Unravelling the Masked Faces

A Project Report Submitted by

Shiksha Mishra

in fulfillment of the requirements for the award of the degree of

MTech

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

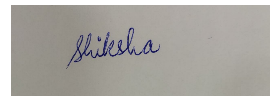
Indian Institute of Technology Jodhpur

Computer Science and Engineering

June, 2021

Declaration

I hereby declare that the work presented in this Project Report titled Unravelling the Masked Faces submitted to the Indian Institute of Technology Jodhpur in partial fulfilment of the requirements for the award of the degree of MTech, is a bonafide record of the research work carried out under the supervision of Dr Mayank Vatsa , Dr Richa Singh. The contents of this Project Report in full or in parts, have not been submitted to, and will not be submitted by me to, any other Institute or University in India or abroad for the award of any degree or diploma.



Signature

Shiksha Mishra

MT19CS013

Certificate

This is to certify that the Project Report titled Unravelling the Masked Faces, submitted by Shiksha Mishra(MT19CS013) to the Indian Institute of Technology Jodhpur for the award of the degree of MTech, is a bonafide record of the research work done by her under my supervision. To the best of my knowledge, the contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Signature

Dr Mayank Vatsa , Dr Richa Singh

Acknowledgements

I am grateful to my family, especially my mother, for always supporting me. I cannot express enough thanks to my guides, Dr. Richa Singh and Dr. Mayank Vatsa, for their invaluable guidance and motivation. I want to thank Muskan Dosi for her invaluable support. I am thankful to Puspita Majumdar for her guidance and to Kartik Thakral for his helping hand. Special thanks to all my IITJ friends for all their support. To all those who have contributed to my data collection, my deepest gratitude. ...

Abstract

In the domain of Biometrics, face recognition has achieved impressive milestones. However, when it comes to occluded faces, not much work has been done on masked faces. Due to the sudden pandemic of covid 19, people worldwide have started wearing the mask that covers almost their entire faces. The state-of-the-art techniques for detecting and recognizing faces do not show good performance on the masked face compared with the performances on the non-occluded face. Face detection with masks and their recognition models are yet to be adequately exploited by researchers. However, some research has been done in the given domain over the limited database that is available for masked faces. Currently, no masked-faced dataset is available for Indian ethnicity. This paper presents a new masked face dataset for Indian ethnicity. This is the first masked face dataset on Indian ethnicity to the best of our knowledge. We present the labeled masked face dataset for Indian ethnicity in detail, which includes the evaluation protocols, baseline results, and performance analysis. We further present a multitask pipeline that performs face detection, mask detection, and gender detection on masked faces. We also show the baseline results of face recognition on new datasets.

Contents

List of Figures

List of Tables

Unravelling the Masked Faces

1 Introduction and background

Face recognition algorithms have achieved tremendous success in handling different covariates such as low resolution, pose, expression, illumination [1, 2, 3, 4]. It is now used in various applications such as surveillance, access control, forensics, and e-payments, and researchers have been enhancing the state-of-the-art performance on challenging tasks [5, 6, 7, 8]. With the advent of deep learning technologies such as [9],[10],[11],[12] the performance on face recognition has improved significantly over the years. However, the problem of recognizing faces that have occlusion is still considered a challenging task. This occlusion occurs when the subject's face is not completely visible, which compromises the recognition accuracy. The occlusion of the face may be intentionally (intended to fool the system) or unintentional; nevertheless still pose a challenge for the system.

As the world is facing the pandemic of COVID-19, people worldwide are wearing masks as a protective measure. Wearing masks all the time has been completely normalized and also a mandatory step at workplaces. This mask on the face occludes a fair amount of the part, leaving only their eyes and forehead visible. Detecting masked faces is a challenging task. Further, when these masked faces are subject to a face recognition system, their performance degrades. Also, due to the normalization of wearing masks, people wear different kinds of masks, from simple to the ones having many graphics in them. In India, wearing "gamchas" and "stoles" in place of the masks is another common trend you can come across. This variety of masks pose another challenge to face detection as well as face recognition, confusing one's identity. Fig. 1.1 shows some of these challenges in unconstrained masked face recognition, including the variation in the type of masks. Though we have an enormous work done in face recognition and the occlusion problem has been addressed by some in [13], little work has been done in resolving the drawback of face recognition with a mask. Also, the limit of availability of masked faces dataset for recognition has constrained such research earlier.

One another challenge that we are facing is the limit of the masked face dataset. Other existing datasets mainly include faces with Caucasian and North Asian demography with limited variations in the masks worn by the subjects. Some of them are collected in controlled settings. Further, none of the existing masked face datasets are built around the attire diversity, typically observed in the Indian context. India is a diverse country, with people wearing different kinds of clothes such as stole or handkerchief as masks which pose new challenges to existing face recognition algorithms. Wearing printed masks with graphics is a common trend in India. This research work introduces two new datasets for detecting and recognizing faces occluded with masks based on Indian ethnicity in a constrained and unconstrained environment.In an unconstrained setting, we collected the dataset Indian Masked faces in the wild. IMFW dataset contains faces of 200 people with and without masked in an unconstrained environment. We have partitioned the dataset into three sets: Celebs, Instagram, and the General public. In a constrained setting, the dataset is collected of 300 subjects using two cameras. Further, both these

Figure 1.1: Illustration of various challenges in masked face recognition: (a) people wearing masks printed with facial images, (b) multiple people wearing same masks, and (c) different types of cloths are used as masks to cover the faces.

dataset is evaluated, and its performance is analyzed. We also present a multitask pipeline that performs face detection mask detection and gender classification. The contribution of the research is as follows:

1. Indian Masked Faces in the Wild (IMFW) that has been designed in keeping in view the problem faced to recognize Indian face during Covid 19 pandemic
2. Indian masked faces in the constrained settings
3. We have evaluated the dataset on multiple standard face detector models
4. The face recognition task is performed on the new Indian Masked Faces in the Wild on baseline models.
5. we also perform the face detection ,mask detection and gender classification using the multitask pipeline

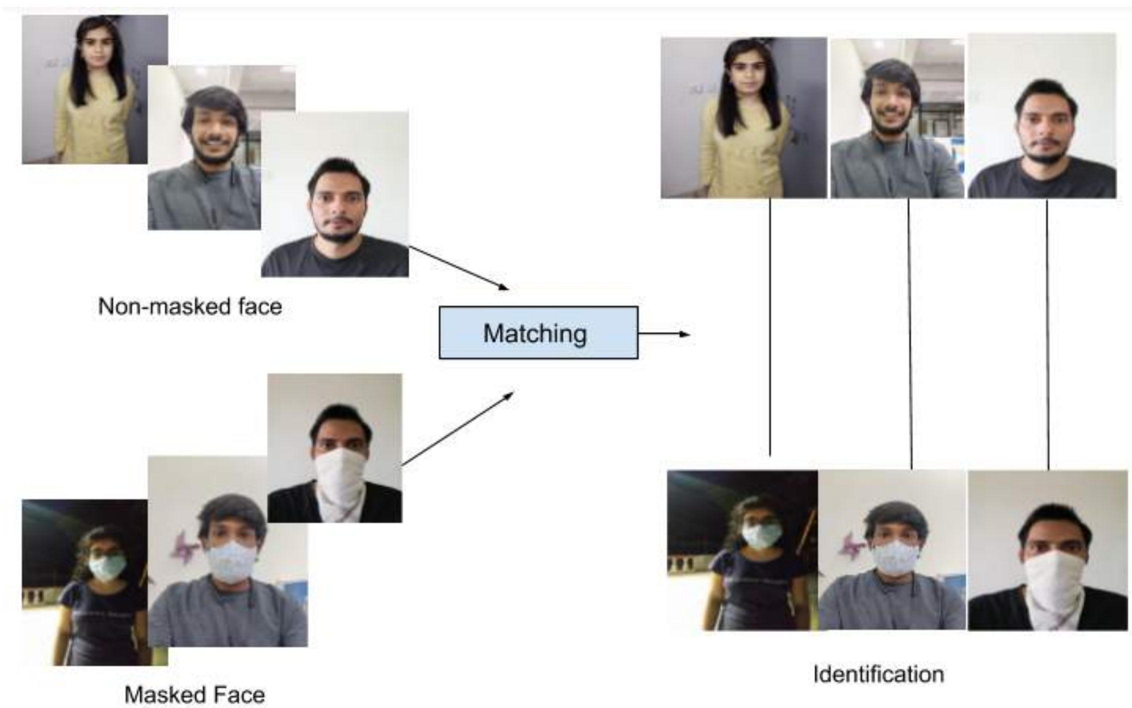


Figure 1.2: Non masked face when matched with masked face should produce the correct Identification

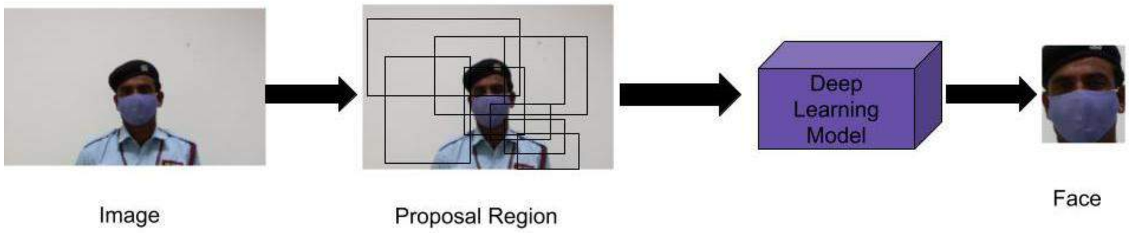


Figure 2.1: Detecting faces with mask

Figure 2.2: Use of same masks by different people confuses the system

2 Problem Formulation

Due to the sudden challenge face by the world, we are forced to wear masks all the time. The challenges due to the presence of masks are:

Occlusion: Wearing masks occlude a major section of the face leaving only eyes and forehead visible. Sometimes people wear eyeglasses/sunglasses that further occludes their faces. Such occlusion leaves the detection and recognition task to be done by the forehead region only. In addition to this, unconstrained settings like the variation in pose, illumination, and resolution further exaggerate the challenges of automatic face recognition.

Inter-class and intra-class variations: As masks have become an integral part of our wardrobe, people accessorize themselves with a variety of masks on a regular basis. This variety of masks used by a subject results in intra-class variation among the samples of the same subject. Also, some masks like surgical masks are very common across the globe. These masks worn by different subjects decrease the inter-class separability. These issues lead to additional challenges in masked face recognition. Figure2.3 and Figure 2.2 shows these challenges

Masks with printed faces: Recently, printed masks with face images have attracted the attention of some sections of the population. Some of these masks either contain the image of the lower half of a face or the complete face image. Such variation in masks further complicates the task of face recognition. Wearing such masks may hamper the performance of face recognition models. Figure 2.4 shows example of people wearing face print masks containing someone's face. This may lead to fooling of the system.

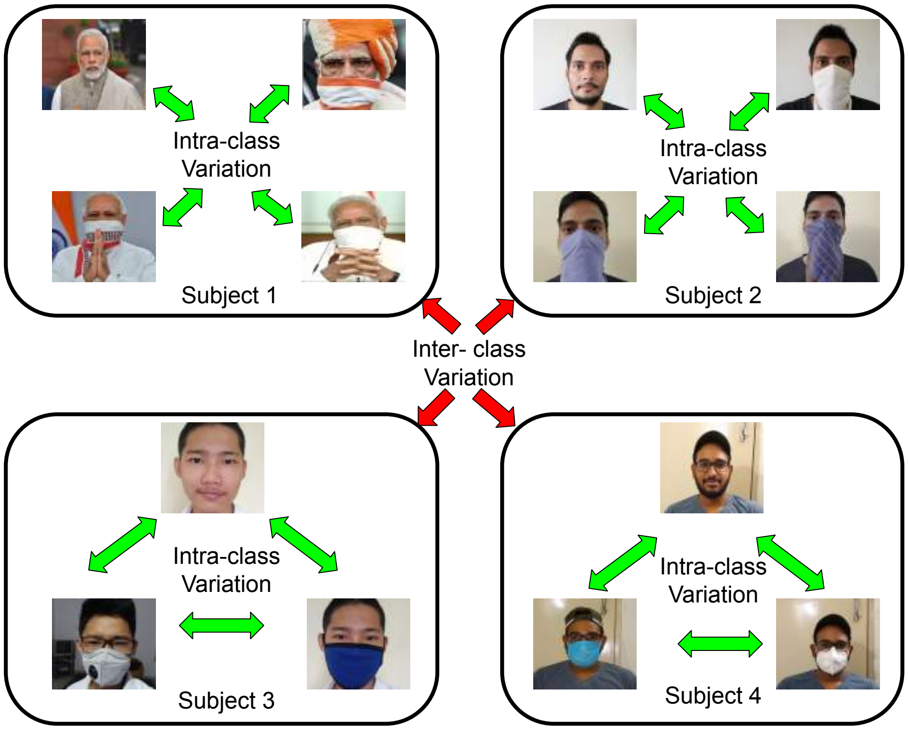


Figure 2.3: Summarizing the challenges of masked face recognition captured in the proposed IMFW dataset. All the subjects are representing diversity in demography and attire. Subject 1 and Subject 2 are showing the use of "gamcha" and "handkerchief" for covering faces.

Figure 2.4: Image showing how face print can pose a challenge for face detection and recognition

3 Literature survey

Face recognition for occluded faces is an arduous task. Different methods have been developed to recognize faces in such cases. Some algorithms first detect the occluded region from the face, and then recognition is done using the unoccluded part of the face. As in most real case scenarios, the occluded part is generally the eye and mouthpart due to the use of specs and scarfs. Some approaches are studied for detecting and recognizing occluded faces. In this section, we will be going to discuss a few of them. In one such method, based on this [14], a Principal Component Anaysis subspace was created using the patches of faces that do not have occlusion, and the nearest neighbour algorithm was used to detect the occlusion free part of the face. Another research used binary classification [15]) for searching obstructed parts, by dividing faces into regions and applying SVM to classify that part as occluded or not. A multitask learning algorithm was also implemented [16] that divides the face into four parts and checks if they are occluded or not.

Other algorithms detect the visible part first, and recognition is done using that visible part. As face occlusion is random, if the eye is occluded, then the mouth part is selected, and if a scarf/mask covers the mouth, then recognition is done using eye region. The methods proposed by [17] Non-negative matrix factorization (NMF) [18] that adaptively estimates the visible face using loss reconstruction. This method does not need any prior information regarding occlusion. A technique [19] was used that adds MaskNet in between CNN layers, that learns image features without the interference of occluded parts.

After detection of the occluded region, we need a partial face recognition algorithm to identify the face using features from the visible part. For partial face recognition, either feature extraction will try to extract full face feature, or in the comparison method, a robust recognition model is developed that can recognize a face in occlusion. For feature extraction, Multiscale Double Supervision Convolutional Neural Network (MDSCNN) [20] model was proposed that was trained on different patches of face to extract features. Different such facial patches training will lead to full-face feature extraction. This algorithm was time-consuming.

To make the recognition system robust and using a model that can match occluded face features with a gallery image. A Lophoscopic PCA method [21] was proposed that can recognize faces in the absence of some part. Different subspaces were created like the left eye, right eye, mouth etc. and were used while training and had different weights. Similarity measures can be used for partial face recognition [22], Local scale-invariant feature transform (SIFT) features were extracted from facial regions and compared with gallery images. Another variant of this is used in [23] that used geometric features along with visual features. Deep learning methods were also proposed for partial face recognition. A sliding window method [24] that matches a face part using a sliding window. Due to recognition of many loss function introduces recently over the years such as triplet loss[25], sphereface loss[26] ,cosface loss[27] and arcface loss[4], recognition accuracy has shown to significantly improved

Till now, we have seen methods over occlusion based detection and recognition. Not much literature is available that specifically discusses masked face detection and recognition methods. Some of the few approach that addresses them are discussed in the following parts. [28] suggests a casscade structure for masked face detection. It consists of three cascade convolutional neural networks (i) Mask-12, followed by (i) Mask-24-1i, and (iii) Mask-24. Each of these CNN is a binary classifier with classification ability from low to high. This framework used WIDER dataset for training which is fine-tuned with MASKED FACE dataset proposed by [28]. The result of testing on the testing set of masked shows an accuracy 86.6% at 87.8% recall.

LLE-CNNs is for masked face detection is another approach proposed by [29] for detecting masked faces. It consists of three modules: (i) proposal module followed by (ii) embedding module and (iii) verification module. The proposal module extracts the candidate region using CNN built by P-net and features are extracted with VGG pre-trained face model. The embedding module refines the descriptors with respect to its nearest neighbours which are obtained by synthesizing faces and non-faces. The verification module is done using deep neural network. It uses MAFA [29] and reaches an overall accuracy of 76%. However, the accuracy decreases as an increase in occlusion degree.

A recent approach for masked face recogniton was given by [30]. HGL method used HSV color channel instead for RGB. The H channel of HSV of images is used for training of CNN to extract feautres and SVM is used for classification. The result when the model is trained using MAFA [29] is 93% for frontal face and 87% for side face.

4 Problem definition and Objective

Wearing masks occludes a significant portion of the face. This occlusion in faces has put forward another set of challenges for masked faces. The occluded faces do not only make face recognition a challenge but also tasks like face detection and gender detection difficult. Further, the problems in the Indian context take another shape as India is a diverse country with different demography and attire worn by the people. People in India wear various kinds of masks, from simple surgical ones to the ones having varied prints as well as masks with face prints. Further, Indians are often seen using clothing like 'gamcha,' 'stoles,' and 'handkerchief' in place of masks. The problems in the Indian context need to be analyzed separately. However, there is a limit to the availability of databases that can be used for such analysis. We divide the problems due to masked faces as:

1. Masked Face Detection: Given an image with the person wearing masks, locate the region of the face. As the masks occlude a major portion of the face, the detection task needs to be done only by the binocular region of the face.

2. Gender Detection: Given a masked face, identify the gender of the person behind the mask. Though gender is an easy problem, however, once we only have half of the face region, this problem also becomes a challenge.

3. Masked Face Recognition: Given a masked face, identify the person behind the masks. In masked faces, the recognition has to be done only by the non-occluded parts. Further, the occlusion can be increased due to the attire of the person. Thus, face recognition is a problem for masked faces.

To analyze the performance of different tasks, we need to have a dataset that addresses all these concerns. With such idea in mind, we define the objective of our project work as follows:

1. To create a dataset that addresses the problems of masked faces in the Indian context.

2. To propose an algorithm based on a multi-task network that performs the multiple tasks of face detection, mask detection, and gender detection simultaneously.

3. To benchmark the performance of existing face recognition models on the proposed masked face dataset.

5 Datasets

Although there are large scale dataset that are publicly available for to train the face recognition methods for instance [31], [32], [33], [34], very few dataset deals with the problem of occlusion. Though sudden demand of face recognition system for masks face has given rise to datasets MAFA[29], and [35]. Table ?? summarize the datasets for masked face detection and recognition. Due to the limited Dataset for masked face, we present two new sets of Dataset in constrained and unconstrained settings : (i)Indian Masked Face Dataset(IMFD) (ii) Indian Masked Faces in the Wild(IMFW)

5.1 IMFW

The proposed IMFW dataset is collected under unconstrained settings with large variations in pose, background, illumination, resolution, and the type of masks worn by the subjects. Images are captured in different lighting conditions with varying backgrounds. Additionally, the images have large pose (-90^0 to $+90^0$) variations. Variation in resolution (low to high) of the images is also considered during data collection. Multiple constraints present in the dataset, along with the diversity in the subjects with respect to demography, attire pose challenges to face recognition algorithms. Further, the masks worn by the subjects increase the difficulty of recognizing faces by automated systems. The proposed IMFW dataset consists of three different sets with different number of subjects collected via different modes. For masked face recognition, face recognition algorithms are required to match the masked faces with the enrolled non-masked images. Therefore, for each subject, both masked and non-masked images are collected. Table 5.2 summarizes the statistics of the proposed dataset. The following subsections discuss the details of the three sets in the proposed dataset.

1. Set 1: Indian Celebrity As shown in Fig. 5.1(a), the Indian celebrity set of the proposed IMFW dataset contains 40 Indian celebrities with 435 images, including Bollywood actors/actresses, television stars, sports personalities, and politicians. These images are downloaded from the Internet. The majority of the downloaded images are taken in an unconstrained environment with varying poses and resolution.

2. Set 2: Instagram This set contains 377 images of 40 subjects downloaded from Instagram. We collected masked and non-masked images of Indian people with a public profile. Similar to Set 1, the majority of the images are taken in unconstrained settings, which includes variation in illumination conditions and backgrounds. Sample images of this set are shown in Fig. 5.1(b).

3. Set 3: Indian Crowd The images of this set are collected from the common people who volunteered to contribute to the dataset. Images are collected in both constrained and unconstrained environments. Images are captured using a mobile phone with a 48-megapixel rear camera and a 13-megapixel front camera. This set contains 120 subjects with 562 images. Apart from the unconstrained settings, variations in the type of masks used by the subjects are also considered during data collection. A wide variety of masks, including surgical, colored, N95, and printed, are used by the subjects.

DATASET	DESCRIPTION
WIDER FACE[34]	This dataset is a standard dataset,for face detection. This is collected from the WIDER datasets which are available to the public. IIt contains about 32,202 images with 393,703 labelled faces, which have variability in pose, occlusion and scale
DFW [36]	It consists of 11157 images of 1000 people. Each subject has disguides, impersonator and validation images along with normal image.
AR [37]	Images of 56 women candidates and 70 men candidates are taken which have variation of illumination conditions, expressions and occlusion by scarf or sunglasses. Total images are more than 4000 in color.
MASKED FACE dataset [28]:	Proposed by [28].The dataset is collected from web images and has 200 images.
IIIT-Delhi face database.[38]	Developed by the team of IIIT Delhi .
MS-Celeb-1M [33]	This dataset contains large scale images obtained from the web for the task of recognising faces. The dataset accomodate images of 10 Million Faces.
MAFA [29]	Proposed by [29]. Collected from internet over 30,811, 35,805 images with masks. The images in this dataset, has different degree of occlusion and orientation.Occlusion is by masks.
MFDD [35]	Proposed by [35]. This built dataset has around 24,771 images of masked face. It can be used for building robust face recognition models.
RMFRD [35]	Proposed by [35]. Images of 525 subjects in masks which are around 5000 images and 90000 images without masks of those 525 people.
SMFRD [35]	Proposed by [35] .500000 images of faces created using simulation of masked faces of 10000 people.

Table 5.1: Summary of Datasets for Masked Face Detection and Recognition

Figure 5.1: Sample images of the proposed IMFW dataset. (a) Set 1: Indian Celebrity, (b) Set 2: Instagram, and (c) Set 3: Indian Crowd.

Table 5.2: Statistics of the proposed IMFW dataset.

	No. of images		No. of subjects
	With Mask	Without Mask	
Set 1: Indian Celebrity	214	221	40
Set 2: Instagram	140	237	40
Set 3: Indian Crowd	276	286	120
Total	630	744	200

Some subjects have used stoles and handkerchiefs as masks during data collection to emulate the real-world scenarios. Fig. 5.1(c) shows sample images of the Indian Crowd set.

5.2 Indian Masked Face Dataset

The dataset contains images of 300 subjects with and without masks. Each subject has worn five different types of masks, of which one mask contains a face print image and the rest of the masks have random color prints. The dataset contains six images per subject, one without a mask and the rest with different varieties of masks. All the images are captured with two cameras: i) a mobile phone camera of 48-megapixel (f/1.79, 1.6-micron) + 5-megapixel and (ii) a DSLR camera of 32.5 megapixels with a sensor size of 22.3 x 14.9 mm. Images are captured from a distance of 4 meters from the subjects.

All the images are taken in controlled settings in a room with a plain light color background with slight illumination variation. The masks worn by the subjects have a lot of variation ranging from plain masks to ones with colorful prints. The dataset contains subjects of different skin tones, demography, and attire. Due to the variation in the attire of the subjects (cap, veils, eyeglasses), the dataset contains varying degrees of occlusion apart from the occlusion due to masks. Table 5.3 shows the statistics of the dataset collected. Figure 5.2 shows the snapshot of the dataset collected (a) dslr and (b) redmi

Figure 5.2: Imgaes captured from two cameras (a) dslr, (b) Phone camera

Table 5.3: Statistics of the Masked Faces in Controlled setting.

	Female	**Male**	**Total**
No. of Subjects	**70**	**230**	**300**
Masked Faces	**350**	**1150**	**1500**
Non Masked faces	**70**	**230**	**300**

6 Methodology

6.1 Face Detection

To perform face Detection, we use a multitask framework inspired by Hyperface[39]. Our Multitask network performs face detection, mask detection, and gender classification. Our network consist of (i) Region Proposal Network, (ii) Classification Network. We use selective search algorithm from [40] to generate proposal regions. Here we have used ArcFace pre-trained on the CASIA-WebFace dataset [41] as the backbone of the network. We add a fully connected layer of which outputs a dimension of 256 feature vector. To this network, we add three layers of a fully connected network of dimension 256. For each task, we add an output layer of size two to each of 256 layer output.

1. Face Detection: Bounding boxes extracted using selective search are labeled as the face if the overlap of the bounding box with the ground truth is greater than 0.75. If this overlap is less than 0.35, then it is labeled as non-face. These bounding boxes are then passed to our classification network, where it is classified as a face or non-face.

2. Mask Detection: For each face region, we have a label as a mask or no mask. During training is the input to the classification is a face region, then it predicts if the mask is present or not, and the loss is propagated to the network; otherwise if the network receives the non-face region in that case, the training loss for mask detection layers will be zero.

3. Gender Classification: For each face region, we have labeled it as male or female. During training is the input to the classification is a face region, then its prediction is the face is male or female, and the loss is propagated to the network, otherwise if the network receives the non-face region in that case, the training loss for gender classification layers will be zero

All the tasks are trained using softmax loss. The pipeline of the proposed algorithm is shown in Figure 6.1

Protocol: Both the Datasets are divided into 70% training and 30% testing sets.

Implementation Details: Resnet18 pretrained modelis used as backbone for multi-task network. The layers of the backbone model are freezed while other layers are trainable. We also replace the the backbone with 3 blocks of densenet and perform the experiments.

6.2 Face Recognition

The performance of existing deep face recognition models is evaluated on the proposed IMFW dataset. For this purpose, two different experiments are performed. The first experiment is performed to evaluate the performance of four pre-trained deep face recognition models namely, VGGFace [42], ResNet50 (trained on the VGGFace2 dataset) [43], LightCNN29 [44], and ArcFace [4] on the proposed IMFW dataset. The second experiment evaluates the performance of existing loss functions, Contrastive loss [1] and triplet loss [3] for masked face recognition. The Piple line used for Masked Face Recognition is shown in Figure 6.2

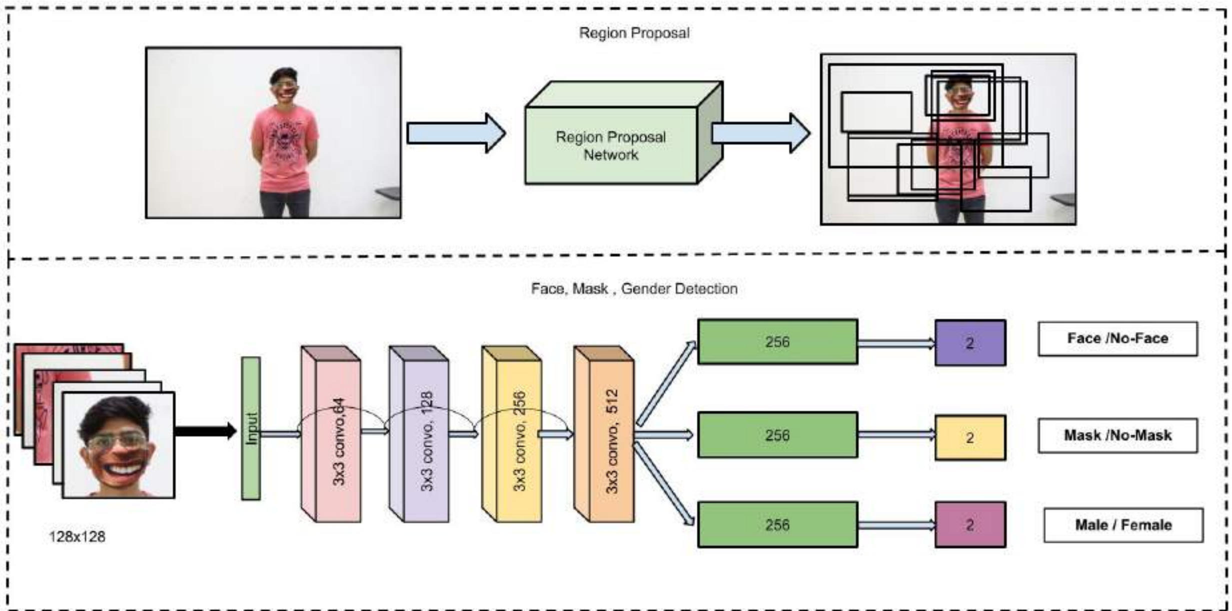


Figure 6.1: Pipeline of the Proposed Multitask network for Face Detection, Mask Detection and Gender Detection

Protocol: Experiments are performed by dividing both the datasets into training and testing sets with non-overlapping subjects. Each of the three sets (Indian Celebrity, Instagram, Indian Crowd) of IMFW dataset is split into training and testing partitions with 70% subjects in the training set and 30% subjects in the testing set. Further, the training and testing sets are divided into gallery and probe. Experiments are performed to emulate the real-world scenario of matching masked face images with non-masked enrolled images. Therefore the gallery contains non-masked images, and the probe contains masked face images of each subject. The gallery contains a single image per subject, while the probe contains multiple images per subject.

Implementation Details: LightCNN29 is used as the base network for model training using contrastive loss and triplet loss. Initial layers of the models are frozen, and the last ten layers are trained by minimizing the existing loss functions. Models are trained for 50 epochs with a learning rate of 0.00001. Adam optimizer is used with a batch size of 50. For model training using contrastive loss, the margin is set to 2. During triplet training, a margin of 0.4 is used.

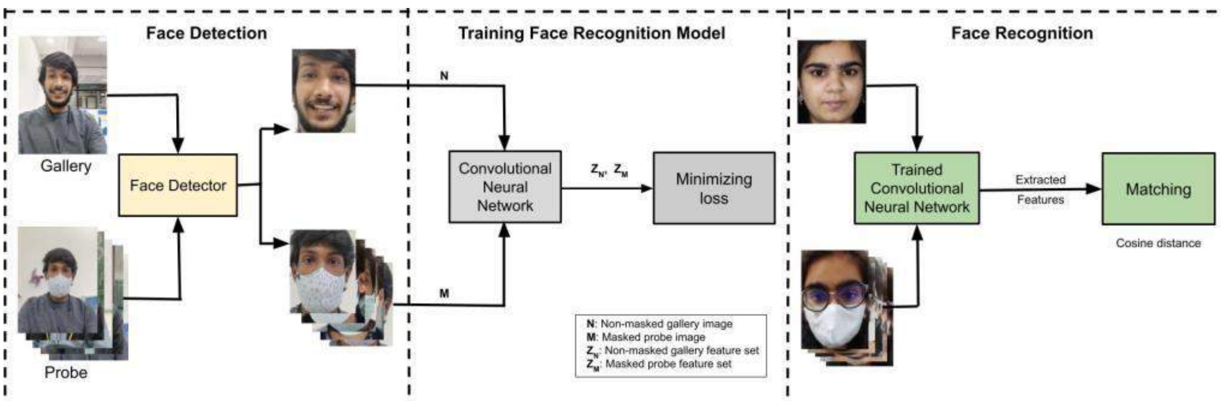


Figure 6.2: Pipeline of masked face recognition. The first block shows face detection. Training of a face recognition model is shown in the second block. Third block shows face recognition using trained model.

7 Experiments

We perform various experiments on the masked face datasets. These can be summarized as:

1. Masked Face Recognition using Pre-trained Models: We used existing deep models to evaluate their performance on both the datasets. The features are extracted from these models of gallery and probe image, and the extracted features are matches using cosine distance. This is done for both the datasets.

2. Masked Face Recognition using Existing Loss Functions : Here we fine-tune one of the existing models using existing loss funtion (contrastive loss and triplet loss) with both our dataset and evaluate their performance.

3. Multitask Network for Face Detection, Mask Detection and Gender Detection : We evaluated our proposed algorithm on both the dataset. We also tested the generalising ability of our model by training the model on one dataset and testing the model on different dataset. We also performed all the three task on a single network independently and compared them with our multi-task network.

7.1 Masked Face Recognition using Pre-trained Models

For establishing the baseline performance on the IMFW dataset and on the Indian Masked Face Dataset, four existing pre-trained deep face recognition models, VGGFace, ResNet50, LightCNN29, and ArcFace are used. LightCNN29 is pre-trained on the MS-Celeb-1M dataset [45]. The dataset contains approximately 10M images of 1M subjects. ArcFace is pre-trained on the CASIA-WebFace dataset [41]. The dataset contains 0.5M images of 10K subjects. Features extracted from the gallery and probe of the testing set are matched using Cosine distance. Table 7.1 shows the identification accuracy at rank 1, rank 5, and rank 10 on IMFW Dataset and Table 7.2 on Indian Masked Faces. For the experiments, we have combined the sets that are created by downloading images from the web in IMFW Dataset. Therefore, results are reported on combined Set 1 and Set 2, Set 3, and the full dataset (IMFW). It is observed that existing deep models do not perform well for recognizing masked face images. ArcFace is one of the state-of-the-art face recognition models on IMFW. However, it achieves only 62.96%, 61.76%, and 57.96% identification accuracy at rank 1 on (Set 1 + Set 2), Set 3, and the full dataset, respectively. Fig. 7.1 shows the cumulative match characteristic (CMC) curves. The low baseline performance indicates the need for sophisticated face recognition algorithms for masked face recognition in unconstrained settings. The Results on Indian Masked Faces were comparatively much better than IMFW Dataset. We can clearly see that arc face model it achieves rank 1 accuracy of 94% on Dslr images and 9% on Mobile Phone images.

7.2 Masked Face Recognition using Existing Loss Functions

This experiment is performed to evaluate the performance of existing loss functions on the proposed IMFW dataset and Indian Masked Face Dataset. Two existing loss functions, contrastive loss and triplet

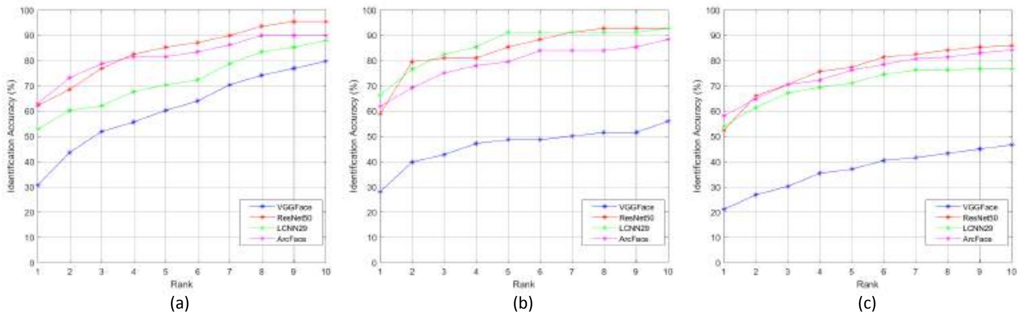


Figure 7.1: CMC curve of the four pre-trained deep face recognition models for masked face recognition on (a) Set 1 + Set 2, (b) Set 3, and (c) complete IMFW.

loss, are used to train the LightCNN-29 model. Here, our aim is to match masked faces with their corresponding non-masked images in the gallery. Therefore, pairs are created by taking one masked image and one non-masked image during training using contrastive loss. Similarly, during model training using triplet loss, triplets are generated by taking non-masked images as the anchor and masked images as positive and negative. Fig. 6.2 shows the pipeline for masked face recognition. In the first step, facial regions are segmented using Tiny Face detector [46] and resized to 128 × 128 resolution. Next, models are trained by minimizing existing loss functions to reduce the intra-class separation and increase the inter-class separability. During testing, features are extracted from the gallery and probe of the trained model. Finally, matching is performed using Cosine distance. Table 7.3 shows the identification accuracy at ranks 1, 5, and 10 using the existing algorithms. It is observed that model training using existing algorithms enhance the performance at least by 10% at rank 1. For instance, the identification accuracy at rank 1 on (Set 1 + Set 2) increases by 10.19% and 14.82% using contrastive loss and triplet loss, respectively, compared to the pre-trained model performance. It is important to observe that the performance of the algorithms is better on Set 3 compared to others. Set 3 contains images captured in both constrained and unconstrained settings, while the majority of the images in Sets 1 and 2 are taken in unconstrained settings. This further highlights the challenges of masked face recognition under unconstrained environmental conditions.

We have further benchmarked the proposed dataset against the LFWA dataset [47]. For the experiment, we have followed the protocol similar to the proposed one. We have sampled 200 subjects with 1374 images from the LFWA dataset and divided the dataset into training and testing sets with non-overlapping subjects (70% subjects in the training set and 30% in the testing set). Using this protocol, the pre-trained LCNN29 model yields an accuracy of 96.11% at rank 1 on the LFWA dataset. The fine-tuned LightCNN29 enhances the performance to 97.22% rank 1 accuracy using triplet loss. However, on the IMFW dataset, LightCNN29 yields 53.97% and 66.47% rank 1 accuracies using pre-trained and triplet loss, respectively. The low performance on the IMFW dataset highlights the challenges of unconstrained masked face recognition in the Indian context.

Table 7.1: Identification accuracy (%) of existing pre-trained deep face recognition models on the proposed IMFW dataset.

Models	Set1 + Set2			Set 3			Complete IMFW		
	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
VGGFace	30.55	60.18	79.62	27.94	48.52	55.88	21.02	36.93	46.59
ResNet50	62.03	85.18	95.37	58.82	85.29	92.64	52.27	77.27	85.79
LCNN29	52.77	73.37	87.96	66.17	91.17	92.64	53.97	71.02	76.70
ArcFace	62.96	81.48	89.81	61.76	79.41	88.23	57.96	76.13	84.09

Table 7.2: Identification accuracy (%) of existing pre-trained deep face recognition models on the proposed Indian Masked Facedataset.

Models	DSLR			Mobile Camera		
	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
VGGFace	52.61	78.78	87.10	49.47	77.35	86.93
LCNN29	90.79	98.08	98.78	93.03	97.56	98.95
ArcFace	94.94	98.08	98.95	94.07	97.73	98.60

Table 7.3: Identification accuracy (%) of existing algorithms on the proposed IMFW dataset using the LightCNN-29 model.

Models	Set1 + Set2			Set 3			Complete IMFW		
	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
Pre-trained	52.77	73.37	87.96	66.17	91.17	92.64	53.97	71.02	76.70
Contrastive Loss	62.96	83.33	92.59	77.94	92.64	95.58	63.06	80.68	85.22
Triplet Loss	67.59	83.33	92.59	84.88	91.17	98.52	66.47	81.25	86.36

Table 7.4: Identification accuracy (%) of existing algorithms on the proposed Indian Masked Face dataset using the LightCNN-29 model.

Models	DSLR			Mobile Camera		
	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
Pre-trained	90.76	98.08	98.78	93.03	97.56	98.95
Contrastive Loss	98.03	100	100	98.60	100	100
Triplet Loss	99.41	100	100	100	100	100

Table 7.5: Results of Multi-task Network using Backbone model as Resnet 18 and Backbone model Densenet when trained and tested on same set.

Task	Resnet18		Densenet	
	IMFD	IMFW	IMFD	IMFW
Face Detection	99.58	99.4	83.73	91.94
Mask Detection	96.00	84.68	85.18	54.50
Gender Detection	89.93	72.67	82.96	52.40

Table 7.6: Results of Multi-task Network using Backbone model as Resnet 18 and Backbone model Densenet when trained and tested on different dataset.

Task	Resnet18		Densenet	
	IMFD	IMFW	IMFD	IMFW
Face Detection	99.81	99.63	98.36	60.55
Mask Detection	69.78	66.67	15.85	46.69
Gender Detection	57.48	64.26	82.37	50.33

7.3 Multitask Network for Face Detection, Mask Detection and Gender Detection

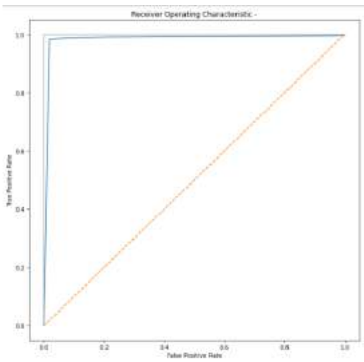
We use Arcface pre-trained on the CASIA-WebFace dataset [41] as our backbone network. The features extracted from the backbone layers are passed into the classification layers. Moreover, the loss of the network for all the output is calculated using cross-entropy loss. This loss is then backpropagated to the network. We also show the performance of the network of three blocks of densenet [48] is used as a backbone network. We evaluate the network by training and testing on the same dataset and training and testing on different datasets. We have also evaluated the performance of all the tasks when performed on a single task network independently.

Table 7.5 shows the performance of the network when evaluated on the same dataset. Table 7.6 shows the performance of the network when evaluated on cross datasets. We can see the performance of face detection reaches 99% for both the dataset whereas for mask detection, it reaches 96% and 89.93% for IMFD and 84.68% and 72.67% for IMFW. This shows that the IMFW dataset is still a challenging dataset for mask and gender detection problems. Figure 7.2 and Figure 7.4 shows the roc curve for the two datasets for same dataset training and testing using Resnet1 backbone. Roc curve for cross-data testing for two datasets are shown in Figure 7.3 for IMFD and Figure 7.5 for IMFW. We can see that when the model is evaluated using a different dataset than it was trained, then the performance of the model degrades. When the backbone of the multi-task network is replaced with three blocks of the Densenet model, the performance reduces significantly. This could be associated with the Densenet not being pretrained on any face dataset as in the case of our Arcface model. The roc curve for models with backbone of Densenet Model is given in Figure 7.6, and Figure 7.8 for same dataset testing and Figure 7.7, and Figure 7.9 for cross-data testing. We see the results of all the tasks performed independently on a single network in table 7.7. The detection accuracy, in this case, is similar to our multi-task network. However, it is clear that a single network performs better than a multi-task for mask detection and gender detection. From all the results, we can see that the IMFW dataset is a tougher dataset compared to the IMFD dataset, especially for mask and gender detection. The results of faces detecting using our proposed multi-task network are shown in Figure 7.10

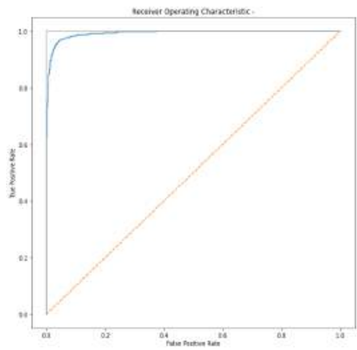
Table 7.7: Results of all the tasks when performed independently.

	IMFD	**IMFW**
Face Detection	98.15	99.53
Mask Detection	98.37	53.3
Gender Detection	94.22	44.44

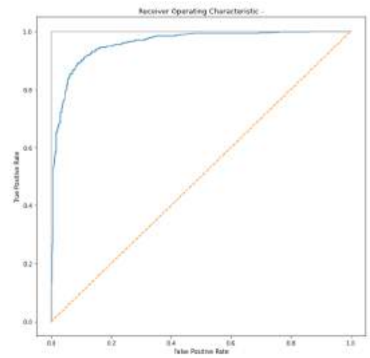
Figure 7.2: ROC curves of all the tasks , when performed on IMFD, when trained and tested on same dataset (backbone network: Resnet18)



(a) Face Detection

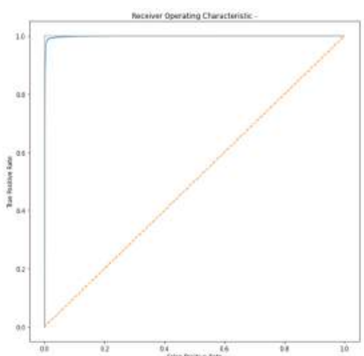


(b) Mask Detection

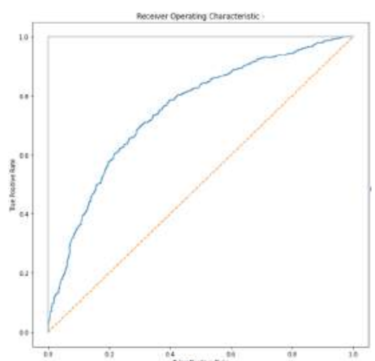


(c) Gender Detection

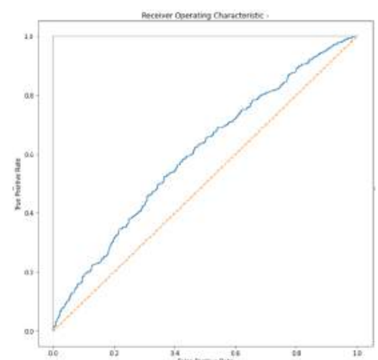
Figure 7.3: ROC curves of all the tasks , when performed on IMFD, when trained and tested on different dataset (backbone network: Resnet18)



(a) Face Detection

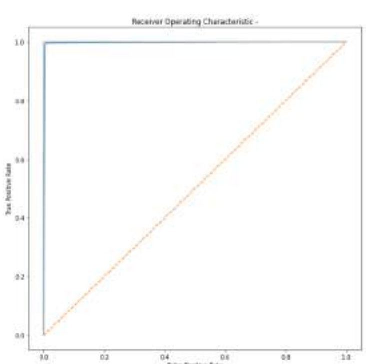


(b) Mask Detection

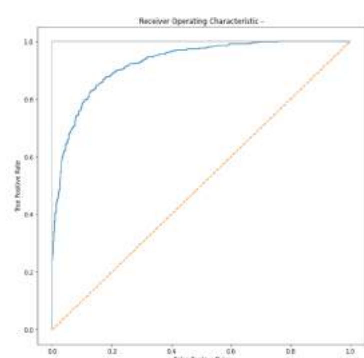


(c) Gender Detection

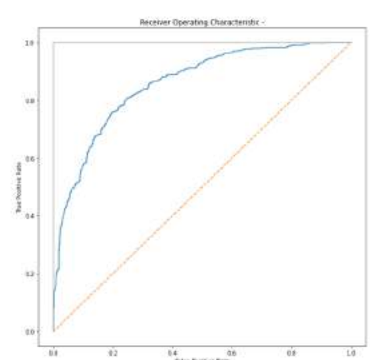
Figure 7.4: ROC curves of all the tasks , when performed on IMFW, when trained and tested on same dataset (backbone network: Resnet18)



(a) Face Detection

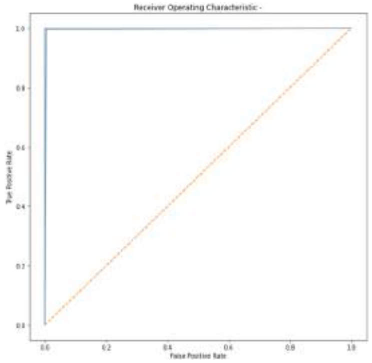


(b) Mask Detection

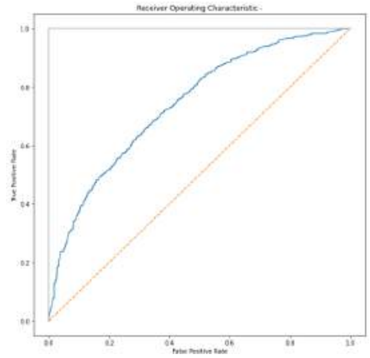


(c) Gender Detection

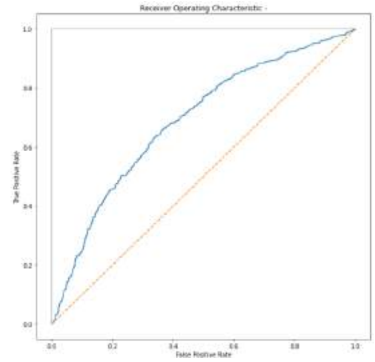
Figure 7.5: ROC curves of all the tasks , when performed on IMFW, when trained and tested on different dataset (backbone network: Resnet18)



(a) Face Detection

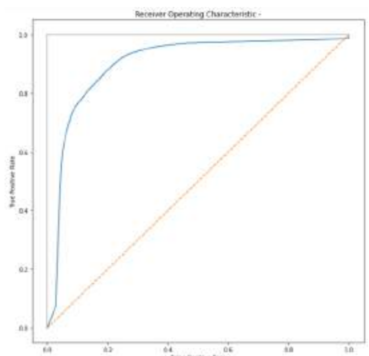


(b) Mask Detection

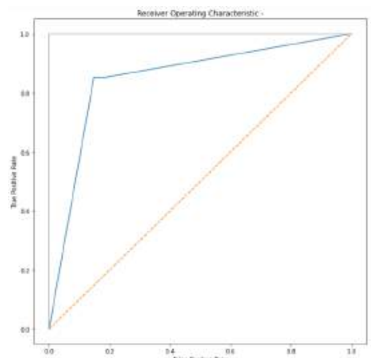


(c) Gender Detection

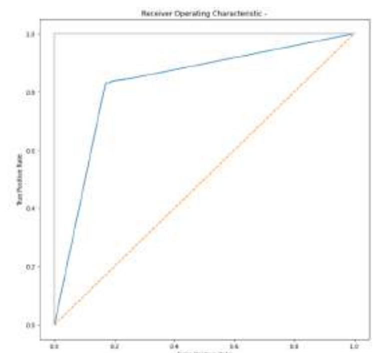
Figure 7.6: ROC curves of all the tasks , when performed on IMFD, when trained and tested on same dataset (backbone network: Densenet)



(a) Face Detection

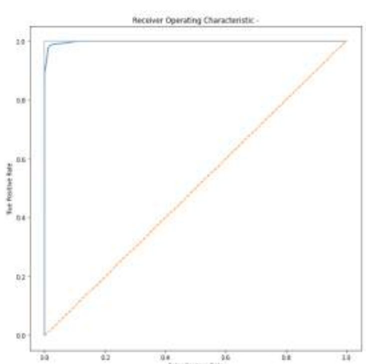


(b) Mask Detection

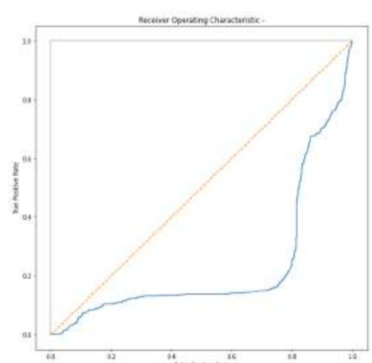


(c) Gender Detection

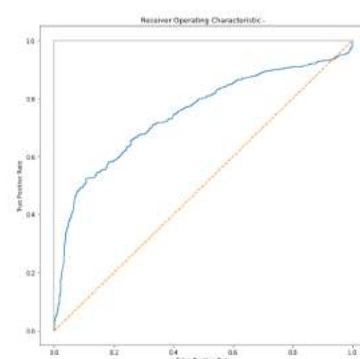
Figure 7.7: ROC curves of all the tasks , when performed on IMFD, when trained and tested on different dataset (backbone network: Densenet)



(a) Face Detection

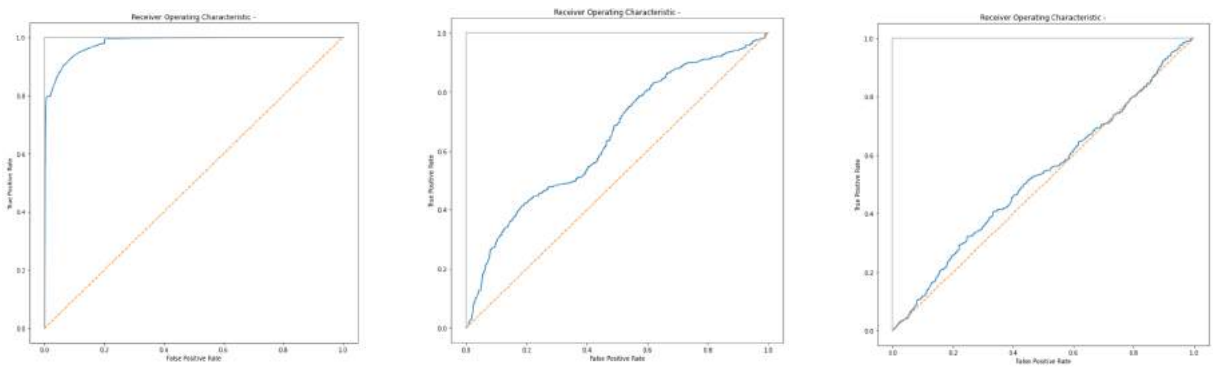


(b) Mask Detection



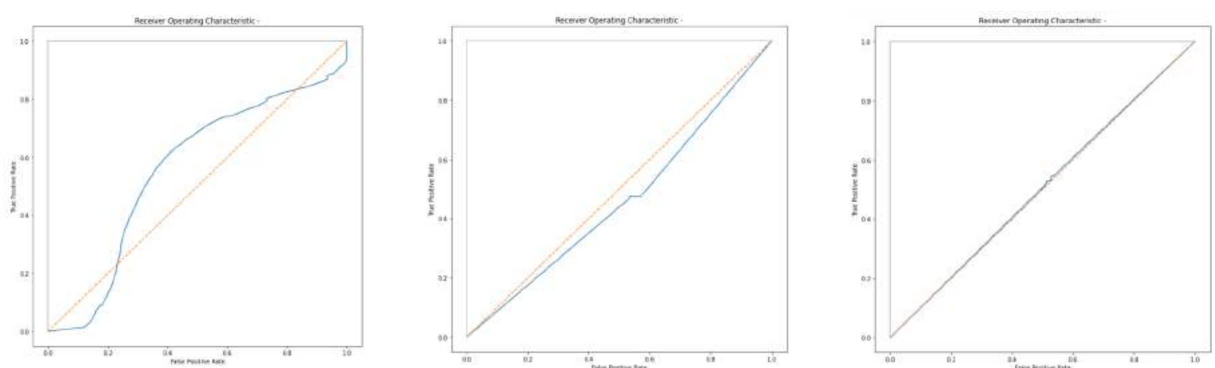
(c) Gender Detection

Figure 7.8: ROC curves of all the tasks , when performed on IMFW, when trained and tested on same dataset (backbone network: Densenet)



(a) Face Detection (b) Mask Detection (c) Gender Detection

Figure 7.9: ROC curves of all the tasks , when performed on IMFW, when trained and tested on different dataset (backbone network: Densenet)



(a) Face Detection (b) Mask Detection (c) Gender Detection

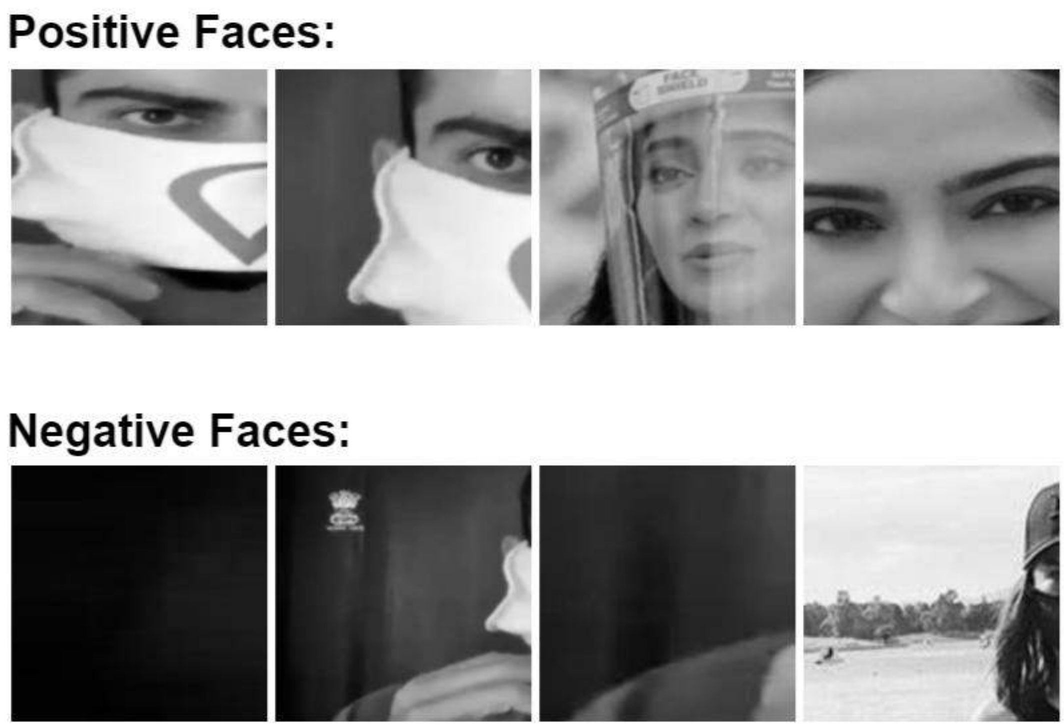


Figure 7.10: Results of Face Detection by our proposed multitask network (i) Images in the top row shows the results of Positive Faces , (ii) Images in the bottom row shows the results of Negative Faces .

8 Summary and Future plan of work

COVID-19 pandemic has posed several challenges including masked face recognition. In the "new normal", faces are covered at workplaces and public-places. In such a scenario, performing face detection and recognition is a major challenge as existing algorithms generally do not perform well when the faces are obfuscated. The problem is further exacerbated when the attire (masks) diversity show very high intra-class and very low inter-class variations. This paper presents the novel Datasets Indian Masked Faces in the Wild database which includes the attire diversity and samples collected in unconstrained settings and Indian Masked Face Dataset collected in constrained settings. We propose a novel multi-task framework based algorithm that performs face detection as well as mask and gender detection. We can see that how enviornment setting can effect the performance of the models. We also show Baseline experiments of the state-of-the-art algorithm on two datasets. Baseline experiments show that the masked face recognition is still an arduous task and require dedicated research efforts however if the images are captured in controlled settings, we can still perform recognition using existing methods with good accuracy.

For Future Work:

1. We will develop an algorithm that performs well for masked face recognition using constrained settings.

2. We will further analyze how masks with face print can hamper the performance of face detection and face recognition or can it be used to attack on the system.

Publications

Mishra, Shiksha, et al. ”Indian Masked Faces in the Wild Dataset.”, 2021 IEEE International Conference on Image Processing.

References

[1] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[2] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha, "Improving cross-resolution face matching using ensemble-based co-transfer learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[4] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[5] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1273–1280, 2017.

[6] H. Fang, W. Deng, Y. Zhong, and J. Hu, "Generate to adapt: Resolution adaption network for surveillance face recognition," in *European Conference on Computer Vision*. Springer, 2020, pp. 741–758.

[7] P. Majumdar, S. Chhabra, R. Singh, and M. Vatsa, "Recognizing injured faces via scifi loss," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2020.

[8] X. Zhang, R. Zhao, Y. Qiao, X. Wang, and H. Li, "Adacos: Adaptively scaling cosine logits for effectively learning deep face representations," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[10] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6738–6746.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[12] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[13] S. R. Ekenel H.K., "Why is facial occlusion a challenging problem?" *Tistarelli M., Nixon M.S. (eds) Advances in Biometrics. ICB 2009. Lecture Notes in Computer Science, vol 5558. Springer, Berlin, Heidelberg.*, 2009.

[14] H. J. Oh, K. M. Lee, S. U. Lee, and C.-H. Yim, "Occlusion invariant face recognition using selective lnmf basis images," in *Computer Vision – ACCV 2006*, P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 120–129.

[15] Zhaohua Chen, Tingrong Xu, and Zhiyuan Han, "Occluded face recognition based on the improved svm and block weighted lbp," in *2011 International Conference on Image Analysis and Signal Processing*, 2011, pp. 118–122.

[16] Yizhang Xia, Bailing Zhang, and F. Coenen, "Face occlusion detection based on multi-task convolution neural network," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015, pp. 375–379.

[17] W. Ou, X. Luan, J. Gou, Q. Zhou, W. Xiao, X. Xiong, and W. Zeng, "Robust discriminative nonnegative dictionary learning for occluded face recognition," *Pattern Recognit. Lett.*, vol. 107, pp. 41–49, 2018.

[18] S. Z. Li, Xin Wen Hou, Hong Jiang Zhang, and Qian Sheng Cheng, "Learning spatially localized, parts-based representation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[19] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 773–782.

[20] L. He, H. Li, Q. Zhang, Z. Sun, and Z. He, "Multiscale representation for partial face recognition under near infrared illumination," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–7.

[21] S.-M. H.-F. Yang, "Robust face recognition under different facial expressions, illumination variations and partial occlusions," *International Conference on Multimedia Modeling*, 2011.

[22] J. Hu, J. Lu, and Y. Tan, "Robust partial face recognition using instance-to-class distance," in *2013 Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.

[23] R. Weng, J. Lu, J. Hu, G. Yang, and Y. Tan, "Robust feature set matching for partial face recognition," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 601–608.

[24] W. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong, "Partial person re-identification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 4678–4686.

[25] X. Dong and J. Shen, "Triplet loss in siamese network for object tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[26] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 212–220.

[27] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5265–5274.

[28] W. Bu, J. Xiao, C. Zhou, M. Yang, and C. Peng, "A cascade framework for masked face detection," in *2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2017, pp. 458–462.

[29] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with lle-cnns," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 426–434.

[30] S. Li, X. Ning, L. Yu, L. Zhang, X. Dong, Y. Shi, and W. He, "Multi-angle head pose classification when wearing the mask for face recognition under the covid-19 coronavirus epidemic," in *2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD IS)*, 2020, pp. 1–5.

[31] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 464–473.

[32] 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)*, 2018, pp. 67–74.

[33] Y. H. X. H. J. G. Yandong Guo, Lei Zhang, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," *Compter Vision and Pattern Recognition*, 2016.

[34] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5525–5533.

[35] B. H. Z. X. Q. H. H. W. P. Y. K. J. N. W. Y. P. H. C. Y. M. Z. H. J. L. Zhongyuan Wang, Guangcheng Wang, "Masked face recognition dataset and application," *Computer Vision and Pattern Recognition*, 2020.

[36] 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.

[37] A. Martinez and R. Benavente, "The ar face database," *CVC Technical Report*, 1998.

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

[39] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 1, pp. 121–135, 2017.

[40] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders, "Segmentation as selective search for object recognition," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1879–1886.

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

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

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

[44] 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.

[45] 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*. Springer, 2016, pp. 87–102.

[46] P. Hu and D. Ramanan, "Finding tiny faces," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1522–1530.

[47] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[48] 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*, 2017, pp. 4700–4708.