Disguised Faces in the Wild

Vineet Kushwaha, Maneet Singh, Richa Singh, Mayank Vatsa,
IIIT-Delhi, India
{maneets, rsingh, mayank}@iiitd.ac.in

Nalini Ratha
IBM TJ Watson Research Center, USA
ratha@us.ibm.com

Rama Chellappa
University of Maryland, College Park, USA
rama@umiacs.umd.ed

Abstract

Existing research in the field of face recognition with variations due to disguises focuses primarily on images captured in controlled settings. Limited research has been performed on images captured in unconstrained environments, primarily due to the lack of corresponding disguised face datasets. In order to overcome this limitation, this work presents a novel Disguised Faces in the Wild (DFW) dataset, consisting of over 11,000 images for understanding and pushing the current state-of-the-art for disguised face recognition. To the best of our knowledge, DFW is a first-of-a-kind dataset containing images pertaining to both obfuscation and impersonation for understanding the effect of disguise variations. A major portion of the dataset has been collected from the Internet, thereby encompassing a wide variety of disguise accessories and variations across other covariates. As part of CVPR2018, a competition and workshop are organized to facilitate research in this direction. This paper presents a description of the dataset, the baseline protocols and performance, along with the phase-I results of the competition.

1. Introduction

Research in face recognition has seen tremendous growth over the past few years. As a result, nowadays utility of face recognition spans from law-enforcement applications such as border control [1] to commercial applications including smartphone unlocking and face tagging in social media [2]. Face recognition algorithms, traditionally, are susceptible to covariates such as pose, illumination, expression, aging, heterogeneity, and disguise. Several of these covariates have been well explored for face recognition, however, face recognition with disguise variations has received limited attention [4, 5, 15, 19].

In real world scenarios, an individual might use disguise both intentionally or unintentionally, to obfuscate themselves or impersonate another person. For instance, facial hair, moustache, and hairstyle might naturally result in obfuscation. On the other hand, easily available disguise accessories such as wigs, hats, beard, moustache, and sunglasses, in conjunction with makeup variations, can help impersonating someone's identity.

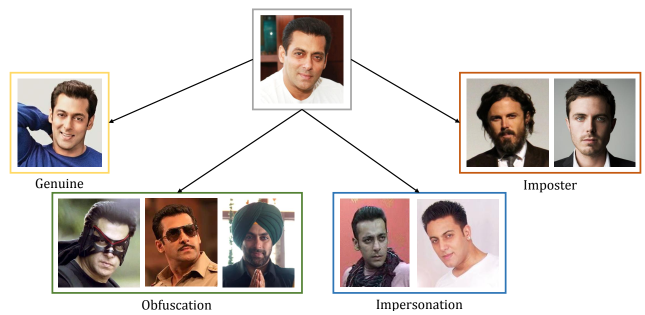

Figure 1: Sample *genuine*, *cross-subject impostor*, *impersonator*, and *obfuscated* face images for a single subject.

As shown in Figure 1, traditionally, an unaltered (normal) face is matched with another unaltered face of the same person, and state-of-the-art face recognition algorithms are able to achieve very high performance. The challenge of disguise introduces two variations: (i) when a person wants to obfuscate his/her own identity, and (ii) another individual impersonates someone else's identity. Obfuscation increases intra-class variations whereas impersonation reduces the inter-class dissimilarity, thereby affecting face recognition/verification task.

In literature, limited research has been undertaken to understand the impact of disguise variations on face recognition [5, 15, 18, 19]. Moreover, most of the existing techniques focus on solving disguise face recognition in controlled scenarios [5]. To fill this gap and to understand the impact of disguise variations on face recognition, this re-

Table 1: Review of existing disguise face datasets.

Name	Year	Controlled	Number of		Publicly Available
			Images	Subjects	
AR Dataset [12]	1998	Yes	3,200	126	Yes
National Geographic Dataset [15]	2004	Yes	46	1	No
Synthetic Disguise Dataset [19]	2009	Yes	4,000	100	No
IIIT-Delhi Disguise V1 Dataset [5]	2014	Yes	684	75	Yes
Disguised and Makeup Faces Dataset [22]	2016	No	2,460	410	Yes
Proposed DFW Dataset	**2018**	**No**	**11,157**	**1,000**	**Yes**

search proposes a novel Disguised Faces in the Wild (DFW) dataset of 1,000 subjects having over 11,000 images. Three benchmark experimental protocols along with baseline results have also been provided. This work also presents the Disguised Faces in the Wild Competition, an ongoing competition to facilitate research in this direction. Phase-I of the competition is complete, however, final results are still awaited.

2. Disguised Faces in the Wild Dataset

Recently, researchers have proposed large scale datasets captured in uncontrolled scenarios for performing face recognition [7, 8, 24]. However, none of these focus on the specific challenge of face recognition under the disguise covariate. Table 1 presents the list of publicly available disguise face datasets. It can be observed that most of the existing datasets in literature contain a controlled set of disguises, and a limited number of subjects. The Disguised Faces in the Wild (DFW) dataset has been prepared in order to address these limitations. The proposed DFW dataset consists of 11,157 images of 1,000 subjects. It includes IIIT-Delhi Disguise Version 1 Face Database (ID V1) [5] containing images of 75 subjects, while images pertaining to the remaining 925 subjects correspond to well known celebrities, taken from the Internet.

Figure 2 presents sample disguised face images of the dataset. Since the images have been taken from the Internet, the dataset contains a broad set of unconstrained disguised faces. The dataset encompasses disguise variations with respect to hairstyles, beard, moustache, glasses, make-up, caps, hats, turbans, veils, masquerades and ball masks. This is coupled with other variations with respect to pose, lighting, expression, background, ethnicity, age, gender, clothing, hairstyles, and camera quality, thereby making the dataset challenging for the task of face recognition. The dataset is made available to the research community[1]. Details regarding the dataset statistics along with experimental protocols are described in the following subsections.

[1]DFW dataset can be downloaded at: http://iab-rubric.org/resources/dfw.html

2.1. Dataset Statistics

The DFW dataset contains 11,157 face images of 1,000 identities, primarily of Indian or Caucasian origin. Each subject has at least five face images, and can have four types of images:

- **Normal Face Image:** Each subject has a single normal face image, which corresponds to a non-disguised frontal face image.
- **Validation Face Image:** 903 subjects have a single validation face image, which corresponds to a non-disguised frontal face image. This image can be used for generating a non-disguised pair within a subject.
- **Disguised Face Image:** All 1,000 subjects have disguised face images in the DFW dataset. For a given subject, disguised faces correspond to face images of the same subject having intentional or unintentional disguise. Each subject has at least 1 and at most 12 disguised face images.
- **Impersonator Face Image:** 874 subjects have images pertaining to their impersonators. An impersonator of a subject refers to an image of any other person (intentionally or unintentionally) pretending to be the subject's identity. Out of the 874 subjects, each subject has at least 1 and at most 21 number of impostor images.

Figure 2 presents the different types of images for three subjects of the DFW dataset. In total, the DFW dataset contains 1,000 normal face images, 903 validation face images, 4,814 disguised face images, and 4,440 impersonator images.

2.2. Protocol

Three verification protocols have been provided with the DFW dataset to understand and evaluate the effect of disguises on face recognition. Subject disjoint training and testing partitions are created in order to imitate real world scenarios. Images pertaining to 400 subjects form the training set, while the remaining 600 subjects constitute the test

Figure 2: Sample images of three subjects from the DFW dataset. Each row corresponds to one subject, containing the normal (gray), validation (yellow), disguised (green), and impersonated (blue) images. Best viewed in color.

Table 2: Number of images in the training and testing partition of the proposed 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

set. Table 2 presents the statistics of the training and testing partition. Details regarding the pre-defined protocols are given below:

- **Protocol-1 (Impersonation):** This is useful for evaluating the performance of a given technique under impersonation only. Pairs corresponding to (normal, validation) images of a given subject form the genuine pairs. Pairs created with the impersonator images of the normal, validation, and disguised images of the same subject form the impostor pairs. No other pairs are considered for this protocol.

- **Protocol-2 (Obfuscation):** In cases where a technique needs to be evaluated for disguises via obfuscation only, this protocol can be used. In this case, pairs corresponding to the (normal, disguise), (validation, disguise), and (disguise$_1$, disguise$_2$) images of the same subject form the genuine set. Here, disguise$_n$ refers to the n^{th} disguised image of a given subject. For the impostor set, cross-subject pairs are generated, wherein the normal, validation, and disguised images of one subject are paired with the normal, validation, and disguised images of another subject. No impersonators are used in this protocol.

- **Protocol-3 (Overall Performance):** The third protocol corresponds to evaluating a given algorithm on the entire dataset. The genuine and impostor pairs for this protocol are created by combining the pairs created in the above two protocols. Genuine pairs consist of (normal, validation), (normal, disguise), (validation, disguise), as well as (disguise$_1$, disguise$_2$) pairs. The impostor pairs are created using the impersonator images with the normal, validation, and disguised images of the same subject, along with cross-subject impostor pairs explained above.

2.3. Nomenclature and Data Distribution

As shown in Figure 2, each subject may have four types of images: normal, validation, disguise, and impersonator. The following nomenclature has been followed for the entire dataset:

- Subject normal images are named as *subjectName.jpg*. For example: *Anna_Nicole.jpg*.

- Subject validation images are named with a postfix of '_a' to form *subjectName_a.jpg*. That is, the validation image of subject Anna Nicole is named as: *Anna_Nicole_a.jpg*.

- Disguised face images are named with a postfix of '_h' as *subjectName_h_number.jpg*. Here, *number* takes value as '001', '002', ... '010', depending upon the number of disguise images for a given subject. For example, the first disguise image of subject Anna Nicole is named as *Anna_Nicole_h_001.jpg*.

- Impersonator images are named with a postfix of '_I' as *subjectName_I_number.jpg*. For example, the first impersonator image of subject Anna Nicole is named as *Anna_Nicole_I_001.jpg*.

The DFW dataset is available as an archived file containing 1,000 folders, one for each subject. Since the images are downloaded from the Internet, in some cases, one image might contain multiple faces as well. To address this issue, face coordinates obtained via Faster RCNN [17] are also provided with the dataset for both training and testing partitions. The face coordinates give the location of the face in the entire image. As part of the dataset, mask matrices are also be provided for all three protocols. A mask matrix can be used for obtaining the genuine/impostor pairs or extracting the relevant scores for a given protocol.

3. Disguised Faces in the Wild Competition

Disguised Faces in the Wild competition is an ongoing competition in conjunction with the *First International Workshop on Disguised Faces in the Wild*[2], at the International Conference on Computer Vision and Pattern Recognition, 2018 (CVPR'18). Participants are required to develop an algorithm addressing disguised face recognition, which will be evaluated using the proposed DFW dataset. The competition is open to all academic and industrial researchers, focused on addressing the given problem. We believe that the availability of the proposed Disguised Faces in the Wild dataset will help in facilitating research in this domain.

As part of the competition, the participants are provided with the entire dataset, with the training and testing splits. They are required to report results for the protocols mentioned in Section 2.2. External training data is allowed, while maintaining subject exclusivity between the training and testing sets. The competition is divided into two phases:

- **Phase-I:** The early stage presents the participants with an opportunity to submit their results and a written paper in the workshop describing their strategy.
- **Phase-II:** In the second phase, participants will have more time to present their results. Some participants may be invited to present their work orally at the CVPR workshop, based on their model's performance.

Participants may select to participate in either or both the phases. Phase-I of the competition is complete, the remainder of the paper presents the performance analysis and details of the submissions.

4. Disguised Faces in the Wild Competition: Phase-I

Nine teams submitted their results and models for phase-I of the DFW competition. Table 3 provides details regarding the teams and their affiliations. This phase saw submissions from all over the World, industry and academia alike. A brief description of each team and their corresponding proposed model is provided below:

[2] http://iab-rubric.org/DFW/dfw.html

Table 3: List of teams participating in Phase-I.

Team	Institution
AEFRL [20]	The Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO)
ByteFace	Bytedance Inc.
DDRNET [9]	West Virginia University
DisguiseNet [14]	Indian Institute of Technology Ropar
LearnedSiamese	Computer Vision Center UAB
MEDC	Northeastern University
MiRA-Face [25]	National Taiwan University
Resnet [3]	The University of Maryland
Tessellation	Tessellate Imaging

(i) Appearance Embeddings for Face Representation Learning (AEFRL) [20]: AEFRL is proposed by a team from the Information Technologies, Mechanics and Optics (ITMO), Russian Federation. It uses MTCNNs [26] for performing face detection, followed by horizontal flipping. An ensemble of five networks is used to obtain features for the original and flipped image. The concatenation of these features is used for performing classification using Cosine similarity.

(ii) ByteFace: The model is presented by a team from Bytedance Inc., China. It consists of an ensemble of three models, scores of each are combined in a linear weighted manner. The three models are built over center loss [23], CNNs and joint Bayesian similarity, and sphereface loss [10]. For different models, either the facial coordinates provided with the dataset are used, or MTCNN [26] is used for performing face detection.

(iii) Deep Disguise Recognizer Network (DDRNET) [9]: Proposed by a team from the West Virginia University, the technique (name updated to 'Deep Disguise Recognizer' by the authors) performs pre-processing in the form of face cropping and whitening. The pre-processed data is provided to an Inception network [21] with Center Loss [23], followed by classification via a similarity metric.

(iv) DisguiseNet [14]: Proposed by the Indian Institute of Technology, Ropar, the model utilizes a pre-trained VGG-Face [13] for the given task. A Siamese-based approach is used along with Cosine similarity for classification.

(v) LearnedSiamese: LearnedSiamese is proposed by

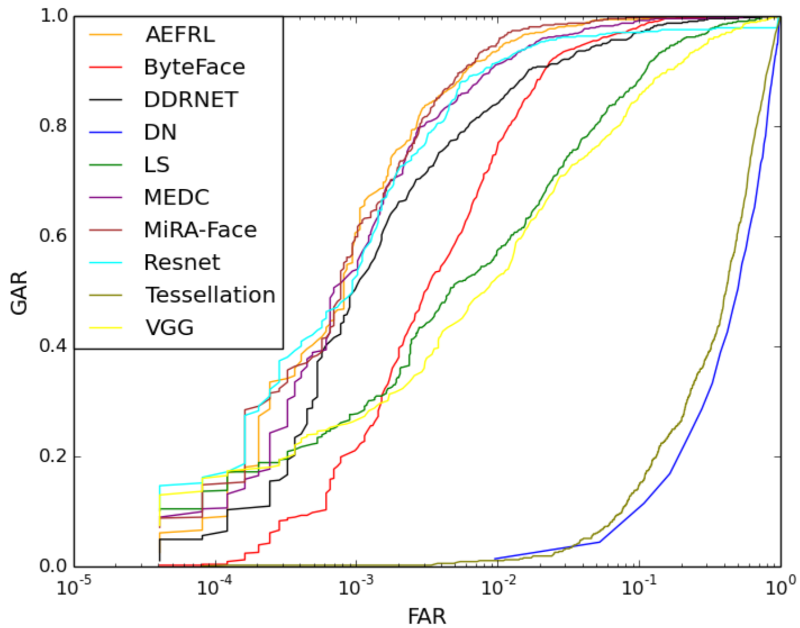

Figure 3: ROC curves obtained on the proposed DFW dataset for protocol-1 (impersonation) by phase-I submissions. 'LS' is LearnedSiamese, 'DN' is DisguiseNet, 'MEDC' is Model Ensemble with Different CNNs and 'AEFRL' is Appearance Embeddings for Face Representation Learning.

the Computer Vision Center, Universitat Autnoma de Barcelona, Spain. Cropped faces are provided as input to a Siamese Neural Network for addressing the given problem.

(vi) Model Ensemble with Different CNNs (MEDC): Proposed by the Northeastern University, USA, the algorithm uses an ensemble of three models for the given task. Multi-task Cascaded Convolutional Networks (MTCNN) [26] is used to detect facial points, which is followed by 2D-alignment to pre-process the faces. Feature extraction is performed by three CNN models: center face model [23], sphereface model [10], and Resnet-18 [6] trained with MS-Celeb-1M. It is ensured that the identities of MS-Celeb-1M do not overlap with the test set. Cosine distance is used to calculate the similarity, which is averaged across the models to obtain the final result.

(vii) MiRA-Face [25]: MiRA-Face is proposed by a team from the National Taiwan University. It consists of two models, built over Convolutional Neural Networks (CNNs), one for aligned input and the other for unaligned input. Alignment is performed by landmark detection and similarity transformation using Multi-task Cascaded Convolutional Networks (MTCNN) [26] and Recurrent Scale Approximation (RSA) [11]. Features learned by the CNN are then used as representations for performing classification.

(viii) Resnet [3]: This model (name updated to 'DCNN-based approach' by the authors) is presented by a team from the University of Maryland. Face images are detected using the All-in-One network [16], which is followed by alignment to a canonical view. The aligned images are used for feature extraction, followed by Cosine similarity based classification.

(ix) Tessellation: A team from Tessellate Imaging, India has proposed the Tessellation model. Nine image channels of the input image are used by a pre-trained model and a Siamese network based triplet loss model. The final layer returns a distance metric between 0-1.

4.1. Results

Results have been computed for the above mentioned submissions for all the protocols, that is, obfuscation (protocol-1), impersonation (protocol-2), and overall (protocol-3). Baseline results are computed with pre-trained VGG-Face descriptor [13] and Cosine similarity. VGG-Face model is used for feature extraction of the test set, which are then compared using Cosine similarity. Details regarding each protocol and corresponding results are provided below:

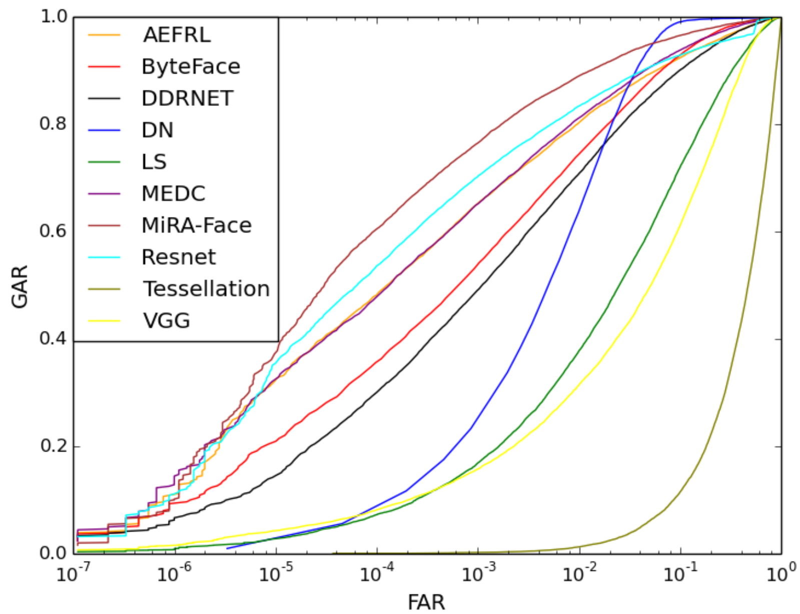

Figure 4: ROC curves obtained on the proposed DFW dataset for protocol-2 (obfuscation) by phase-I submissions. 'LS' is LearnedSiamese, 'DN' is DisguiseNet, 'MEDC' is Model Ensemble with Different CNNs and 'AEFRL' is Appearance Embeddings for Face Representation Learning.

Table 4: Verification accuracy (%) obtained on the proposed DFW dataset for Protocol-1 (Impersonation). The table summarizes the performance of participants and baseline results using VGG-Face.

Algorithm	**GAR**	
	@1%FAR	**@0.1%FAR**
AEFRL	93.78	**61.17**
Baseline (VGG-Face)	52.77	27.05
ByteFace	75.53	55.11
DDRNET	84.20	51.26
DisguiseNet	1.34 [3]	1.34 [4]
LearnedSiamese	57.64	27.73
MEDC	91.26	55.46
MiRA-Face	**95.12**	59.83
Resnet	91.59	53.78
Tessellation	1.00	0.16

Results on Protocol-1 (Impersonation): Figure 3 presents the Receiver Operating Characteristics (ROC) curves for protocol-1 (impersonation) for all submissions, along with the baseline results (VGG-Face) [13]. Table 4 presents the Genuine Acceptance Rate (GAR) corresponding to two values of False Acceptance Rate (FAR): 1% FAR and 0.1% FAR. It can be observed that the performance of MiRA-Face algorithm from National Taiwan University outperforms other techniques, at 1%FAR, by achieving an accuracy of 95.12%. On a stricter and lower FAR of 0.1%, AEFRL outperforms other techniques by achieving a performance of 61.71%, which is at least 1.3% better than other reported results. MiRA-Face algorithm follows closely with a performance of 59.83%.

[3]GAR@0.95%FAR

[4]The smallest FAR value is 0.95%FAR for DisguiseNet.

Results on Protocol-2 (Obfuscation): Figure 4 presents the ROC curves obtained on protocol-2, and Table 5 presents the accuracies for all the submissions (along with the baseline results). MiRA-Face achieves the best performance by reporting a GAR of 88.94% on 1% FAR. At a lower FAR of 0.1%, MiRA-Face continues to outperform other techniques by achieving an accuracy of 76.46%. It is interesting to observe that while the performance at 1%FAR is lower for protocol-2 as compared to protocol-1, however, the reverse happens at a stricter FAR of 0.1%. This shows that when the system accepts lesser false accepts, genuine accept rate under the case of impersonation suffers more as compared to obfuscation.

Results on Protocol-3 (Overall): Figure 5 presents the ROC curves of protocol-3, and the corresponding GAR values have been tabulated in Table 6. Consistent with previous protocols, it can be observed that MiRA-Face achieves

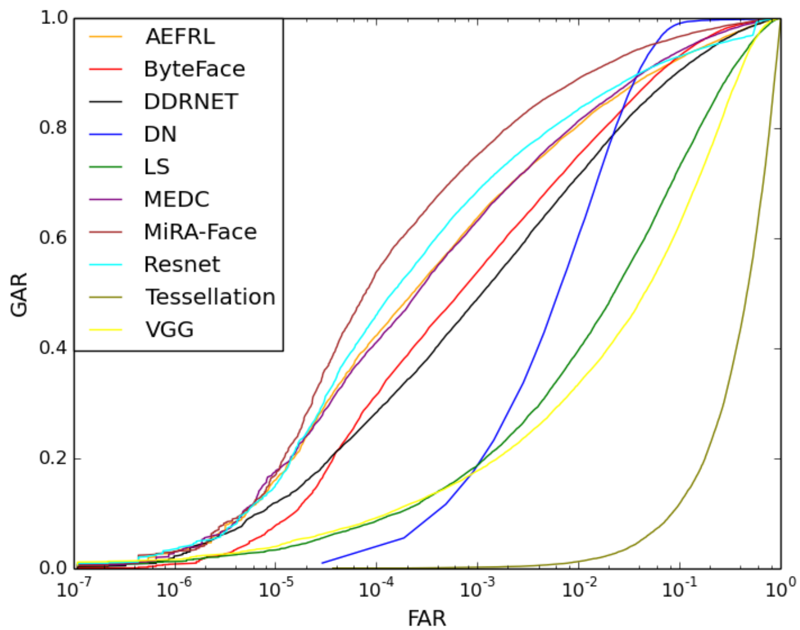

Figure 5: ROC curves obtained on the proposed DFW dataset for protocol-3 (overall) by phase-I submissions. 'LS' is LearnedSiamese, 'DN' is DisguiseNet, 'MEDC' is Model Ensemble with Different CNNs and 'AEFRL' is Appearance Embeddings for Face Representation Learning.

Table 5: Verification accuracy (%) obtained on the proposed DFW dataset for Protocol-2 (Obfuscation). The table summarizes the performance of participants and baseline results using VGG-Face.

Algorithm	GAR	
	@1%FAR	**@0.1%FAR**
AEFRL	80.37	65.23
Baseline (VGG-Face)	31.52	15.72
ByteFace	76.97	21.51
DDRNET	71.04	49.28
DisguiseNet	66.32	28.99
LearnedSiamese	37.81	16.95
MEDC	81.25	65.14
MiRA-Face	**88.94**	**76.46**
Resnet	83.37	70.42
Tessellation	1.23	0.18

Table 6: Verification accuracy (%) obtained on the proposed DFW dataset for Protocol-3 (Overall). The table summarizes the performance of participants and baseline results using VGG-Face.

Algorithm	GAR	
	@1%FAR	**@0.1%FAR**
AEFRL	80.59	63.52
Baseline (VGG-Face)	33.76	17.73
ByteFace	75.53	54.16
DDRNET	71.43	49.08
DisguiseNet	60.89	23.25
LearnedSiamese	39.73	18.79
MEDC	81.31	63.22
MiRA-Face	**89.04**	**75.08**
Resnet	83.49	68.52
Tessellation	1.23	0.17

the best performance of 89.04% at 1%FAR, while achieving 75.08% at 0.1%FAR. MiRA-Face showcases an improvement of at most 5% as compared to other algorithms for both the cases.

Overall, it can be observed that all the submissions of phase-I were Deep Learning based algorithms, with a majority of them being CNN based. Figures 6 - 8 present sample images that have been correctly classified and misclassified by all the algorithms. While the proposed models are able to achieve a good performance, however, cases involving high level of obfuscation such as large eye masks or sunglasses still appear challenging (Figure 7). Obfuscation in the form of makeup or natural facial variations such as facial hair, moles, or scars further render the problem challenging. It is also interesting to observe that in case of impersonation (Figure 6) presence of similar accessories (cowboy hat) or

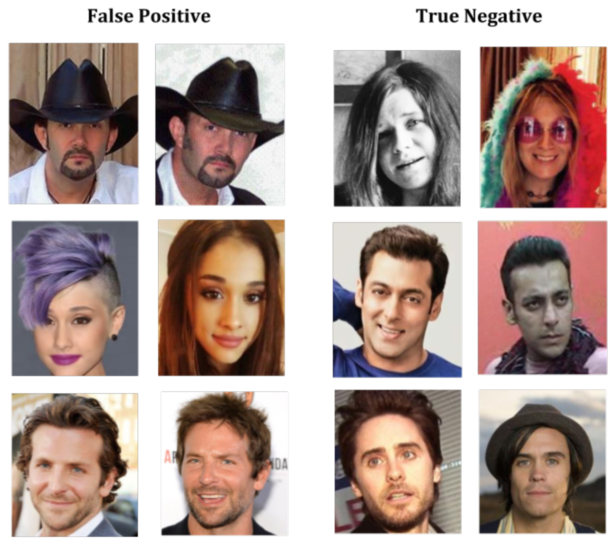

Figure 6: Sample pairs correctly classified and misclassified by almost all algorithms for protocol-1 (impersonation). A false positive pair is a pair that has been falsely classified as same identity by all algorithms. A true negative pair is a pair which has been correctly classified as the different identities by all algorithms.

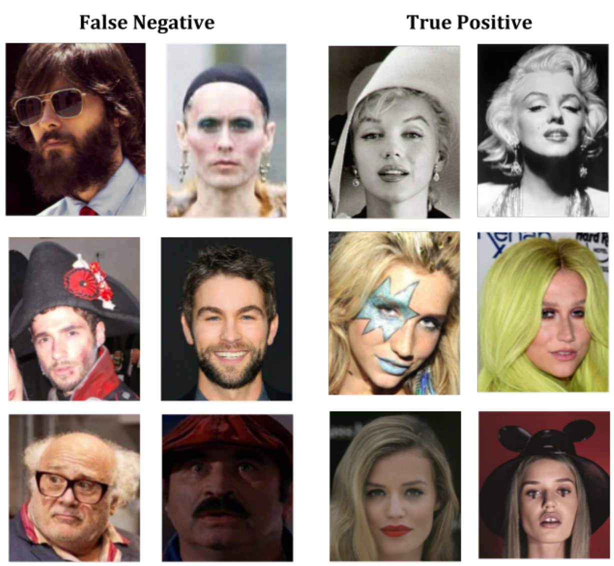

Figure 7: Sample pairs correctly classified and misclassified by all algorithms for protocol-2 (obfuscation). A false negative pair is a pair that has been falsely classified as different identities by all algorithms. A true positive pair is a pair which has been correctly classified as same identity by all algorithms.

similar hairstyle often results in false positives. The large gap between the performance at 1% FAR and 0.1% FAR

Figure 8: Sample pairs correctly classified and misclassified by all algorithms for protocol-3 (overall). A false positive pair is a pair that has been falsely classified as same identity by all algorithms. A true positive pair is a pair which has been correctly classified as the same identity by all algorithms.

further demands development of sophisticated algorithms for deployment in real world scenarios.

5. Conclusion

This paper presents the Disguised Faces in the Wild (DFW) dataset containing over 11,000 face images. The proposed dataset showcases challenging disguise variations such as headgear, face masks, and make-up. Three protocols and corresponding baseline results have been presented on the proposed dataset. This paper also presents the phase-I results of Disguised Faces in the Wild competition. The first phase saw nine submissions from across the World, and results with respect to all three protocols have been reported. We believe that the availability of Disguised Faces in the Wild dataset, will help the research community in advancing the current state-of-the-art on disguised face recognition.

References

[1] CBP Deploys Facial Recognition Biometric Technology at 1 TSA Checkpoint at JFK Airport, 11 Oct 2017. Available at https://www.cbp.gov/newsroom/national-media-release/cbp-deploys-facial-recognition-biometric-technology-1-tsa-checkpoint. 1

[2] Facebook can now find your face, even when it's not tagged, 19 Dec 2017. Available at: https://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/. 1

[3] 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. 4, 5

[4] 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. 1

[5] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLOS ONE*, 9(7):1–16, 2014. 1, 2

[6] 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*, pages 770–778, 2016. 5

[7] 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. Technical report, University of Massachusetts, Amherst, 2007. 2

[8] 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*, pages 4873–4882, 2016. 2

[9] 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. 4

[10] 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. 4, 5

[11] 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. 5

[12] A. M. Martinez. The AR face database. *CVC Technical Report*, 1998. 2

[13] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 4, 5, 6

[14] 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. 4

[15] N. Ramanathan, R. Chellappa, and A. R. Chowdhury. Facial similarity across age, disguise, illumination and pose. In *International Conference on Image Processing*, pages 1999–2002, 2004. 1, 2

[16] 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*, pages 17–24, 2017. 5

[17] 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*, pages 91–99, 2015. 4

[18] G. Righi, J. J. Peissig, and M. J. Tarr. Recognizing disguised faces. *Visual Cognition*, 20(2):143–169, 2012. 1

[19] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. *Image and Vision Computing*, 27(3):245–257, 2009. 1, 2

[20] 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. 4

[21] 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. 4

[22] T. Y. Wang and A. Kumar. Recognizing human faces under disguise and makeup. In *IEEE International Conference on Identity, Security and Behavior Analysis*, 2016. 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*, pages 499–515, 2016. 4, 5

[24] 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*, pages 529–534, 2011. 2

[25] K. Zhang, Y.-L. Chang, and W. Hsu. Deep Disguised Faces Recognition . In *CVPR Workshop on Disguised Faces in the Wild*, 2018. 4, 5

[26] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 4, 5