images	1,000
Validation Images	903
Impersonator Images	4,440
Range of Images per Subject	[5,26]

The availability of such datasets facilitate research in real world scenarios, however, they do not focus on the aspect of disguised face recognition.

Disguised face recognition presents the challenge of matching faces under both intentional and unintentional distortions. It is interesting to note that both forms of disguise can result in either genuine or imposter pairs. For instance, a criminal may intentionally attempt to conceal his identity by using external disguise accessories, thereby resulting in a genuine match for an authentication system. On the other hand, an individual might intentionally attempt to impersonate another person, resulting in an imposter pair for the face recognition system. Similarly, in case of unintentional disguises, use of casual accessories such as sunglasses or hats results in a genuine disguised pair, while individuals who look alike are imposter pairs for the recognition system. The combination of different disguise forms along with the intent makes the given problem more challenging.

To the best of our knowledge, no existing disguise dataset captures the wide spectrum of intentional and unintentional disguises. To this effect, we prepared and released the DFW dataset. The DFW dataset simulates the real world scenario of unconstrained disguise variations, and provides multiple impersonator images for almost all subjects. The presence of impersonator face images enables the research community to analyze the performance of face recognition models under physical adversaries. The dataset was released as part of the DFW workshop, where researchers from all over the world were encouraged to evaluate their algorithms against this challenging task. Inspired by the presence of disguise intent in real world scenarios, algorithms were evaluated on three protocols: (i) Impersonation, (ii) Obfuscation, and (iii) Overall. Impersonation focuses on disguise variations where an individual either attempts to impersonate another individual intentionally, or looks like another individual unintentionally. In both cases, the authentication system should be able to detect an (imposter) unauthorized access attempt. The second protocol, obfuscation, focuses on intentional or unintentional disguise variations across genuine users. In this case, the authentication system should be able to correctly identify genuine users even under varying disguises. The third protocol evaluates a face recognition model

on the entire DFW dataset. As mentioned previously, it is our hope that the availability of DFW dataset along with the three pre-defined protocols would enable researchers to develop state-of-the-art algorithms robust to different physical adversaries.

3 Disguised Faces in the Wild (DFW) Dataset

As shown in Table 1, most of the research in the field of disguised face recognition has focused on images captured in controlled settings, with limited set of accessories. In real world scenarios, the problem of disguised face recognition extends to data captured in uncontrolled settings, with large variations across disguise accessories. Combined with the factor of *disguise intent*, the problem of disguise face recognition is often viewed as an exigent task. The DFW dataset simulates the above challenges by containing 11,157 face images belonging to 1,000 identities with uncontrolled disguise variations. It is the first dataset which also provides impersonator images for a given subject. The DFW dataset contains the IIIT-Delhi Disguise Version 1 Face Database (ID V1) [15] having 75 subjects, and images corresponding to the remaining 925 subjects have been taken from the Internet. Since the images have been taken from the Web, most of the images correspond to famous personalities and encompass a wide range of disguise variations. The dataset contains images with respect to unconstrained disguise accessories such as hair-bands, masks, glasses, sunglasses, caps, hats, veils, turbans, and also variations with respect to hairstyles, mustache, beard, and make-up. Along with the disguise variations, the images also demonstrate variations across illumination, pose, expression, background, age, gender, and camera quality. The dataset is publicly available for research purposes and can be downloaded from our website [2]. The following subsections present the dataset statistics, protocols for evaluation, and details regarding data distribution.

3.1 Dataset Statistics

As mentioned previously, the DFW dataset contains images pertaining to 1,000 identities, primarily collected from the Internet. Most of the subjects are adult famous personalities of Caucasian or Indian ethnicity. Each subject contains at least five and at most twenty six images. The dataset comprises of 11,157 face images including different kinds of images for a given subject, that is, *normal*, *validation*, *disguised*, and *impersonator*. Detailed description of each type is given below:

- **Normal Face Image:** Each subject has a frontal, non-disguised, good quality face image, termed as the normal face image.
- **Validation Face Image:** Other than the normal face image, 903 subjects have another non-disguised face image, referred to as the validation image. This can help in evaluating a proposed model for matching non-disguised face images.
- **Disguised Face Image:** For each subject, disguised face images refer to images having intentional or unintentional disguise of the same subject. For the 1,000 identities present in the dataset, every subject has at least one and at most 12 disguised images. These images form genuine pairs with the normal and validation face images, and can help in evaluating the true positive rate of an algorithm.

2. http://iab-rubric.org/resources/dfw.html

TABLE 3: Statistics of the training and testing sets of the DFW dataset.

Number of	Training Set	Testing Set
Subjects	400	600
Images	3,386	7,771
Normal Images	400	600
Validation Images	308	595
Disguised Images	1,756	3,058
Impersonator Images	922	3,518

- **Impersonator Face Image:** Impersonators refer to people who intentionally or unintentionally look similar to another person. For a given subject, impersonator face images belong to different people, thereby resulting in imposter pairs which can be used to evaluate the true negative rate of an algorithm. The images were collected from the Internet using different relevant keywords on Google Images, news articles, and popular entertainment blogs and later manually verified by human examiners. In the DFW dataset, 874 subjects have images corresponding to their impersonators, each having at least 1 and at most 21 images.

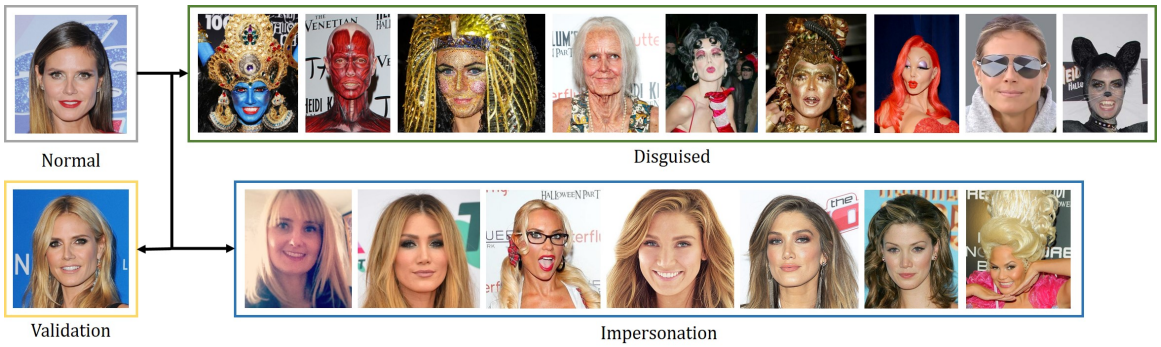
Statistics of the DFW dataset are presented in Table 2, and Fig. 2 demonstrates sample images of two subjects. It can be observed that disguised face images result in increased intra-class variations for a given subject, while the impersonator images render lower inter-class variability. Overall, the DFW dataset contains 1,000 and 903 normal and validation face images, respectively, along with 4,814 disguised face images, and 4,440 impersonator images.

3.2 Protocols for Evaluation

The DFW dataset has been released with three protocols for evaluation. A fixed training and testing split is provided which ensures mutual exclusion of images and subjects. Images from four hundred subjects are used to create the training set, and the remaining six hundred subjects form the test set. Table 3 presents the statistics of the testing and training sets. All three protocols correspond to verification, where a face recognition module is expected to classify a pair of images as genuine or imposter. Detailed description of each protocol on the pre-defined training and testing partitions is given below:

Protocol-1 (Impersonation) evaluates a face recognition model for its ability to distinguish impersonators from genuine users with high precision. A combination of a normal image with a validation image of the same subject corresponds to a genuine pair for this protocol. For imposter pairs, the impersonator images of a subject are partnered with the normal, validation, and disguised images of the same subject.

Protocol-2 (Obfuscation) is useful for evaluating the performance of a face recognition system under intentional or unintentional disguises, wherein a person attempts to hide his/her identity. The genuine set contains pairs corresponding to the (normal, disguise), (validation, disguise), and (disguise_1, disguise_2) images of a subject. Here, disguise_n corresponds to the n^{th} disguised image of a subject. That is, all pairs generated using the normal and validation images with the disguise images, and the pairs generated between the disguise images of the same subject, constitute the genuine pairs. The imposter set is created by combining the normal, validation, and disguised images of one subject with the normal,



(a) Subject A

(b) Subject B

Fig. 2: Images pertaining to two subjects of the DFW dataset. The dataset contains at most four types of images for each subject: Normal, Validation, Disguised, and Impersonator.

validation, and disguised images of a different subject. This results in the generation of cross-subject imposter pairs. The impersonator images are not used in this protocol.

Protocol-3 (Overall Performance) is used to evaluate the performance of a given face recognition algorithm on the entire DFW dataset. The genuine and imposter sets created in the above two protocols are combined to generate the data for this protocol. For the genuine set, pairs are created using the (normal, validation), (normal, disguise), (validation, disguise), and (disguise_1, disguise_2) images of the same subject. For the imposter set, cross-subject imposter pairs are considered, wherein the normal, validation, and disguised face images of one subject are combined with normal, validation, and disguised face images of another subject. Apart from the cross-subject imposter pairs, the impersonators of one subject are also combined with normal, validation, and disguised face images of the same subject to further supplement the imposter set.

3.3 Nomenclature and Data Distribution

The DFW dataset is available for download as an archived file containing one folder for each subject. Each of the 1,000 folders is named with the subject's name and may contain the four types of images discussed above: normal, validation, disguise, and impersonator. In order to ensure consistency and eliminate ambiguity, the following nomenclature has been followed across the dataset:

- Each subject has a single normal face image, which has been named as *firstName_lastName.jpg*. For instance, for the subject Alicia Keys, the subject's normal image is named *Alicia_Keys.jpg*.
- As mentioned previously, a given subject contains only a single validation face image. Therefore, the validation image is named with a postfix '_a', that is, *firstName_lastName_a.jpg*. For the example of Alicia Keys, the subject validation image is stored as *Alicia_Keys_a.jpg*.
- For disguised face images, a postfix of '_h' is adopted, along with a number for uniquely identifying the disguised face image of a given subject. That is, *firstName_lastName_h_number.jpg*. Here, *number* can take values such as '001', '002', ... '010'. For example, the first disguise image of subject Alicia Keys can be named as *Alicia_Keys_h_001.jpg*, while the third disguised face image can be named as *Alicia_Keys_h_003.jpg*.
- Similar to the disguised image nomenclature, a postfix of '_I' is used to store the impersonator images of a subject. That is, impersonator images are named as *firstName_lastName_I_number.jpg*. For example, the first impersonator image of subject Alicia Keys can be named as *Alicia_Keys_I_001.jpg*.

In order to correctly follow the protocols mentioned above, and report corresponding accuracies, training and testing mask matrices are also provided along with the dataset. Given the entire training or testing partition, the mask matrix can be used to extract relevant genuine and imposter pairs or scores for a given protocol. The DFW dataset also contains face co-ordinates obtained via faster RCNN [33]. Given an image of the dataset, the co-ordinates provide the face location in the entire image.

TABLE 4: List of teams which participated in the DFW competition.

Model	Affiliation	Brief Description
AEFRL [28]	The Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO), Russia	MTCNN + 4 networks for feature extraction + Cosine distance
ByteFace	Bytedance Inc., China	Weighted linear combination of ensemble of 3 CNNs
DDRNET [29]	West Virginia University, USA	Inception Network with Center Loss
DisguiseNet [30]	Indian Institute of Technology Ropar, India	Siamese network with VGG-Face having a weighted loss
DR-GAN	Michigan State University, USA	MTCNN + DR-GAN + Cosine distance
LearnedSiamese	Computer Vision Center UAB, Spain	Cropped faces + Siamese Neural Network
MEDC	Northeastern University, USA	MTCNN + Ensemble of 3 CNNs + Average Cosine distance
MiRA-Face [31]	National Taiwan University, Taiwan	MTCNN + RSA + Ensemble of CNNs
OcclusionFace	Zhejiang University, China	MTCNN + Fine-tuned ResNet-28
Tessellation	Tessellate Imaging, India	Siamese network with triplet loss model
UMDNets [32]	The University of Maryland, USA	All-In-One + Average across scores obtained by 2 networks
WVU_CVL	West Virginia University, USA	MTCNN + CNN + Softmax

4 Disguised Faces in the Wild Competition

Disguised Faces in the Wild competition was conducted as part of the *First International Workshop on Disguised Faces in the Wild*[3], at the International Conference on Computer Vision and Pattern Recognition, 2018 (CVPR'18). Participants were required to develop a disguised face recognition algorithm, which was evaluated on all three protocols of the DFW dataset. The competition was open world-wide, to both industry and academic institutions. The competition saw over 100 registrations from across the world.

All participating teams were provided with the DFW dataset, including the training and testing splits, face co-ordinates, and mask matrices for generating the genuine and imposter pairs. Evaluation was performed based on the three protocols described in Section 3.2. No restriction was enforced in terms of utilizing external training data, except ensuring mutual exclusion with the test set. The remainder of this section presents the technique and performance analysis of all the submissions, including the baseline results.

4.1 Baseline Results

Baseline results are computed using the VGG-Face descriptor [34], which is one of the top performing deep learning models for face recognition. A pre-trained VGG-Face model is used for feature extraction (trained on the VGG-Face dataset [34]). Baseline results were also provided to the participants. Baseline results have also been computed with the ResNet-50 architecture trained on the MS-Celeb-1M and VGGFace2 datasets [26]. The extracted features are compared using Cosine distance, followed by classification into genuine or imposter. Both the models achieve high recognition performance on challenging face datasets.

4.2 DFW Competition: Submissions

The DFW competition received 12 submissions from all over the world, having both industry and academic affiliations. Table 4 presents the list of the participating teams, along with their affiliation. Details regarding the technique applied by each submission are provided below:

(i) Appearance Embeddings for Face Representation Learning (AEFRL) [28]: AEFRL is a submission from the Information Technologies, Mechanics and Optics (ITMO), Russian Federation. Later in the competition, it was renamed to Hard Example Mining with Auxiliary Embeddings. Faces are detected, aligned, and cropped using Multi-task Cascaded Convolutional Networks (MTCNN) [35]. This is followed by horizontal flipping, and feature extraction by four separate networks. Feature-level fusion is performed by concatenation of features obtained for the original and flipped image, followed by concatenation of all features from different networks. l_2 normalization is performed on the concatenated feature vector, followed by classification using Cosine distance. The CNN architecture used in the proposed model is given in Fig. 3(a).

(ii) ByteFace: Proposed by a team from Bytedance Inc., China, ByteFace uses an ensemble of three CNNs for performing disguised face recognition. For detection and alignment, the algorithm uses a mixture of co-ordinates provided with the DFW dataset and MTCNN. Three CNNs are trained with (i) modified center loss and Cosine similarity [36], (ii) joint Bayesian similarity, and (iii) sphere face loss [37] with joint Bayesian similarity, respectively. A linear weighted combination of scores obtained via the three models is used for performing the final classification. The CASIA WebFace [38] dataset is also used for training the proposed model.

(iii) Deep Disguise Recognizer Network (DDRNET) [29]: A team from West Virginia University, USA presented the DDRNET model. The name of the model was later changed to Deep Disguise Recognizer by the authors. Faces are cropped using the co-ordinates provided with the dataset, which is followed by pre-processing via whitening. An Inception network [39] along with Center loss [36] is trained on the pre-processed images, followed by classification using a similarity metric.

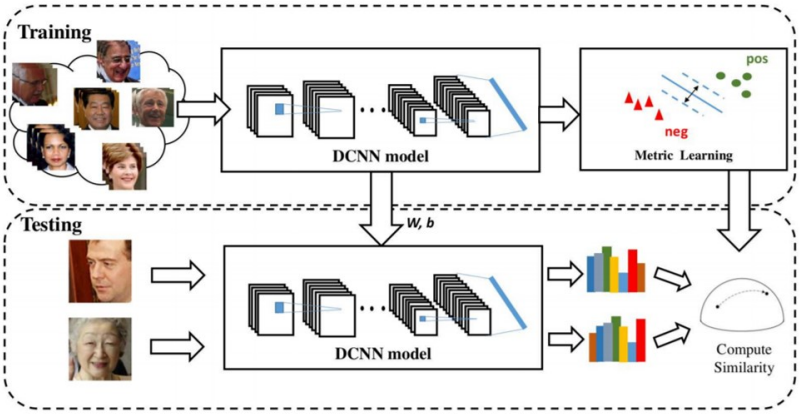
(iv) DisguiseNet (DN) [30]: Submitted by a team from the Indian Institute of Technology, Ropar, DisguiseNet performs face detection using the facial co-ordinates provided with the dataset. A Siamese network is built using the pre-trained VGG-Face [34], which is fine-tuned with the DFW dataset. Cosine distance is applied for performing classification of the learned features.

(v) DR-GAN: Proposed by a team from Michigan State University, USA, the framework performs face detection and alignment on the input images using MT-CNN [35]. This is followed by feature extraction using the Disentangled Representation learning-Generative Adversarial Network (DR-GAN) [40]. Classification is performed using Cosine distance.

3. http://iab-rubric.org/DFW/dfw.html

Description	Output
input image	$255 \times 255 \times 3$
$5 \times 5 \times 32$ Conv, stride 2	$126 \times 126 \times 32$
$3 \times 3 \times 64$ Conv	$124 \times 124 \times 64$
2×2 MaxPool, stride 2	$62 \times 62 \times 64$
$3 \times 3 \times 64$ ResBlock	$62 \times 62 \times 64$
$3 \times 3 \times (96 + 64 + 32)$ Conv, d=1,2,3	$60 \times 60 \times 192$
2×2 MaxPool, stride 2	$30 \times 30 \times 192$
$(3 \times 3 \times 192$ SE-ResBlock, r=16$)\times 2$	$30 \times 30 \times 192$
$3 \times 3 \times (336 + 112)$ Conv, d=1,2	$28 \times 28 \times 448$
2×2 MaxPool, stride 2	$14 \times 14 \times 448$
$(3 \times 3 \times 448$ SE-ResBlock, r=16$)\times 5$	$14 \times 14 \times 448$
$3 \times 3 \times 1024$ Conv	$12 \times 12 \times 1024$
2×2 MaxPool, stride 2	$6 \times 6 \times 1024$
$(3 \times 3 \times 1024$ SE-ResBlock, r=16$)\times 5$	$6 \times 6 \times 1024$
$512 + 512$, fc + maxout, group = 2	512
L_2-normalization	512

(a) AEFRL



(b) UMDNets

Fig. 3: Diagrammatic representation of (a) AEFRL [28], and (b) UMDNets [32]. Images have been taken from their respective publications.

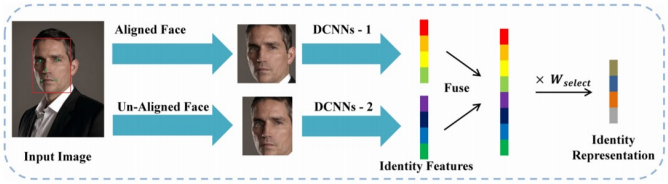


Fig. 4: Diagrammatic representation of MiRA-Face [31]. Image has directly been taken from their publication.

(vi) LearnedSiamese (LS): A team from the Computer Vision Center, Universitat Autnoma de Barcelona, Spain proposed LearnedSiamese. Facial co-ordinates provided with the dataset are used for performing face detection, followed by learning a Siamese Neural Network for disguised face recognition.

(vii) Model Ensemble with Different CNNs (MEDC): MEDC is proposed by a team from the Northeastern University, USA. Face detection is performed using MTCNN followed by 2-D alignment. An ensemble of three CNNs is used for performing the given task of disguised face recognition. The algorithm utilizes a Center face model [36], Sphere face model [37], and a ResNet-18 model [41] trained on the MS-Celeb-1M dataset [42]. Since MS-Celeb-1M dataset also contains images taken from the Internet, mutual exclusion is ensured with the test set of the DFW dataset. Classification is performed using Cosine distance for each network, the average of which is used for computing the final result.

(viii) MiRA-Face [31]: Submitted by a team from the National Taiwan University, MiRA-Face uses a combination of two CNNs for performing disguised face recognition. It treats aligned and unaligned images separately, thereby using a context-switching technique for a given input image. Images are aligned using the co-ordinates provided with the dataset along with MTCNN and Recurrent Scale Approximation (RSA) [43]. Features learned by the CNNs are directly used for classification. Fig. 4 presents a diagrammatic representation of the proposed model.

(ix) OcclusionFace: A team from ZJU, China proposed the OcclusionFace framework. MT-CNN [35] is used to perform face landmark detection and alignment based on five facial landmarks. ResNet-28 [41] is used for performing classification. The model is first pre-trained on the CASIA Webface dataset [38] followed by fine-tuning on the DFW dataset.

(x) Tessellation: Proposed by a team from Tessellate Imaging, India, Tessellation uses a Siamese network with triplet loss. Facial co-ordinates provided with the dataset are used for performing pre-processing, followed by training of the Siamese network. The final layer of the model learns a distance metric which returns a score between 0-1 for a given pair of images.

(xi) UMDNets [32]: Proposed by a team from University of Maryland, USA, its name was later modified to 'DCNN-based approach'. Face detection is performed by the All-in-One network [44], followed by alignment using the detected keypoints. Feature extraction is performed using two networks, followed by independent score computation. Classification is performed by averaging the scores obtained via the two feature sets. Fig. 3(b) presents the training and testing pipeline of the proposed model.

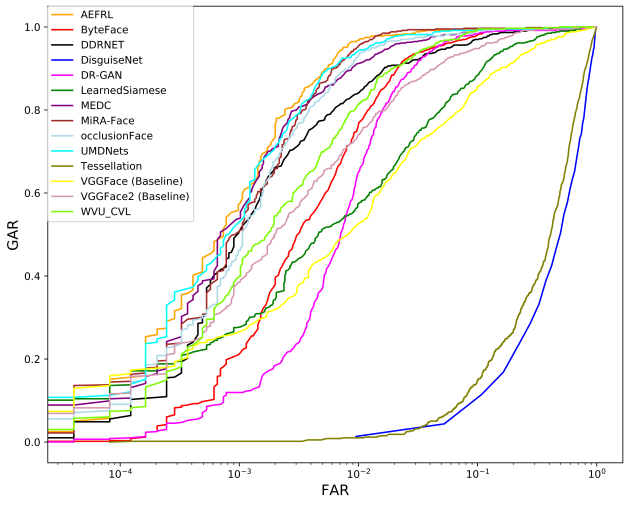
(xii) WVU_CL: Submitted by a team from West Virginia University, USA, WVU_CL uses the face co-ordinates provided with the dataset along with MT-CNN [35] for face alignment. The aligned images are provided to a CNN architecture for performing classification using a softmax classifier.

4.3 Results

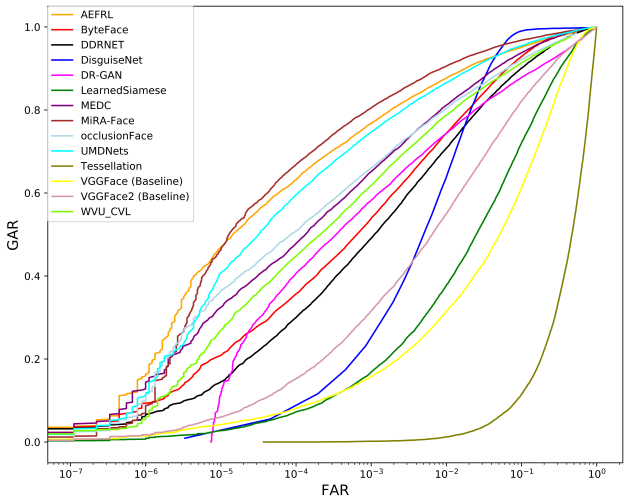
Tables 5-7 and Fig. 5 present the Receiver Operating Characteristic (ROC) curves of the above mentioned models for all three protocols. Along with the submissions, the performance of VGG-Face [34] with Cosine distance is also tabulated as baseline. The performance of each model is reported in terms of Genuine Acceptance Rate (GAR) at 1% False Acceptance Rate (FAR) and 0.1% FAR. Results for each protocol are given in detail below:

Results on Protocol-1 (Impersonation): Fig. 5(a) presents the ROC curves for all the submissions, and Table 5 presents the GAR corresponding to two FAR values. It can be observed that for the task of impersonation, AEFRL outperforms other algorithms at both the FARs by achieving 96.80% and 57.64% at 1% FAR and

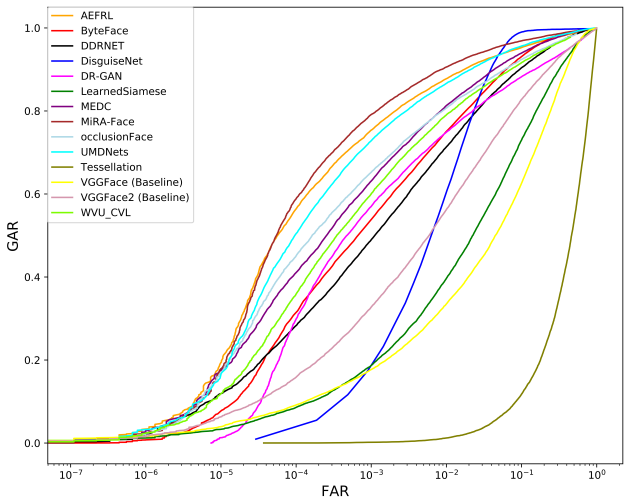
4. Not part of DFW competition
5. GAR@0.95%FAR
6. The smallest FAR value is 0.95%FAR for DisguiseNet.



(a) Protocol-1



(b) Protocol-2



(c) Protocol-3

Fig. 5: ROC curves of all participants along with the baseline results on protocol-1 (impersonation), protocol-2 (obfuscation), and protocol-3 (overall) of the DFW dataset.

TABLE 5: Verification accuracy (%) of the participants and baseline performance on protocol-1 (impersonation).

Algorithm	GAR	
	@1%FAR	@0.1%FAR
AEFRL	**96.80**	**57.64**
ByteFace	75.53	**55.11**
DDRNET	84.20	51.26
DenseNet + COST [4]	92.10	**62.20**
DisguiseNet	1.34 [5]	1.34 [6]
DR-GAN	65.21	11.93
LearnedSiamese	57.64	27.73
MEDC	91.26	**55.46**
MiRA-Face	**95.46**	51.09
OcclusionFace	93.44	46.21
Tessellation	1.00	0.16
UMDNets	**94.28**	53.27
VGGFace (Baseline)	52.77	27.05
VGGFace2 (Baseline)	73.94	38.48
WVU_CL	81.34	40.00

0.1%FAR, respectively. A difference of around 40% is observed between the accuracies at both the FARs, which suggests that for scenarios having stricter authorized access, further improved performance is required. The second best performance is reported by MiRA-Face which presents a verification accuracy of 95.46% and 51.09%, respectively. At 0.1%FAR, MEDC performs second best and achieves an accuracy of 55.46%. All three algorithms utilize MT-CNNs for face detection and alignment before feature extraction and classification.

Results on Protocol-2 (Obfuscation): Fig. 5(b) presents the ROC curves for the obfuscation protocol, and Table 6 summarizes the verification accuracies for all the models, along with the baseline results. MiRA-Face achieves the best accuracy of 90.65% and 80.56% for the two FARs. It outperforms other algorithms by a margin of at least 2.8% for GAR@1%FAR and 2.5% for GAR@0.1%FAR. As compared to the previous protocol (impersonation), the difference in the verification accuracy at the two FARs is relatively less. While further improvement is required, however, this suggests that recognition systems suffer less in case of obfuscation, as compared to impersonation at stricter FARs.

Results on Protocol-3 (Overall): Table 7 presents the GAR values of all the submissions, and Fig. 5(c) presents the ROC curves for the third protocol. As with the previous protocol, MiRA-Face outperforms other algorithms by a margin of at least around 3%. An accuracy of 90.62% and 79.26% is reported by the model for 1% and 0.1%FAR.

Other than the DFW competition submissions, Suri *et al.* [45] proposed a novel COST (Color (CO), Shape (S), and Texture (T)) based framework for performing disguised face recognition. COST learns different dictionaries for Color, Shape, and Texture, which are used for feature extraction, along with the deep learning based model, DenseNet [46]. Final output is computed via classifier level fusion of the deep learning and dictionary learning models. The performance of the proposed DenseNet + COST algorithm has also been tabulated in Tables 5 - 7.

Figs. 6 - 7 demonstrate sample images of the DFW dataset correctly classified or misclassified by almost all the submissions. Fig. 6 presents False Positive and True Negative samples for protocol-1 (impersonation). Upon analyzing the False Positive samples, it

TABLE 6: Verification accuracy (%) of the participants and baseline performance on protocol-2 (obfuscation).

Algorithm	GAR	
	@1%FAR	@0.1%FAR
AEFRL	**87.82**	**77.06**
ByteFace	76.97	21.51
DenseNet + COST [4]	**87.10**	72.10
DDRNET	71.04	49.28
DisguiseNet	66.32	28.99
DR-GAN	74.56	58.31
LearnedSiamese	37.81	16.95
MEDC	81.25	65.14
MiRA-Face	**90.65**	**80.56**
OcclusionFace	80.45	66.05
Tessellation	1.23	0.18
UMDNets	**86.62**	**74.69**
VGGFace (Baseline)	31.52	15.72
VGGFace2 (Baseline)	54.86	31.55
WVU_CL	78.77	61.82

TABLE 7: Verification accuracy (%) of the participants and baseline performance on protocol-3 (overall).

Algorithm	GAR	
	@1%FAR	@0.1%FAR
AEFRL	**87.90**	**75.54**
ByteFace	75.53	54.16
DenseNet + COST [4]	**87.60**	71.50
DDRNET	71.43	49.08
DisguiseNet	60.89	23.25
DR-GAN	74.89	57.30
LearnedSiamese	39.73	18.79
MEDC	81.31	63.22
MiRA-Face	**90.62**	**79.26**
OcclusionFace	80.80	65.34
Tessellation	1.23	0.17
UMDNets	**86.75**	**72.90**
VGGFace (Baseline)	33.76	17.73
VGGFace2 (Baseline)	56.22	32.68
WVU_CL	79.04	60.13

can be observed that all pairs have similar lower face structure, which might result in algorithms incorrectly classifying them as the same subject. Moreover, external disguises such as the cowboy hat (first pair) might also contribute to the misclassification. For protocol-2 (obfuscation), Fig. 7 presents sample False Negative and True Positive pairs common across almost all submissions. It is interesting to observe that in the False Negative pairs, disguise results in modification of face structure and textural properties. Coupled with obfuscation of face and pose variations, the problem of disguised face recognition is rendered further challenging.

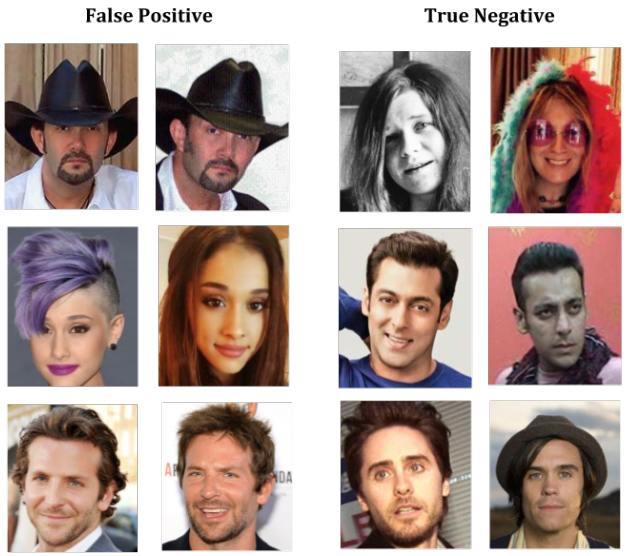


Fig. 6: Sample False Positive and True Negative pairs reported by a majority of submissions for protocol-1 (impersonation). False Positive refers to the case where an algorithm incorrectly classifies a pair as genuine, and True Negative refers to the case where two samples of different identities are correctly classified as imposters.

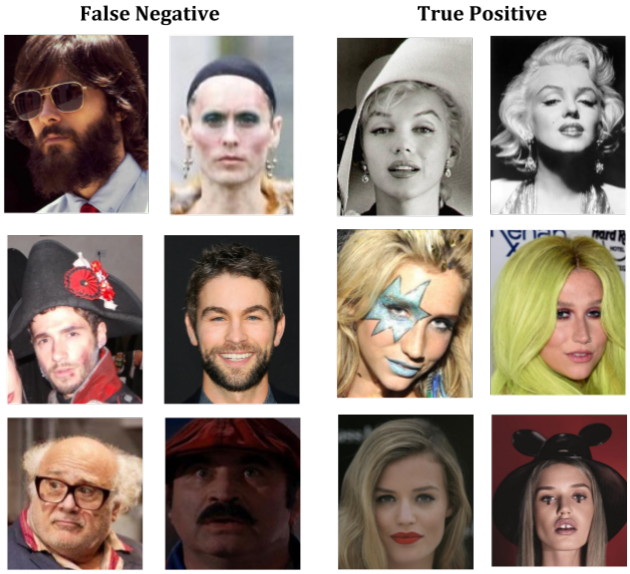


Fig. 7: Sample False Negative and True Positive pairs reported by a majority of submissions for protocol-2 (obfuscation). False Negative refers to the case where a pair of images are incorrectly classified as an imposter pair, while True Positive refers to the scenario where a pair of images are correctly classified as a genuine pair.

5 Degree of Difficulty: Easy, Medium, and Hard

In order to further analyze the DFW dataset, and study the problem of disguised faces in the wild, the DFW dataset has been partitioned into three sets: (i) easy, (ii) medium, and (iii) hard. The *easy* partition contains pairs of face images which are relatively easy to classify by a face recognition system, the *medium* set contains pairs of images which can be matched correctly by a majority of face recognition systems, while the *hard* partition contains image pairs with high matching difficulty. In literature, a similar partitioning was performed for the Good, Bad, and Ugly (GBU) face recognition challenge [47], where a subset of FRVT 2006 competition data [48] was divided into the three sets. The GBU challenge contained data captured over an academic year, in constrained settings with frontal face images having minimal pose or appearance variations. This section analyzes the DFW dataset containing data captured in unconstrained scenarios with variations across disguise, pose, illumination, age, and acquisition

(a) *Easy* Genuine Samples

(b) *Hard* Genuine Samples

(c) *Hard* Imposter Samples

Fig. 8: Sample *easy* and *hard* pairs of the DFW dataset.

TABLE 8: Number of *easy*, *medium*, and *hard* pairs for 1% and 0.1% FAR. TP and TN refer to True Positive and True Negative, respectively.

FAR	Number of								
	Easy			Medium			Hard		
	Genuine (TP)	Imposter (TN)	Total	Genuine (TP)	Imposter (TN)	Total	Genuine (TP)	Imposter (TN)	Total
1%	11,544	8,878,599	8,890,143	789	106,398	107,187	1,564	67,435	68,999
0.1%	9,461	9,034,109	9,043,570	1,138	11,534	12,672	3,298	6,789	10,087

device.

The top-3 performing algorithms of the DFW competition have been used for partitioning the dataset, that is, AERFL, MiRA-Face, and UMDNets. The performance of the three algorithms is used for dividing the test set of the DFW dataset into three components: (i) easy, (ii) medium, and (iii) hard. *Easy* samples correspond to those pairs which were correctly classified by all three algorithms, and are thus easy to classify. *Medium* samples were correctly classified by any two algorithms, while the *hard* samples were correctly classified by only one algorithm, or misclassified by all the algorithms, and thus are the most challenging component of the dataset. It is ensured that the partitions are disjoint, and samples belonging to one category do not appear in another category.

Table 8 presents the number of easy, medium, and hard pairs at different False Accept Rates of 1% and 0.1%. At 1%FAR, 11,544 genuine pairs are correctly classified as True Positive, while 8,878,599 imposter pairs are correctly classified as True Negative by all three techniques. This results in a total of 8,890,143 *easy* pairs, signifying that the total number of *easy* samples are highly dominated by the imposter pairs. In comparison, at 0.1%FAR, the total number of *easy* pairs increase to 9,043,570. It is interesting to observe that this increase is primarily due to the increased number of *easy* imposters at the lower FAR. Since at lower FARs, more pairs are classified as imposters, it leads to an increased number of *easy* pairs. Intuitively, at a stricter threshold of 0.1%FAR, one would expect the number of *easy* genuine samples to reduce. This trend is observed in Table 8, where the number of genuine pairs reduces from 11,544 at 1%FAR to 9,461 at 0.1%FAR.

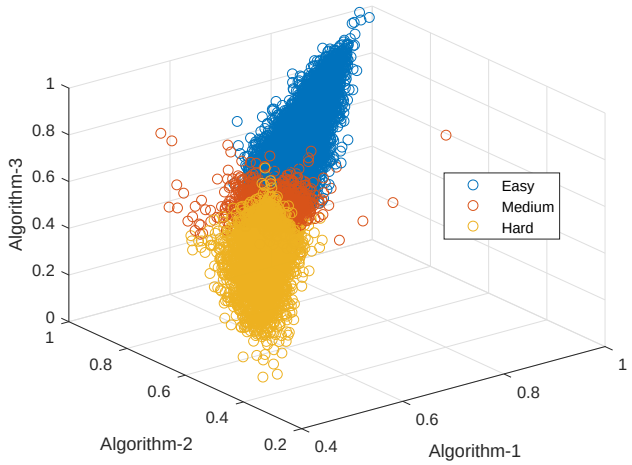


Fig. 9: Score distribution of the genuine pairs at 0.01% FAR, in terms of three levels of difficulty: easy, medium, and hard.

The opposite trend is observed for the *hard* partition, where the total number of *hard* pairs reduces at 0.1%FAR, as compared to 1%FAR, however, the number of genuine samples increases. The last three columns of Table 8 can be analyzed in order to observe this effect. At 1%FAR, the number of *hard* genuine pairs, that is, samples which are classified correctly by at most one algorithm is 1,564, while at 0.1%FAR it is 3,298. This implies that at a stricter FAR of 0.1%, more genuine samples were misclassified by all

three algorithms. However, the number of *hard* imposter samples drops from 67,435 to 6,789 at a lower FAR. A similar trend is observed for the *medium* partition, wherein a total of 107,187 and 12,672 samples were correctly classified by any two algorithms at 1% and 0.1%FAR, respectively.

Fig. 9 presents the score distribution of the genuine samples across the three categories of *easy*, *medium*, and *hard* at 0.1% FAR. The *easy* and *hard* samples occupy opposite ends of the distribution, while the *medium* category corresponds to a dense block between the two. Fig. 8 presents sample *easy* and *hard* pairs of the DFW dataset at 0.1% FAR. The first row corresponds to *easy* genuine pairs, that is, genuine pairs correctly classified by all three top performing algorithms. Most of these pairs contain images with no pose variations ((i)-(ii)) or *similar* pose variations across images of the pairs ((iii)-(iv)). It can also be observed that most of these pairs are of the normal and validation images of the dataset, with minimal or no disguise variations. Images which involve disguise in terms of hair variations or hair accessories with minimal change in the face region are also viewed as *easy* pairs by the algorithms. Since in such cases, the face region remains unchanged, algorithms are often able to correctly classify such samples with ease. This observation is further substantiated by the *hard* genuine samples (Fig. 8(b)). Most of the samples which were not correctly classified by any of the top algorithms contain occlusions in the face region. A large majority of genuine samples misclassified have occlusions near the eye region. All the pairs demonstrated in Fig. 8(b) have at least one sample with occluded eye region. Effect of occlusion can also be observed in the *hard* imposter samples (Fig. 8(c)), that is, imposters which were not correctly classified by either of the top-3 performing algorithms. Large variations due to heavy make-up, similar hair style or accessories, coupled with covariates of pose, occlusion, illumination, and acquisition device further make the problem challenging. It is our belief that in order to develop robust face recognition systems invariant to disguises, research must focus on addressing the *hard* pairs, while ensuring high performance on the *easy* pairs as well.

6 Conclusion and Future Work

This research presents the Disguised Faces in the Wild (DFW) dataset containing 11,157 images pertaining to 1,000 identities with variations across different disguise accessories. A given subject may contain four types of images: normal, validation, disguised, and impersonator. Out of these, normal and validation images are non-disguised frontal face images. Disguised images of a subject contain genuine images of the same subject with different disguises. Impersonator images correspond to images of different people who try to impersonate (intentionally or unintentionally) another subject. To the best of our knowledge, this is the first disguised face dataset to provide impersonator images for different subjects. Three evaluation protocols have been presented for the DFW dataset, along with the baseline results. The dataset has also been analyzed in terms of three degrees of difficulty: (i) easy, (ii) medium, and (iii) hard. The dataset was released as part of the DFW competition held in conjunction with the First International Workshop on DFW at CVPR'18. Details regarding the submissions and their performance evaluation have also been provided. It is interesting to observe that most of the submissions utilized traditional face recognition architectures, with limited *task-specific* inclusions in the entire pipeline, which might have resulted in lower verification performance at the stringent FAR of 0.1%. Dedicated research in the direction of disguised face recognition could help in the development of robust face recognition systems, imperative for several real world applications. It is our hope that the DFW dataset would help facilitate research in this important yet less explored domain of face recognition.

As future research directions, we plan to extend the database to include more challenging cases of obfuscation and impersonation. Additionally, DFW database may be extended to analyze the performance of presentation attack detection algorithms. As per the ISO standards (ISO/IEC CD 30107-1), presentation attack refers to the "presentation of an artifact or human characteristic to the biometric capture subsystem in a manner that could interfere with the intended policy of the biometric system". Given the above definition, presentation attacks could encompass variations due to intentional disguise and impersonation. Since the DFW dataset presents a separate set of impersonators, it may thus be used for evaluating and developing robust presentation attack detection algorithms as well.

7 Acknowledgement

Maneet Singh, Richa Singh, and Mayank Vatsa were partially supported through Infosys CAI at IIIT-Delhi. Mayank Vatsa was also partially supported by the SwarnaJayanti Fellowship from Government of India. The authors acknowledge V. Kushwaha for his help in data collection. Rama Chellappa was supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

[1] I. Masi, A. T. an Trãn, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in *European Conference on Computer Vision*, 2016.

[2] G. Goswami, M. Vatsa, and R. Singh, "Face verification via learned representation on feature-rich video frames," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1686–1698, 2017.

[3] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1139–1153, 2018.

[4] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041–2056, 2015.

[5] G. Goswami, N. K. Ratha, A. Agarwal, R. Singh, and M. Vatsa, "Unravelling robustness of deep learning based face recognition against adversarial attacks," in *AAAI Conference on Artificial Intelligence*, 2018.

[6] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[7] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[8] F. Juefei-Xu, D. K. Pal, K. Singh, and M. Savvides, "A preliminary investigation on the sensitivity of cots face recognition systems to forensic analyst-style face processing for occlusions," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 25–33.

[9] 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*, vol. 39, no. 1, pp. 156–171, 2017.

[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.

[11] S. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *International Conference on Computer Vision and Pattern Recognition*, 2017, pp. 86–94.

[12] S. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: A simple and accurate method to fool deep neural networks," in *International Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2574–2582.

[13] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: a comprehensive survey," *ACM Computing Surveys*, vol. 50, no. 1, 2017.

[14] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, "Disguised faces in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.

[15] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, "Recognizing disguised faces: Human and machine evaluation," *PLOS ONE*, vol. 9, no. 7, pp. 1–16, 2014.

[16] A. M. Martinez, "The AR face database," *CVC Technical Report*, 1998.

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

[18] 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.

[19] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna, "Using Kinect for face recognition under varying poses, expressions, illumination and disguise," in *IEEE Workshop on Applications of Computer Vision*, 2013, pp. 186–192.

[20] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *International Conference on Biometrics*, 2013.

[21] T. Y. Wang and A. Kumar, "Recognizing human faces under disguise and makeup," in *IEEE International Conference on Identity, Security and Behavior Analysis*, 2016.

[22] R. Raghavendra, N. Vetrekar, K. B. Raja, R. Gad, and C. Busch, "Detecting disguise attacks on multi-spectral face recognition through spectral signatures," in *International Conference on Pattern Recognition*, 2018.

[23] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep., 2007.

[24] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4873–4882.

[25] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 529–534.

[26] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2018, pp. 67–74.

[27] A. Nech and I. Kemelmacher-Shlizerman, "Level playing field for million scale face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3406–3415.

[28] E. Smirnov, A. Melnikov, A. Oleinik, E. Ivanova, I. Kalinovskiy, and E. Lukyanets, "Hard Example Mining with Auxiliary Embeddings," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[29] N. Kohli, D. Yadav, and A. Noore, "Face Verification with Disguise Variations via Deep Disguise Recognizer," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[30] S. V. Peri and A. Dhall, "DisguiseNet : A Contrastive Approach for Disguised Face Verification in the Wild," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[31] K. Zhang, Y.-L. Chang, and W. Hsu, "Deep Disguised Faces Recognition ," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[32] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa, "Deep Features for Recognizing Disguised Faces in the Wild," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[34] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[35] 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, 2016.

[36] 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.

[37] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[38] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014.

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich *et al.*, "Going deeper with convolutions," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2015.

[40] L. Tran, X. Yin, and X. Liu, "Representation learning by rotating your faces," *CoRR*, vol. abs/1705.11136, 2017.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[42] 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.

[43] Y. Liu, H. Li, J. Yan, F. Wei, X. Wang, and X. Tang, "Recurrent scale approximation for object detection in CNN," in *IEEE International Conference on Computer Vision*, 2017.

[44] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, "An all-in-one convolutional neural network for face analysis," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2017, pp. 17–24.

[45] S. Suri, A. Sankaran, M. Vatsa, and R. Singh, "On matching faces with alterations due to plastic surgery and disguise," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2018.

[46] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." in *International Conference on Computer Vision and Pattern Recognition*, 2017.

[47] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, "An introduction to the good, the bad, and the ugly face recognition challenge problem," in *Face and Gesture*, 2011, pp. 346–353.

[48] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831–846, 2010.

Maneet Singh received her Bachelors of Technology in Computer Science degree from the Indraprastha Institute of Information Technology - Delhi, India, in 2015 with an undergraduate thesis on "Recognizing Face Images with Weight and Age Variations". She is currently pursuing her PhD with the Image Analysis and Biometrics (IAB) Lab at the same Institute. Her research interests are machine learning and deep learning with applications in face recognition. Her research received the "Best Poster Award" in IEEE International Joint Conference on Biometrics 2017. She is also a recipient of the Google Anita Borg Memorial Scholarship, 2015, and has received the "Best Teaching Assistant Award" for two consecutive years in 2014 and 2015. She has authored over twenty publications including journals and peer-reviewed conferences. She has also served as a reviewer for several journals such as Pattern Recognition, Information Fusion, Computer Vision and Image Understanding, and IEEE Transactions on Information Forensics and Security.

Richa Singh received the Ph.D. degree in computer science from West Virginia University, Morgantown, USA, in 2008. She is currently an Associate Dean of Alumni and Communications, an Associate Professor with the IIIT-Delhi, India, and an Adjunct Associate Professor with West Virginia University. She has co-edited book Deep Learning in Biometrics and has delivered tutorials on deep learning and domain adaptation in ICCV 2017, AFGR 2017, and IJCNN 2017. Her areas of interest are pattern recognition, machine learning, and biometrics. She is a fellow of IAPR and a Senior Member of IEEE and ACM. She was a recipient of the Kusum and Mohandas Pai Faculty Research Fellowship at the IIIT-Delhi, the FAST Award by the Department of Science and Technology, India, and several best paper and best poster awards in international conferences. She has also served as the Program Co-Chair of BTAS 2016 and IWBF 2018, and a General Co-Chair of ISBA 2017. She is currently serving as a Program Co-Chair of AFGR 2019 and IJCB 2020. She is serving as the Vice President (Publications) of the IEEE Biometrics Council. She is an Editorial Board Member of Information Fusion (Elsevier), an Associate Editor of Pattern Recognition, Computer Vision and Image Understanding, and the EURASIP Journal on Image and Video Processing (Springer).

Mayank Vatsa received the M.S. and Ph.D. degrees in computer science from West Virginia University, USA, in 2005 and 2008, respectively. He is currently the Head of the Infosys Center for Artificial Intelligence, an Associate Professor with the IIIT-Delhi, India, and an Adjunct Associate Professor with West Virginia University, USA. He has co-edited a book Deep learning in Biometrics and co-authored over 250 research papers. His areas of interest are biometrics, image processing, machine learning, computer vision, and information fusion. He is a Senior Member of IEEE and ACM. He was a recipient of A. R. Krishnaswamy Faculty Research Fellowship at the IIIT-Delhi, the FAST Award Project by DST, India, and several Best Paper and Best Poster Awards at international conferences. He is also the recipient of the prestigious Swarnajayanti fellowship award from Government of India. He is an Area Chair of the Information Fusion (Elsevier), General Co-Chair of IJCB 2020, and the PC Co-Chair of the ICB 2013 and IJCB 2014. He has served as the Vice President (Publications) of the IEEE Biometrics Council where he started the IEEE Transactions on Biometrics, Behavior, And Identity Science.

Nalini K. Ratha is a Research Staff Member at IBM Thomas J. Watson Research Center, Yorktown Heights, NY. He received his B. Tech. in Electrical Engineering from Indian Institute of Technology, Kanpur, M.Tech. degree in Computer Science and Engineering also from Indian Institute of Technology, Kanpur and Ph. D. in Computer Science from Michigan State University. He has authored more than 100 research papers in the area of biometrics and has been co-chair of several leading biometrics conferences and served on the editorial boards of IEEE Trans. on PAMI, IEEE Trans. on SMC-B, IEEE Trans. on Image Processing and Pattern Recognition journal. He has co-authored a popular book on biometrics entitled Guide to Biometrics and also co-edited two books entitled Automatic Fingerprint Recognition Systems and Advances in Biometrics: Sensors, Algorithms and Systems. He has offered tutorials on biometrics technology at leading IEEE conferences and also teaches courses on biometrics and security. He is Fellow of IEEE, Fellow of IAPR and an ACM Distinguished Scientist. He has been an adjunct professor at IIIT Delhi, Cooper Union and NYU. During 2011-2012 he was the president of the IEEE Biometrics Council. At IBM, he has received several awards including a Research Division Award, Outstanding Innovation Award and Outstanding Technical Accomplishment Award along with several patent achievement awards. Recently he was designated as an IBM Master Inventor. His research interests include biometrics, pattern recognition and computer vision.

Rama Chellappa is a Distinguished University professor and a Minta Martin professor of engineering, and a professor in the ECE Department at the University of Maryland. He received the K. S. Fu Prize from the International Association of Pattern Recognition (IAPR). He is a recipient of the Society, Technical Achievement and Meritorious Service Awards from the IEEE Signal Processing Society and four IBM faculty Development Awards. He also received the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society. At UMD, he received college and university level recognitions for research, teaching, innovation and mentoring of undergraduate students. In 2010, he was recognized as an Outstanding ECE by Purdue University. He was also recognized as the Distinguished Alumnus of the Indian Institute of Science, Bengaluru, India. He served as the editor-in-chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence. He is a golden core member of the IEEE Computer Society, served as a Distinguished lecturer of the IEEE Signal Processing Society and as the president of the IEEE Biometrics Council. He is a fellow of the IEEE, IAPR, OSA, AAAS, ACM and the AAAI and holds six patents.