Deep Learning Approach for Face Recognition with Disguise Variations

By

Mohit Chawla

Under supervision of *Dr. Richa Singh* and *Dr. Mayank Vatsa*

Department of Computer science

Indraprastha Institute of Information Technology

July, 2019

Deep Learning Approach for Face Recognition with Disguise Variations

By

Mohit Chawla

Submitted

in partial fulfillment of the requirements for the degree of
Master of Technology
to

Indraprastha Institute of Information Technology
July, 2019

Certificate

This is to certify that the Thesis titled "**Deep Learning Approach for Face Recognition with Disguise Variations**" being submitted by **Mohit Chawla MT17028** to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree.

May, 2019

Dr. Richa Singh and Dr. Mayank Vatsa
Department of Computer Science
Indraprastha Institute of Information Technology Delhi
New Delhi 110020

Acknowledgements

I sincerely thank my advisors, **Dr. Richa Singh** and **Dr. Mayank Vatsa**, along with my mentor **Maneet Singh** who have supported me throughout with their patience and knowledge. They were always available whenever I needed them and always showed me the right direction to proceed in order to achieve the goal. I attribute the level of my Masters thesis to their encouragement and effort. One simply could not wish for a better or friendlier supervisors. Also, I must express my profound gratitude to my parents and to my friends for providing me with unfailing support and continuous encouragement throughout my study and through the process of researching and working on this project. This accomplishment would not have been possible without them.

Abstract

With the increased interest in face recognition across different applications, the research in this area has flourished over the past few decades. However, face recognition with disguise variations has gained little attention. Faces in unconstrained settings with disguise as an additional covariate makes this problem challenging. It includes alterations in facial appearance using disguise accessories. In this thesis, we propose deep learning based transfer learning approach to handle the problem of disguise, with the network being fine-tuned on the proposed loss function termed as "*Disguised Loss*". We have evaluated our network on Disguised Faces in Wild (DFW) 2018 dataset [1] where the proposed algorithm is able to produce competitive results. We have also introduced a new dataset termed as "DFW2019" which is an extension of DFW2018 dataset [1]. Apart from the addition of 250 subjects with 3140 images, 250 plastic surgery image pairs and 100 bridal image pairs have also been added. Additional protocols for plastic surgery face recognition have also been introduced. We have presented baselines for all the protocols along with the results of the proposed approach.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Face recognition has seen tremendous growth in research as well as from an application point of view. In the deep learning era, we have witnessed the great success of deep convolutional neural networks (DCNN) based state-of-the-art models such as Residual Networks (ResNet) [2] and OpenFace [3]. Even though these algorithms are becoming better at their task, still face recognition systems are vulnerable to covariates such as disguises.

Several covariates such as pose, illumination, expression, ageing and heterogeneity have been well explored for face recognition. However, face recognition with disguise variations has received limited attention. Here disguise denotes hiding own's identity or impersonating someone else's using different types of facial accessories (e.g. glasses, beard, different hairstyles, scarfs or caps) and makeup. There are intentional disguises as well as unintentional disguises. In intentional disguise, a subject tries to hide his/her identity using disguise with the purpose of fooling face recognition systems whereas, in case of unintentional disguise, a subject casually uses different types of disguises such as glasses and scarf and is able to fool the system unintentionally. Due to these variations, the inter-class distance between subjects reduces and intra-class variations increase, which

makes the problem more challenging.

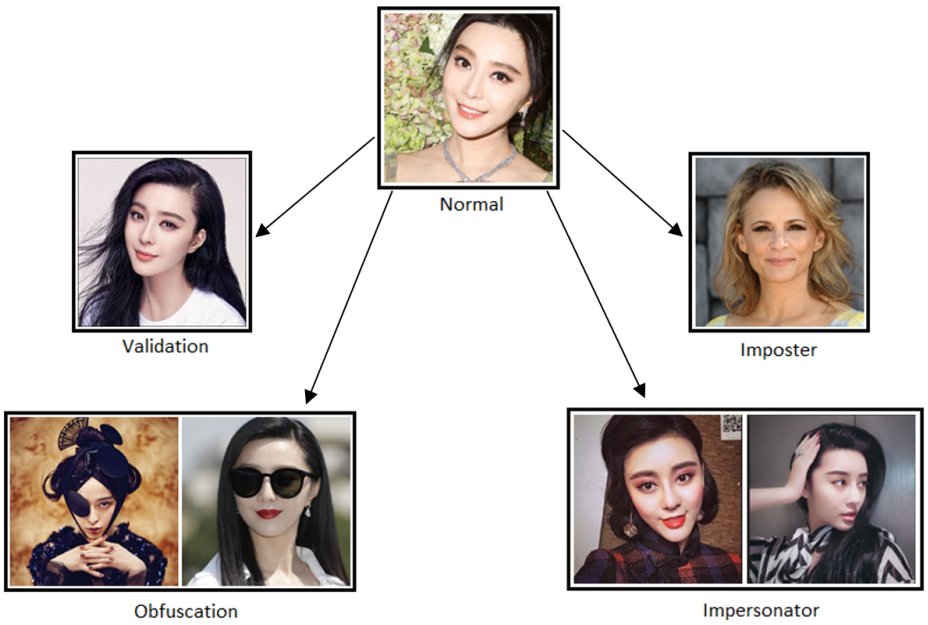


Figure 1.1: Normal, validation, impersonator, obfuscated and cross-subject imposter images for a subject.

Face recognition with disguises is still an open research problem because of the challenges it poses due to unconstrained environments and covariates. Therefore, development in this field would encourage effective applications of this technology. Some of these applications are:

- Access control in sensitive areas
- Identify people on social media
- Secure payment transaction
- Law enforcement protection such as border control
- Unlocking phones

1.1 Literature Review

Compared to research in traditional face recognition, there has been little work to address the effect of disguise variations on the recognition performance. Initial research in the problem domain focused on occlussion for example, Martinez et al. [4] divided each face image into r local regions and projected these regions on eigenspaces. Mahalanobis distance is then used to find the importance of these local regions. The combined local score of these regions is then used for face recognition under occlusion [4]. Yoon et al. [5] used support vector machine (SVM) for detection of partially occluded face. Ramanathan et al. [6] used principal component analysis (PCA) and matched disguised faces with the help of mahalanobis cosine distance [7]. Wright et al. [8] tried to find occlusion representation by utilizing the fact that it should be sparse in comparison to the whole image representation. They used L1 minimization for this purpose. Further, they tried to retrieve clean image representation from the whole image representation. They used dictionary learning in order to find the sparse representation. L2 distance is then used for classification. In [9], Singh et al. verified the faces with disguises accurately by the use of 2D log-polar gabor features. There have been different approaches to detect disguises that use PCA [10] and other texture descriptors [11] [12].

Dhamecha et al. [13] proposed a pipeline in which SVM is used to find patches that were biometric/non-biometric. It is done by the threshold score given by SVM, which is then used to classify the type of patch. Further, on biometric patches, local binary pattern (LBP) is used with eight sampling points. Distance between the probe image and gallery image is calculated using L2 distance. Smirnov et al. [14] proposed auxiliary embedding and used hard mining to make batch contain a hard example. It is done to make network learn important discriminating features while training. Every training example has an embedding

associated with it. The idea is to have more difference between two hard positive examples and less for two hard negatives. These embeddings are then used for mini-batch generation. Hard positives are those examples which have minimum cosine similarity but are similar to each other whereas, hard negatives are those examples which have maximum cosine similarity but are dissimilar to each other in mini-batch. These deep embedded features are then fed to Adaptive Rational Fraction Activation (ARFA) based ARFANet, which is their proposed architecture. Zhang et al. [15] proposed to use two DCNN for feature extraction for generic faces and then applied a transformation matrix to adapt these features for the disguised dataset. PCA is used to find the transformation matrix. This can be more clearly understood as the projection of identity features extracted by DCNN into principal components with the highest variance of disguised face dataset.

Bansal et al. [16] used DCNN based approach where in training phase two DCNN are trained on large datasets with loss function as L2-constrained softmax [17]. These two network features are then combined and used for recognizing disguised faces and impersonators. All-in-one CNN is used for face detection and alignment. Two ResNet architectures are used for learning feature with different parameters. Average of both scores obtained by the two CNNs is then used as the final score for each pair of images. Kohli et al. [18] used transfer learning by using Inception-Net [19] based features. They used centre loss [20], which tries to embed features of same class closer and different classes apart. The total loss is the sum of cross-entropy loss and weighted centre loss. The predicted similarity is computed using the cosine similarity formula. Suri et al. [21] used dictionary learning to transfer fundamental visual features, which are learnt from a generic image dataset. They also used transfer learning with the help of DenseNet [22] and fine-tuned it. They applied classifier fusion of visual features scores and

scores computed using fine-tuned model.

Last year, as part of Conference on Computer Vision and Pattern Recognition (CVPR) 2018 workshop and competition[1], the largest publicly available dataset on disguised faces "Disguised Faces in Wild (DFW) 2018" was released. This dataset contains variations in the form of facial features such as beard, moustache, hairstyles, also but not limited to apparel items like turban, hats, cap, glasses, masks and makeup. An example of images in the DFW2018 dataset can be seen in figure 1.2.

Figure 1.2: Example subject for the DFW2018 dataset.

The dataset is mainly collected with the help of the internet in an unconstrained environment. This, combined with above-mentioned variations, makes the dataset and the problem challenging for face recognition. There were 12 submissions for the DFW competition [1] from all over the world, which included both industry and academic affiliations. The results of top 5 algorithms for each protocol including ours has been mentioned in the tables 4.3, 4.4 and 4.5. A summary of the datasets used in various research works are provided in table 1.1

[1]http://cvpr2018.thecvf.com/program/workshops

[1]`http://vision.ucsd.edu/content/yale-face-database`

[2]Intensity and Texture Encode (ITE)

[3]Spatial convolutional neural network (SCNN)

Table 1.1: Summary of disguised face datasets in literature.

Dataset Name	Controlled Settings	Plastic Surgery Images	Bridal Make-Up Images	Subjects	Total Images	Availability of Impersonators
AR Dataset [23]	Yes	No	No	126	3200	No
National Geographic Dataset [6]	Yes	No	No	1	46	No
Curtin Faces Dataset [24]	Yes	No	No	52	5000	No
Disguised and Face Makeup Dataset [25]	No	No	No	410	2460	No
Spectral Disguise Face Dataset [26]	Yes	No	No	54	6480	No
Simple and Complex Face Disguise Datasets [27]	Yes	No	No	25	4000	No
Disguised Faces in Wild 2018 [1]	No	No	No	1000	11157	Yes
Disguised Faces in Wild 2019 (Addition)	No	Yes	Yes	600	3840	Yes

and algorithms used in literature have been summarized in table 1.2.

1.2 Research Motivation

Primary challenge to build algorithms for face recognition with disguise variations is the lack of corresponding datasets. Most of the datasets are prepared with controlled settings and thus fail to capture the real world scenario. Moreover, majority of the databases do not contain both intentional and unintentional unconstrained disguises which are usually encountered by a face recognition system. Also, many research works before 2014 used AR dataset [23], which was released in 1998 and only contains limited disguise (sunglass, scarves) in controlled settings. In year 2018, a new dataset was released called DFW2018 dataset [1]. This dataset offered a more realistic view where images are collected from unconstrained settings, containing disguised and impersonator images. Research in this area can be very helpful for law enforcement applications like one where a suspect's disguised face can be matched to existing database to know if he/she

Table 1.2: Summary of different algorithms used in literature.

Algorithm	Year	Approach	Similarity/ Distance Measure	Database
Martinez et al. [4]	2002	Probabilistic matching	Mahalnobis distance	AR [23]
Yoon et al. [5]	2002	PCA + SVM	-	AR [23]
Ramanathan et al. [6]	2004	PCA	Mahalnobis cosine distance	National Geographic [6]
Kim et al. [28]	2005	Independent component analysis (ICA)	L1, L2 and cosine distance	AR [23], FERET [29]
Kim et al. [10]	2005	PCA + SVM	-	AR [23]
Wright et al. [8]	2008	Sparse representation-based classification (SRC)	Euclidean distance	AR [23], Yale B[1]
Singh et al. [9]	2009	2D log polar Gabor	Hamming distance	AR [23]
Yang et al. [30]	2010	Gabor SRC	-	AR [23],Yale B[1]
Choi et al. [31]	2010	AdaBoost + modified census transform-based features	-	AR [23]
Min et al. [11]	2011	Gabor + PCA + SVM, LBP	Chi-square distance	AR [23]
Dhamecha et al. [12]	2013	ITE[2], LBP with disguise detection	L2 distance	I2BVSD [12]
Peng et al. [32]	2014	Dictionary learning + KNN detection	Cityblock distance metric	AR [23]
Wang et al. [25]	2016	LBP with disguise detection	Chi-Square dist. based similarity metric	Disguise and Makeup [25]
Singh et al. [27]	2017	Facial keypoint detection with SCNN[3]	L1 distance based on orientation	Simple and Complex Face Disguise [27]
Hung et al. [33]	2018	Disguise detection + CNN	-	IIIT-Delhi Disguise Version1 [13]
Smirnov et al. [14]	2018	Ensemble of 4 CNN + hard example mining	Cosine distance	DFW2018 [1]
Kohli et al. [18]	2018	Inception network + center loss	Cosine similarity	DFW2018 [1]
Zhang et al. [15]	2018	Ensemble of 2 CNN	Cosine similarity	DFW2018 [1]
Bansal et al. [16]	2018	Ensemble of 2 CNN	Cosine similarity	DFW2018 [1]
Suri et al. [21]	2019	Dictionary learning (color, shape and texture) + finetuned DenseNet	L2 distance	DFW2018 [1]

is a known criminal or to know his/her true identity. Therefore, in an attempt to make face recognition systems more robust and secure, we choose this as our research problem.

1.3 Research Contributions

Disguise being an important covariate, can hinder the performance of a face recognition system. Handling this covariate can help a face recognition system to be robust, usable and more secure. In order to tackle different aspects involved with disguise, exposure to different types of disguises can help the recognition systems to perform better. In this thesis, we present:

- a novel loss function termed as *Disguised Loss*.
- using this loss function, we have proposed an approach to address disguised face recognition problem.
- a new dataset termed as DFW2019, which is an extension of DFW2018 [1] dataset. It has an additional protocol for plastic surgery along with disguise variations based protocols. This dataset will further motivate researchers to study this challenging problem. Baselines for this dataset have also been reported using standard models for face recognition such as Residual Networks (ResNet-50) [2], Squeeze-and-Excitation (SE) Networks (SeNet-50) [34] and LCNN29 v2 [35].

Chapter 2

Proposed Algorithm

This chapter explains in detail the algorithm proposed for the given problem statement. A detailed overview of the proposed loss and the framework used is provided in the upcoming sections.

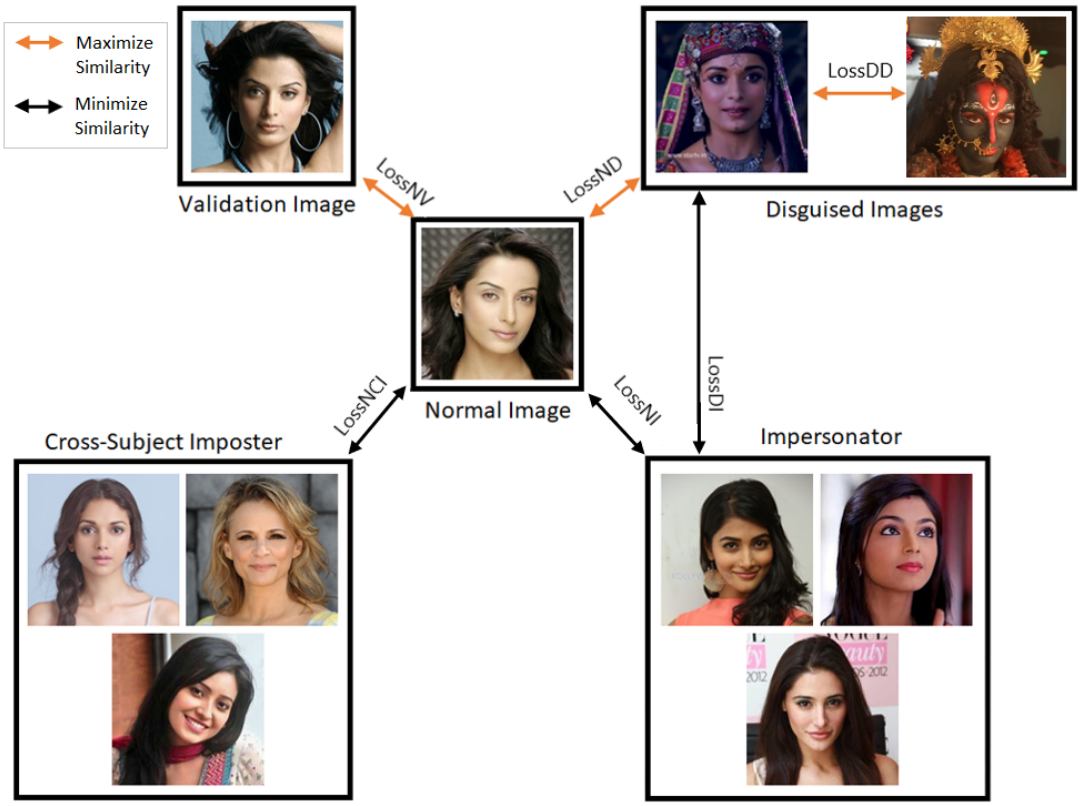
2.1 Disguised Loss

To tackle the disguised face recognition problem, a novel loss is proposed and the LCNN29 v2 network is fine-tuned with it. It is called *Disguised Loss*.

The proposed loss is based on the exponential cosine similarity. In simpler terms, let X_1 *and* X_2 *be the two images and* x_1 *and* x_2 are vectors which are the representation of these two images in the embedding space. Then "C" is a function which takes two input variables and compute the cosine similarity between them as shown in equation 2.1

$$C(x_1, x_2) = \frac{x_1.x_2}{|x_1||x_2|} \tag{2.1}$$

For a subject S_i, where i is in the range 1 to n (n is the number of subjects in training set), its normal image is N_i, validation image is V_i, disguised images are



.

Figure 2.1: Pictorial representation of *Disguised Loss*.

denoted as D_i^k where k has a range from 1 to K (K is the number of disguised images for subject S_i), impersonator images are denoted as I_i^l where l goes from 1 to L (L is number of impersonator images for S_i) and cross-subject imposter images are CI_i^m where m goes from 1 to M (M is the number of cross-subject imposter images selected randomly from subjects other than S_i). Now, the image pairs for subject S_i can be denoted as,

(i) $P(N_i,V_i)$,

(ii) $\forall_{k=1 \text{ to } K}\ P(N_i,D_i^k)$,

(iii) $\forall_{k_1=1 \text{ to } K}\ \forall_{k_2=k_1+1 \text{ to } K}\ P(D_i^{k_1},D_i^{k_2})$,

(iv) $\forall_{l=1 \text{ to } L}\ P(N_i,I_i^l)$,

(v) $\forall_{\mathrm{k=1\ to\ K}}\ \forall_{\mathrm{l=1\ to\ L}}\ \mathrm{P(D_i^k,I_i^l)}$

(vi) $\forall_{\mathrm{m=1\ to\ M}}\ \mathrm{P(N_i,CI_i^m)}$,

where $\mathrm{P(X_1,X_2)}$ is a pair containing 2 images $\mathrm{X_1}$ and $\mathrm{X_2}$. The pairs corresponding to (i), (ii) and (iii) belongs to set_1 and pairs corresponding to (iv), (v) and (vi) belongs to set_2. Let "$Pairs_i$" be the set of all the pairs defined above for subject S_i, then *Disguised Loss* can be defined as shown in eq 2.2

$$L_{\mathrm{D}} = \begin{cases} \sum_{i=1}^{n} \sum_{p=1}^{|Pairs_{\mathrm{i}}|} e^{-\mathrm{C(Pairs_i^P)}} & if Pairs_{\mathrm{i}}^{\mathrm{p}} \in set_1 \\ \sum_{i=1}^{n} \sum_{p=1}^{|Pairs_{\mathrm{i}}|} e^{\mathrm{C(Pairs_i^P)}} & if Pairs_{\mathrm{i}}^{\mathrm{p}} \in set_2 \end{cases} \tag{2.2}$$

This loss exploits the properties of the dataset in such a way that it tries to bring closer the embeddings of normal-disguised, normal-validation and disguised$_1$ - disguised$_2$ pairs of same subject along with trying to move apart the embeddings of disguised - impersonator, normal-impersonator and normal-cross imposter pairs. *Disguised Loss* penalizes more as similarity value of a pair belonging to set_1 tends towards -1 and for set_2, as it tends towards +1. Many pairs are getting implicitly away from each other using the proposed loss function. For example, for the pair validation-impersonator, there is no additional term, but as normal-validation are trying to come closer (becoming more similar) and normal-impersonator are trying to go far away (becoming less similar), automatically, validation-impersonator pair goes away. Many pairs come closer in embedding space automatically such as validation-disguise.

2.2 Preprocessing

Image preprocessing can be highly beneficial depending on the problem statement. It can better detect the local and global features and can help increase the efficiency of a problem. It is an important and effective approach to remove

the undesired distortions and increase the amount of relevant information in order to enhance the performance of the machine learning systems that work on images. For the disguised face recognition problem, the chances that the image is correctly recognized depends highly on the quality of the images. Thus, the preprocessing of images in such cases can give better results. The techniques used in this particular problem are as follows:

Figure 2.2: Pre-processing pipeline used.

2.2.1 Face Detection and Cropping

The images used in this problem vary widely. They do not just contain faces but the whole body or even different objects as well. Since our problem concerns with just the faces of the subjects, everything else in the image can be considered as noise. Thus removing this noise is not going to harm the efficiency, rather it can make it more efficient. To detect the faces in the images in the DFW2019 dataset, TinyFace [36] has been used which is a deep CNN architecture. The faces detected in the above step are then cropped. TinyFace is not able to detect some faces, thus those images are cropped manually.

2.2.2 Face Alignment and Resizing

One major preprocessing technique which helped the results to improve is face alignment. One potential reason we think that this was able to cause a significant difference as compared to its absence is that, the dataset contains a lot of unaligned images due to which features are not comparable for different images. After applying face alignment, those images get aligned, and the embeddings from LightCNN29 v2 are more comparable than before. Whenever the image data is collected in raw form, each might not be of the same size since they have been collected from very different sources [37]. Thus to eliminate these differences that occurred in the data, the re-sizing of the images is done to make all the images of a standard size and normalize the data. The images obtained after face alignment are then resized to 128x128.

2.2.3 Feature Extraction

In early face recognition systems, feature extraction was usually done using histogram of oriented gradient (HOG) based features, LBP and many more but recently transfer learning based feature extraction is used. For our proposed approach, several pretrained models such as ResNet-50 [2] trained on VGGFace2 dataset [38], LCNN29 v2 [35] are used to extract features by using their embedding output from the last layer before the softmax layer. LCNN29 v2 is finally used to produce the results. The output vector is of size 256 for every image and cosine similarity between the two images is computed on vector embeddings of images.

2.3 Proposed Framework for Disguised Face Recognition

The LCNN29 v2 [35] architecture is used as the network to be fine-tuned with the proposed loss, and this network has been fine-tuned according to figure 2.3. LCNN29 v2 is a 29-layer convolution network which is inspired from ResNet. This architecture uses residual blocks with Maximum Feature Map(MFM) which is a variation of maxout activation. The residual block for LCNN29 v2 contains two 3 x 3 convolution layers and two MFM operations without batch normalization. A total of 12,637K trainable parameters are present. It has been trained using CASIA-WebFace and MS-celeb Datasets. MFM has been clearly explained in [35].

In the proposed approach, faces are detected from the original images first and then faces are aligned. After alignment of the faces, data augmentation is performed. Once the network has been fine-tuned, it is tested as described in figure 2.4, where both original aligned faces and mirrored aligned faces are sent as an input to the fine-tuned network separately and produce output features. These output features are then used to calculate a cosine similarity score for both type of images respectively, and that score is fused to get the final results.

Figure 2.3: Training using proposed framework.

2.3.1 Data Augmentation

As there is a lot of class imbalance in the dataset, data augmentation was also used while training. Data augmentation techniques such as horizontal flipping

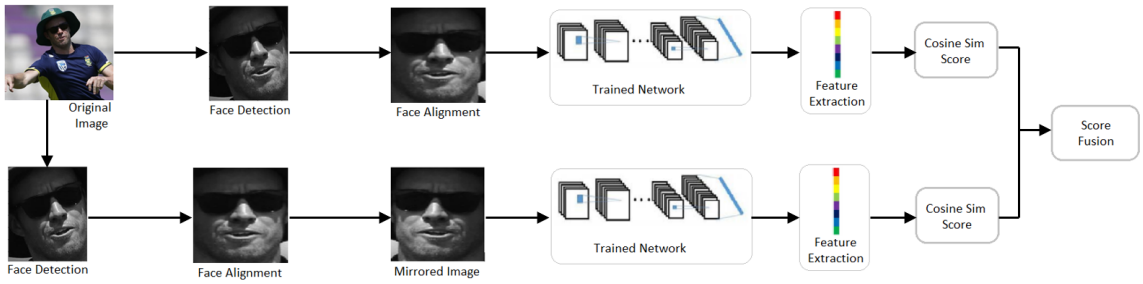


Figure 2.4: Testing using proposed framework.

random cropping have been used for this purpose. The value of p was set as 0.5 for horizontal flipping, where p denotes the probability of input being converted to its mirror image. Augmentation is done on the fly. The network has been trained in such a way that the batch size is variable. The batch making approach is as follows: three folders for every mini iteration within an epoch are selected, and all required combinations of pairs of images for the proposed loss are made. Further, these batches which can be of variable length get transferred to network for training.

2.3.2 Score Fusion

After the network is trained, score fusion is applied using scores of aligned images and scores of mirrored images of aligned images. It is applied using different methods as well, such as fusion of score matrix generated from pairs when non-aligned and aligned images are used. In all score fusion techniques that we used, mirrored images aligned, and non-mirrored images aligned worked the best. This can be because, by fusing the scores of both the matrix and then normalizing, some examples which got misclassified earlier, got correctly classified due to their direction of the pose.

2.3.3 Implementation Details

The proposed loss is used for fine-tuning the pretrained LCNN29 v2 network. It's last classification layer is removed and the last two layers of the updated network are used and their weights are fine-tuned. The learning rate for the training process is set to 0.00001 with weight decay value as 10^{-9}. Adam optimizer is used to train the network and the best model in all epochs is saved.

Chapter 3

Disguised Faces in Wild (DFW)

This chapter talks about the Disguised Faces in Wild dataset which was released by IAB lab at IIIT Delhi. The version released in 2018 i.e., DFW2018 dataset and the new version DFW2019 dataset have been described in detail along with their protocols. The baseline results for both datasets have also been reported.

3.1 DFW2018 Dataset

Datasets that have been initially used for studying disguise variations like AR dataset [23] were released in 1998. Some datasets with a moderate size of face images have also been released, but with controlled disguises. The major problem with such datasets is that although they do provide a more realistic picture of the world, they fail to capture the face when disguises are involved. The challenge presented by the disguised face recognition problem is to match the faces under both intentional and unintentional disguises. Genuine or imposter pairs can be a result of both forms of disguises. For example, in intentional disguise, a thief might try to hide his/her identity by using some disguise accessories and is successful in his attempt to fool the authentication system. Another case could be

when someone tries to pass as a genuine face, i.e., try to impersonate someone but is recognized as an impersonator pair. Similarly, in unintentional disguise, if a person is using sunglasses or any other accessory, this can result in genuine disguise pair. Also, twins or people who look similar will be imposter pairs for the system. To study this topic and for further research, IAB lab at IIIT Delhi released the DFW2018 [1] dataset. A wide range of both intentional and unintentional disguises have been captured in this dataset. A detailed description of the dataset has been given in the upcoming section.

3.1.1 Description of the DFW2018 Dataset

This dataset has been collected from the internet for most of the subject images, which is the reason that it provides unconstrained variations of disguises. In 2018, DFW2018 dataset was released for the workshop in Conference on Computer Vision and Pattern Recognition (CVPR). This dataset contains a total of 11,157 face images that belong to 1000 subjects having uncontrolled disguise variations. Out of these 1000 subjects, 400 subjects belong to the training set, and rest 600 belong to the testing set. Majorly every subject in the database contains their impersonator images and disguise images. It was the first dataset to contain impersonator images for a subject. An aspect of physical adversaries can also be analyzed with the existing models using these impersonator images. Most of the images contain famous personalities of Caucasian or Indian ethnicity thus giving way to a wide spectrum of disguise variation.

3.1.2 Baseline Results

Three protocols are specifically designed for the evaluation by focusing on disguise as a major covariate, namely (i) Impersonation, (ii) Obfuscation and (iii) Overall. Baseline results calculated on these three protocols are reported in Singh et al. [1]

on VGGFace (VGG-Face model pre-trained on the VGG-Face dataset [39]) and VGGFace2 (ResNet-50 model trained on the MS-Celeb-1M [40] and VGGFace2 [41] dataset) architectures have been summarized in Table 3.1

Table 3.1: Baseline GAR for the DFW2018 dataset with VGGFace and VGGFace2.

	Algorithm	**Protocol1**	**Protocol2**	**Protocol3**
At 1% FAR	VGGFace	52.77	31.52	33.76
	VGGFace2	73.94	54.86	56.22
At 0.1% FAR	VGGFace	27.05	15.72	17.73
	VGGFace2	38.84	31.55	32.68

3.2 DFW2019 Dataset

Following the same criterion as the DFW2018 dataset, an additional 600 subjects have been added to the current dataset. It contains a total of 3840 images that are collected from the Internet with a focus on disguises and plastic surgery. These subjects are divided into two sets, where the first set includes 350 subjects, and the second set contains 250 subjects focusing on the plastic surgery aspect of the faces.

Among set one, 250 subjects have the same properties as of DFW2018 dataset where every subject has its disguise and impersonator images with normal and validation images of the same. These 250 subjects contribute 3140 images in total. The other 100 subjects have disguises related to bridal makeup specifically. It contains before and after pictures of brides with a specific focus on bridal makeup. Set two contains both before and after plastic surgery pictures for each subject.

Table 3.2: Statistics of the DFW2019 dataset.

	Number of Subjects	Number of Images
Total	600	3840
Bridal	100	200
Plastic surgery	250	500
Disguised	250	3140

All images for DFW2019 dataset have been collected manually by searching for subjects and their images with the purpose of finding a variation with respect to disguise and plastic surgery. It contains four different kinds of face images namely,

Normal Face Image: Frontal, high quality, and non-disguised subject image

Validation Face Image: Non-disguised images other than normal face images which can help for the evaluation of a model for matching non disguised images

Disguise Face Images: Image having either intentional or unintentional disguise

Impersonator Face Images: Image of a person that intentionally or unintentionally looks similar to the subject's genuine image

3.2.1 Protocols for Evaluation

There are four protocols specifically designed for the evaluation by focusing on the disguise as a major covariate namely (i) Impersonation, (ii) Obfuscation, (iii) Plastic Surgery and (iv) Overall. Two of these protocols, namely Obfuscation and Plastic surgery, are different from DFW2018 dataset. A detailed explanation of these protocols is provided below.

3.2.1.1 Protocol 1 - Impersonation

This protocol has been defined by Khushvaha et al. [42]. In this protocol, the main emphasis is on identifying intentional and unintentional variations of impersonator where it looks like the subject in question. In both cases, we want to have an authentication system that should be able to detect any kind of attempt which is unauthorized. Here genuine pairs consist of normal and validation image of the same subject and the imposter set consists of impersonator-normal pair, impersonator-disguise pair and impersonator-validation pair of the same subject.

3.2.1.2 Protocol 2 - Obfuscation

This protocol has been defined by Khushvaha et al. [42]. It focuses on intentional or unintentional disguise variations across genuine users. In this case, the authentication system should be able to identify genuine users even under varying disguises correctly. Additionally, for bridal subjects, where makeup and other forms of disguises have been added in the DFW2019 dataset, every subject has it's before and after image. In case of the same subject, ground truth for it's before and after image pair is the genuine set whereas cross subject pairs constitute of imposter set. It tries to capture one specific type of disguise where a face recognition system might not be able to correctly identify a person due to heavy (bridal) makeup.

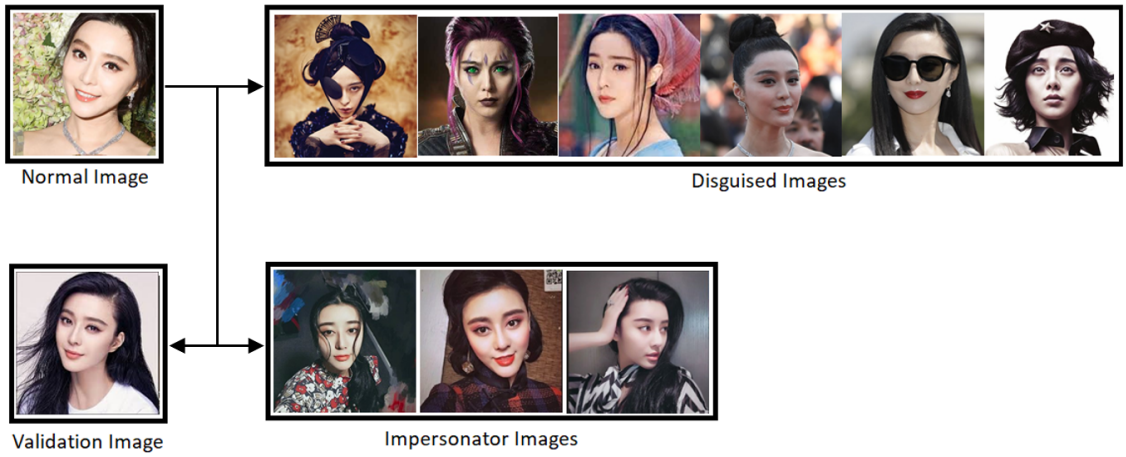
3.2.1.3 Protocol 3 - Plastic Surgery

This protocol is specifically targeted towards change in facial features due to plastic surgery and how the face recognition system performs in the presence of it. Most of the times, due to these changes, even humans are also not able to recognise the other person who has undergone surgery. This protocol is designed to study this aspect of the face recognition system. Here, for every subject, it's

before and after image is provided, which is the genuine set whereas other cross subject pairs are a part of imposter set.

3.2.1.4 Protocol 4 - Overall

This protocol has been defined by Khushvaha et al. [42]. It evaluates a face recognition model on the entire DFW2019 dataset. Its genuine set contains genuine pairs of protocol 1, 2 and 3 combined and the same goes with the imposter set.



(a) Subject with corresponding normal, validation, disguised and impersonator images.

(b) Before and after images for plastic surgery and bridal makeup.

Figure 3.1: Sample images for the DFW2019 dataset.

Table 3.3: Baseline TPR for the DFW2019 dataset.

	Algorithm	**Protocol1**	**Protocol2**	**Protocol3**	**Protocol4**
At 1% FAR	ResNet-50	80.00	61.78	72.40	62.12
	SeNet-50	78.40	64.74	72.40	64.55
	LCNN29 v2	**92.40**	**76.69**	**88.00**	**76.97**
At 0.1% FAR	ResNet-50	46.40	35.23	46.40	36.05
	SeNet-50	46.80	38.22	37.20	36.92
	LCNN29 v2	**70.00**	**55.87**	**69.60**	**55.75**

3.2.2 Baseline Results

For baseline accuracy calculations, we have cropped the faces from images and used pretrained networks such as ResNet-50 [2], SeNet-50 [34] and LCNN29 v2 [35]. ResNet-50 is trained on the MS-Celeb-1M dataset and then fine-tuned on the VGG-Face2 [41] dataset. SeNet-50 is trained like ResNet-50. LCNN29 v2 is trained on CASIA-WebFace [43], VGG-Face and MS-Celeb-1M datasets. We used these three models as these are state of the art for many face recognition challenges. We passed these images through all three networks to find corresponding embeddings. Embedding size for ResNet-50 and SENet-50 is a vector of size 2048, and in case of LCNN29 v2, it is 256. All of these vectors are extracted from the second last layer, which is before the softmax layer. Further, cosine similarity is calculated for every pair of image. We reported results for all the four protocols mentioned in the DFW2019 dataset description. The results have been calculated as TPR at x% FPR where x has a value of 1 and 0.1. The results are shown in table 3.3. As we can see, for baseline results LCNN29 v2 [35] is able to perform better than ResNet-50 [2] and SeNet-50 [34], therefore, we have used it for training and for further reporting of results. As a part of the International Conference on Computer Vision (ICCV) 2019, this dataset will be used in competition and workshop to facilitate research in this direction.

Chapter 4

Experimental Results

Results for all the protocols discussed above have been calculated for both DFW2018 and DFW2019 datasets. The models used as baseline for the disguised face recognition problem are ResNet-50 [2], SeNet-50 [34] and LCNN29 v2 [35] for DFW2019 dataset (as explained in 3.2.2) and they are compared with the proposed approach where LCNN29 v2 has been fine-tuned with the proposed loss function *Disguised Loss*, and a score fusion of mirrored and aligned images is used. Further details about the results have been explained in the upcoming sections.

4.1 Results on the DFW2019 Dataset

Table 4.1 summarizes the results on the DFW2019 dataset with ResNet-50, SeNet-50, LCNN29 v2 and the proposed approach on protocols 1, 2, 3 and 4. Apart from these four algorithms, the table shows how the results change as different intermediate techniques are introduced. Here ResNet-50, SeNet-50 and LCNN29 v2 are pretrained networks that are used as it is and the intermediate approaches are as follows:

- Fine tuned LCNN29 v2 w/o Disguised Loss: It is the approach where simple

cosine similarity loss is used to fine-tune the LCNN29 v2 pretrained model.

- Fine tuned LCNN29 v2: It is the approach where proposed loss is used to fine-tune the LCNN29 v2 pretrained model
- Fine tuned LCNN29 v2 + aligned images: It is the approach that along with being fine tuned using proposed loss, also uses aligned images
- Fine tuned LCNN29 v2 + mirrored aligned images: It is the approach that is fine-tuned using proposed loss and uses mirrored images of aligned images

4.1.1 Protocol 1 - Impersonation

Figure 4.3a contains the receiver operating characteristics (ROC) curves for protocol 1. Table 4.1 presents the TPR at 1% FAR and 0.1% FAR. As can be seen from the results, the proposed algorithm where mirrored and aligned score fusion technique is used along with fine-tuning on the proposed loss, has outperformed all other baselines by a minimum margin of 3.6% at 1%FAR.

Use of aligned images along with fine-tuned LCNN29 v2 with proposed loss, has shown a 2% increase in the accuracy at 1% FAR over the model that did not use aligned images. However, at 0.1% FAR, fine-tuned LCNN29 v2 without aligned images shows the best performance. Use of mirrored images also showed a little improvement but maximum improvement was with score fusion method i.e., the proposed approach.

4.1.2 Protocol 2 - Obfuscation

Figure 4.3b contains the receiver operating characteristics (ROC) curves for protocol 2. Table 4.1 presents the TPR at 1% FAR and 0.1% FAR and as is evident from the results, the proposed approach where aligned and mirrored score fusion

Table 4.1: TPR for protocols 1,2,3 and 4 on the DFW2019 dataset with different approaches along with the proposed approach.

	Algorithm	**Protocol1**	**Protocol2**	**Protocol3**	**Protocol4**
At 1% FAR	ResNet-50	80.00	61.78	72.40	62.12
	SeNet-50	78.40	64.74	72.40	64.55
	LCNN29 v2	92.40	76.69	88.00	76.97
	Fine tuned LCNN29 v2 w/o Disguised Loss	92.40	75.51	86.80	75.98
	Fine tuned LCNN29 v2	93.20	76.69	87.60	77.05
	Fine tuned LCNN29 v2 + aligned image	95.20	80.58	**89.20**	80.65
	Fine tuned LCNN29 v2 + mirrored aligned image	95.60	80.14	88.40	80.28
	Proposed approach	**96.00**	**80.89**	88.40	**80.97**
At 0.1% FAR	ResNet-50	46.40	35.23	46.40	36.05
	SeNet-50	46.80	38.22	37.20	36.92
	LCNN29 v2	70.00	55.87	69.60	55.75
	Fine tuned LCNN29 v2 w/o Disguised Loss	68.80	54.42	71.20	55.33
	Fine tuned LCNN29 v2	**71.20**	55.72	70.40	55.89
	Fine tuned LCNN29 v2 + aligned image	68.80	62.50	74.40	62.06
	Fine tuned LCNN29 v2 + mirrored aligned image	67.20	62.46	72.80	62.13
	Proposed approach	68.8	**63.79**	**75.20**	**62.96**

technique was used, has outperformed all other baselines by a minimum margin of 4% at 1% FAR and by 8% at 0.1% FAR. It can be seen that on fine-tuning the LCNN29 v2 model by the proposed loss, the TPR has increased for both 1% FAR and 0.1% FAR. There is an increase of almost 4% accuracy at 1% FAR when aligned images are used with the LCNN29 v2 model, fine-tuned on the proposed loss and an increase of almost 7% at 0.1% FAR but use of mirrored images did not give significant improvements.

4.1.3 Protocol 3 - Plastic Surgery

For Plastic surgery, the proposed approach has shown improvements over baselines for 0.1% FAR resulting in almost 6% increase over the baseline. Figure 4.3c contains receiver operating characteristics (ROC) curves for the same. Here, use of aligned images with LCNN29 v2 fine-tuned on proposed loss gave the maximum accuracy at 1% FAR showing an increase of 1.2% while the TPR for proposed approach i.e., score fusion of mirrored and aligned images was 88.40% .

4.1.4 Protocol 4 - Overall

Figure 4.3d contains the receiver operating characteristics (ROC) curves for protocol 4. The proposed algorithm (mirrored + aligned) has outperformed all other baselines by a minimum margin of 4% at 1%FAR as well as resulted in an 8% increase in TPR at 0.1% FAR from LCNN29 v2. Each intermediate step at 0.1% FAR has shown improvement where the use of aligned images with fine-tuned LCNN29 v2 has shown maximum improvement of almost 7%.

Figure 4.1 and figure 4.2 shows some examples of false positive, false negative and true positive, true negative images pairs respectively for the DFW2019 dataset on 0.1% FAR for protocol 1. False positive refers to the pair of images where the

algorithm predicts that the two images are of the same person but actually the two images are of two different persons. False negative refers to the pair of images where the algorithm predicts that the two images are of the different persons but actually the two images are of same person. Similarly True positive refers to the pair of images where algorithm correctly predicts that the two images are of the same person and true negative is where algorithm correctly predicts that the two images are of different persons.

Figure 4.1: False positive and false negative pairs example for the DFW2019 dataset on 0.1% FAR for protocol 1.

Figure 4.2: True positive and true negative pairs example for the DFW2019 dataset on 0.1% FAR for protocol 1.

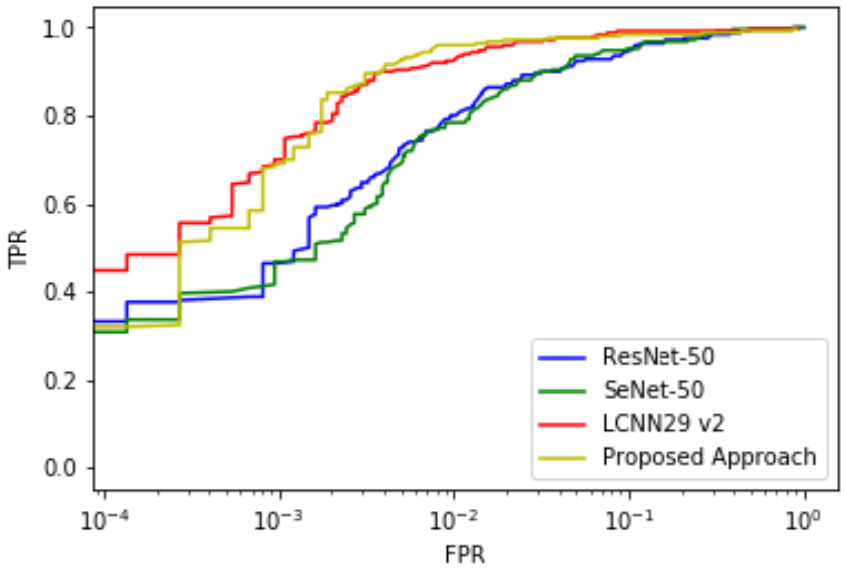
4.1.5 Significance Testing

Significance testing has been done between two pairs of algorithms, pair1 consists of fine-tuned LCNN29 v2 + aligned images and proposed approach, pair2 is fine-tuned LCNN29 v2 + mirrored aligned images and proposed approach. McNemar test is a statistical test for paired nominal data and it has been used to evaluate the predictive accuracy of the two models. The p-value and statistic for all four protocols of the DFW2019 dataset have been reported at both 1% and 0.1% FAR as shown in table 4.2. Here, the statistic is the McNemar test statistic. The significance threshold, alpha is used as 0.05, which means that if p-value for a pair of two models is less than this threshold, we can reject the null hypothesis which states that the performance of the two models is same.

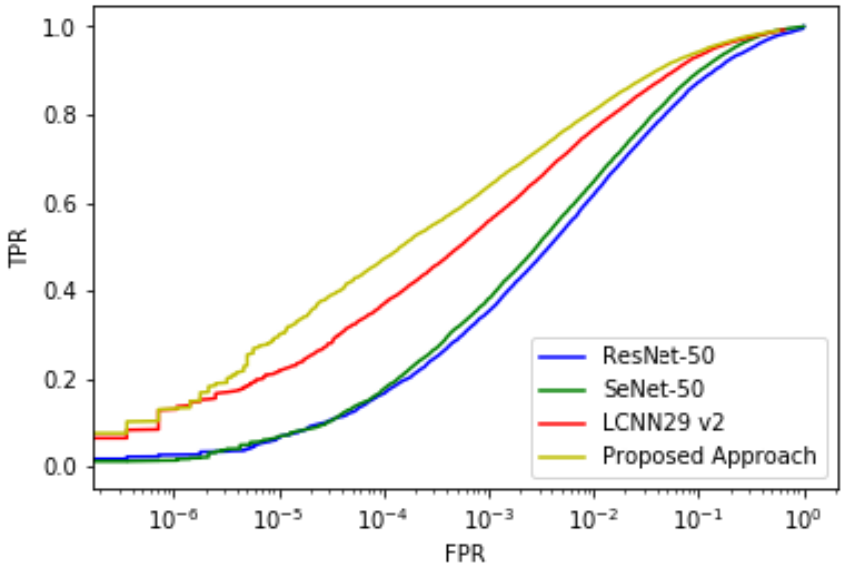
For protocol 4, the p-value at 1% FAR and 0.1% FAR is very close to 0. This suggests that there is significant difference between the performance of pair1 classifiers. Same pattern is observed for pair2. For protocol 3, the p-value at both the 1% and 0.1% FAR is more than 0.05 which means that we can not reject the null hypothesis and there is not much difference between the performance of pair1 and pair2 classifiers. For protocol 1 and 2, there is statistical difference for both of the pairs at 0.1% FAR as can be seen from the table 4.2. Given all these observations, we can say that applying fusion of the two pipelines gave us results with significant statistical difference.

Table 4.2: McNamer test results

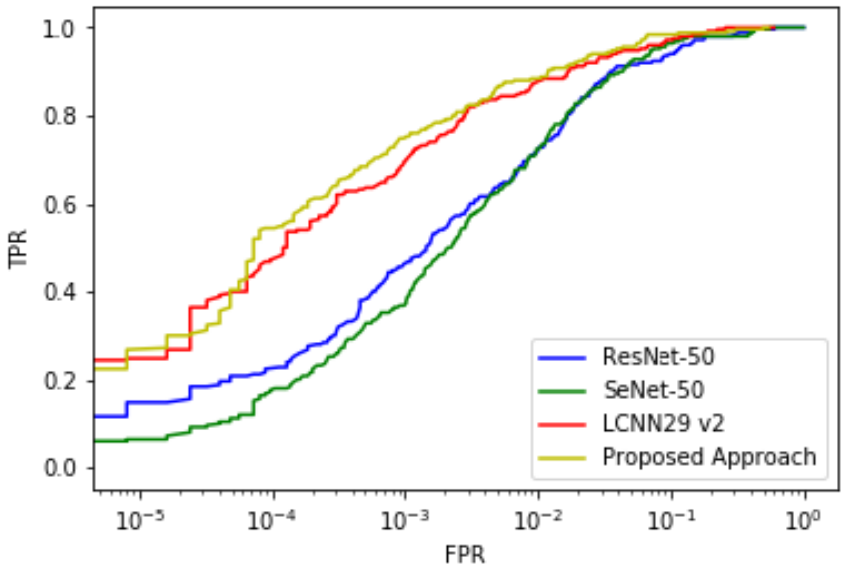
McNamer Test between Models	**Protocol**	**At 1% FAR**		**At 0.1% FAR**	
		p-value	**Statistic**	**p-value**	**Statistic**
Fine-tuned LCNN29 v2 + aligned images and proposed approach	**P1**	0.700	0.148	0.027	4.920
	P2	0.881	0.022	0.009	6.914
	P3	0.682	0.168	0.880	0.023
	P4	0.000	432.71	0.000	93.82
Fine-tuned LCNN29 v2 + mirrored aligned images and proposed approach	**P1**	0.831	0.045	0.070	3.273
	P2	0.155	2.023	0.000	12.93
	P3	0.765	0.089	1.000	0.000
	P4	0.000	520.44	0.000	84.03



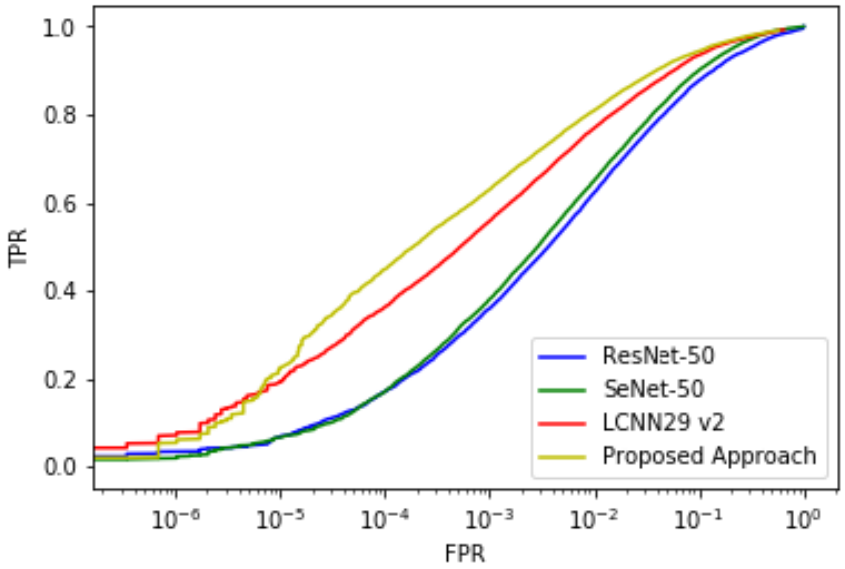
(a) ROC for Protocol 1, Impersonation



(b) ROC for Protocol 2, Obfuscation



(c) ROC for Protocol 3, Plastic Surgery



(d) ROC for Protocol 4, Overall

Figure 4.3: ROC on the DFW2019 dataset for protocols 1,2,3 and 4 with ResNet-50, SeNet-50, LCNN29 v2 and proposed approach.

4.2 Results on the DFW2018 dataset

We have also evaluated our approach on the DFW2018 dataset and we got competitive results for protocols 1, 2 and 3. Table 4.3 summarizes the top five results including the proposed approach which is ranked 3^{rd} on protocol 1. Similarly, table 4.4 and table 4.5 summarizes the top five results including the proposed approach on protocols 2 and 3 respectively. The proposed approach is at 5^{th} position for both of them.

Table 4.3: Top five results on the DFW2018 dataset for protocol 1.

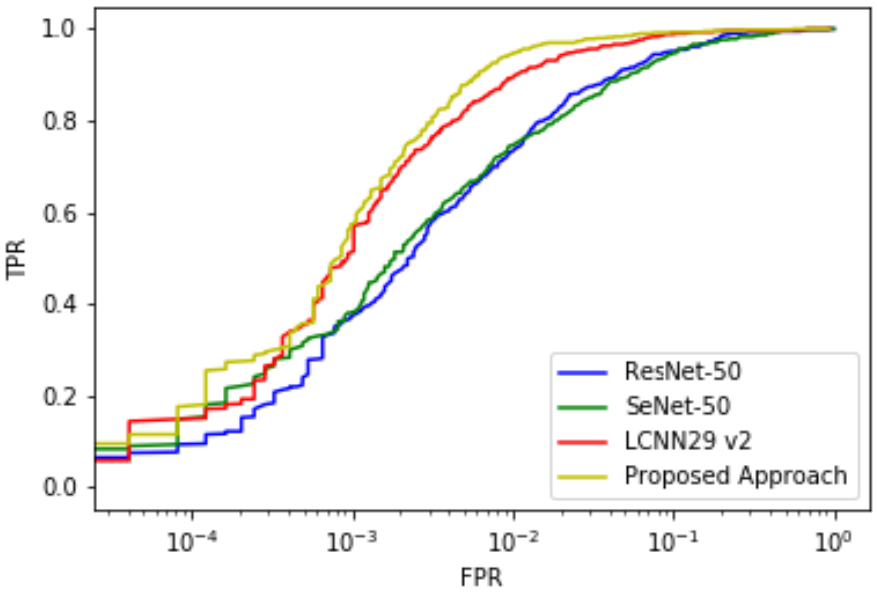
Algorithm	**GAR @ 0.1% FAR**
DenseNet + COST [21]	62.20
AEFRL [14]	57.64
Proposed approach	**57.31**
MEDC	55.46
ByteFace	55.11

Table 4.4: Top five results on the DFW2018 dataset for protocol 2.

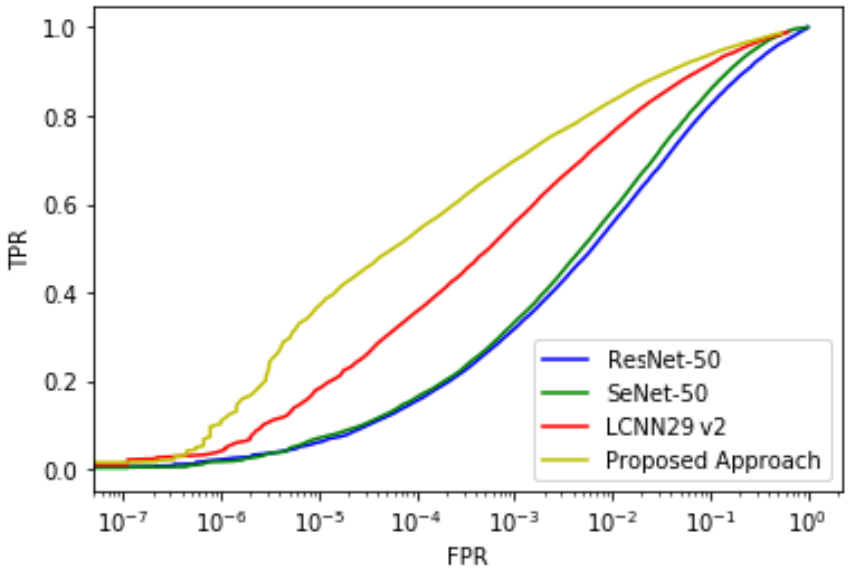
Algorithm	**GAR @ 0.1% FAR**
MiRA-Face [15]	80.56
AEFRL [14]	77.06
UMDNets [16]	74.69
DenseNet + COST [21]	72.10
Proposed Approach	**69.52**

Table 4.5: Top five results on the DFW2018 dataset for protocol 3.

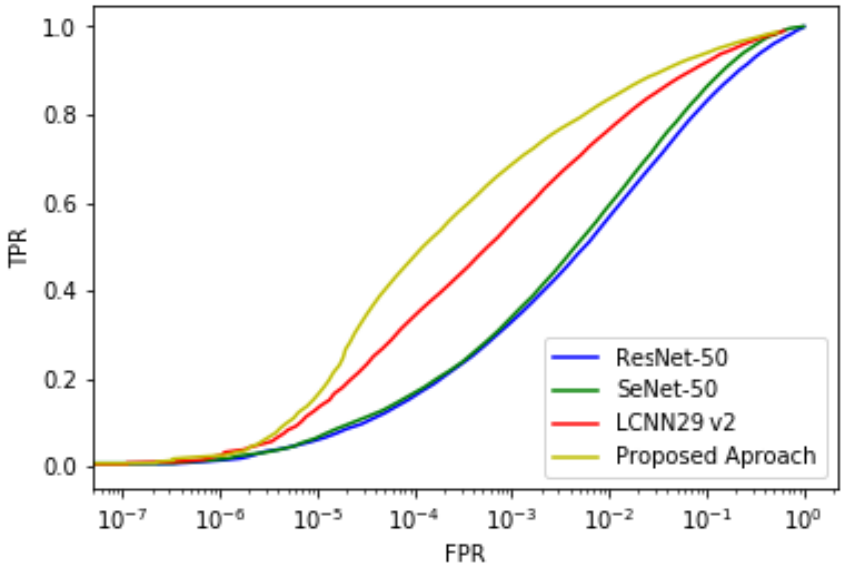
Algorithm	**GAR @ 0.1% FAR**
MiRA-Face [15]	79.26
AEFRL [14]	75.54
UMDNets [16]	72.90
DenseNet + COST [21]	71.50
Proposed approach	**69.14**



(a) ROC for protocol 1, Impersonation



(b) ROC for protocol 2, Obfuscation



(c) ROC for protocol 3, Overall

Figure 4.4: ROC on the DFW2018 dataset for the protocols 1, 2 and 3 with ResNet-50, SeNet-50, LCNN29 v2 and proposed approach.

Chapter 5

Conclusion

Face recognition under disguise variations is a challenging yet less explored problem. Disguised faces are often characterized by obfuscation, wherein a part of the face is hidden or occluded; along with variations due to make-up, and changes in the facial structure. Existing research has primarily focused on disguised face recognition in relatively constrained environments with limited disguise accessories. On the contrary, in real world scenarios, face recognition systems often operate in unconstrained settings and encounter face images with large variations. To this effect, this research focuses on addressing the challenging task of recognizing Disguised Faces in the Wild (DFW).

The contributions of the research are two-fold: (i) proposing a novel Disguised Faces in the Wild dataset (DFW2019 dataset), and (ii) proposing a novel disguised face recognition pipeline using the proposed Disguised Loss. The proposed DFW2019 dataset contains images pertaining to 600 subjects, including variations due to obfuscation, impersonation, plastic surgery, and bridal make-up. The images are collected from the Internet, therefore also containing variations across pose, illumination, ethnicity, gender, and acquisition device. Four protocols have also been presented for evaluating the proposed DFW2019 dataset, along with the baseline results using three state-of-the-art deep learning based

face recognition models. A novel Disguised Loss has also been presented which promotes learning features useful for disguised face recognition. The proposed loss explicitly minimizes the intra-class variations and maximizes the inter-class variations during feature learning. A face recognition pipeline is presented with the proposed Disguise Loss which demonstrates improvement over the baseline results across all protocols for the DFW2019 dataset, and presents comparable results on the DFW2018 dataset.

The DFW2019 dataset will be made publicly available to the research community, in order to encourage development of face recognition models robust to disguises. While the proposed face recognition pipeline presents improved performance, we believe that there is still scope for further improvement. As part of future work, the proposed Disguise Loss can be used in conjunction with a attention network framework. The attention network module can help in selecting regions of interest from the given input face image, while the proposed Disguised Loss will enable learning discriminative features.

References

[1] M. Singh, R. Singh, M. Vatsa, N. K. Ratha, and R. Chellappa, "Recognizing disguised faces in the wild," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 2, pp. 97–108, 2019. iii, 5, 6, 7, 8, 18

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. 1, 8, 13, 23, 24

[3] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, vol. 6, 2016. 1

[4] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 748–763, 2002. 3, 7

[5] S. M. Yoon and S.-C. Kee, "Detection of partially occluded face using support vector machines.," in *MVA*, pp. 546–549, 2002. 3, 7

[6] N. Ramanathan, R. Chellappa, and A. R. Chowdhury, "Facial similarity across age, disguise, illumination and pose," in *International Conference on Image Processing, 2004.*, vol. 3, pp. 1999–2002. 3, 6, 7

[7] J. R. Beveridge, D. Bolme, B. A. Draper, and M. Teixeira, "The csu face

identification evaluation system," *Machine vision and applications*, vol. 16, no. 2, pp. 128–138, 2005. 3

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2008. 3, 7

[9] 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. 3, 7

[10] J. Kim, Y. Sung, S. M. Yoon, and B. G. Park, "A new video surveillance system employing occluded face detection," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 65–68, Springer, 2005. 3, 7

[11] R. Min, A. Hadid, and J. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Face and Gesture*, pp. 442–447, March 2011. 3, 7

[12] 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*, pp. 1–8, 2013. 3, 7

[13] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, "Recognizing disguised faces: Human and machine evaluation," *PloS one*, vol. 9, no. 7, p. e99212, 2014. 3, 7

[14] E. Smirnov, A. Melnikov, A. Oleinik, E. Ivanova, I. Kalinovskiy, and E. Luckyanets, "Hard example mining with auxiliary embeddings," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 37–46, 2018. 3, 7, 32

[15] K. Zhang, Y.-L. Chang, and W. Hsu, "Deep disguised faces recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 32–36, 2018. 4, 7, 32

[16] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa, "Deep features for recognizing disguised faces in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 10–16, 2018. 4, 7, 32

[17] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," *arXiv preprint arXiv:1703.09507*, 2017. 4

[18] N. Kohli, D. Yadav, and A. Noore, "Face verification with disguise variations via deep disguise recognizer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 17–24, 2018. 4, 7

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015. 4

[20] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European conference on computer vision*, pp. 499–515, Springer, 2016. 4

[21] S. Suri, A. Sankaran, M. Vatsa, and R. Singh, "On matching faces with alterations due to plastic surgery and disguise," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–7, IEEE, 2019. 4, 7, 32

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017. 4

[23] A. M. Martinez, "The AR face database," *CVC Technical Report24*, 1998. 6, 7, 17

[24] B. Y. 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*, pp. 186–192, 2013. 6

[25] T. Y. Wang and A. Kumar, "Recognizing human faces under disguise and makeup," in *IEEE International Conference on Identity, Security and Behavior Analysis*, pp. 1–7, 2016. 6, 7

[26] 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*, pp. 3371–3377, 2018. 6

[27] A. Singh, D. Patil, M. Reddy, and S. Omkar, "Disguised face identification (dfi) with facial keypoints using spatial fusion convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1648–1655, 2017. 6, 7

[28] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ica for face recognition robust to local distortion and partial occlusion," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 1977–1981, 2005. 7

[29] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, "The feret evaluation methodology for face-recognition algorithms," in *Proceedings of IEEE*

Computer Society Conference on Computer Vision and Pattern Recognition, pp. 137–143, IEEE, 1997. 7

[30] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *European conference on computer vision*, pp. 448–461, Springer, 2010. 7

[31] I. Choi and D. Kim, "Facial fraud discrimination using detection and classification," in *International symposium on visual computing*, pp. 199–208, Springer, 2010. 7

[32] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognition*, vol. 47, no. 9, pp. 2794–2806, 2014. 7

[33] K.-M. Hung, J.-A. Wu, C.-H. Wen, and L.-M. Chen, "A system for disguised face recognition with convolution neural networks," in *Proceedings of the International Conference on Digital Medicine and Image Processing*, pp. 65–69, 2018. 7

[34] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018. 8, 23, 24

[35] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018. 8, 13, 14, 23, 24

[36] P. Hu and D. Ramanan, "Finding tiny faces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 951–959, July 2017. 12

[37] K. Dharavath, F. A. Talukdar, and R. H. Laskar, "Improving face recognition rate with image preprocessing," *Indian Journal of Science and Technology*, vol. 7, no. 8, pp. 1170–1175, 2014. 13

[38] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 67–74, IEEE, 2018. 13

[39] O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.*, "Deep face recognition.," in *bmvc*, vol. 1, p. 6, 2015. 19

[40] 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*, pp. 87–102, Springer, 2016. 19

[41] 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*, pp. 67–74, 2018. 19, 23

[42] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, "Disguised faces in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9, 2018. 21, 22

[43] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014. 23

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

Mohit Chawla

MT17028, Email: mohit17028@iiitd.ac.in
DOB: February 16th , 1995
Address: 715/2, Pardhana Mohalla, near Bara Bazar, Rohtak

Education

Indraprastha Institute of Information Technology Delhi CGPA:7.77
M.Tech(CSE-Data Engg)
2017 – Present

Maharaja Agrasen Institute of Technology, GGSIPU Percentage:78.07
B.Tech(IT)
2012 – 2016

Shiksha Bharti School, Rohtak Percentage: 84.4
CBSE
2010 – 2012

Shiksha Bharti School, Rohtak CGPA: 9.2
CBSE
2009 – 2010

Skills

Area of Interest	Deep learning, Machine Learning, Computer Vision, Natural Language Processing, Information Retrieval, Parallel Programming, Algorithms, Operating System
Programming Language	Python, Java, C, C++, SQL
Tools and Technologies	PyTorch, TensorFlow, Keras, SciPy, Scikit-learn, Numpy, Git, Android Studio
Technical Electives	Advanced Machine Learning, Statistical Machine Learning, Machine Learning, Natural Language Processing, Information Retrieval, Mobile Computing, Foundations of Parallel Programming

Internship

Teaching Assistant(TA) IIITD (May'18-Aug'18)

- **Competitive Programming I**
  Teaching Fellow: Kushagra Arora
- **Statistical Computing** (Jan'18-Apr'18)
  Guide: Prof. Vibhor Kumar
- **Introduction to Programming** (Aug'17-Dec'17)
  Guide: Prof. Vikram Goyal
- **Statstical Machine Learning** (Jan'19-Apr'19)
  Guide: Prof Richa Singh
- **Natural Language Processing** (Aug'18-Dec'18)
  Guide: Prof Tanmoy Chakraborty

Sparklin Innovations Pvt. Ltd (Aug'17-Dec'17)

- Worked as **Backend Developer** and made Web Site in PHP.

Projects

M.Tech Thesis: Disguised Face Recognition (Aug'18–present)
Guide: Prof. Richa Singh and Prof Mayank Vatsa

- Working on problem of matching faces under both impersonation and obfuscation (intentional and unintentional distortion) in disguised face recognition. Proposed Novel Loss based On Exponential Cosine Similarity and used Transfer Learning by using Pre-trained Models such as RESNet with VGGFACE, Light CNN. Contributed a new Version of Disguised Face Dataset consisting of approx. 4k images

Speaker Recognition with Adversarial Attacks and Detection (Dec'18–Apr'19)
Guide: Prof. Richa Singh and Prof Mayank Vatsa

- Worked on problem of Robustness of Speaker Recognition System with Non-targeted attacks such as FGSA, DeepFool etc and Targeted Attacks such as L0, L1, L2 attacks and then developing system to detect these attacks.

Hindi Chatbot (Aug'18–Dec'18)
Guide: Prof. Richa Singh Team size-2

- Built a working prototype of Hindi ChatBot by using RNN and Attention Networks. Several Other approaches using FASTText vectors and BERT have also been implemented in project. .

Age and Gender Detection System (Jan'18–Apr'18)
Guide: Prof. Richa Singh Team Size-2

- Classified images of Adience Bench Mark Dataset on the basis of Age and Gender by using different Feature extraction techniques(PCA, LDA etc) and applied Machine Learning Models such as CNN, SVM, ensemble methods models to train classifier.

Context Sensitive spell Corrector for OCR predicted text (Jan'18–Apr'18)
Guide: Prof. Tanmoy Chakraborty Team size-2

- Trained machine learning model (RNN using LSTM) to correctly identify words which are not correctly predicted by OCR on Research Papers.

Virtual-I : Medicine Information Retrieval App using OCR (Jan'18–Apr'18)
Guide: Prof. Pushpendra Singh Team size-6

- Made an android app specifically for Blind People using OCR to get information of medicine and other Meta data.
- Incorporated various features such as voice notes, reminders for medicine.
- Used Speech to Text and Text to Speech for Explicit Search using voice.

Energy Efficient Work Stealing Language Runtime (Jan'18–Apr'18)
Guide: Prof. Vivek Kumar Team size-2

- Made a Work Stealing Runtime using DVFS in linux.

Click Through Rate Prediction (Aug'17-Dec'17)
Guide: Prof. Anubha Gupta Team size-4

- Trained Machine Learning Model to predict whether an Advertisement Will be clicked or not using Decision Tree, Adaptive Boosting and Ensemble Learning methods.

Facility Management System (Aug'17-Dec'17)
Guide: Prof. Chetan Arora Team size-2

- Used Facebook API to build a FMS portal for IIITD residents.

Sentiment Extraction and analysis of Movie Reviews (Aug'17-Dec'17)
Guide: Prof. Tanmoy Chakraborty Team size-2

- Used *Rotten Tomatoes* dataset from Kaggle and applied NLP techniques such as N-gram model to analyze and predict sentiments of movie reviews.

Online Help Desk System (2016)

- Made online portal in PHP where user can come and request for the assets and the respective authority will assign it using OHD without going to manual process.

Positions of Responsibility

- Sponsorship Team for Oddysey'18 (Oct'17 – Jan'18)
- Teaching Assistant(TA) IIITD
- Organized annual cultural and sports events that had participation from over 12 schools across Haryana as Team head in school.
- Volunteered in NGO "Youth For Seva".

Awards and Achievements

- Secured AIR 473 in GATE'17.
- Secured 2nd prize in "BHARAT JANO PRATIYOGITA", an interschool competition out of 24 teams.
- Awarded 3rd position out of 12 teams in state level singing competition organized by "BHARAT VIKAS PARISHAD".(India)

Interests and Hobbies

- Singing
- Quizzing
- Applied Psychology

Declaration: The above information is correct to the best of my knowledge.
Mohit Chawla
Date: May'20, 2019