Supplementary Materials for Hierarchical Face Aging with Disentangled Adversarial Learning

Peipei Li[1,3⋆], Huaibo Huang[1,3,4⋆], Yibo Hu[1], Xiang Wu[1], Ran He[1,2,3∗], and Zhenan Sun[1,2,3]

[1] Center for Research on Intelligent Perception and Computing, NLPR, CASIA
[2] Center for Excellence in Brain Science and Intelligence Technology, CAS
[3] School of Artificial Intelligence, University of Chinese Academy of Sciences
[4] Artificial Intelligence Research, CAS, Jiaozhou, Qingdao, China
{peipei.li, huaibo.huang}@cripac.ia.ac.cn,
{huyibo871079699,alfredxiangwu}@gmail.com,{rhe, znsun}@nlpr.ia.ac.cn

1 Network Architecture

Layer	Input	Output Size
	Encoder	
E_conv0	x	$128 \times 128 \times 96$
E_ResBlock0	E_conv0	$64 \times 64 \times 192$
E_ResBlock1	E_ResBlock0	$32 \times 32 \times 384$
E_ResBlock2	E_ResBlock1	$16 \times 16 \times 768$
E_ResBlock3	E_ResBlock2	$8 \times 8 \times 1536$
E_ResBlock4	E_ResBlock3	$4 \times 4 \times 1536$
E_flatten,E_fc	E_ResBlock4	792
E_split	E_fc	128,128,140,140,256
E_reparal1	E_split[0,1]	128,128
E_reparal2	E_split[2,3]	140,140
E_reparal3	E_split[4]	256
	Age Embedding	
A_fc0	E_reparal1	256
A_relu	A_fc0	256
A_fc1	A_relu	384
	Identity Embedding	
I_fc0	E_reparal3	320
I_relu	I_fc0	320
I_fc1	I_relu	384

Table 1. Network architecture of the encoder, age embedding and identity embedding of DAAE. The filter/stride of E_conv0 is $5 \times 5/2$. The architecture of E_ResBlock are shown in Fig. 1.

⋆ Equal contribution

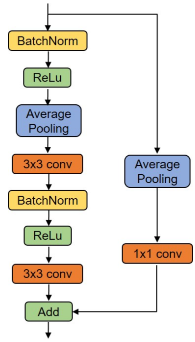

Fig. 1. A Residual Block in DAAE's encoder.

For convenience of calculations, all images are resized to 256×256 before being sent to DAAE. The detailed structures of the encoder, age embedding and identity embedding in the DAAE models are shown in Table 1. The Residual Block in the encoder is shown in Fig. 1. The generator architecture of DAAE is the same as BigGAN [1]. In DAAE, we use the disentangled age and identity embeddings instead of the single shared class embedding in BigGAN [1] as the condition for face aging. Specifically, we split the extraneous information along the channel dimension into 7 chunks $K_E[i], i \in [0,1,2,3,4,5,6]$ and send the first chunk $K_E[0]$ into the input layer of the generator. In addition, we split the age information and identity information along the channel dimension into 3 chunks $K_A[i]$ and $K_I[i], i \in [0,1,2]$, respectively. When $1 \leq i \leq 3$, the extraneous chunk $K_E[i]$ is concatenated with the identity chunk $K_I[i-1]$ and passed to the residual block $R[i-1]$ as conditioning vectors. When $4 \leq i \leq 6$, the extraneous chunk $K_E[i]$ is concatenated with the age chunk $K_A[i-4]$ and passed to the residual block $R[i-4]$ as conditioning vectors.

2 Evaluation Metric of Age Estimation

We evaluate the performance of DAAE on age estimation tasks with the Mean Absolute Error(MAE). MAE is widely used to evaluate the performance of age estimation models,which is defined as the average of the distances between the inferred age $\hat{y}$ and the ground-truth y and can be written as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \quad (1)$$

3 Face Aging on AgeDB

By manipulating the mean value μ_A and sampling from the age distribution, DAAE can generate arbitrary age based on the input. Fig. 2 presents the age regression and progression results on AgeDB. Obviously, with the age label decreasing, the skin is firmer and the face is more round. While with the age label

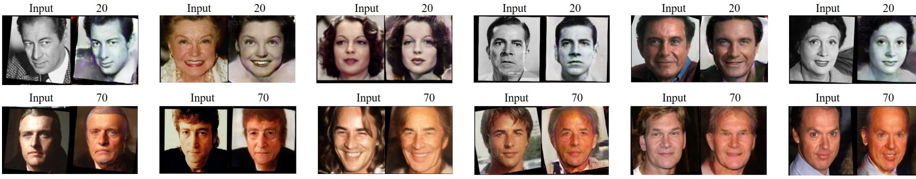

Fig. 2. Face Aging results on AgeDB. The first row shows the age regression results, while the second row shows the age progression results.

Fig. 3. Face Aging results on Morph. Note that the target age (13 years old) is not existing in Morph.

increasing, the hair becomes white from black and wrinkles are significantly increased.

4 Generalized Abilities of DAAE

Benefiting from the introduced age distribution prior and the adopted disentangling manner, DAAE has the ability to predict the underlying age representations in the latent space, which further empowers DAAE potential to deal with the data-poor ages or missing ages. For example, as shown in Fig. 3, when performing face aging with 13 years old on Morph (16 to 77 years old), it is amazing that DAAE can synthesize photo-realistic aging images. This observation demonstrates the generalized abilities of our framework and indicates that DAAE can efficiently predict the underlying age distribution even with long-tailed dataset.

5 Ablation Study

In this subsection, we conduct the ablation study on Morph to evaluate the effectiveness of the disentangled adversarial learning L_{adv}, the two regularization terms $L_{reg}^{(age)}$, $L_{reg}^{(id)}$ and the hierarchical conditional generation. We report the qualitative visualization and the quantitative results for a comprehensive comparison as the ablation study.

To verify the effectiveness of the disentangled adversarial learning mechanism, we remove the disentangled adversarial loss and use the Eq.(2) as the total objective function. To verify the effectiveness of the two regularization terms, we remove the loss function of Eq.(8) or Eq.(9). To verify the effectiveness of

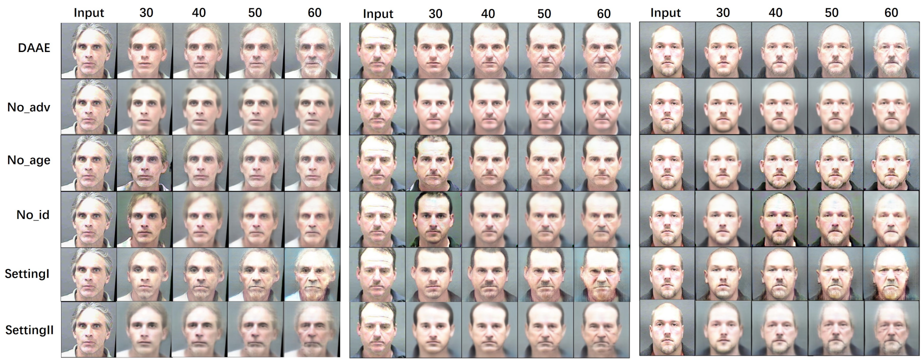

Fig. 4. Model comparisons: the face aging results of DAAE and its variants.

the hierarchical conditional generation, we design two variants by passing the age, identity and extraneous conditions at different layers of our generator. For the setting I, when $1 \leq i \leq 3$, the extraneous chunk $K_E[i]$ is concatenated with the age chunk $K_A[i-1]$ and passed to the residual block $R[i-1]$ as conditioning vectors. When $4 \leq i \leq 6$, the extraneous chunk $K_E[i]$ is concatenated with the identity chunk $K_I[i-4]$ and passed to the residual block $R[i-4]$ as conditioning vectors. For the setting II, we concatenate the age, identity and extraneous vectors along the channel dimensions and send it to the input layer of the generator.

Figure 4 shows visual comparisons between DAAE and its five incomplete variants. We observe that the proposed DAAE is visually better than its variants across all ages. Without the disentangled adversarial loss, the generated faces are blur. Without the age regularization term, the age characteristics are not obvious enough. Without the identity regularization term, part of the identity of the generated faces are lost. With the generator's architecture changed, the generated faces lack of detailed texture information and the identity information is not well preserved. The visual results demonstrate that the disentangled adversarial learning, the two regularization terms and the hierarchical conditional architecture are essential to the proposed DAAE.

We report face verification results of DAAE and its five variants for a comprehensive comparison as the ablation study. Table 2 presents the comparison results. Similar to the visualization results, we observe the face verification accuracy will decrease when the disentangled adversarial loss is removed or the one of the two regularization terms is removed or the generator's architecture is changed. These phenomena indicate that each component in our method is essential for face aging.

Testing Face	AG1	AG2	AG3
w/oL_{adv}	95.74	95.61	95.60
$w/oL_{reg}^{(age)}$	97.84	97.89	97.89
$w/oL_{reg}^{(id)}$	96.21	93.01	93.09
Setting I	89.04	87.35	87.35
Setting II	76.10	72.20	72.15
Ours	99.48	99.36	99.36

Table 2. Face verification results(%) of the ablation study on Morph.

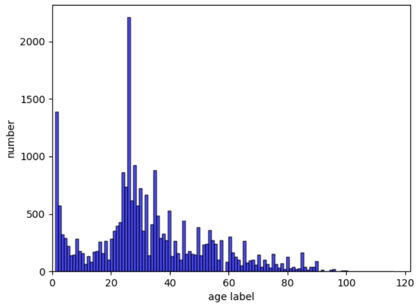

Fig. 5. Age distribution in UTKFace

6 More Face Aging Results on UTKFace

In this section, we display the face aging results on UTKFace in Figure 6, which is from 0 to 118 years old with 1-year-old aging interval. Since the images in UTKFace ages from 0 to 116, it is amazing that UVA can synthesize photorealistic aging images with 117 and 118 years old. Besides, we discover that UTKFace exhibits a long-tailed distribution where a large part of age classes only have few images. As shown in Figure 5, the number of face images over 90 years old is in the low single digits, while the number of images in 25 years old is more than 2000. Our DAAE can achieve face aging even with few training images. The face aging results on UTKFace indicate that DAAE can efficiently and accurately estimate the age-related distribution from the facial images, even if the training set performs a long-tailed age distribution.

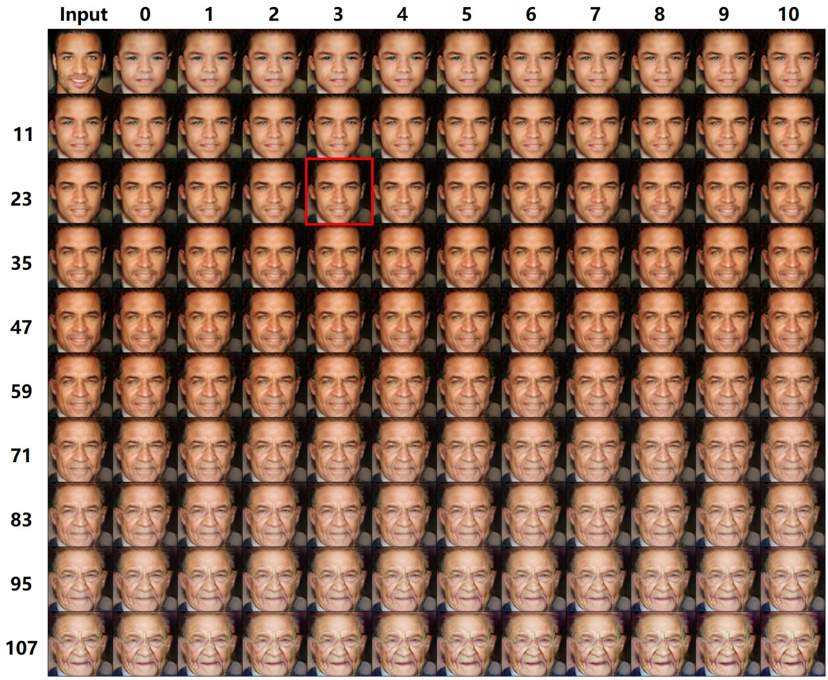

Fig. 6. Face aging results on UTKFace from 0 to 118 years old with 1-year-old interval. The facial age in the red box is equal to the input (27 years old).

References

1. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: ICLR (2019)