

INDIAN MASKED FACES IN THE WILD DATASET

Shiksha Mishra¹ Puspita Majumdar^{1,2} Richa Singh¹ Mayank Vatsa¹

¹ IIT Jodhpur, India ² IIIT-Delhi, India

ABSTRACT

Due to the COVID-19 pandemic, wearing face masks has become a mandate in public places worldwide. Face masks occlude a significant portion of the facial region. Additionally, people wear different types of masks, from simple ones to ones with graphics and prints. These pose new challenges to face recognition algorithms. Researchers have recently proposed a few masked face datasets for designing algorithms to overcome the challenges of masked face recognition. However, existing datasets lack the cultural diversity and collection in the unrestricted settings. Country like India with attire diversity, people are not limited to wearing traditional masks but also clothing like a thin cotton printed towel (locally called as “gamcha”), “stoles”, and “handkerchiefs” to cover their faces. In this paper, we present a novel **Indian Masked Faces in the Wild (IMFW)** dataset which contains images with variations in pose, illumination, resolution, and the variety of masks worn by the subjects. We have also benchmarked the performance of existing face recognition models on the proposed IMFW dataset. Experimental results demonstrate the limitations of existing algorithms in presence of diverse conditions.

Index Terms— Masked Face Recognition, COVID-19

1. INTRODUCTION

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]. However, the problem of recognizing faces under partial or heavy occlusion is still considered a challenging task [9]. 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. Masks occlude a fair amount of facial region, thereby making recognition difficult by face recognition algorithms. People wear different kinds of masks, from simple to the ones having many graphics that pose new challenges to automatic face recognition. Fig. 1 shows some of these challenges in unconstrained masked face recognition, including the variation in the type of masks.

Researchers have proposed several algorithms to overcome the challenges of occlusion in face recognition [10, 11, 12, 13]. However, very limited work has been done towards addressing the challenges of masked face recognition due to scarcity of masked face datasets. Ge et al. [14] have proposed MAsked FAcEs dataset (MAFA) containing around 35,806 images of masked faces with varying pose and degree of occlusion. The dataset is created by downloading images from the Internet. MaskedFace-Net is a recent dataset proposed by Cabani et al. [15] that focuses on detecting



Fig. 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.

masked faces as well as detecting whether the masks are worn correctly or not on the faces. Wang et al. [16] have proposed three datasets, namely, MFDD, RMFRD, and SMFRD. MFDD is created for building robust detection algorithms. RMFRD is created by using web images, and SMFRD is prepared by simulating masks on the LFW [17], and Webface [18] datasets for recognition purposes. Naser et al. [19] have also proposed a dataset of 20 subjects to test the performance of face recognition algorithms.

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 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 motivated us to propose a novel masked face dataset with Indian ethnicity, named as **Indian Masked Faces in the Wild (IMFW)** dataset. This dataset contains 200 subjects with both masked and non-masked images captured in an unconstrained environment. We believe that the proposed dataset will enable researchers to design sophisticated algorithms to overcome the challenges of masked face recognition in the Indian context. The following sections discuss the details of the proposed IMFW dataset, challenges in the IMFW dataset, and results of existing face recognition algorithms on the IMFW dataset.

2. INDIAN MASKED FACES IN THE WILD

We present the Indian Masked Faces in the Wild (IMFW) dataset of 200 subjects. The aim of this research is to overcome the challenges of unconstrained masked face recognition in the Indian context. Wearing “gamchas” and “stoles” as masks is a common trend in India. Variation in masks followed by the difference in skin tone variations pose new sets of challenges to face recognition as shown

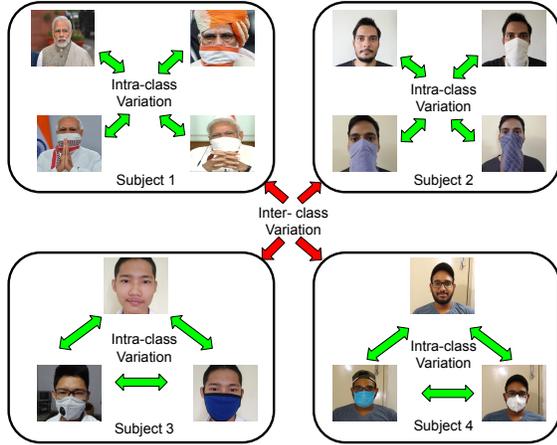


Fig. 2. Summarizing the challenges of masked face recognition captured in the proposed IMFW dataset. All the subjects are representing diversity in demography and attire. Subject 1 and Subject 2 are showing the use of "gamcha" and "handkerchief" for covering faces.

Table 1. 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

in Fig 2. These variations are included in the proposed dataset to cover the diversity in India¹. 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 1 summarizes the statistics of the proposed dataset. The following subsections discuss the details of the three sets in the proposed dataset.

2.1. Set 1: Indian Celebrity

As shown in Fig. 3(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.

2.2. 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. 3(b).

2.3. 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. Some subjects have used stoles and handkerchiefs as masks during data collection to emulate the real-world scenarios. Fig. 3(c) shows sample images of the Indian Crowd set.

2.4. Challenges in the IMFW dataset

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 challenges of face recognition in 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 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.

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.

3. EXPERIMENTAL RESULTS

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 [20], ResNet50 (trained on the VGGFace2 dataset) [21], LightCNN29 [22], and ArcFace [4] on the

¹The dataset and baseline models are available at: <http://www.iabrubric.org/resources/imfw.html>

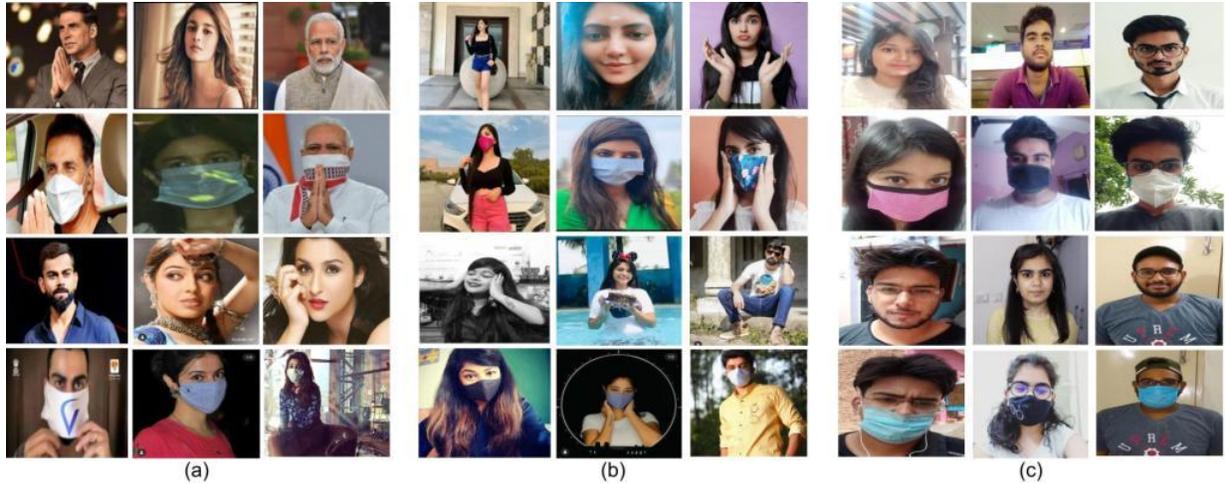


Fig. 3. Sample images of the proposed IMFW dataset. (a) Set 1: Indian Celebrity, (b) Set 2: Instagram, and (c) Set 3: Indian Crowd.

Table 2. Identification accuracy (%) of existing pre-trained deep face recognition models on the proposed IMFW dataset.

Models	Set 1 + Set 2			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

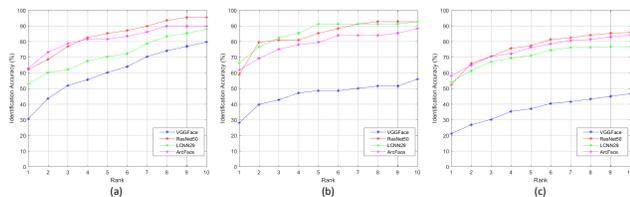


Fig. 4. CMC curve of the four pre-trained deep face recognition models for masked face recognition on (a) Set 1 + Set 2, (b) Set 3, and (c) complete IMFW.

proposed IMFW dataset. The second experiment evaluates the performance of existing loss functions, Contrastive loss [1] and triplet loss [3] for masked face recognition.

Protocol: Experiments are performed by dividing the IMFW dataset into training and testing sets with non-overlapping subjects. Each of the three sets (Indian Celebrity, Instagram, Indian Crowd) 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. Code is implemented in PyTorch. All the experiments are performed on a DGX station with Intel Xeon CPU, 256 GB RAM, and four 32 GB Nvidia V100 GPU cards.

3.1. Masked Face Recognition using Pre-trained Models

For establishing the baseline performance on the IMFW 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 [23]. The dataset contains approximately 10M images of 1M subjects. ArcFace is pre-trained on the CASIA-WebFace dataset [18]. 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 2 shows the identification accuracy at rank 1, rank 5, and rank 10. For the experiments, we have combined the sets that are created by downloading images from the web. 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. 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. 4 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.

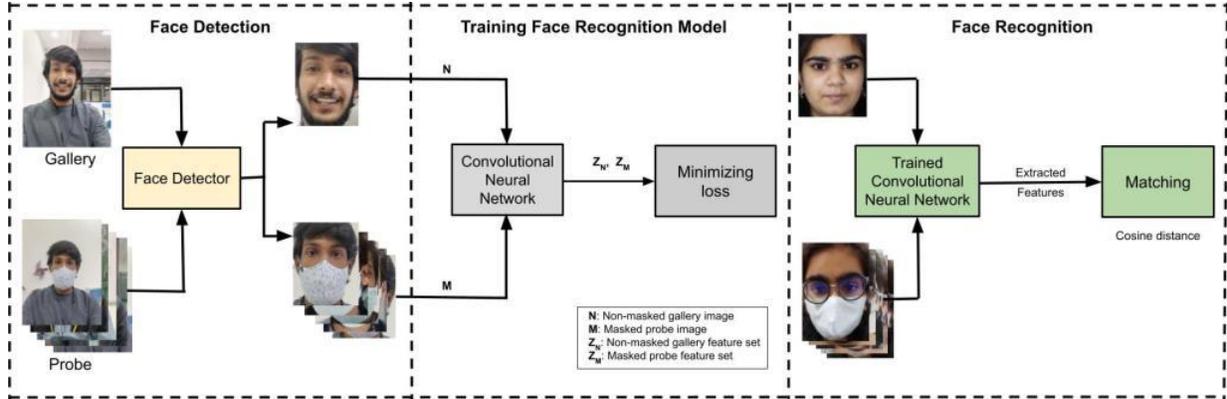


Fig. 5. Pipeline of masked face recognition. The first block shows face detection using Tiny Face detector. Training of a face recognition model is shown in the second block. Third block shows face recognition using trained model.

Table 3. Identification accuracy (%) of existing algorithms on the proposed IMFW dataset using the LightCNN-29 model.

Algorithms	Set 1 + Set 2			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	80.88	91.17	98.52	66.47	81.25	86.36

3.2. Masked Face Recognition using Existing Loss Functions

This experiment is performed to evaluate the performance of existing loss functions on the proposed IMFW dataset. Two existing loss functions, contrastive loss and triplet 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. 5 shows the pipeline for masked face recognition. In the first step, facial regions are segmented using Tiny Face detector [24] 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 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 [17]. 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.

4. CONCLUSION

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 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 Indian Masked Faces in the Wild database which includes the attire diversity and samples collected in unconstrained settings. Baseline experiments show that the masked face recognition is still an arduous task and require dedicated research efforts.

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. The authors would like to express their gratitude towards Muskan Dosi for helping in the process of data collection.

5. REFERENCES

- [1] Raia Hadsell, Sumit Chopra, and Yann LeCun, "Dimensionality reduction by learning an invariant mapping," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1735–1742.
- [2] Himanshu S. Bhatt, Richa Singh, Mayank Vatsa, and Nalini 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] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.
- [4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.
- [5] Angshul Majumdar, Richa Singh, and Mayank 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] Han Fang, Weihong Deng, Yaoyao Zhong, and Jiani Hu, "Generate to adapt: Resolution adaption network for surveillance face recognition," in *European Conference on Computer Vision*. Springer, 2020, pp. 741–758.
- [7] Puspita Majumdar, Saheb Chhabra, Richa Singh, and Mayank Vatsa, "Recognizing injured faces via SCIFI loss," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 1, pp. 112–123, 2021.
- [8] Xiao Zhang, Rui Zhao, Yu Qiao, Xiaogang Wang, and Hongsheng Li, "Adacos: Adaptively scaling cosine logits for effectively learning deep face representations," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10823–10832.
- [9] Dan Zeng, Raymond Veldhuis, and Luuk Spreeuwers, "A survey of face recognition techniques under occlusion," *arXiv preprint arXiv:2006.11366*, 2020.
- [10] Lingshuang Du and Haifeng Hu, "Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes," *Neurocomputing*, vol. 340, pp. 133–144, 2019.
- [11] Xiaowei Yuan and In Kyu Park, "Face de-occlusion using 3d morphable model and generative adversarial network," in *IEEE International Conference on Computer Vision*, 2019, pp. 10062–10071.
- [12] Xiao-Xin Li, Pengyi Hao, Lin He, and Yuanjing Feng, "Image gradient orientations embedded structural error coding for face recognition with occlusion," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 6, pp. 2349–2367, 2020.
- [13] Shiming Ge, Chenyu Li, Shengwei Zhao, and Dan Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3387–3397, 2020.
- [14] Shiming Ge, Jia Li, Qiting Ye, and Zhao Luo, "Detecting masked faces in the wild with LLE-CNNs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2682–2690.
- [15] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi, "MaskedFace-Net—a dataset of correctly/incorrectly masked face images in the context of covid-19," *Smart Health*, vol. 19, pp. 100144, 2020.
- [16] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al., "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093*, 2020.
- [17] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric 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.
- [18] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.
- [19] Naser Damer, Jonas Henry Grebe, Cong Chen, Fadi Boutros, Florian Kirchbuchner, and Arjan Kuijper, "The effect of wearing a mask on face recognition performance: an exploratory study," in *IEEE International Conference of the Biometrics Special Interest Group*, 2020, pp. 1–6.
- [20] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al., "Deep face recognition," in *British Machine Vision Conference*, 2015.
- [21] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2018, pp. 67–74.
- [22] Xiang Wu, Ran He, Zhenan Sun, and Tieniu 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.
- [23] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 87–102.
- [24] Peiyun Hu and Deva Ramanan, "Finding tiny faces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 951–959.