When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework —Supplmentary Material—

Zhizhong Huang[1] Junping Zhang[1] Hongming Shan[2,3*]

[1] Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China

[2] Institute of Science and Technology for Brain-inspired Intelligence and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China

[3] Shanghai Center for Brain Science and Brain-inspired Technology, Shanghai 200031, China

{zzhuang19, jpzhang, hmshan}@fudan.edu.cn

Abstract

This supplement provides more details on the collected dataset, training, and quantitative comparison for FAS.

A. Dataset

The collection of our dataset is summarized in the following three steps. First, we first use the public Azure Facial API [2] to estimate the ages and genders of faces from the clean MS-Celeb-1M dataset provided by [4]. Second, we randomly select faces from a total of 5M faces to check whether the faces are correctly labeled, and try our best to manually correct them if any apparent mistakes; we mainly focus on the young ages under 20 that are often mislabeled by the API [2]. Last, a large-scale balanced age dataset is constructed by balancing both age and gender.

Fig. A.1 presents example images and dataset statistics of SCAF.

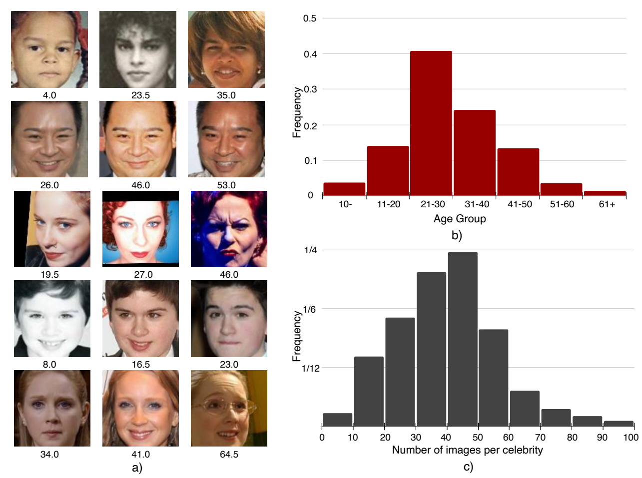

Figure A.1: Sample faces (a) and dataset statistics (b and c) on SCAF.

B. Training Details

We adopted ResNet-50 similar to [4] as the encoder E. In the decoder D, the identity age condition is bilinearly upsampled and processed with multi-level high-resolution features extracted from E by two ResBlocks [5], each of which follows the instance normalization [8] and ReLU activation, the synthesized faces of size 112×112 were produced by one 1×1 convolutional layer. There are four ICBs in ICM. In the discriminator D_{img}, six convolutional layers with strides of 2, 2, 2, 2, 1, 1 follow the spectral normalization [6] and leaky ReLU activation except the last one, outputting 7×7 confidence map. The AIFR is optimized by SGD with an initial learning rate of 0.1 and momentum of 0.9 while the ICM, decoder D, and D_{img} are trained by Adam with a fixed learning rate of 10^{-4}, β_1 of 0.9 and β_2 of 0.99 for the face age synthesis. We trained all models with a batch size of 512 on 8 NVIDIA GTX 2080Ti GPUs, 110K iterations for LCAF and 36K iterations for SCAF. The learning rate of AIFR was warmed up linearly from 0 to 0.1, reduced by a factor of 0.1, at iterations 5K, 70K, and 90K on LCAF and 1K, 20K, 23K on SCAF, respectively. The hyper-parameters in the loss functions were empirically set as follows: $\lambda^{\text{AIFR}}_{\text{age}}$ was 0.001, $\lambda^{\text{AIFR}}_{\text{id}}$ was 0.002, $\lambda^{\text{FAS}}_{\text{adv}}$ was 75, $\lambda^{\text{FAS}}_{\text{id}}$ was 0.002, and $\lambda^{\text{FAS}}_{\text{age}}$ was 10. The multiplicative margin and scale factor of CosFace loss [9] were set to 0.35 and 64, respectively. All images were aligned to 112×112,

*Corresponding author

with five facial landmarks detected by MTCNN [11], and linearly normalized to $[-1, 1]$.

C. Quantitative Comparison for FAS

We quantitatively evaluate all models by the following two metrics:

1) age accuracy: we trained a ResNet-100 model on 80% faces of LCAF using ℓ_{AE} as the loss function to predict the ages of all synthesized faces and the proportion of the predicted ages falling into the target age groups is the age accuracy; and

2) identity preservation: an external well-trained face recognition model, ResNet-100 network pre-trained on MS-Celeb-1M dataset provided by [4], is used for fair comparisons to compute the cosine similarity between the input and synthesized faces.

We trained all models on the SCAF dataset for fair comparisons, which are then directly applied to three external cross-age datasets: MORPH [7], FG-NET [1] and CACD [3]. Table 2 presents the quantitative results of different face aging/rejuvenation methods, including CAAE [12], IPCGAN [10], our proposed MTLFace and its variant (w/o ICM). It can be observed that MTLFace outperforms CAAE and IPCGAN by a clear margin; this is a direct results of the AIFR task and ICM. On the other hand, without ICM, MTLFace reduces to a common cGANs-based method that uses one-hot encoding to control face aging/rejuvenation at the group level. Remarkably, the MTLFace without ICM still outperforms these two baseline methods, implying that our multi-learning framework with attention-based feature decomposition is effective in improving the age accuracy and identity preservation.

References

[1] FG-NET aging database. https://yanweifu.github.io/FG_NET_data, 2014. 2

[2] Microsoft Azure. Microsoft azure cognitive services facial recognition. https://azure.microsoft.com/en-us/services/cognitive-services/face/. 1

[3] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans. Multimedia*, 17(6):804–815, 2015. 2

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4690–4699, 2019. 1, 2

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 770–778, 2016. 1

[6] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *Proc. Int. Conf. Learn Represent.*, pages 1–26, 2018. 1

[7] K. Ricanek and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. In *Proc. Int. Conf. Autom. Face Gesture Recognit.*, pages 341–345, 2006. 2

[8] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 1

[9] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. CosFace: Large margin cosine loss for deep face recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5265–5274, 2018. 1

[10] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7939–7947, 2018. 2

[11] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 2

[12] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5810–5818, 2017. 2