Dual Sensor Indian Masked Face Dataset

Shiksha Mishra[1], Puspita Majumdar[1,2], Muskan Dosi[1], Mayank Vatsa[1], Richa Singh[1]
[1]IIT Jodhpur, India, [2]IIIT-Delhi, India

Abstract— **With the advancements in deep learning technologies, real-world applications like face detection, gender prediction, and face recognition have achieved human-level performance. However, the emergence of the COVID-19 pandemic brought new challenges to existing deep learning algorithms. People are forced to wear a mask to limit the spread of COVID-19. These face masks occlude a significant portion of the face, thereby posing multiple challenges to existing algorithms. Images captured using surveillance cameras have a low resolution which hinders the model performance. Along with this, skin tone, ethnicity and attire also play a significant role in detection and recognition performance. India is a large country with huge diversity in skin tone and attire of the people. To address the challenges due to masks in the Indian context, we propose a novel Dual Sensor Indian Masked Face (DS-IMF) dataset, which contains images captured in constrained environmental settings with a variety of masks and degrees of occlusion. Multiple experiments are performed on the DS-IMF dataset at different resolutions. Experimental results demonstrate the limitations of existing algorithms on low-resolution masked face images. The proposed dataset can be found at http://www.iab-rubric.org/resources/dsimf.html.**

I. Introduction

The advancements in deep learning technologies have led to the development of sophisticated algorithms for various tasks, including face detection, gender prediction, and face recognition [11], [13], [21]. Deep learning algorithms have achieved tremendous success and are used in various real-world applications, such as surveillance, authentication, and automation. However, the emergence of the COVID-19 pandemic has put a bar on progress and brought new challenges to the deep learning community. In COVID-19, wearing face masks is mandatory in public places worldwide. Thus, automated systems are required to check the presence of masks on people's faces in regulated areas to limit the spread of COVID-19. Nevertheless, face masks occlude a significant portion of the face, thereby posing multiple challenges to existing algorithms and automated systems. For instance, traditional face recognition systems at public places like airports and railway stations can not effectively recognize faces in the presence of masks. Further, the images captured using surveillance cameras are generally of low resolution that poses additional challenges. Fig. 1 illustrates various challenges due to masks.

In recent years, several masked face datasets have been proposed to perform various studies, including the analysis of existing face recognition algorithms on masked faces and designing algorithms to detect masked faces [2], [6], [19].

Fig. 1. Illustrating various challenges due to masks. (a) Same subject wearing different masks, (b) occlusion due to attire along with the mask, and (c) low resolution masked faces.

However, most of the datasets contain faces of Caucasian and North Asian demography. The literature shows that skin tone plays a vital role in recognition performance [10]. Thus, datasets from multiple demographic groups with ethnic variations are required to thoroughly analyze existing algorithms (in the presence of masks). Recently, Mishra et al. [14] proposed the Indian Masked Faces in the Wild (IMFW) dataset, containing masked faces of Indian people. The IMFW dataset is limited in size (only 1374 images) and captured in unconstrained environmental settings with significant variations in pose, illumination conditions, and backgrounds, making it challenging to analyze the performance variation due to masks. The dataset also lacks other demographic information such as gender. India is a large country with massive diversity in skin tone, facial features, and attire. Such diversity among people of the same country poses additional challenges to existing algorithms.

To promote the research on masked faces in the Indian context, in this research, we propose a novel **Dual Sensor Indian Masked Face (DS-IMF)** dataset. The dataset contains masked and non-masked images of 300 subjects captured using two devices in constrained environmental settings. The dataset is also annotated with gender information. Multiple experiments are performed to benchmark the performance of existing algorithms for masked face detection, gender prediction, mask detection, and masked face recognition in the Indian context. All the experiments are performed at varied resolutions to analyze the effect of resolution combined with the challenges of masked faces on existing algorithms.

II. Related Work

In the literature, researchers have proposed few masked face datasets to analyze the effect of occlusion on different algorithms. Earlier in 2017, Ge et al.[6] proposed the MAsked FAces dataset (MAFA) to address the problem of

978-1-6654-3176-7/21/$31.00 ©2021 IEEE

(a) (b)

Fig. 2. Sample images of the proposed DS-IMF dataset. (a) Set 1 and (b) Set 2. The subjects represent the diversity in attire, facial features, and skin tone. The variations in attire lead to different degrees of occlusion.

face detection in the presence of occlusion due to masks. The dataset contains 30,811 Internet images and 35,806 masked faces with variations in pose and the degree of occlusion. The authors further manually annotated six attributes for each masked face. With the onset of the COVID-19 pandemic, the creation of masked face datasets has gained momentum, and several datasets have been proposed in the past two years. Wang et al.[19] proposed three different datasets, Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD) for masked face detection and recognition. MFDD is created for building robust detection algorithms that contain 24,771 masked face images downloaded from the Internet. All the images are labeled and annotated for the presence of masks, along with the position coordinates of the masked faces. RMFRD is created by crawling online resources for the frontal face images of public figures. The dataset contains 5,000 masked and 90,000 non-masked images of 525 people. SMFRD is prepared by simulating masks on previous non-masked faces datasets such as LFW [8], and WebFace [22] datasets to increase the volume and diversity of the masked face recognition dataset. The dataset contains 500,000 face images of 10,000 subjects.

Recently, Cabani et al.[2] proposed three types of masked face datasets for masked face detection, namely, Correctly Masked Face Dataset (CMFD), Incorrectly Masked Face Dataset (IMFD), and global masked face detection (MaskedFace-Net), created by combining CMFD and IMFD datasets. Flickr-Faces-HQ (FFHQ) [9] dataset, containing 70,000 high-quality face images, has been used to create the MaskedFace-Net dataset by defining a mask-to-face deformable model. The dataset is created for (i) detecting whether people have worn a mask or not and (ii) whether the mask is worn correctly or not. Thermal-Mask Dataset is proposed by Queiroz et al. [16] containing 153,360 images in both visible and thermal spectrum. The dataset is created from SpeakingFaces dataset [1] by generating masks on the unmasked images. A small, realistic fabric face mask dataset is proposed by lionnie et al. [12] containing 176 images of 8 subjects (22 images per subject) with variations in the pose. Most of the datasets contain faces of Caucasian and North Asian demography. The first masked dataset in the Indian context, Indian Masked Faces in the Wild (IMFW), is proposed by Mishra et al. [14]. The dataset contains masked and non-masked images of 200 Indian subjects. The dataset is limited in size and does not contain other demographic information.

III. Dual Sensor Indian Masked Face Dataset

This research presents a novel masked face dataset of 300 subjects belonging to Indian ethnicity. As opposed to the existing IMFW dataset [14], this dataset is collected in constrained environmental settings to analyze the variation in model performance due to masks. During data collection, the diversity in attire, facial features, and skin tone of different subjects are considered to represent the Indian population. The dataset is collected using two different sensors, and the details of the sensors are given below.

- a DSLR camera of 32.5 megapixels with a sensor size of 22.3 x 14.9 mm and,
- a mobile phone camera of 48-megapixel (f/1.79, 1.6-micron) + 5-megapixel(used for depth perception).

We have divided the dataset into two sets: (i) **Set 1** and (ii) **Set 2**, depending on the sensor used to capture the

TABLE I
STATISTICS OF THE PROPOSED DS-IMF DATASET ('M' DENOTES *male* AND 'F' DENOTES *female*).

No. of	Set 1			Set 2		
	M	F	Total	M	F	Total
Subjects	229	71	300	229	71	300
Non-Masked Images	229	71	300	229	71	300
Masked Images	1145	355	1500	1145	355	1500
Total Images	1374	426	**1800**	1374	426	**1800**

images. The images captured using DSLR camera are kept in **Set 1** while the images captured using mobile phone camera are kept in **Set 2**. Figure 2 shows sample images of the proposed Dual Sensor Indian Masked Face (DS-IMF) dataset. The proposed dataset can be found at http://www.iab-rubric.org/resources/dsimf.html.

A. Details of Data Collection

Images using both sensors are captured at a distance of 1.5 to 2 meters from the subjects. All the images are collected in controlled settings in a room during daylight with a plain light color background and slight illumination variation. The subjects are made to stand straight in front of the camera to simulate the real-life framework of voluntary photo submission. The dataset contains subjects of different skin tones, facial features, and attire. Due to the variations in the subjects' attire, such as caps, veils, and eyeglasses, the dataset contains varying degrees of occlusion apart from the occlusion due to masks.

A total of 3600 images corresponding to 300 subjects were collected. During masked face recognition or mask detection, deep learning algorithms require both masked and non-masked images. Therefore, one non-masked image and five masked images are collected corresponding to each sensor for each subject. Thus, the DS-IMF dataset consists of 600 non-masked and 3000 masked images. Table I shows the dataset statistics. The masks used to collect the data have different colors, random prints, with some having face prints as well.

B. Annotation Details

Multiple annotations have been provided corresponding to each image for different experimental purposes. Thus, each image is annotated with subject id, gender, face coordinates, binocular region coordinates, and mask/non-mask information. All the annotations are done manually.

To anonymize the identity of the subjects, we have assigned a unique id to each subject. During the process of data collection, the gender information of each subject is recorded. Additionally, we have annotated the face and binocular region coordinates for each image with the help of the online VGG-Image annotator[1]. It records the x and y coordinate along with the width and height of the annotated region. Furthermore, the mask/non-mask information for each image is provided. We have annotated the type of mask worn by the subject (random printed or face printed masks).

To segregate the images captured using different sensors, the images in **Set 1** and **Set 2** are stored into two separate folders. Images of all the subjects follow the same naming convention and are labeled as **x_yy.JPG** where,

- **x** represent the subject id from 1 to 300 and,
- **yy** represent the mask/non-mask information as **nm**, **fp** or **rpn**. Here, **nm** represents non-masked images. Masked images are represented as **fp** or **rpn**, where **fp** represent images with face printed masks and **rpn** represent images with random printed masks. **n** in **rpn** denotes the number of the random printed mask whose value lies between 1 and 4.

The gender information is provided in a separate file with ID and Gender columns, where the ID column represents the subject id, and the corresponding gender column provides the gender information as 'f' for female and 'm' for male. Face and binocular region coordinates corresponding to each image are provided in two separate files for the two sets. Each file has three columns: (i) image label, (ii) region id with '0' for face coordinates and '1' for binocular region coordinates, and (iii) coordinates. The proposed DS-IMF dataset with multiple annotations can be used to explore different challenges of masked faces.

IV. EXPERIMENTAL SETUP

The performance of existing models is evaluated for various tasks on the proposed Dual Sensor Indian Masked Face (DS-IMF) dataset. For this purpose, four different experiments have been performed. All the experiments are performed by varying the resolution of the images to analyze the robustness of deep models in the presence of masks at varied resolutions. The details of the experiments are summarized below.

- The first experiment is performed to evaluate the performance of existing face detectors for the task of **masked face detection**. We used the MTCNN [23], S3FD [24], and RetinaFace [5] face detectors in our experiments.
- In the second experiment, the performance of existing models are evaluated for the task of **gender prediction**. For this purpose, four deep models namely, VGGFace [15], ResNet50 (trained on the VGGFace2 dataset) [3], LightCNN29 [20], and ArcFace [4] are trained to classify an input image into *male* or *female*.
- The third experiment evaluates the performance of the models for the task of **mask detection**. Models are trained to classify whether the subject in the input image has worn a mask or not.
- The fourth experiment is performed to evaluate the model's performance for **masked face recognition**. Experiments are performed under two settings: (i) pre-trained model evaluation and (ii) model training using existing loss functions. In the first setting, models are evaluated without training, while in the second setting, models are evaluated by training them using existing loss functions on the proposed DS-IMF dataset.

[1]https://www.robots.ox.ac.uk/ vgg/software/via/via.html

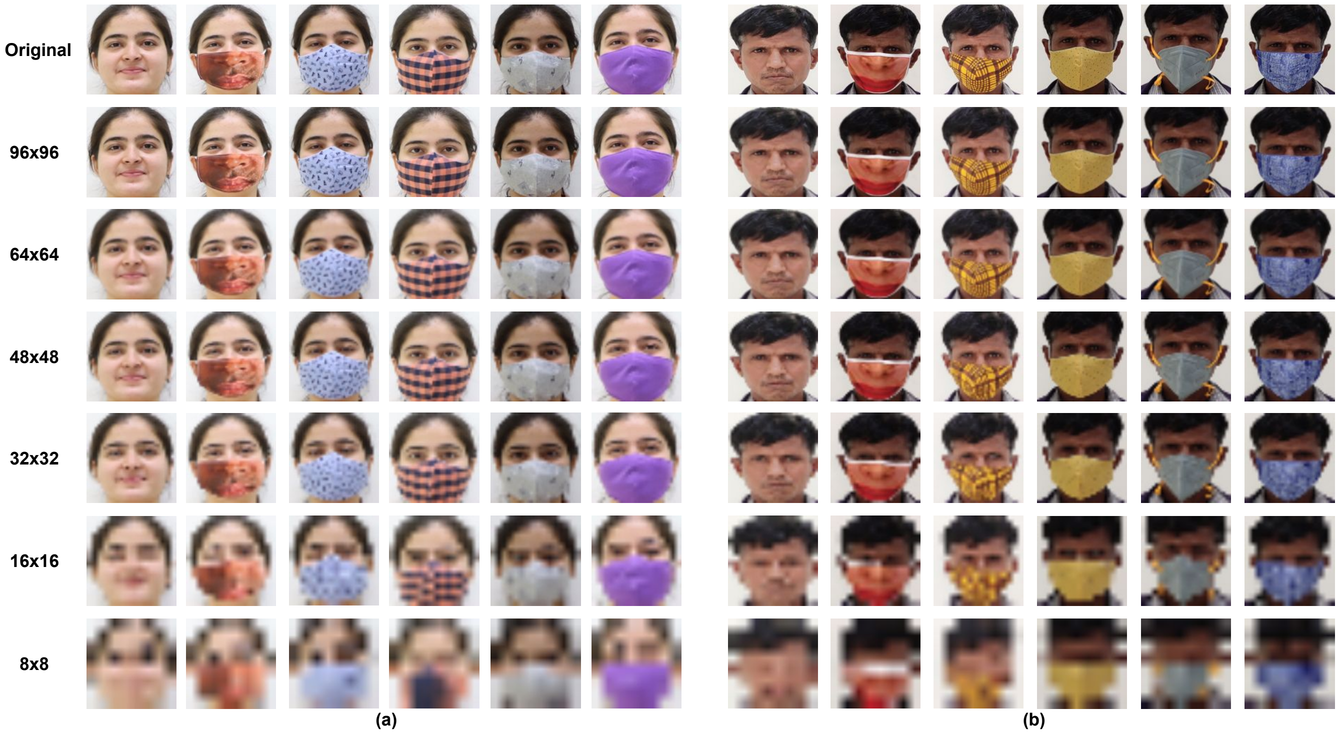

Fig. 3. Face images of the DS-IMF dataset at different resolutions corresponding to (a) Set 1 and (b) Set 2. Downsampled images have been bicubically interpolated to 128×128 resolution.

Protocol: All experiments have been performed on both sets with the same training and testing splits. The training and testing sets contain non-overlapping subjects with 70% subjects in the training set and 20% in the testing set. For the task of masked face recognition, we additionally divided the training and testing sets into gallery and probe, where the gallery contains only non-masked face images and the probe contains only masked face images. Experiments are performed at different resolutions following the same protocol. For masked face recognition experiments, the resolution of only the probe images is varied while keeping the gallery images at original resolution. This is done to emulate the real-world scenario of matching low-resolution real-time images with high-resolution enrolled images.

Implementation Details: All the experiments are performed on Nvidia GeForce GTX 1080 Ti. Implementation details of the experiments are discussed below.

Masked Face Detection: Face detectors are used to obtain the bounding box containing faces for an input image. We compute Intersection over Union (IOU) of the ground truth and the regions predicted by face detectors. For each image, if the IOU is greater than a threshold (0.4) then we consider it as a face.

Gender Prediction and Mask Detection: After the final convolutional layer for each task, all the models are trained separately by adding two dense layers. The dimensions of the dense layers for the VGGFace [15], ResNet50 [3], and LightCNN29 [20] models are 128 and 64. For the ArcFace [4] model, dense layers are of dimensions 512 and 256. Each dense layer is followed by ReLU activation. The models are trained using the cross-entropy loss function. The lightCNN29 model is trained for 30 epochs using Stochastic Gradient Descent optimizer with a learning rate of 0.0001, the momentum of 0.9, weight decay of 1e-4. Other models are trained for 20 epochs using Adam optimizer with a learning rate of 0.0001 for the first ten epochs, which is reduced by 0.1 after every five epochs. During training, the last few convolutional layers are trained along with the dense layers. For the ArcFace model, only the dense layers are trained while keeping the convolutional layers frozen.

Masked Face Recognition: LightCNN29 is used as the base network for model training using existing loss functions. Models are trained by freezing the initial convolutional layers and updating the last ten convolutional layers to minimize loss. The models are trained for 10 epochs for contrastive and triplet loss using Adam optimizer with 50 batch size and 0.00001 learning rate. The margin is set to 2.0 and 0.4 for the contrastive and triplet loss, respectively. For model training using ArcFace and CosFace loss, a batch size of 64 is used. The models are trained for 30 epochs using a Stochastic Gradient Descent optimizer with a learning rate of 0.01, the momentum of 0.9, weight decay of 5e-4.

V. Results and Analysis

Multiple experiments have been performed to analyze the challenges induced by masks coupled with the effect of resolution on existing algorithms in the Indian context. Fig. 3 shows sample images at different resolutions. The pipeline of different tasks performed on the DS-IMF dataset is shown

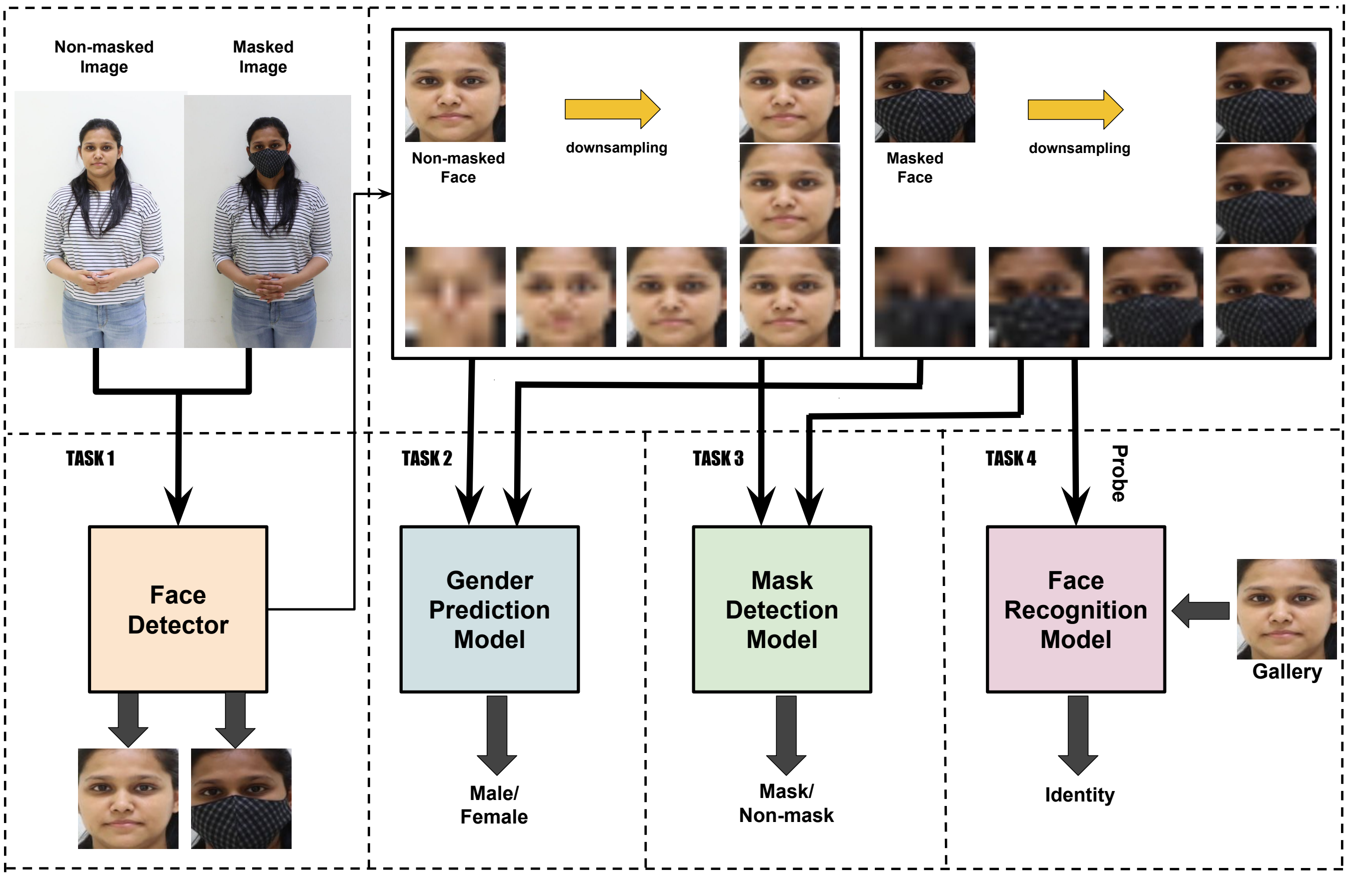

Fig. 4. Pipeline of different tasks performed on the DS-IMF dataset. The first task is to detect faces from the input images. The detected faces are given as input to perform other tasks. All the tasks are performed at different resolutions by downsampling the detected faces to required resolutions.

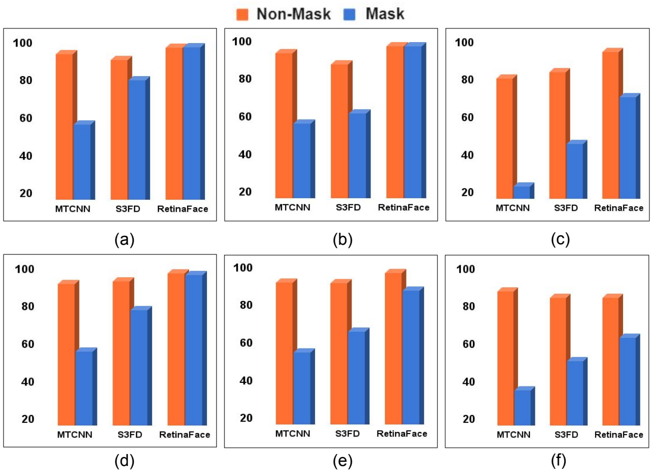

Fig. 5. Masked face detection on the DS-IMF dataset corresponding to (a-c) Set 1 and (d-f) Set 2. (a) and (d) show the results on the original resolution. (b) and (e) show the results on the images downsampled by a factor of 0.25. (c) and (f) show the results on the images downsampled by a factor of 0.06.

in Fig. 4. The following subsections provide a detailed experimental evaluation and analysis of all the experiments.

A. Masked Face Detection

We evaluated the performance of three face detectors on the proposed DS-IMF dataset. We observed that the regions predicted by the detectors contain non-facial regions along with the facial regions. Thus, we calculated the IOU of the regions predicted by the detectors with the ground truth region (obtained using the VGG-Image annotator). The regions for which the IOU is greater than a threshold are considered as faces. Multiple experiments are performed to select the threshold to get the desired result. The IOU is computed as:

$$IOU = \frac{Intersection(predicted\ region, ground\ truth)}{Union(predicted\ region, ground\ truth)} \quad (1)$$

Fig. 5 shows the detection accuracy on masked and non-masked faces obtained using the three detectors at different resolutions on the DS-IMF dataset. It is observed that most of the detectors fail to detect masked faces. The popular MTCNN detector could only detect 58.99% and 58.33% masked faces from Set 1 and Set 2 at original resolution. S3FD detector also shows low detection accuracy for masked faces. Face detection is the first step of the recognition pipeline. Other tasks such as gender prediction and mask detection also require the detected faces, as shown in Fig. 4. Low performance of face detectors on masked faces

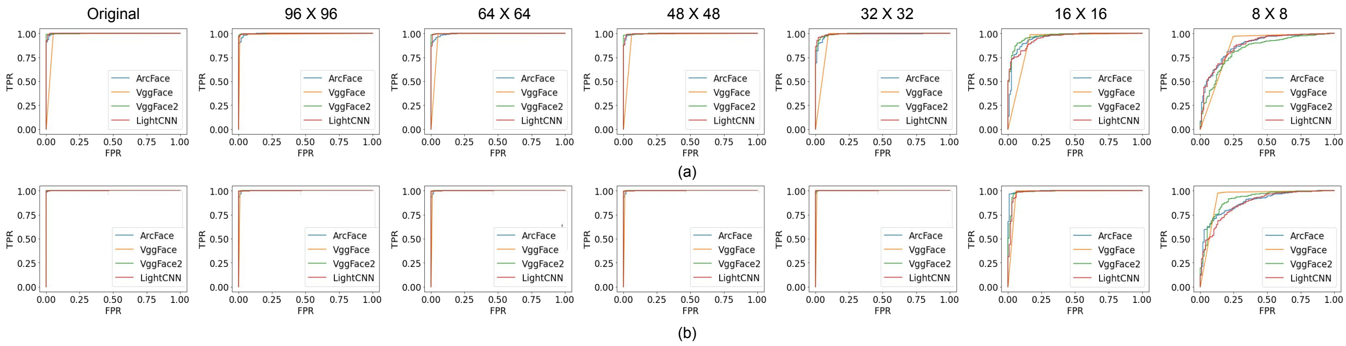

Fig. 6. ROC curves for (a) gender prediction and (b) mask detection using four models at different resolutions on Set 1 (VggFace2 denotes ResNet50 model trained on VggFace2 dataset).

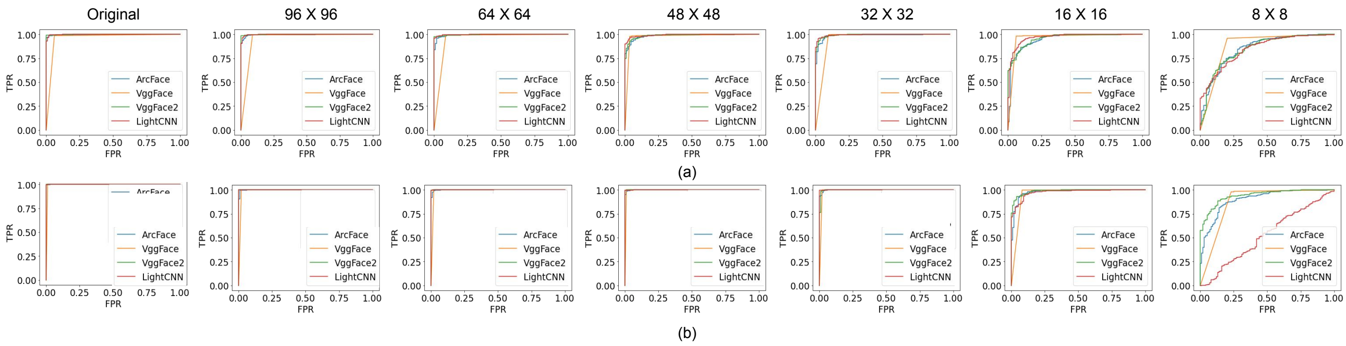

Fig. 7. ROC curves for (a) gender prediction and (b) mask detection using four models at different resolutions on Set 2 (VggFace2 denotes ResNet50 model trained on VggFace2 dataset).

adversely affects the performance of other tasks as well.

RetinaFace is one of the most recent face detectors that achieves high detection accuracy of more than 99% on both masked and non-masked faces at original resolution. However, significant degradation in the accuracy of all the models is observed at low resolution on masked images corresponding to both sets. For instance, when the images are downsampled by a factor of 0.06, the accuracy of RetinaFace detector reduces to 73.13% and 65.60% on masked faces corresponding to Set 1 and Set 2, respectively. This highlights that low resolution masked face detection is still an open problem.

B. Gender Prediction & Mask Detection

The models are trained separately for the task of gender prediction and mask detection. Fig. 6 shows the receiver operating characteristic (ROC) curve for the two tasks on Set 1. It is observed that all the models perform well at higher resolutions. However, as the resolution of the images decreases to 16×16 and 8×8 resolution, the model performance deteriorates. For instance, all the models achieve more than 97% accuracy upto 32×32 resolution for gender prediction. But the accuracy reduced to 93.63% and 88.14% at 16×16 and 8×8 resolutions, respectively, corresponding to the ArcFace model. A similar set of observations are drawn for the task of mask detection. The results on Set 2 are shown in Fig 7. Similar to Set 1, the performance of the models decreases at low resolution for the images of Set 2. This indicates that low resolution gender prediction (in the presence of mask) and low resolution mask detection is a challenging task.

C. Masked Face Recognition

Two different experiments are performed to analyze the effect of masked faces on recognition performance. The following discusses the results of the two experiments.

Pre-trained model evaluation: For baselining, the performance of the four pre-trained face recognition models are evaluated for the task of masked face recognition. To emulate the real-world scenario of matching masked images with enrolled non-masked images, we kept one non-masked image per subject in the gallery and the rest of the masked images of each subject in the probe. We extract the features from the gallery and probe images using the pre-trained models. Extracted features are then matched using cosine distance. We repeated the experiments for both sets of the DS-IMF dataset at different resolutions.

Table II shows the rank-1 identification accuracy of different models at varied resolution. It is observed that except VGGFace model, which achieves 52.61% and 49.47% rank-1 accuracy corresponding to Set 1 and Set 2, respectively, all other models achieve more than 90% rank-1 accuracy at original resolution. At low resolution, specifically beyond 48×48 resolution, the performance of all the models degrade

TABLE II

RANK-1 IDENTIFICATION ACCURACY (%) OF EXISTING PRE-TRAINED DEEP FACE RECOGNITION MODELS ON THE PROPOSED DS-IMF DATASET.

Model	Set 1							Set 2						
	Original	96×96	64×64	48×48	32×32	16×16	8×8	Original	96×96	64×64	48×48	32×32	16×16	8×8
VGGFace [15]	52.61	50.34	50.17	47.37	40.38	13.28	1.74	49.47	48.43	44.42	39.72	29.61	6.96	1.91
ResNet50 [3]	94.76	94.40	92.83	88.46	78.35	22.33	2.44	94.94	94.07	92.85	89.37	75.43	22.99	2.96
LightCNN29 [20]	90.76	89.70	89.00	87.95	80.10	20.24	2.44	93.03	92.16	90.59	88.85	76.30	17.59	2.26
ArcFace [4]	94.94	94.24	92.67	90.05	77.66	13.96	1.91	94.07	93.90	93.03	89.91	71.77	10.97	1.74

TABLE III

RANK-1 IDENTIFICATION ACCURACY (%) OF EXISTING LOSS FUNCTIONS ON THE PROPOSED DS-IMF DATASET USING THE LIGHTCNN29 MODEL.

Loss	Set 1							Set 2						
	Original	96×96	64×64	48×48	32×32	16×16	8×8	Original	96×96	64×64	48×48	32×32	16×16	8×8
Contrastive [7]	98.60	98.25	98.25	97.38	94.06	56.63	10.12	99.47	99.47	98.95	98.78	95.12	53.65	11.14
Triplet [17]	98.42	97.90	97.73	97.73	93.54	55.32	12.91	99.47	99.30	99.30	98.60	93.55	48.08	14.45
CosFace [18]	98.42	98.08	98.08	97.90	95.46	54.45	7.15	99.65	99.82	99.47	98.78	93.90	52.61	5.74
ArcFace [4]	98.42	98.25	98.25	97.20	95.46	68.76	8.20	99.65	99.65	99.30	98.43	96.16	54.18	7.31

significantly. A drop of more than 10% rank-1 accuracy is observed at 32×32 resolution compared to the original resolution, and all the models achieve approximately 2-3% accuracy at 8×8 resolution. This shows the ineffectiveness of existing face recognition models towards recognizing low resolution masked face images.

On comparing the model performance for Set 1 and Set 2, it is observed that at lower resolutions (beyond 48×48), the performance of the models is lower on Set 2 compared to Set 1. As mentioned earlier, the images of Set 1 are captured using DSLR camera, while the images of Set 2 are captured using mobile phone camera. The difference in model performance across different sets highlights the role of different sensors used to capture the images on recognition.

Model training using existing loss functions: We evaluated the existing loss functions, namely contrastive loss, triplet loss, cosface loss, and arcface loss on the proposed DS-IMF dataset. We trained the LightCNN29 model using these loss functions. These experiments are performed for each set of the DS-IMF dataset and for different resolutions. For model training using contrastive loss, pairs are created by taking one non-masked image and one masked image. Similarly, we created the triplets by considering the non-masked image as the anchor, one masked image of the same subject as positive, and one masked image of a different subject as negative. This type of pairing is followed with the aim of optimizing the model to match the masked probe images with the enrolled non-masked images.

The results are shown in Table III. It is observed that the performance of the LightCNN29 model improves after model training on the DS-IMF dataset using existing loss functions. All the models achieve more than 98% and 99% rank-1 accuracy for Set 1 and Set 2, respectively. The performance of the models (at higher resolution) on the DS-IMF dataset is comparable to the performance on other non-masked face datasets [8], owning to the fact that the images are captured in controlled settings. Table III further shows a significant drop in model performance beyond 32×32 resolution. An average drop of around 90% accuracy is observed at 8×8 resolution compared to the original resolution for all the models on both sets. It is clear from the results that even after training the models using state-of-the-art loss functions, masked face recognition is challenging at lower resolution.

For all the experiments, we observe that the models perform well at high resolution. It is important to note that the dataset is collected in constrained environmental settings to analyze the challenges due to masks. The results indicate that the existing algorithms are able to handle the challenges of occlusion due to masks. But when the resolution of the images decreases, the performance of the models degrade, especially for the task of face recognition. This highlights the challenges of low resolution masked face recognition. We believe that the problem requires the attention of the research community and focused research efforts to develop robust algorithms.

VI. CONCLUSION

In the events of unexpected challenges, people have always learned to upgrade themselves and modify their way of living. The worldwide crisis of COVID-19 is one such predicament that completely reformed our lives. As masks have become an essential part of daily wear and a compulsion to wear them in workspaces, where face authentication is required, we need to update our systems to detect and recognize masked faces. With this motivation, we propose a new dataset Dual Sensor Indian Masked Face (DS-IMF) dataset, to capture the variation in ethnicity, attire, and skin tone of the Indian population. In this paper, we explored different problems of masked faces on the DS-IMF dataset at different resolutions. Baseline experiments show that while state-of-the-art algorithms improvise well on these problems, some reform is still required to deal with lower resolution images.

ACKNOWLEDGEMENTS

This work is supported through research grants from RAKSHAK (iHub Drishti Foundation and DST) and MEITY. P. Majumdar is partly supported by DST Inspire Ph.D. Fellowship. M. Vatsa is also partially supported through Swarnajayanti Fellowship.

REFERENCES

[1] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis, and H. A. Varol. Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams. *Sensors*, 21(10):3465, 2021.

[2] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi. MaskedFace-Net–a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144, 2021.

[3] 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*, pages 67–74, 2018.

[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*, pages 4690–4699, 2019.

[5] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*, 2019.

[6] S. Ge, J. Li, Q. Ye, and Z. Luo. Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2682–2690, 2017.

[7] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1735–1742, 2006.

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

[9] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[10] K. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer. Issues related to face recognition accuracy varying based on race and skin tone. *IEEE Transactions on Technology and Society*, 1(1):8–20, 2020.

[11] X. Li, S. Lai, and X. Qian. Dbcface: Towards pure convolutional neural network face detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[12] R. Lionnie, C. Apriono, and D. Gunawan. Face mask recognition with realistic fabric face mask data set: A combination using surface curvature and glcm. In *IEEE International IOT, Electronics and Mechatronics Conference*, pages 1–6, 2021.

[13] P. Majumdar, S. Chhabra, R. Singh, and M. Vatsa. Recognizing injured faces via SCIFI loss. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1):112–123, 2021.

[14] S. Mishra, P. Majumdar, R. Singh, and M. Vatsa. Indian masked faces in the wild dataset. In *IEEE International Conference on Image Processing*, pages 884–888, 2021.

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

[16] L. Queiroz, H. Oliveira, and S. Yanushkevich. Thermal-mask–a dataset for facial mask detection and breathing rate measurement. In *IEEE International Conference on Information and Digital Technologies*, pages 142–151, 2021.

[17] 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*, pages 815–823, 2015.

[18] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018.

[19] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.

[20] 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*, 13(11):2884–2896, 2018.

[21] X. Yang, X. Jia, D. Gong, D.-M. Yan, Z. Li, and W. Liu. Larnet: Lie algebra residual network for face recognition. In *International Conference on Machine Learning*, pages 11738–11750. PMLR, 2021.

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

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

[24] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li. S3fd: Single shot scale-invariant face detector. In *Proceedings of the IEEE international conference on computer vision*, pages 192–201, 2017.