Technical Details of our submissions of MFR-ICCV2021 Challenge, WebFace260M Track

Taewan Ethan Kim
Face Group
AI Lab.
Kakao Enterprise
ethan.y@kakaoenterprise.com

Abstract

In this report, we describe the technical details of our submissions. We used the WebFace42M dataset provided. All samples from the dataset are augmented via two off-the-shelf masked-face image generation tools. In compliance with the FRUITS-1000 protocol, most BatchNorm (BN) weights and running statistics are absorbed in the previous convolution weights manually so that it makes us possible to employ a larger model. As a result, we used a ResNet240 as our feature-extraction network. An angle-based feature distillation algorithm was employed to achieving further improvement. To train this model, we applied data-parallelism and model-parallelism on feature-extraction network and classification weight, respectively. Other distributed training techniques are also employed in terms of total training time and highest-peak memory consumption. ***Finally, we placed 2nd (out of 203 participants) in Masked Face Recognition (MFR).***

1. Dataset

We used the provided WebFace42M dataset [10] only for training.

Masked-Face Augmentation: To deal with MFR, we used two off-the-shelf masked-face synthesis tool [1, 4]. Examples of synthesized masked-face images are shown in Fig. 1. The reason we employed the two masked-face generation tools simultaneously is that these tools show different characteristics of masked-face synthesis in the shape and texture of template-masks (See Fig. 1). The combination of these two methods enriched the variation of the masked-face images.

2. Model

In the first round, we adopted the Mixture of Expert (MoE) [6] for dealing with masked and non-masked face images simultaneously. In this model, we explicitly used two *output layers*[1] for each masked and non-masked face image. However, we observed the plain ResNet [2] variants (single *output layer* as usual) give higher performance than that of our MoE variants. Consequently, we describe the ResNet variants we used in the next.

Figure 1. Examples of synthesized masked-face images from the WebFace42M [10]. Samples in the first and second row are generated via [1], and [4], respectively. From [1] we can augment synthesized images with various shapes and textures of face masks, but resulting images are slightly unnatural. In contrast, [4] generates relatively natural masked-face images, however, limited to synthesis with various commodity face masks.

ResNet variants: Fig. 2 depicts our feature-extraction network (i.e. mostly inspired by [2]). The number of *Blocks* of the ResNet variants is listed in the Table 1. Since we did not know the largest model acceptable under the FRUITS-1000 protocol [10], we gradually increased the trainable parameters of our model. Finally, ResNet240 was employed in our feature-extraction network. In this model, most BN weights and running statistics are absorbed into the previous convolution weights manually after finishing a training process.

3. Our Approach

Knowledge distillation (KD) [5] is a popular algorithm to improve the performance of a given network that contains a limited number of trainable parameters. Many research works make use of this method for their lightweight face recognition [3, 8]. However, there is little use of

[1]The *output layer* is depicted in Fig. 2.

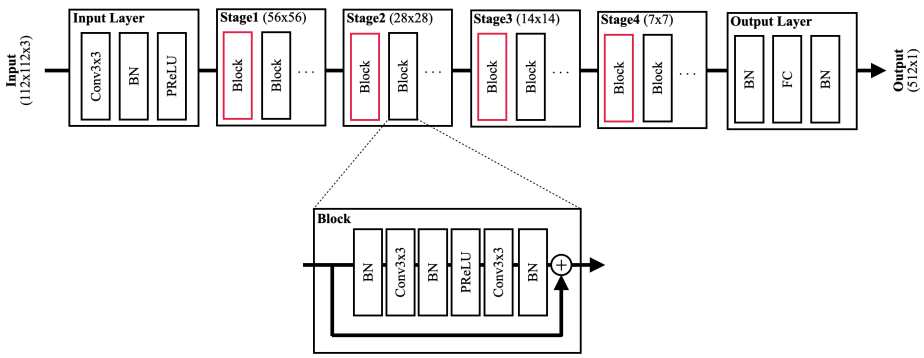

Figure 2. Block diagram of the network we used in this challenge. The channel-size for each *Stage* was set to 64, 128, 256, and 512, respectively. In the first *Blocks* (denoted in red rectangles), we set the stride of 2 for each of the first *Conv3x3* layers.

Architecture	Number of Blocks			
	Stage1	Stage2	Stage3	Stage4
ResNet120	3	13	40	3
ResNet140	3	15	48	3
ResNet200	6	26	60	6
ResNet240	3	25	88	3
ResNet600	3	70	220	3
ResNet1.2K	3	80	512	3

Table 1. ResNet variants we used in this challenge. The ResNet240 is adapted for the final submission. We utilized the ResNet600 and ResNet1.2K as a teacher model for the knowledge distillation [5, 3, 8]

.

it to improve performance for the relatively larger models. In this challenge, we employed the KD and observed a performance improvement. Specifically, we trained the ResNet600 with the CosFace [7] on the mask-augmented WebFace42M as a teacher network in advance. To train a student model (ResNet240 in our case), we adopted a convex combination of the CosFace and KD loss (See [8] for details) on the same train-dataset. It is worth to note we did not use any additional data or pre-trained model at all.

4. Results

We summarize our submission results in Table 2. Both Masked Face Recognition (MFR) score and the Standard Face Recognition (SFR) score are measured from the evaluation server. Details for the test set can be found in [9].

Submission	Score	
	MFR	SFR
ResNet240	0.0803	0.0235
ResNet240-Distill	**0.0769**	0.0241

Table 2. The last two submission results. We used the ResNet600 as our teacher network in the *ResNet240-Distill* submission.

References

[1] Aqeel Anwar. Mask the face. `https://github.com/aqeelanwar/MaskTheFace`, 2020. Online; accessed 05-Oct-2021.

[2] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *CoRR*, abs/1801.07698, 2018.

[3] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Ngan Le. Shrinkteanet: Million-scale lightweight face recognition via shrinking teacher-student networks. *CoRR*, abs/1905.10620, 2019.

[4] Jia Guo, Jiankang Deng, Xiang An, and Jack Yu. Insightface. `https://github.com/deepinsight/insightface/tree/master/recognition/_tools_`, 2020. Online; accessed 05-Oct-2021.

[5] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015.

[6] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *CoRR*, abs/1701.06538, 2017.

[7] 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. *CoRR*, abs/1801.09414, 2018.

[8] Mengjia Yan, Mengao Zhao, Zining Xu, Qian Zhang, Guoli Wang, and Zhizhong Su. Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition. *CoRR*, abs/1910.04985, 2019.

[9] 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. *CoRR*, abs/2108.07189, 2021.

[10] 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. *CoRR*, abs/2103.04098, 2021.