Masked Face Recognition Challenge: The InsightFace Track Report

Jiankang Deng
Imperial College London
j.deng16@imperial.ac.uk

Jia Guo
InsightFace
guojia@gmail.com

Xiang An
InsightFace
anxiangsir@gmail.com

Zheng Zhu
Tsinghua University
zhengzhu@ieee.org

Stefanos Zafeiriou
Imperial College London
s.zafeiriou@imperial.ac.uk

Abstract

During the COVID-19 coronavirus epidemic, almost everyone wears a facial mask, which poses a huge challenge to deep face recognition. In this workshop, we organize Masked Face Recognition (MFR) challenge [1] and focus on bench-marking deep face recognition methods under the existence of facial masks. In the MFR challenge, there are two main tracks: the InsightFace track and the WebFace260M track [38]. For the InsightFace track, we manually collect a large-scale masked face test set with 7K identities. In addition, we also collect a children test set including 14K identities and a multi-racial test set containing 242K identities. By using these three test sets, we build up an online model testing system, which can give a comprehensive evaluation of face recognition models. To avoid data privacy problems, no test image is released to the public. As the challenge is still under-going, we will keep on updating the top-ranked solutions as well as this report on the arxiv.

1. Introduction

Recently, great progress has been achieved in face recognition with large-scale training data [14, 24, 1, 39], sophisticated network structures [26, 16] and advanced loss designs [28, 29, 26, 25, 8, 20, 31, 30, 4, 3, 18, 6]. However, existing face recognition systems are presented with mostly non-occluded faces, which include primary facial features such as the eyes, nose, and mouth. During the COVID-19 coronavirus epidemic, almost everyone wears a facial mask, which poses a huge challenge to existing face recognition systems. Traditional face recognition systems may not effectively recognize the masked faces, but removing the mask for authentication will increase the risk of virus infection.

To cope with the above-mentioned challenging scenarios arising from wearing facial masks, it is crucial to improve the existing face recognition approaches[2]. Generally, there are two kinds of methods to overcome masked face recognition: (1) recovering unmasked faces for feature extraction and (2) producing direct occlusion-robust face feature embedding from masked face images.

Based on Generative Adversarial Network (GAN) [13], there are many identity-preserved masked face restoration methods [9, 11]. In [9], masked face images are first segmented and then impainted with fine facial details while retaining the global coherency of face structure. Ge *et al.* [11] propose identity-preserved inpainting to facilitate occluded face recognition. The core idea is integrating GAN with an optimized pre-trained CNN model which serves as the third player to compete with the generator by enabling the inpainted faces to be close to their identity centers.

Since occlusion recovery methods [9, 11] are more complicated to set up the online evaluation toolkit, we focus on occlusion-robust face feature embedding in this challenge. In [36], a new partial face recognition approach is proposed by using local texture set matching to recognize persons of interest from their partial faces. In [15], a masked-aware face feature embedding is proposed by extracting deep features from the unmasked regions (mostly eyes and forehead regions). In [22], masked face augmentation and extra mask-usage classification loss is proposed to train mask robust facial feature embedding. In [19, 35], visual attention mechanism is employed to enhance feature learning from non-occluded face regions.

Even though there are some existing explorations for occluded (masked) face recognition, there is yet no publicly available large-scale masked face recognition benchmark due to the sudden outbreak of the epidemic. In this report, we make a significant step further and propose a new com-

[1]https://ibug.doc.ic.ac.uk/resources/masked-face-recognition-challenge-workshop-iccv-21/

[2]https://pages.nist.gov/frvt/html/frvt_facemask.html

Dataset	# Identities	# Images
MS1M	93K	5.1M
Glint360K	360K	17M

Table 1. Statistics of the training data of the masked face recognition challenge (the InsightFace track).

prehensive benchmark for masked face recognition as well as non-masked face recognition. To this end, we have collected a real-world masked test set, children test set, multi-racial test set (*i.e*. African, Caucasian, South Asian and East Asian [37, 12, 34, 33]). We define different sub-tracks with fixed training data, and each sub-track has strict constraints on computational complexity and model size. Therefore, the performance comparison between different models can be fair.

By using the proposed test data, we organized the InsightFace track in Masked Face Recognition Challenge (ICCV 2021). This report presents the details of this track, including the training data, the test set, evaluation protocols, baseline solutions, performance analysis of the top-ranked submissions received as part of the competition, and effective strategies for masked face recognition. The report of another WebFace260M track is available in [38].

2. Datasets of InsightFace Track

2.1. Training Dataset

As given in Tab. 1, we employ two existing datasets (*i.e*. MS1M [14] and Glint360K [1]) as the training data.

MS1M: The MS1M training dataset is cleaned from the MS-Celeb-1M [14] dataset. All face images are pre-processed to the size of 112×112 by the five facial landmarks predicted by RetinaFace [5]. Then, a semi-automatic refinement is conducted by employing the pre-trained ArcFace [4] model and ethnicity-specific annotators [7]. Finally, the refined MS1M dataset contains 5.1M images of 93K identities.

Glint360K: The Glint360K training dataset is cleaned from the MS-Celeb-1M [14] and Celeb-500k [2] datasets. All face images are downloaded from the Internet and pre-processed to the size of 112×112 by the five facial landmarks predicted by RetinaFace [5]. Then, an automatic refinement is conducted by employing the pre-trained ArcFace [4] model for intra-class and inter-class cleaning. Finally, the released Glint360K dataset contains 17M images of 360K individuals, which is one of the largest and cleanest training datasets [39] in academia.

The training data (*i.e*. MS1M and Glint360K) are fixed to facilitate performance reproduction and fair comparison. Detailed requirements:

- No external dataset is allowed and no pre-trained model is allowed.
- All participants must use the predefined training dataset for a particular challenge track. Data augmentation for the facial mask is allowed but the augmentation method needs to be reproducible.

2.2. Test Dataset

As shown in Tab. 2 and Fig. 1, we manually collected the following three test sets for the comprehensive evaluation of different algorithms. Unlike existing face recognition test sets (*e.g*. LFW [17], CFP-FP [27], AgeDB [23], and IJB-C [21]), our test sets are not collected from celebrities, thus we can naturally avoid the identity-overlapping problem. The pre-processing step for the test set is the same as that on the training data. All of the faces are normalized into 112×112 by using RetinaFace [5]. We also employ a semi-automatic method to strictly ensure that (1) most of the test sets are noise-free and (2) there is no identity overlap between our training data and the test set.

Masked Test Set: The masked test set contains 6,964 masked facial images and 13,928 non-masked facial images of 6,964 identities. In total, there are 13,928 positive pairs and 96,983,824 negative pairs for the verification evaluation.

Children Test Set: The children test set contains 157,280 images of 14,344 identities aging between 2 and 16. There are totally 1,773,428 positive pairs and 24,735,067,692 negative pairs for the verification evaluation.

Multi-racial Test Set: Participants will also have their algorithms tested on the multi-racial test set for fairly evaluating the performance on different demographic groups. The multi-racial test set consists of four demographic groups [3]: African, Caucasian, South Asian and East Asian [37, 12, 34, 33]. In total, there are 1.6M images of 242K identities.

3. Evaluation Protocols of InsightFace Track

The test is aimed to determine whether, and to what degree, face recognition performance differs when they process photographs of masked faces, child faces, and individuals from various demographic groups (*e.g*. African, Caucasian, South Asian and East Asian). All pairs between the gallery and probe sets will be used for evaluation. We employ the 1:1 face verification as the evaluation metric.

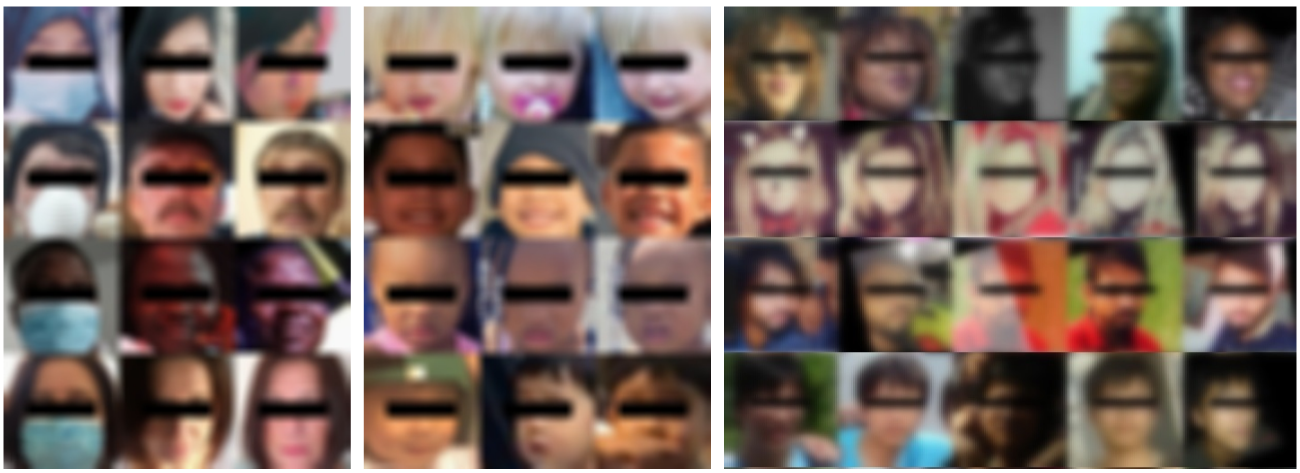
Masked Test Set: We report True Positive Rate (TPR) @ False Positive Rate (FPR) = 1e-4 given 13,928 positive pairs and 96,983,824 negative pairs.

Children Test Set: We report True Positive Rate (TPR) @ False Positive Rate (FPR) = 1e-4 given 1,773,428 positive pairs and 24,735,067,692 negative pairs.

[3]Here, we refer to the NIST standard: `https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf`

	# Identities	# Images	# Positive Pairs	# Negative Pairs
Masked Test Set	6,964	20,892	13,928	96,983,824
Children Test Set	14,344	157,280	1,773,428	24,735,067,692
Multi-racial Test Set	242,143	1,624,305	4,689,037	2,638,360,419,683
African	43,874	298,010	870,091	88,808,791,999
Caucasian	103,293	697,245	2,024,609	486,147,868,171
South Asian	35,086	237,080	688,259	56,206,001,061
East Asian	59,890	391,970	1,106,078	153,638,982,852

Table 2. Statistics of the test sets of the masked face recognition challenge (the InsightFace track).



(a) Masked Test Set (b) Children Test Set (c) Multi-racial Test Set

Figure 1. Exemplar blurred face images of Masked Test Set, Children Test Set and Multi-racial Test Set (*i.e.* African, Caucasian, South Asian, East Asian). To ensure data privacy, we intentionally decrease the quality of the exemplar facial images. On our test server, all of the test images are still in high quality.

Multi-racial Test Set: We assess accuracy by demographic groups (*e.g.* African, Caucasian, South Asian and East Asian) and report True Positive Rate (TPR) @ False Positive Rate (FPR) = 1e-6. The number of positive pairs for each demographic group is of the order of million and the number of negative pairs for each demographic group is of the order of billion.

InsightFace Track Ranking Rules: To protect data privacy and ensure fairness in the competition, we withhold all images as well as labels of the test data. Participants can submit their models in the ONNX format to our evaluation server and get their results from the leader-board after the online evaluation (usually several hours). Participants are only allowed to use the training data we provided for a particular challenge track. On the widely used V100 GPU, we set an upper bound of inference time ($<$ 10 ms/image for the MS1M sub-track and $<$ 20 ms/image for the Glint360 sub-track) to control the model complexity and the submitted model size should be smaller than 1GB in the format of float32. On our online test server, we employ cosine similarity for the verification test. The feature dimension of the MS1M sub-track should be smaller than 512 and the feature dimension of the Glint360K sub-track should be smaller than 1024. All challenge submissions are ordered in terms of weighted TPRs across two test sets (*i.e.* Masked Test Set and Multi-racial Test Set) by the formula of 0.25 * TPR@Masked + 0.75 * TPR@MR-All.

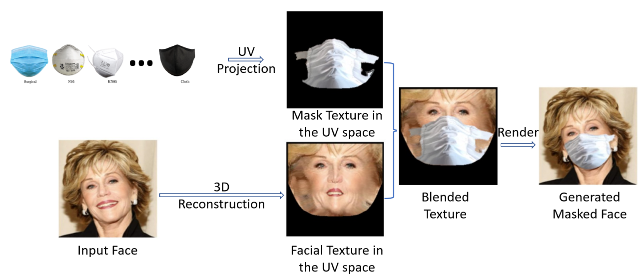


Figure 2. Masked face augmentation through texture blending in the UV space. Demo images and the pipeline are from the JDAI-CV toolkit [32].

layer name	R18	R34	R50	R100	output size
					112×112×3
stem	3×3, 64, s=1	3×3, 64, s=1	3×3, 64, s=1	3×3, 64, s=1	112×112×64
Conv1_x	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	56×56×64
Conv2_x	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times13$	28×28×128
Conv3_x	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times14$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times30$	14×14×256
Conv4_x	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	7×7×512
FC					1×1×512
Flops (G)	2.62	4.48	6.33	12.12	
# Params (M)	24.03	34.14	43.59	65.16	

Table 3. The network configuration, computation complexity and model size of baseline models. Convolutional building blocks are shown in brackets with the numbers of blocks stacked. Down-sampling is performed by the second conv in conv1_1, conv2_1, conv3_1, and conv4_1 with a stride of 2.

Data	Backbone	Loss	LFW	CFP-FP	AgeDB	IJB-C (1e-4)	IJB-C (1e-5)
MS1M	R18	ArcFace	99.77	97.73	97.77	94.66	92.07
MS1M	R34	ArcFace	99.80	98.67	98.10	95.90	94.10
MS1M	R50	ArcFace	99.83	98.96	98.35	96.46	94.79
MS1M	R100	ArcFace	99.85	99.06	98.48	96.81	95.31
Glint360K	R18	ArcFace	99.77	97.73	97.72	95.33	93.16
Glint360K	R34	ArcFace	99.82	98.78	98.33	96.56	95.16
Glint360K	R50	ArcFace	99.83	99.20	98.38	96.97	95.61
Glint360K	R100	ArcFace	99.82	99.29	98.48	97.32	95.88

Table 4. Baseline performance on the public test benchmarks (*e.g*. LFW, CFP-FP, AgeDB and IJB-C).

4. Baseline Solutions of InsightFace Track

Training details of baseline models are released before the challenge to facilitate participation. We re-implement a simple online masked face augmentation function [32], customize the ResNet [16] for the baseline models and employ ArcFace [4] as our loss function, which is one of the top-performing methods for deep face recognition.

4.1. Masked Face Augmentation

As shown in Fig. 2, we follow the JDAI-CV toolkit [32] [4] to implement our online masked face generation function [5]. After 3D face reconstruction [10] on the input 2D face image, we obtain the UV texture map, the face geometry and the camera pose. Then, we randomly select one facial mask from the collected mask dataset and project it into the UV space. Based on a simple texture blending, we can easily get the masked facial UV texture. Finally, we combine the masked facial UV texture and the face geometry, and render the masked face into a 2D face image.

4.2. Implementation Details

During training, we follow ArcFace [4] to set the feature scale to 64 and choose the angular margin at 0.5. As shown in Tab. 3, we customize the ResNet [16] as our base-

[4] https://github.com/JDAI-CV/FaceX-Zoo/tree/main/addition_module/face_mask_adding/FMA-3D

[5] https://github.com/deepinsight/insightface/tree/master/recognition/_tools_

line models (*i.e*. R18, R34, R50 and R100). More specifically, we only employ the basic residual block instead of the bottleneck residual block following ArcFace [4]. The baseline models are implemented by PyTorch with parallel acceleration on both features and centres[6]. We set batch size as $1,024$ and train models on eight NVIDIA V100(32GB) GPUs. The learning rate starts with 0.1, drops by 0.1 at 10, 16, 22 epochs, and the whole training procedure finishes at 24 epochs. We set the momentum to 0.9 and the weight decay to $5e-4$. During testing, we only keep the feature embedding network without the fully connected layer and extract the 512-D features for each normalized face crop. We use the cosine similarity metric for each feature pair.

4.3. Baseline Performance

As shown in Tab. 4, we first test our baseline models on public benchmarks, including LFW [17], CFP-FP [27], AgeDB [23], and IJB-C [21]. By increasing the computation complexity from R18 to R100, the performance rises steadily across all test sets. After changing the training data from MS1M to Glint360K, TPR@FPR=1e-4 on IJB-C significantly increases from 96.81% to 97.32% for R100. On CFP-FP, R100 trained on Glint360K outperforms the counterpart model trained on MS1M by 0.23%, indicating that frontal-to-profile face verification can benefit from more training data. By contrast, the verification accuracy on LFW and AgeDB is almost the same, which indicates that LFW and AgeDB are saturated to distinguish high performing models.

In Tab. 5, we report the performance on the challenge benchmarks. As we can see from these results, the verification accuracy benefits from more training data (from MS1M to Glint360K) and heavier backbone structures (from R18 to R100) across all testing scenarios (*i.e*. masked, children and multi-racial test sets). Compared to the performance gaps on public test datasets (*i.e*. LFW [17], CFP-FP [27], AgeDB [23], and IJB-C [21]), the performance gaps on the proposed masked test set, children test set and multi-racial test set are more obvious. In addition, we also conduct experiments with masked face augmentation. When 10% of training data wear facial masks during training, the verification accuracy on the masked test set significantly increases from 69.091% to 77.325% by using MS1M, and increases from 75.567% to 83.710% by using Glint360K. However, masked face augmentation is slightly harmful for non-masked face verification, as the TPR on the MR-All dataset drops by 0.484% for the MS1M sub-track and decreases by 0.644% for the Glint360K sub-track. We leave the balance of masked face augmentation for the challenge participants.

[6]`https://github.com/deepinsight/insightface/tree/master/recognition/arcface_torch`

5. Leader-board Results of InsightFace Track

The masked face recognition competition (InsightFace track) is conducted as part of the *Masked Face Recognition Challenge & Workshop*[7], at the International Conference on Computer Vision 2021 (ICCV 2021). Participants can freely select different sub-tracks to develop a face feature embedding model, which is automatically evaluated on our test server based on the above-mentioned protocols. The competition has been opened worldwide, to both industry and academic institutions. By 16th August 2021, the InsightFace track has received hundred of registrations from across the world. More specifically, the competition has received 123 valid submissions for the MS1M sub-track and 69 valid submissions for the Glint360K sub-track. Here, multi-submissions for one sub-track from the same participant is only counted once.

As we postpone the leader-board submission to 1st October 2021, we can not collect the final top-ranked solutions before the camera-ready deadline. After the competition, we will close the test server and select the valid top-3 solutions for each track. We will collect the training code from these top-ranked participants and re-train the models to confirm whether the performance of each submission is valid or not. We will update the challenge report through arxiv with detailed team information and detailed top-ranked solutions.

By 16th August 2021, we have found the best model of the MS1M sub-track has achieved 84.169% on the masked test set, and 90.452% on the MR-All test set. As given in Tab. 6, we list the top-15 submissions from the leader-board. Comparing with the baseline models in Tab. 5, there are around 7% absolute improvements on the masked test set and the MR-All test set. For the Glint360K sub-track, the best model has achieved 88.972% on the masked test set, and 93.512% on the MR-All test set as shown in Tab. 7. Comparing with the baseline models in Tab. 5, there is around 6% absolute improvement on the masked test set and about 3.5% absolute improvement on the MR-All test set. Therefore, there is huge space for the training optimization to improve masked face recognition without the accuracy drop on the non-masked face recognition.

6. Ethical Considerations

Face recognition has been a controversial topic recently. There have been questions over ethical concerns about invasion of privacy, alongside how well face recognition systems recognize darker shades of skin (known as the bias problem). In the InsighFace track of this challenge, we employ existing academic data as the training datasets. Most of the identities inside the training data are well-known

[7]`https://ibug.doc.ic.ac.uk/resources/masked-face-recognition-challenge-workshop-iccv-21/`

Data	Backbone	Loss	Mask	Children	African	Caucasian	South Asian	East Asian	MR-All	Size(MB)	Time (ms)
MS1M	R18	ArcFace	47.853	41.047	62.613	75.125	70.213	43.859	68.326	91.658	1.856
MS1M	R34	ArcFace	58.723	55.834	71.644	83.291	80.084	53.712	77.365	130.245	3.054
MS1M	R50	ArcFace	63.850	60.457	75.488	86.115	84.305	57.352	80.533	166.305	4.262
MS1M	R100	ArcFace	69.091	66.864	81.083	89.040	88.082	62.193	84.312	248.590	7.031
MS1M+MA-0.1	R100	ArcFace	77.325	67.053	80.247	88.706	87.583	61.410	83.828	248.590	7.032
Glint360K	R18	ArcFace	53.317	48.113	68.230	80.575	75.852	47.831	72.074	91.658	2.013
Glint360K	R34	ArcFace	65.106	65.454	79.907	88.620	86.815	60.604	83.015	130.245	3.044
Glint360K	R50	ArcFace	70.233	69.952	85.272	91.617	90.541	66.813	87.077	166.305	4.340
Glint360K	R100	ArcFace	75.567	75.202	89.488	94.285	93.434	72.528	90.659	248.590	7.038
Glint360K+MA-0.1	R100	ArcFace	83.710	75.894	88.919	94.038	92.882	71.137	90.015	248.590	7.036

Table 5. The baseline performance of the masked face recognition challenge (the InsightFace track). "MR-All" denotes the verification accuracy on all multi-racial images. Inference time is evaluated on Tesla V100 GPU using onnxruntime-gpu==1.6. "MA-0.1" means masked face augmentation with a specific probability of 10%.

Rank	Participant	Mask	Children	African	Caucasian	South Asian	East Asian	MR-All	Size(MB)	Time (ms)	Feat Dim
1	agir	84.169	75.003	88.322	93.396	93.349	72.623	90.452	317.665	9.764	512
2	Rhapsody	83.831	64.152	86.516	93.459	92.461	72.616	90.098	327.618	9.083	512
3	paradox	84.183	75.105	88.436	93.374	92.398	71.127	89.710	357.488	9.520	512
4	mayidong	84.312	73.936	86.258	92.227	91.244	70.042	88.897	295.954	9.762	512
5	jerrysunnn	82.201	57.467	85.395	92.124	91.270	71.501	89.252	327.624	9.036	512
6	mind_ft	84.528	68.303	86.820	92.251	88.326	68.595	88.355	250.145	9.318	512
7	upupup	82.352	53.794	85.069	92.061	91.044	71.159	89.000	327.618	9.071	512
8	hammer_hk	81.706	58.097	84.853	91.917	91.163	70.783	88.894	327.618	9.078	512
9	unitykd0701	83.522	71.915	84.158	91.172	89.093	68.684	87.239	322.265	9.656	512
10	kisstea	83.831	71.050	83.828	90.866	90.054	67.108	87.046	288.849	9.079	512
11	Hello	79.308	67.126	86.012	92.168	92.603	68.694	88.529	250.145	9.253	512
12	JulieXU	82.209	70.465	83.823	90.734	89.889	68.194	87.236	235.505	7.909	512
13	hjgw	82.115	70.463	83.813	90.689	89.956	68.039	87.155	235.505	7.892	512
14	xuyang1	76.163	77.104	87.962	93.256	92.580	68.774	89.080	302.869	9.654	512
15	webill	78.123	72.833	85.942	92.099	91.151	69.273	88.333	253.756	9.688	512

Table 6. Top-15 submissions of the MS1M sub-track by 16 August 2021.

celebrities [14]. The pre-processed training data are compressed into the binary record and only released to relevant researchers to facilitate the reproducible training. Our private test data will not be released to the public to avoid the data privacy problem and ensure fairness for all participants. For the bias concern, we follow the most authoritative evaluation set up by NIST-FRVT [8]. We wish to promote fairness among deep face recognition and thus set up the multi-racial verification benchmark.

7. Conclusions and Future Works

In this InsightFace track report, we introduce our new benchmark for the evaluation of masked face recognition as well as non-masked face recognition. Based on our baseline solutions, we confirm the effectiveness of the naive masked face augmentation. As the challenge is still under-going, we will keep on updating the top-ranked solutions as well as this report on arxiv.

Besides the InsightFace track, there is also a parallel WebFace260M track [9] in the Masked Face Recognition challenge. The WebFace260M track is organized by Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jia Guo, Jiwen Lu, Dalong Du and Jie Zhou. Detailes can be found in the arxiv report [38], which will be also updated in the future.

[8] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[9] https://www.face-benchmark.org/challenge.html

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.

Rank	Participant	Mask	Children	African	Caucasian	South Asian	East Asian	MR-All	Size(MB)	Time (ms)	Feat Dim
1	jerrysunnn	88.972	86.628	93.064	96.278	95.578	77.969	93.512	728.879	18.981	512
2	mayidong	86.933	84.545	93.043	96.529	95.361	77.554	93.358	541.023	17.837	1024
3	derron	87.636	84.679	92.289	95.564	95.109	77.512	92.815	566.625	19.153	1024
4	mind_ft	86.962	81.321	92.566	96.100	95.417	76.024	92.705	456.505	17.801	512
5	DongWang	83.487	82.619	92.182	95.655	94.791	76.392	92.743	564.831	17.101	1024
6	helloface	87.881	82.423	90.337	94.382	93.272	73.818	91.252	248.590	7.025	512
7	didujustfart	89.130	83.165	90.373	95.181	93.557	72.638	90.776	453.391	13.584	512
8	yossi_avram	81.469	82.848	92.974	96.162	95.739	76.584	93.005	248.583	7.472	512
9	tinytan	84.657	81.438	91.139	95.183	94.238	75.369	91.823	497.807	14.875	1024
10	deepcam	84.391	82.781	90.871	94.573	94.176	75.821	91.834	284.102	11.276	512
11	sgglink	84.966	78.323	90.067	94.738	93.251	73.284	91.162	605.258	19.071	512
12	suanying	77.642	82.362	92.874	96.305	95.484	77.582	93.477	456.505	17.813	512
13	dingweichao	82.223	79.652	90.042	94.772	93.438	73.479	91.263	453.391	13.559	512
14	HYL_Dave	78.116	79.770	91.670	95.575	94.690	74.872	92.098	453.391	13.602	512
15	EvilGeniusFeng	76.249	80.140	91.042	94.860	94.300	75.468	92.011	250.390	9.363	512

Table 7. Top-15 submissions of the Glint360K sub-track by 16 August 2021.

[2] Jiajiong Cao, Yingming Li, and Zhongfei Zhang. Celeb-500k: A large training dataset for face recognition. In *ICIP*, 2018.

[3] Jiankang Deng, Jia Guo, Tongliang Liu, Mingming Gong, and Stefanos Zafeiriou. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *ECCV*, 2020.

[4] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 2019.

[5] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *CVPR*, 2020.

[6] Jiankang Deng, Jia Guo, Jing Yang, Alexandros Lattas, and Stefanos Zafeiriou. Variational prototype learning for deep face recognition. In *CVPR*, 2021.

[7] Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi. Lightweight face recognition challenge. In *ICCV Workshop*, 2019.

[8] Jiankang Deng, Yuxiang Zhou, and Stefanos Zafeiriou. Marginal loss for deep face recognition. In *CVPR Workshop*, 2017.

[9] Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi. A novel gan-based network for unmasking of masked face. *IEEE Access*, 2020.

[10] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *ECCV*, 2018.

[11] Shiming Ge, Chenyu Li, Shengwei Zhao, and Dan Zeng. Occluded face recognition in the wild by identity-diversity inpainting. *TCSVT*, 2020.

[12] Sixue Gong, Xiaoming Liu, and Anil K Jain. Jointly debiasing face recognition and demographic attribute estimation. *arXiv:1911.08080*, 2019.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014.

[14] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV*, 2016.

[15] Walid Hariri. Efficient masked face recognition method during the covid-19 pandemic. *arXiv:2105.03026*, 2021.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[17] 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.

[18] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *CVPR*, 2020.

[19] Yande Li, Kun Guo, Yonggang Lu, and Li Liu. Cropping and attention based approach for masked face recognition. *Applied Intelligence*, 2021.

[20] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *CVPR*, 2017.

[21] 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 *ICB*, 2018.

[22] David Montero, Marcos Nieto, Peter Leskovsky, and Naiara Aginako. Boosting masked face recognition with multi-task arcface. *arXiv:2104.09874*, 2021.

[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 *CVPR Workshop*, 2017.

[24] Aaron Nech and Ira Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *CVPR*, 2017.

[25] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *BMVC*, 2015.

[26] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

[27] 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 *WACV*, 2016.

[28] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *NeurIPS*, 2014.

[29] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, 2014.

[30] Feng Wang, Weiyang Liu, Haijun Liu, and Jian Cheng. Additive margin softmax for face verification. *SPL*, 2018.

[31] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Zhifeng Li, Dihong Gong, Jingchao Zhou, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *CVPR*, 2018.

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

[33] Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In *CVPR*, 2020.

[34] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *ICCV*, 2019.

[35] Qiangchang Wang, Tianyi Wu, He Zheng, and Guodong Guo. Hierarchical pyramid diverse attention networks for face recognition. In *CVPR*, 2020.

[36] Renliang Weng, Jiwen Lu, and Yap-Peng Tan. Robust point set matching for partial face recognition. *TIP*, 2016.

[37] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. Investigating bias and fairness in facial expression recognition. *arXiv:2007.10075*, 2020.

[38] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jia Guo, Jiwen Lu, Dalong Du, and Jie Zhou. Masked face recognition challenge: The WebFace260M track report. *arXiv:2108.07189*, 2021.

[39] 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 *CVPR*, 2021.