Masked Face Recognition Challenge: The WebFace260M Track Report

Zheng Zhu
Guan Huang
Jiankang Deng
Yun Ye
Junjie Huang
Xinze Chen
Jiagang Zhu
Tian Yang
Jia Guo
Jiwen Lu
Dalong Du
Jie Zhou
Jinjie Huang
Xinze Chen

1Tsinghua University 2XForwardAI 3Imperial College London 4InsightFace

{zhengzhu,lujiwen}@tsinghua.edu.cn {guan.huang,dalong.du}@xforwardai.com
j.deng16@imperial.ac.uk

Abstract

According to WHO statistics, there are more than 204,617,027 confirmed COVID-19 cases including 4,323,247 deaths worldwide till August 12, 2021. During the coronavirus epidemic, almost everyone wears a facial mask. Traditionally, face recognition approaches process mostly non-occluded faces, which include primary facial features such as the eyes, nose, and mouth. Removing the mask for authentication in airports or laboratories will increase the risk of virus infection, posing a huge challenge to current face recognition systems. Due to the sudden outbreak of the epidemic, there are yet no publicly available real-world masked face recognition (MFR) benchmark. To cope with the above-mentioned issue, we organize the Face Bio-metrics under COVID Workshop and Masked Face Recognition Challenge in ICCV 2021. Enabled by the ultra-large-scale WebFace260M benchmark and the Face Recognition Under Inference Time conStraint (FRUITS) protocol, this challenge (WebFace260M Track) aims to push the frontiers of practical MFR. Since public evaluation sets are mostly saturated or contain noise, we adopt the Face Recognition Under Inference Time conStraint (FRUITS) protocol in WebFace260M Track in this workshop. Besides, a new test set is gathered consisting of elaborated 2,478 celebrities and 60,926 faces. Meanwhile, we collect the world-largest real-world masked test set. In the first phase of WebFace260M Track, 69 teams (total 833 solutions) participate in the challenge and 49 teams exceed the performance of our baseline. There are second phase of the challenge until October 1, 2021 and on-going leaderboard. We will actively update this report in the future.

1. Introduction

Due to the boom of CNNs, standard face recognition (SFR) systems have achieved a remarkable success, which usually work with mostly non-occluded faces. However, there are a number of circumstances where faces are occluded by facial masks, rising the masked face recognition (MFR) problem.

During the global COVID-19, people are encouraged to wear masks in public areas, making primary facial features invisible. Few SFR systems can work well with this situation, but removing the mask for authentication will increase the risk of virus infection. Recently, some commercial vendors [27] have developed face recognition algorithms capable of handling face masks, and an increasing number of research publications [13, 15, 14, 18, 5] have surfaced on this topic. However, due to the sudden outbreak of the epidemic, there are yet no publicly available large-scale MFR benchmark.

To address the above-mentioned issue, we organize the Face Bio-metrics under COVID Workshop and Masked Face Recognition Challenge in ICCV 2021. Face benchmarks empower researchers to train high-performance face recognition systems. Enabled by the ultra-large-scale WebFace260M benchmark [41], this challenge aims to push the frontiers of practical MFR. On the other hand, evaluation protocols and test set play an essential role in analysing face recognition performance. Since public evaluation sets are mostly saturated or contain noise, we adopt the Face Recognition Under Inference Time conStraint (FRUITS) protocol in WebFace260M Track in this workshop. Besides, a new test set is gathered consisting of elaborated 2,478 celebrities and 60,926 faces. Meanwhile, we collect the world-largest real-world masked test set.

This paper is the official report of WebFace260M Track in the MFR workshop and challenge. We detail the training data, evaluation protocols, submission rules, test set and metric in SFR and MFR, ranking criterion, baseline solution, and preliminary competition results. The challenge was launched on June 7, 2021. In the first phase of WebFace260M Track, 69 teams from academia and industry participate in the challenge and 49 teams exceed the performance of our baseline. Total 833 solutions are submitted, covering various network designs and training strategies. There are second phase of the challenge till October 1, 2021 and on-going leaderboard. We will actively update this report in the future.
2. Training Data and Evaluation Protocols

2.1. WebFace260M Data

WebFace260M [41] is currently the largest public face recognition dataset, covering noisy 4M identities/260M faces and cleaned 2M identities/42M faces. With such large data size, this dataset takes a significant step towards closing the data gap between academia and industry as shown in Table 1. The celebrity name list consists of two parts: the first one is borrowed from MS1M, the second one is collected from the IMDB database. Based on the name list, celebrity faces are searched and downloaded via Google image search engine [2]. The WebFace42M training set is obtained by automatically utilizing Self-Training (CAST) pipeline. Noise ratio of WebFace26M is lower than 10% (similar to CASIA-WebFace [37] and Glint360K [4]) based on the sampling estimation. After CAST, duplicates of each subject are removed when their cosine similarity is higher than 0.7. Furthermore, the feature center of each subject is compared with popular benchmarks (e.g., the test set in this challenge, LFW families [19, 40, 39], FaceScrub [26], IJB-C [22] et al.), and overlaps are removed if the cosine similarity is higher than 0.7.

2.2. FRUITS Protocol

Most existing face recognition evaluation protocols [19, 30, 23, 40, 39, 20, 22, 36, 21] target the pursuit of accuracy. However, face recognition in real-world application scenarios is always restricted by inference time. Lightweight face recognition challenge [11] takes a step toward this goal by constraining the FLOPs and model size of submissions, which is not a straightforward solution. Besides, it neglects the face detection and alignment module cost. Strict submission policy of NIST-FRVT [3] hinders researchers to freely evaluate their algorithms. In WebFace260M Track of this challenge, we follow the Face Recognition Under Inference Time constraint (FRUITS) protocol. Referring to [41], inference time is measured on a single core of an Intel Xeon CPU E5-2630-v4@2.20GHz processor (no GPU hardware is provided), and the constraint of 1000 milliseconds is adopted.

2.3. Submission Rules

The WebFace260M Track of our challenge has two phases. For the first phase, the number of max submissions per day is 5. For the second phase, the number of max submissions per day is 3. The full WebFace260M data has been open for all applicants, as long as their agreements are qualified. Mask data-augmentation is allowed, for example this method. The applied mask augmentation tool should be reproducible. External dataset and pre-trained models are both prohibited in the second phase. Participants could submit their submission package to the submission server and get scores by our online evaluation. Participants should run the code for verification on the provided docker file to ensure the correctness of feature and time constraints.

### Table 1: Training data for deep face recognition. The cleaned WebFace42M is the largest public training set in terms of both # identities and # images.

| Dataset           | # Identities | # Images | Images/ID | Cleaning   | # Attributes | Availability | Publications |
|-------------------|--------------|----------|-----------|------------|--------------|--------------|--------------|
| CASIA-WebFace [37]| 10 K         | 0.5 M    | 47        | Auto       | -            | Public       | Arxiv 2014   |
| CelebFaces [31]   | 10 K         | 0.2 M    | 20        | Manual     | 40           | Public       | ICCV 2015    |
| UMDFaces [6]      | 8 K          | 0.3 M    | 45        | Semi-auto  | 4            | Public       | IJCB 2017    |
| VGGFace [28]      | 2 K          | 2.6 M    | 1,000     | Semi-auto  | -            | Public       | BMVC 2015    |
| VGGFace2 [8]      | 9 K          | 3.3 M    | 363       | Semi-auto  | 11           | Public       | FG 2018      |
| MS1M [17]         | 0.1 M        | 10 M     | 100       | No         | -            | Public       | ECCV 2016    |
| MS1M-IBUG [12]    | 85 K         | 3.8 M    | 45        | Semi-auto  | -            | Public       | CVPRW 2017   |
| MS1MV2 [10]       | 85 K         | 5.8 M    | 68        | Semi-auto  | -            | Public       | CVPR 2019    |
| MS1M-Glint [1]    | 87 K         | 3.9 M    | 44        | Semi-auto  | -            | Public       | -            |
| MegaFace2 [24]    | 0.6 M        | 4.7 M    | 7         | Auto       | -            | Public       | CVPR 2017    |
| IMDB-Face [34]    | 50 K         | 1.7 M    | 29        | Manual     | -            | Public       | ECCV 2018    |
| Facebook [32]     | 4 K          | 4.4 M    | 1,100     | -          | -            | Private      | CVPR 2014    |
| Facebook [33]     | 10 M         | 500 M    | 50        | -          | -            | Private      | CVPR 2015    |
| Google [29]       | 8 M          | 200 M    | 25        | -          | -            | Private      | CVPR 2015    |
| MillionCelebs [38]| 0.6 M        | 18.8 M   | 30        | Auto       | -            | Private      | CVPR 2020    |
| WebFace260M       | 4 M          | 260M     | 65        | No         | -            | Public       | CVPR 2021    |
| WebFace42M        | 2 M          | 42M      | 21        | Auto       | 7            | Public       | CVPR 2021    |
Both the models for face detection and recognition should be converted to ONNX format. Participants must package the code directory for submission and see the results on the leaderboard. Test images are invisible during challenges.

3. Standard Face Recognition

3.1. Test Set

To compare deep CNN face matchers utilizing FRUITS protocol, we manually construct a elaborated test set. It is well known that recognizing strangers (especially when they are similar-looking) is a difficult task even for experienced researchers. So we choose to select our familiar celebrities, which ensure the high-quality of the test set.

The collecting pipeline of test set is illustrated in Figure 1. Each annotator is asked to write about 100 names of celebrities including stars, politics, business, athletes et al. Different name lists are exclusive, and a identities pool is maintained by merging lists. It is noting that annotators are encouraged to collect gender-balanced and race-balanced identities lists.

The final identities pool consists of 2,748 names. For each celebrity, responsible annotator need to collect his/her faces from reliable sources of information. Targeting at analysing recognition performance in different application scenarios, 3 subclass is defined to guide collection:

- **Controlled image**: Controlled faces collection targets to evaluate ID photo such as visa and driving license. Controlled face in our test set is defined as: near frontal, five landmarks are visible, normal expression, not low-resolution.

- **Wild image**: Wild subclass aims to collect faces in unconstrained scenarios, including large pose, partial occlusion, resolution variation, illumination variation et al.

- **Cross-age image**: This subclass collects faces whose age is obviously different from Controlled and Wild faces, including Cross-age-10 (more than 10-years gap) and Cross-age-20 (more than 20-year gap).

For each subclass of a certain celebrity, we gather about 7 faces (i.e. 20 faces/identity) to construct test set as shown in Figure 1. Furthermore, we design a interface to assist annotators to judge the difficulty and quality of the collected faces. In Figure 2, collected images for a celebrity, its intra-class scores, its highest inter-class scores with other identities are illustrated respectively. Based on scores indicator...
from this system, annotators are encouraged to gather hard case (For example, less than 0.5 similarity score for intra-class, more than 0.5 for inter-class). Besides, noisy faces could be effectively filtered (For example, if the score of one face is less than 0.5 compared with all other intra-class images, it should be paid more attention).

The statistics of final test is listed in Table 2. There are elaborately constructed 2,478 identities and 57,715 faces in total for SFR. 3,328,950,920 impostors and 2,070,305 genuine pairs could be constructed. Protected attributes (gender, race) as well as different scenarios (Controlled, Wild, Cross-age) are accurately annotated for each subject. Cross-age and Cross-scene comparisons are also conducted in corresponding subset.

### 3.2. Metric

Based on the FRUITS protocol and the new test set, we perform 1:1 face verification across various attributes for SFR evaluation. Table 2 shows numbers of imposter and genuine in different verification settings. All means impostors are paired without attention to any attribute, while later comparisons are conducted on age and scenario subsets. Cross-age refers to cross-age (more than 10 and 20 years) verification, while Cross-scene means pairs are compared between controlled and wild settings. Different algorithms are measured on False Non-Match Rate (FNMR) [3], which is defined as the proportion of mated comparisons below a threshold set to achieve the False Match Rate (FMR) specified. FMR is the proportion of impostor comparisons at or above that threshold. It is worth noting that Lower FNMR at the same FMR is better.

### 4. Masked Face Recognition

#### 4.1. Test Set

In contrast with simulated [27, 25] or relatively small [5, 9, 7, 35] masked face test sets, a real-world comprehensive benchmark for evaluating MFR is developed in this challenge. Based on the SFR identities, we further collect masked faces for these celebrities. Specifically, as shown in Table 2, there are carefully selected 3,211 masked faces among 862 identities. Subjects with real-world mask are illustrated in Figure 4.

#### 4.2. Metric

For MFR, assessment is performed with Mask-Nonmask comparisons. Specifically, there is one face with mask in imposter and genuine, while another face is from standard face sets. According to the attribute of face without mask, we evaluate the performance of algorithms under Controlled-Masked, Wild-Masked, and All-Masked settings listed in Table 2.

#### 4.3. Competition Ranking

The competition is ranked according to both MFR and SFR metrics. To reduce a tendency that models overfit on masked or standard face recognition, the main series of evaluation metrics are designed to show a weighted sum to consider both masked and standard faces at the same time. As shown in Table 3, the overall ranking is ascend ordered by the $\text{All (MFR&SFR)} = 0.25 \times \text{All-Masked} + 0.75 \times \text{All (SFR)}$. At the same time, Wild and Controlled metrics can also be computed as: $\text{Wild (MFR&SFR)} = 0.25 \times \text{Wild-Masked} + 0.75 \times \text{Wild (SFR)}$; $\text{Controlled (MFR&SFR)} = 0.25 \times \text{Controlled-Masked} + 0.75 \times \text{Controlled (SFR)}$. It is worth noting that
scores of different metrics is computed at corresponding FNR@FMR=10^{-5}.

Table 3: The leaderboard of first phase. Results outperforming the baseline (Participant: titian1045) are shown.

5. Baseline Solutions

5.1. Implementation Details

In order to fairly evaluate the performance of different face recognition models, we reproduce representative algorithms in one Gluon codebase with the hyper-parameters referred to the original papers. Default batch size per GPU is set as 64 unless otherwise indicated. Learning rate is set as 0.05 for a single node (8 GPUs), and follows the linear scaling rule [16] for the training on multiple nodes (i.e., 0.05\times\# machines). We decrease the learning rate by 0.1× at 8, 12, and 16 epochs, and stop at 20 epochs for all models. During training, we only adopt the flip data augmentation. Note that other data augmentations such as adding simulated mask are encouraged to boost MFR performance.
5.2. Baseline Model and Results

The configuration of the baseline model is ResNet-50 backbone, ArcFace loss, with WebFace12M (30%) training data. The backbone architecture is shown in Table 4. The evaluation results of the baseline model (Participant: lititian1045) are shown in the last row of Table 3. The FNMR@FMR=1e-5 across different attributes (All (MFR&SFR), Wild (MFR&SFR), Controlled (MFR&SFR), All (SFR), Wild (SFR), Controlled (SFR)) is 0.2342, 0.2816, 0.1483, 0.0871, 0.1341, 0.0094, respectively. The time cost of detection, recognition and total is 157 ms, 172 ms, 329 ms, respectively.

6. Preliminary Results of First Phase

First phase was held between June 7, 2021 to August 11, 2021. During this phase, 69 teams submit a total of 833 effective solutions to the challenge. Table 3 gives a detailed result for each participant with their ranks. As shown in Table 3 and Figure 4, Ethan.y ranks the 1st place in main All (MFR&SFR) metric with 0.0980 as well as Controlled (MFR&SFR) with 0.0500, while victor-2021 wins Wild (MFR&SFR) with a score of 0.1222. For SFR metrics, victor-2021 ranks the 1st place among All (SFR), Wild (SFR) and Controlled (SFR) with 0.0162, 0.0270 and 0.0018. Since the FNMR of MFR is much higher than that in SFR, the MFR performance dominates the final ranks. It is worth noting that during first phase, a model is allowed to be evaluated if its Total time (which is a sum of Detection time and Recognition time) is no more than 2000ms. In the final ranks (second phase), we only rank models with a Total time less than 1000ms. As shown in Table 3, there are 9 participants who have submitted models to the leaderboard with Total time more than 1000ms. Since the challenge is still going on, more details of top-ranked solution would be updated in the future.

7. Conclusion

To address the MFR problem during epidemic, we organize the Face Bio-metrics under COVID Workshop and Masked Face Recognition Challenge in ICCV 2021. Enabled by the WebFace260M and FRUITS, this challenge (WebFace260M Track) aims to push the frontiers of practical MFR. This report details the training data, evaluation protocols, submission rules, test set and metric in SFR and MFR, ranking criterion, baseline solution, and preliminary competition results. In the first phase of WebFace260M Track, 69 teams (total 833 solutions) participate in the challenge and 49 teams exceed the performance of our baseline. We will actively update this report in the future.

References

[1] http://trillionpairs.deepglint.com/overview
[2] https://images.google.com
[3] https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing
[4] Xiang An, Xuhuan Zhu, Yang Xiao, Lan Wu, Ming Zhang, Yuan Gao, Bin Qin, Debing Zhang, and Ying Fu. Partial FC: Training 10 million identities on a single machine. arXiv:2010.05222, 2020.
[5] Aqeel Anwar and Arijit Raychowdhury. Masked face recognition for secure authentication. *arXiv preprint arXiv:2008.11104*, 2020. 1, 4

[6] Ankan Bansal, Anirudh Nanduri, Carlos D Castillo, Rajeev Ranjan, and Rama Chellappa. UMD Faces: An annotated face dataset for training deep networks. *arXiv:1611.01484v2*, 2016. 2

[7] Fadi Boutros, Naser Damer, Jan Niklas Kolf, Kiran Raja, Florian Kirchbuchner, Raghavendra Ramachandra, Arjan Kuiper, Pengcheng Fang, Chao Zhang, Fei Wang, et al. MFR 2021: Masked face recognition competition. *International Joint Conference on Biometrics Workshops*, 2021. 4

[8] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*, 2018. 2

[9] Naser Damer, Fadi Boutros, Marius Süßmilch, Florian Kirchbuchner, and Arjan Kuiper. Extended evaluation of the effect of real and simulated masks on face recognition performance. *IET Biometrics*, 2021. 4

[10] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[11] Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi. Lightweight face recognition challenge. In *IEEE International Conference on Computer Vision Workshop*, 2019. 2

[12] Jiankang Deng, Yuxiang Zhou, and Stefanos Zafeiriou. Marginal loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2017. 2

[13] Feifei Ding, Peixi Peng, Yangru Huang, Mengyue Geng, and Yonghong Tian. Masked face recognition with latent part detection. In *ACM International Conference on Multimedia*, 2020. 1

[14] Hang Du, Hailin Shi, Yinglu Liu, Dan Zeng, and Tao Mei. Towards nir-vis masked face recognition. *IEEE Signal Processing Letters*, 2021. 1

[15] Mengyue Geng, Peixi Peng, Yangru Huang, and Yonghong Tian. Masked face recognition with generative data augmentation and domain constrained ranking. In *ACM International Conference on Multimedia*, 2020. 1

[16] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch SGD: Training imagenet in 1 hour. *arXiv:1706.02677*, 2017. 5

[17] 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*, 2016. 2

[18] Walid Hariri. Efficient masked face recognition method during the COVID-19 pandemic. *arXiv preprint arXiv:2105.03026*, 2021. 1

[19] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, 2007. 2

[20] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The MegaFace benchmark: 1 million faces for recognition at scale. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[21] Yuanliu Liu, Peipei Shi, Bo Peng, He Yan, Yong Zhou, Bing Han, Yi Zheng, Chao Lin, Jianbin Jiang, and Yin Fan. IQIYI-VID: A large dataset for multi-modal person identification. *arXiv:1811.07548*, 2018. 2

[22] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, and Jordan Cheney. IARPA Janus Benchmark C: Face dataset and protocol. In *IAPR International Conference On Biometrics*, 2018. 2

[23] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: The first manually collected in-the-wild age database. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2017. 2

[24] Aaron Nech and Ira Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2

[25] Alok Negi, Krishan Kumar, Prachi Chauhan, and RS Rajput. Deep neural architecture for face mask detection on simulated masked face dataset against COVID-19 pandemic. In *2021 International Conference on Computing, Communication, and Intelligent Systems*, 2021. 4

[26] Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *International Conference on Image Processing*, 2014. 2

[27] Mei L Ngan, Patrick J Grother, Kayee K Hanaoka, et al. On-going Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms. 2021. 1, 4

[28] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 2

[29] 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. 2

[30] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, 2016. 2

[31] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 2

[32] Yaniv Taigman, Ming Yang, Marc’ Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 2

[33] Yaniv Taigman, Ming Yang, Marc’ Aurelio Ranzato, and Lior Wolf. Web-scale training for face identification. In *IEEE
[34] Fei Wang, Liren Chen, Cheng Li, Shiyou Huang, Yanjie Chen, Chen Qian, and Chen Change Loy. The devil of face recognition is in the noise. In European Conference on Computer Vision, 2018. 2

[35] 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. 4

[36] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In IEEE Conference on Computer Vision and Pattern Recognition, 2011. 2

[37] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. arXiv:1411.7923, 2014. 2

[38] Yaobin Zhang, Weihong Deng, Mei Wang, Jiani Hu, Xian Li, Dongyue Zhao, and Dongchao Wen. Global-local GCN: Large-scale label noise cleansing for face recognition. In IEEE Conference on Computer Vision and Pattern Recognition, 2020. 2

[39] Tianyue Zheng and Weihong Deng. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. Technical report, 2018. 2

[40] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. arXiv:1708.08197, 2017. 2

[41] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, and Jie Zhou. WebFace260M: A benchmark unveiling the power of million-scale deep face recognition. In IEEE Conference on Computer Vision and Pattern Recognition, 2021. 1, 2