This ICCV workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

MaskOut: A Data Augmentation Method for Masked Face Recognition

Weiqiu Wang [a] Zhicheng Zhao [a,b] Hongyuan Zhang [a] Zhaohui Wang [a] Fei Su [a,b]
[a] School of Artificial Intelligence
[b] Beijing Key Laboratory of Network System and Network Culture
Beijing University of Posts and Telecommunications, Beijing, China
{wangweiqiu, zhaozc, buptzhy, Zhaohui_Wang, sufei}@bupt.edu.cn

Abstract

Deep learning methods have achieved great performances in face recognition. However, the performances of deep learning methods deteriorate in case of wearing a mask. Recently, due to the world-wide COVID-19 pandemic, masked face recognition attracts more attention. It is non-trivial and urgent to improve the performances in masked face recognition. In this work, a simple and effective data augmentation method, named MaskOut, is proposed. MaskOut replaces a random region below the nose of a face with a random mask template to mask out original face features. Our method is computing and memory efficient and convenient to combine with other methods. The experimental results show that the performances in masked face recognition are improved by a large margin with MaskOut. Besides, we construct a real-life masked face dataset, named MCPRL-Mask, to evaluate the performance of masked face recognition models.

1. Introduction

Inspired by DeepFace [21], research focus of face recognition has shifted to deep learning based methods [19, 20, 18, 16, 2, 17, 5] and the performances on the unconstrained condition are dramatically boosted to approach even surpass human. Therefore, face recognition is widely accepted as the prominent biometric technique for identity authentication and applied on various fields, such as finance, military, public security and daily life. However, since the outbreak of world-wide COVID-19 pandemic, wearing masks has become crucial to protecting public health. They have become a part of the new normal and their requirements in many aspects of daily life present a unique challenge for current face recognition systems which rely on data points from all aspects of the face. Due to the lower half face shielded by a mask, nearly half key features to discriminate face for different people are lost and the similarity of masks shorten inter-class distances increasing the difficulty to recognize.

There exist various public available non-masked face datasets that are of vital importance for both model training and testing, such as CASIA-WebFace [26], VGGFace2 [2], MS-Celeb-1M [10], Glint360K [1], IJB-A/B/C [13, 25, 15], LFW [11], and Megaface [12]. However, the performances of deep learning models trained on non-masked datasets deteriorate heavily on masked face recognition. While there are few masked datasets collected from real life scenarios due to the prevention of COVID-19 pandemic.

We address the above problems by introducing an augmentation strategy **MaskOut** and constructing a real-world masked face dataset to evaluate the performance of models for masked face recognition. Instead of simply removing the lower half face, our proposed MaskOut replaces the removed regions with mask templates, which is closer to reality. Besides, MaskOut incurs only negligible additional cost for training compared to the complex 3D projection methods [22]. It is complementary to previous non-masked face recognition methods because it operates on the data level, without changing internal representations or architecture.

2. Related Work

We present extensive experiments of MaskOut, and the results show that our method markedly improves the performance of masked face recognition.

Regional Data Augmentation: CutOut [6] and Random Erasing [28] augment data by removing random regions in images. Cutmix [27] fills the removed regions with patches from another training images. DropBlock [8] has generalized regional data augmentation to the feature space by dropout random regional features. All the above works give us a significant insight that it is effective to enhance the deep convolutional neural networks(CNNs) with regional data augmentation. For masked face recognition task, it is vital to generalize those excellent non-masked face recognition method to masked faces. Therefore, MaskOut is proposed to enhance the generalizability of existing non-masked face

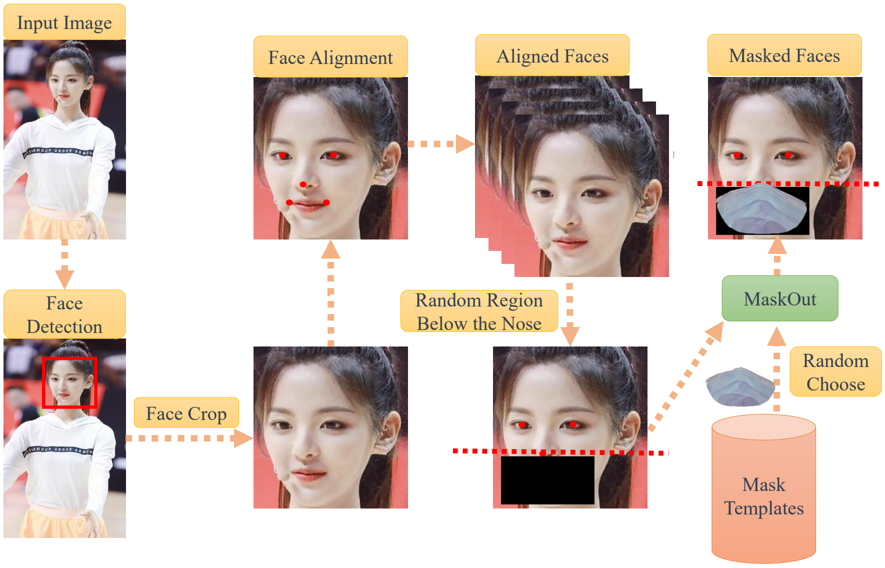

Figure 1. The pipeline of data prepocessing and MaskOut operation.

recognition methods by replacing random region below the nose in a face with a random mask template.

Masked face Recognition: Due to the outbreak worldwide COVID-19 pandemic and the regular epidemic prevention and control, masked face recognition becomes a crucial application demand in many scenarios. Facexzoo [22] propose FMA-3D to add virtual mask to the existing face images. FMA-3D synthesizes a photo-realistic masked face image requiring extra 106 facial landmarks detection and 3D face reconstruction method PRNet [7]. The procedure to add virtual masks is complex and timecomsuming, especially for large scale non-masked face datasets. Besides, the sythesized masked faces double the occupation of storage space and the time for training models. However, MaskOut incurs only negligible additional memory to store mask templates data and cost for training.

Masked face Dataset: Unlike non-masked face recognition, few masked face datasets are available for training and evaluation. RMFD [24] is a real-world masked face dataset by crawling the samples from the website. It contains 5,000 masked faces of 525 people and 90,000 normal faces. The number of masked faces is limited. Therefore, we construct a real-life masked face dataset, named MCPRL-Mask, including 14,023 masked images of 246 individuals, to evaluate masked face recognition methods. We recruit hundreds of volunteers wearing masks to record videos. Masked samples are captured from face videos full of variations. Details will be described in Section 4.

3. MaskOut

3.1. Motivation

The proper way to wear a mask is making sure that the mask fits to cover the nose, mouth and chin. Therefore, for the masked face recognition, the features below nose are all shielded by masks. Moreover, wearing similar masks, the similarities between faces of different people increase dramatically. Motivated by these issues, we introduce a simple and model-agnostic data augmentation routine, termed MaskOut. In a nutshell, random regional features below the nose of a face are replaced with a mask template drawn at random from mask templates data. MaskOut guides the recognition model to recognize a face from the partial features above the nose and introduces the noises of masks to improve the roubustness and generalization of the recognition model to masked face. Besides, MaskOut incurs a negligible computational overhead and can be efficiently utilized to train any network architecture.

3.2. Algorithm

Let $x \in \mathbb{R}^{W \times H \times C}$ denotes a training image. For the input training image, MaskOut replaces the data below the nose in x with a mask template data $t_i \in R^{W \times H \times C}$ drawn at random from mask templates data $\boldsymbol{T} \in R^{N \times W \times H \times C}$, where N is the number of mask templates. Face detection and alignment are the common data prepocessing procedure for face recognition task. After that, the training images are all aligned with standard face, which means the facial landmark's coordinates (e.g. left eye, right eye, nose, left mouth corner, and right mouth corner) are mapped to the same locations with the standard face. We denote the coordinates of the nose in a standard face as h_{nose}. Therefore, the random bounding box coordinates of the facial region below the nose for each aligned training image are $\boldsymbol{B} = (x_c, y_c, w, h)$. The coordinates of box are uniformly sampled by:

$$\begin{aligned} &x_c \sim Uinf(0, W), w = W\sqrt{1-\lambda}, \\ &y_c \sim Uinf(h_{nose}, H), h = (H - h_{nose})\sqrt{1-\lambda} \end{aligned} \quad (1)$$

where x_c, y_c, w and h are the center coordinates, the width and height of the boungding box, respectively. The scale factor λ is sampled from uniform distribution (0,1). The region $\boldsymbol{B}$ in x is removed and filled in with the random mask template data $\widetilde{t_i} \in R^{w \times h \times C}$, which is reshaped from t_i. MaskOut is formulated as:

$$\widetilde{x} = \boldsymbol{M} \odot x + (1 - \boldsymbol{M}) \odot (\textit{zero-padding}(\widetilde{t_i})), \quad (2)$$

where $\boldsymbol{M} \in \{0,1\}^{W \times H}$ is the binary mask filled with 0 within the region $\boldsymbol{B}$, otherwise 1, and *zero-padding*$(\cdot)$ means padding variate with zero to target shape, here target shape is (W, H, C). The pipeline of data prepocessing and MaskOut operation are illustrated in Fig. 1.

4. MCPRL Masked Datasets

As described before, we construct a real-life masked face dataset to evaluate masked face recognition models on real-life masked faces. Hundreds of volunteers wear masks to record videos. When recording videos, volunteers adjust their face poses continuously and change the distance to the camera without any constraints. We capture abundant frames from these vidieos. Furthermore, we use RetinaNet-50 [14] and PFLD [9] to detect face and 106 facial landmarks for face alignment. After cleaning and face alignment, we finally obtained 14,023 images of 246 individuals. Due to the unconstrained condition in data collecting phase, the dataset is full of variability in pose, lighting, focus, age, gender, accessories, make-up, occlusions and background. The sample images in our dataset are exhibited in Fig. 2, which are manually blurred for privacy protection. In order to quantitatively evaluate the performance of masked face recognition methods, we match pairs in the database following LFW [11]. Finally, we construct a real-life masked faces dataset, termed MCPRL-Mask, for 1:1 masked face verification with 5,000 matched and mismatched pairs, respectively.

5. Experiments

5.1. Datasets

We separately employ CASIA-WebFace [26] and MS-Celeb1M-v1c [10] as our training data. To make comparisons of FMA-3D [22], we also conduct experiments on CASIA-WebFace-Mask [22] and MS-Celeb1M-v1c-Mask [22]. To evaluate our method on both non-masked and masked face recognition task, we choose LFW [11] dataset of non-masked faces, LFW-Mask-Only dataset of sythesized masked faces, and MCPRL-Mask dataset of real-life masked faces as our test datasets and report the mean accuracy of 10-fold cross validation for each test dataset. Moreover, we also report our test result in InsightFace Track of ICCV21-MFR [4, 29].

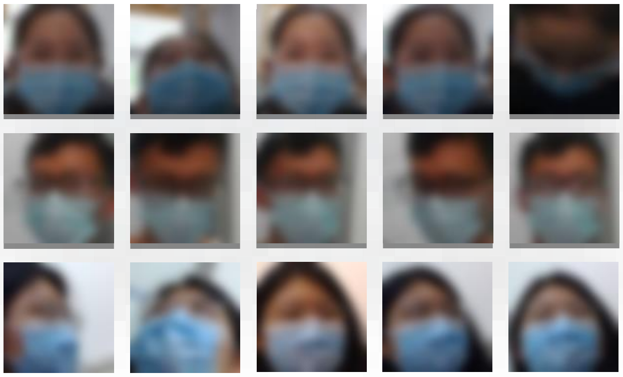

Figure 2. Sample images in MCPRL-Mask dataset (mannually blurred for privacy protection).

CASIA-WebFace [26]: The dataset is collected from Internet by a semi-automatical way. It contains 10,575 individuals and 494,414 images.

MS-Celeb1M-v1c [10]: MS-Celeb1M-v1c dataset has 3,923,399 aligned images of 86,876 individuals cleaned from MS-Celeb-1M [10] dataset. Besides, following the work [22], we clean the dataset further remaining 3,284,503 images of 72,778 individuals.

LFW [11]: It contains 13,233 web-collected images of 5,749 individuals with the pose, expression and illumination variations.

We seperately apply FMA-3D [22] method on LFW, CASIA-WebFace and MS-Celeb1M-v1c to synthesize the masked data from original datasets. CASIA-WebFace-Mask [22], and MS-Celeb1M-v1c-Mask [22] are the corresponding mask version of CASIA-WebFace and MS-Celeb1M-v1c, respectively. They include the original face images of each identity in original datasets, as well as the masked face images corresponding to the original ones. Therefore, CASIA-WebFace-Mask has $494,414 \times 2$ images of 10,575 individuals and MS-Celeb1M-v1c-Mask contains $3,284,503 \times 2$ images of 72,778 individuals. LFW-Mask-Only dataset only contains the masked face images corresponding to the original ones.

5.2. Implementation Details

We adopt the same exprimental settings with Facex-Zoo [22]. The backbone is MobileFaceNet [3] and the supervisory head is MV-Softmax [23]. We train models for 18 epochs with a batch size of 512. The learning rate is initialized as 0.1, and divided by ten at the epoch 10, 13 and 16. For InsightFace Track of ICCV21-MFR, in order to make a fair comparison with baseline, we directly apply our MaskOut method on baseline which is ArcFace [5] model trained on MS1M-V3 [4] or Glint360k [1]. Besides, we train the model for 25 epochs with a batch size of 512. The learning rate is initialized as 0.1, and divided by ten at the epoch 10, 16 and 21. For all the experiments, we set the MaskOut

Method	LFW	LFW-Mask-Only	MCPRL-Mask
MobileFaceNet(Baseline)	98.72	89.08	88.58
MobileFaceNet + Mask [22]	98.60	97.88	90.31
MobileFaceNet +MaskOut	98.63	92.32	**92.52**

Table 1. Results of models trained with CASIA-WebFace [26] dataset. We report the mean accuracies(%) on the evaluation datasets: LFW, LFW-Mask-Only, MCPRL-Mask.

Method	LFW	LFW-Mask-Only	MCPRL-Mask
MobileFaceNet(Baseline)	99.61	94.33	92.71
MobileFaceNet + Mask [22]	99.40	98.68	92.50
MobileFaceNet + MaskOut	99.57	94.72	**93.88**

Table 2. Results of models trained with MS-Celeb1M-v1c [10] dataset. We report the mean accuracies(%) on the evaluation datasets: LFW, LFW-Mask-Only, MCPRL-Mask.

ratio to the whole dataset as 0.4.

5.3. Evaluation Results

As shown in Table 1 and Table 2, trained with CASIA-WebFace and MS-Celeb1M-v1c datasets, respectively, MaskOut improves the model performance on masked face recognition by a large margin compared to baseline. The performances on LFW dataset show that MaskOut slightly sacrifices the accuracy of non-masked face recognition. And the models trained with CASIA-WebFace-Mask and MS-Celeb1M-v1c-Mask are denoted as “MobileFaceNet + Mask” in Table 1 and Table 2. They outperform our method trained on original datasets by a large margin on LFW-Mask-Only evaluation dataset. The main reason is probably that the fake masked faces in LFW-Mask-Only evaluation dataset are sythesized by the same way with masked training data in CASIA-WebFace-Mask and MS-Celeb1M-v1c-Mask. When evaluating on real-life masked faces dataset, MCPRL-Mask, even though CASIA-WebFace-Mask and MS-Celeb1M-v1c-Mask cost double images and double training iterations than our method, they are still inferior to our method by a large margin. Therefore, to make a fair and more actual evaluation of masked face rocognition methods, it is better to test on real-life masked face datasets instead of sythesized fake masked face datasets.

For InsightFace Track of ICCV21-MFR, the results are shown in Table 3. The evaluation metrics are: (1) for Mask set, TAR is measured on mask-to-nonmask 1:1 protocal, with FAR less than 0.0001(e-4), denoted as Mask, (2) for Children set, TAR is measured on all-to-all 1:1 protocal, with FAR less than 0.0001(e-4), (3) for other sets, TAR is measured on all-to-all 1:1 protocal, with FAR less than 0.000001(e-6). Mask denotes the TAR on Mask set with the above evaluation metric. MR-ALL denotes the average TAR on Children set and other sets. Our proposed method improves the performance on the Mask set significantly and slightly decreases MR-ALL.

Sub-Track	Method	Mask	MR-All
ms1m	arcface_torch_r100(Baseline)	69.091	84.312
	Baseline + MaskOut	78.583	83.297
glint360k	arcface_torch_r100(Baseline)	73.047	90.219
	Baseline + MaskOut	81.038	88.238

Table 3. Results in InsightFace Track of ICCV21-MFR.

6. Visualization

We visualize the input data of models before and after MaskOut data augmentation in Fig 3, which shows that MaskOut can effectively shield random regions below the nose of a face with random masked templates data.

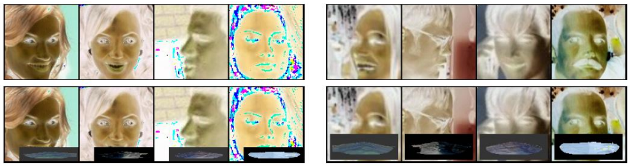

Figure 3. Visualization of data with MaskOut.

7. Conclusion

In this work, a simple and model-agnostic data augmentation routine, named MaskOut, is proposed to improve the roubustness and generalizability of masked face recognition models. And a real-life masked face dataset is constructed to evaluate the masked face recognition methods. In the future, we aim to extend our masked dataset to more masked face applications and improve the performance of masked face recognition persistantly.

8. Acknowledgement

This work is supported by Chinese National Natural Science Foundation under Grants 62076033, U1931202.

References

[1] Xiang An, Xuhan Zhu, Yang Xiao, Lan Wu, Ming Zhang, Yuan Gao, Bin Qin, Debing Zhang, and Fu Ying. Partial fc: Training 10 million identities on a single machine. In *Arxiv 2010.05222*, 2020.

[2] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[3] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Chinese Conference on Biometric Recognition*, pages 428–438. Springer, 2018.

[4] Jiankang Deng, Jia Guo, Xiang An, Zheng Zhu, and Stefanos Zafeiriou. Masked face recognition challenge: The insightface track report. In *Proceedings of the International Conference on Computer Vision Workshops*, 2021.

[5] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[6] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

[7] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 534–551, 2018.

[8] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Dropblock: A regularization method for convolutional networks. *arXiv preprint arXiv:1810.12890*, 2018.

[9] Xiaojie Guo, Siyuan Li, Jinke Yu, Jiawan Zhang, Jiayi Ma, Lin Ma, Wei Liu, and Haibin Ling. Pfld: A practical facial landmark detector. *arXiv preprint arXiv:1902.10859*, 2019.

[10] 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*, pages 87–102. Springer, 2016.

[11] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric 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.

[12] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4873–4882, 2016.

[13] Brendan F Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, and Anil K Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1931–1939, 2015.

[14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[15] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165. IEEE, 2018.

[16] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. 2015.

[17] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[18] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*, 2015.

[19] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1891–1898, 2014.

[20] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2892–2900, 2015.

[21] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[22] Jun Wang, Yinglu Liu, Yibo Hu, Hailin Shi, and Tao Mei. Facex-zoo: A pytorch toolbox for face recognition. *arXiv preprint arXiv:2101.04407*, 2021.

[23] Xiaobo Wang, Shifeng Zhang, Shuo Wang, Tianyu Fu, Hailin Shi, and Tao Mei. Mis-classified vector guided softmax loss for face recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12241–12248, 2020.

[24] 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.

[25] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 90–98, 2017.

[26] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[27] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable fea-

tures. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6023–6032, 2019.

[28] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13001–13008, 2020.

[29] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, et al. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10492–10502, 2021.