Hierarchical Face Aging through Disentangled Latent Characteristics

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

Abstract. Current age datasets lie in a long-tailed distribution, which brings difficulties to describe the aging mechanism for the imbalance ages. To alleviate it, we design a novel facial age prior to guide the aging mechanism modeling. To explore the age effects on facial images, we propose a Disentangled Adversarial Autoencoder (DAAE) to disentangle the facial images into three independent factors: age, identity and extraneous information. To avoid the "wash away" of age and identity information in face aging process, we propose a hierarchical conditional generator by passing the disentangled identity and age embeddings to the high-level and low-level layers with class-conditional BatchNorm. Finally, a disentangled adversarial learning mechanism is introduced to boost the image quality for face aging. In this way, when manipulating the age distribution, DAAE can achieve face aging with arbitrary ages. Further, given an input face image, the mean value of the learned age posterior distribution can be treated as an age estimator. These indicate that DAAE can efficiently and accurately estimate the age distribution in a disentangling manner. DAAE is the first attempt to achieve facial age analysis tasks, including face aging with arbitrary ages, exemplar-based face aging and age estimation, in a universal framework. The qualitative and quantitative experiments demonstrate the superiority of DAAE on five popular datasets, including CACD2000, Morph, UTKFace, FG-NET and AgeDB.

Keywords: Facial age analysis; Variational autoencoder

1 Introduction

Facial age analysis, including face aging, exemplar-based face aging and age estimation, is one of the crucial components in modern face analysis for entertainment and forensics. Face aging aims to aesthetically render the facial appearance

[⋆] Equal contribution

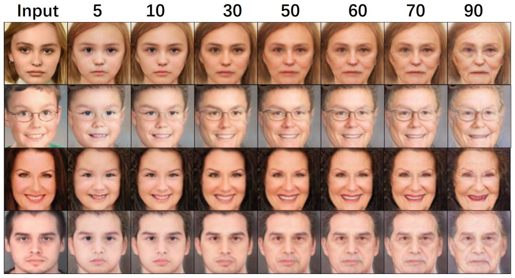

Fig. 1. Continuous face aging results on UTKFace. The first column are the inputs and the rest columns are the synthesized faces from 5 to 90 years old.

based on the given age, while exemplar-based face aging aims to render the facial appearance according to the age of the exemplar face. In recent years, with the developments of the generative adversarial network (GAN) [8], impressive progress [40, 35, 32, 18, 17, 20] has been made on face aging. Since all the existing age datasets, such as CACD2000, Morph and UTKFace, perform a long-tailed age distribution, it is difficult for current methods to describe the aging mechanism for the imbalanced age distribution. Researchers often employ the time span of 10 years as the age clusters for face aging. This age cluster strategy potentially limits the diversity of aging patterns, especially for the younger with the large inter-class appearance variance.

Recently, Variational Auto-Encoder (VAE) [15] shows the promising ability in discovering the underlying data distribution in the latent space [2], [12], [34]. Conditional Adversarial Autoencoder (CAAE) [40] proposes to learn a face manifold in the latent space. With the given age label and the learned face manifold, CAAE achieves continuous face aging. However, there are still four limitations with CAAE: 1) It only focuses on the learning of the image representation and ignores the age representation. The possibility of utilizing it to handle more age-related analysis, such as age estimation and exemplar-based face aging, is limited. 2) CAAE conducts face aging on the cropped faces and reckons without the extraneous information, such as hair. 3) The identity and age embeddings are only injected into the input layer of the generator, which leads to the “wash away” of them during generation. 4) It still adopts a group-based training strategy.

To address the mentioned issues in CAAE, we propose a Disentangled Adversarial Autoencoder (DAAE). We first design a facial age distribution as the facial age prior to directly learn the age representation from the image. A well-trained identity classifier is also employed to supervise the identity feature learning in a knowledge distilling way [10]. Besides, we discover it is also important to disentangle the extraneous information, including hairstyle and pose, from the given image. Following most VAE based methods, we introduce the Gaussian distribution as the prior for extraneous information. Supervised by the variational evidence lower bound (ELBO) [15, 25] and identity knowledge distillation, the facial images are disentangled into three independent factors: age, identity and extraneous information. This disentangling manner makes DAAE more flexible and controllable for facial age analysis. To avoid the "wash away" of identity and age information during generation, we propose a hierarchical aging architecture by passing the disentangled identity and age vectors to the high-level and low-level layers with class-conditional BatchNorm. To synthesize photo-realistic facial images, a disentangled adversarial learning mechanism is introduced to optimize the inference network and the generator jointly and adversarially. Inspired by IntroVAE [12], DAAE is expected to have the capability to self-estimate the age accuracy, identity preserving and generation quality of the synthesized images.

Finally, by manipulating the mean value of age distribution, DAAE can easily realize facial face aging with arbitrary ages, whether the age exists or not in the training dataset. Further, by extracting age information from an exemplar facial image, DAAE can achieve exemplar-based face aging. Moreover, given a face image as input, we can easily obtain its age representation, which indicates the ability of DAAE to achieve age estimation. As stated above, we can implement three different facial age analysis tasks in a universal framework. To the best of our knowledge, DAAE is the first attempt to achieve facial age analysis, including face aging, exemplar-based face aging and age estimation, in a universal framework. The main contributions of DAAE are as follows:

- We propose a novel Disentangled Adversarial Autoencoder (DAAE) for facial age analysis tasks, including face aging, exemplar-based face aging and age estimation. We design two different priors as well as identity knowledge distillation to assist disentangling the facial images into three independent factors: age, identity and extraneous information.
- To fully utilize the disentangled low-level age, high-level identity and mixed-level extraneous information, we propose a hierarchical conditional generator with class-conditional BatchNorm.
- We propose a disentangled adversarial learning mechanism by training the encoder and generator with age preserving regularization and identity knowledge distillation in an introspective adversarial manner.
- Extensive qualitative and quantitative experiments demonstrate that DAAE successfully formulates the facial age prior in a disentangling manner, obtaining state-of-the-art results on the five popular datasets.

2 Related Work

2.1 Face Aging

Recently, deep conditional generative models have shown considerable ability in face aging [40, 32, 18, 17]. Zhang et al. [40] propose a Conditional Adversarial Autoencoder(CAAE) to transform an input facial image to the target age. Yang et al. [35] propose a pyramid GAN to simulate the aging effects in a finer manner. To capture the rich textures in the local facial parts, Li et al. [18]

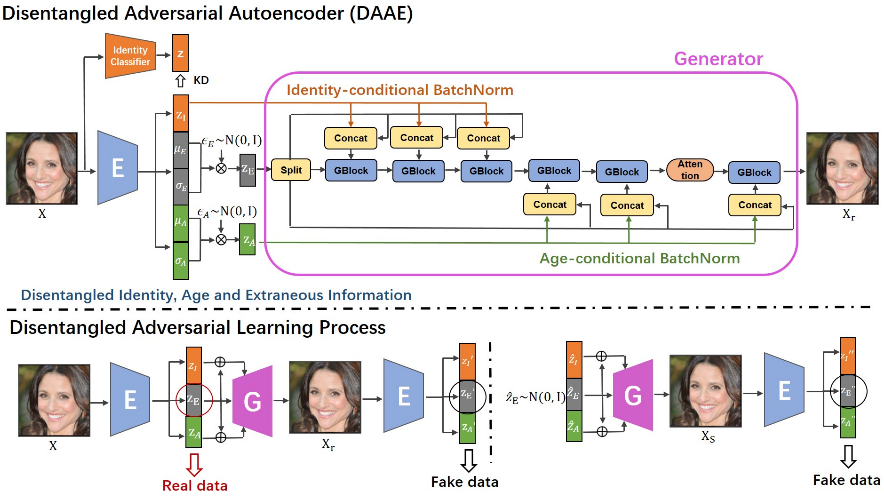

Fig. 2. Overview of the architecture and training flow of our approach. Our model contains two components, the inference network E and the hierarchical generative network G. X, X_r and X_s denote the real sample, the reconstruction sample and the new sample, respectively. Please refer to Section 3 for more details.

propose a Global and Local Consistent Age Generative Adversarial Network (GLCA-GAN). Meanwhile, Identity-Preserved Conditional Generative Adversarial Networks (IPCGAN) [32] introduces an identity-preserved term and an age classification term into face aging. Liu et al. [20] imposes attribute information to guide the aging process and proposes a wavelet-based discriminator to encourage generation quality. Although these methods have achieved promising visual results, they have limitations in discovering the disentangled factors in face aging. Besides, the image and age label are only injected into the input layer of the generator, leading to the "wash away" of image and age information during generation. For non-deep learning methods, [28] demonstrates that disentangling the common and individual components in facial images is crucial for face aging. [22] proposes Multi-Attribute Robust Component Analysis (MA-RCA) that incorporates knowledge from age and identity for age progression. In this paper, benefiting from the proposed VAE-based method, we focus on disentangling facial images into three independent factors: age, identity and extraneous information.

2.2 Variational Autoencoder

Variational Autoencoder (VAE) [15, 25] consists of two networks: an inference network $q_\phi(z|x)$ maps the data x to the latent variable z, which is assumed as a gaussian distribution, and a generative network $p_\theta(x|z)$ reversely maps the latent variable z to the visible data x. The object of VAE is to maximize the variational lower bound (or evidence lower bound, ELBO) of $\log p_\theta(x)$:

$$\log p_\theta(x) \geq E_{q_\phi(z|x)} \log p_\theta(x|z) - D_{KL}(q_\phi(z|x)||p(z)) \tag{1}$$

VAE has shown promising ability to generate complicated data, including faces [12], natural images [9], text [29] and segmentation [11, 30]. Inspired by IntroVAE [12], we propose a disentangled adversarial learning mechanism, which self-estimates the age accuracy, identity preserving and generation quality of the synthesized images.

2.3 Age Estimation

Age estimation aims to automatically label a given face with an exact age or age group [19]. Ranking-CNN and label distribution learning (LDL) based age estimation achieve state-of-the-art performance. Ranking-CNN consists of a series of CNNs, each of which learns a binary classification. Then these binary values are aggregated for the final result [4]. To model the correlations among different ages, label distribution learning [6] utilizes a specific distribution to formulate the aging mechanism. Inspired by it, we design a new age distribution as the facial age prior to directly learn the age information from the image. In this way, the mean value of the learned age posterior distribution can be treated as an age estimator.

3 Approach

In this paper, we propose a Disentangled Adversarial Autoencoder (DAAE) for face aging, examplar-based aging and age estimation. The key idea is to disentangle the facial image into three independent factors, i.e., age, identity and extraneous information. A hierarchical aging generator is introduced to produce photo-realistic images by transferring the identity and age information sequentially. As depicted in Fig. 2, two different priors as well as identity distillation are assigned to regularize the inferred representations. The inference network E and the generator network G are trained in an introspective disentangling manner.

3.1 Disentangled Variational Representations

In the original VAE [15], a probabilistic latent variable model is learned by maximizing the variational lower bound to the marginal likelihood of the observable variables. However, the latent variable z is difficult to interpret and control, since each element of z is treated equally in training. To alleviate this, we manually split z into three parts, i.e., z_A representing the age information, z_I representing the identity information, and z_E representing the extraneous information.

Assume that z_A, z_I and z_E are independent on each other, then the posterior distribution can be written as: $q_\phi(z|x) = q_\phi(z_A|x)\, q_\phi(z_I|x)\, q_\phi(z_E|x)$. The prior distribution $p(z) = p_A(z_A)\, p_I(z_I)\, p_E(z_E)$, where $p_A(z_A)$, $p_I(z_I)$ and $p_E(z_E)$

are the prior distributions for z_A, z_I and z_E, respectively. According to Eq. (1), the optimization objective for the modified VAE is to maximize the lower bound of $\log p_\theta\left(x\right)$:

$$\begin{aligned} \log p_\theta\left(x\right) \geq & E_{q_\phi\left(z_A, z_I, z_E|x\right)} \log p_\theta\left(x|z_A, z_I, z_E\right) \\ & - D_{KL}\left(q_\phi\left(z_A|x\right)||p_A\left(z_A\right)\right) \\ & - D_{KL}\left(q_\phi\left(z_I|x\right)||p_I\left(z_I\right)\right) \\ & - D_{KL}\left(q_\phi\left(z_E|x\right)||p_E\left(z_E\right)\right), \end{aligned} \tag{2}$$

where the first item regularizes the reconstruction accuracy, and the last three regularize the latents, i.e., z_A, z_I and z_E, to learn different types of facial information.

To capture facial aging characteristics, the prior $p_A\left(z_A\right)$ for z_A is designed as a facial aging-specific distribution. It is set to be a centered isotropic multivariate Gaussian, i.e., $p_A\left(z_A\right) = \mathbb{N}\left(\mathbf{y}, \mathbf{I}\right)$, where $\mathbf{y}$ is a vector filled by the age label y of x. We assume the posterior $q_\phi\left(z_A|x\right)$ also follows a centered isotropic multivariate Gaussian, i.e., $q_\phi\left(z_A|x\right) = \mathbb{N}\left(z_A; \mu_A, \sigma_A^2\right)$. As depicted in Fig. 2, μ_A and σ_A are the output vectors of the inference network E. The input z_A for the generator G is sampled from $\mathbb{N}\left(z_A; \mu_A, \sigma_A^2\right)$ using a reparameterization trick, i.e., $z_A = \mu_A + \epsilon_A \odot \sigma_A$, where $\epsilon_A \sim \mathbb{N}\left(\mathbf{0}, \mathbf{I}\right)$. The negative version of the second term in Eq. (2) can be formed as

$$L_{kl}^{(age)} = \frac{1}{2}\sum_{i=1}^{C_A}((\mu_A^i - y)^2 + (\sigma_A^i)^2 - \log((\sigma_A^i)^2) - 1), \tag{3}$$

where y is the age label of the input x and C_A denotes the dimension of z_A. Noted that $(\mu_A^i - y)^2$ in Eq. (3) can be viewed as an L2 constraint between the predicted μ_A and the age label y, which leads to the capability of the proposed method to estimate facial age.

For the difficulty in modeling the identity space with a simple distribution, the prior $p_I\left(z_I\right)$ is obtained through a well-pretrained identity classifier C. Assume that for each $z \in p_I\left(z_I\right)$, there exists a facial image x satisfying that $z = F(x)$, where $F(x)$ is the extracted feature before the softmax layer in C. We employ a paired L1 loss function between the predicted z_I and the extracted feature $F(x)$ to describe the relations between the posterior $q_\phi(z_I|x)$ and the prior $p_A(z_A)$. The paired L1 loss distills the identity knowledge from the identity classifier C to the inference network E, where C and E can be regarded as the teacher net and student net in knowledge distilling [10], respectively. This loss function is formed as

$$L_{kd}^{(id)} = \sum_{i=1}^{C_I}|z_I^i - F(x)|, \tag{4}$$

where z_I and $F(x)$ are extracted from the same image x by the inference network E and the identity classifier C, and C_I denotes the dimension of z_I.

Following the original VAE [15], we set the prior $p_E\left(z_E\right) = \mathbb{N}\left(\mathbf{0}, \mathbf{I}\right)$ and the posterior $q_\phi\left(z_E|x\right) = \mathbb{N}\left(z_E; \mu_E, \sigma_E^2\right)$. Similar to Eq. (3), the negative version of the last term in Eq. (2) can be reformed as

$$L_{kl}^{(ext)} = \frac{1}{2}\sum_{i=1}^{C_E}((\mu_E^i)^2 + (\sigma_E^i)^2 - \log((\sigma_E^i)^2) - 1), \tag{5}$$

where μ_E and σ_E are the output vectors of E, C_E denotes the dimension of z_E. The reconstruction term in Eq. (2) can be optimized by the following form

$$L_{rec} = \frac{1}{2}\|x - x_r\|_F^2, \tag{6}$$

where x and x_r are the input and output images, respectively.

In summary, the optimization object in Eq. (2) can be rewritten in the negative version:

$$L_{vae} = L_{rec} + L_{kl}^{(age)} + L_{kd}^{(id)} + L_{kl}^{(ext)}. \tag{7}$$

3.2 Hierarchical Conditional Generator

In order to effectively utilize the disentangled low-level age, high-level identity and mixed-level extraneous information (e.g., pose, skin color), we propose a hierarchical conditional generator, borrowing from Conditional Batch Normalization (CBN) [5] literature. We regard all of the age, identity and most extraneous information as the conditions for face aging.

We first split the extraneous information into several parts and the first part is regarded as the input of the generator. Since extraneous information contains both high-level (e.g., pose) and low-level (e.g., skin color) information, the rest parts are concatenated with identity or age information and used as the condition information at each residual block. As shown in Fig. 2, identity with extraneous information is passed into the first few layers for high-level identity generation, while age with another extraneous information is passed into the last few layers for low-level texture generation. With the proposed hierarchical conditional generator, the proposed DAAE enables higher intuition and interpretability.

3.3 Disentangled Adversarial Learning

To further disentangle the inferred representations, i.e., z_A, z_I and z_E, and improve the quality of generation, a disentangled adversarial learning mechanism is proposed to optimize the inference network E and the generator network G jointly and adversarially. Inspired by IntroVAE [12], the model is expected to have the capability to self-estimate the age accuracy, identity preserving and image quality of the produced images.

As illustrated in Fig. 2, there exist two types of generated images, i.e., the reconstructed image $x_r = G(z_A, z_I, z_E)$ and the sampled image $x_s = G(\hat{z}_A, \hat{z}_I, \hat{z}_E)$. z_A, z_I and z_E are the inferred representations of the input x, while $\hat{z}_A$, $\hat{z}_I$ and $\hat{z}_E$ are sampled from three marginal product distribution, i.e., $p_A(\hat{z}_A) = p_A(z_A)q_\phi(z_A|x)$, $p_I(\hat{z}_I) = p_I(z_I)q_\phi(z_I|x)$, $p_E(\hat{z}_E) = p_E(z_E)q_\phi(z_E|x)$, respectively. This sampling strategy makes a random combination of age, identity and extraneous information from different sources. Along with the introduced constraints in the following, the learned representations z_A, z_I and z_E can be well disentangled.

To preserve the aging and identity characteristics accurately, two regularization terms are introduced for the generator G. They are computed as

$$L_{reg}^{(age)} = \frac{1}{C_A}\sum_{i=1}^{C_A} \| z_A^{'i} - z_A \| + \frac{1}{C_A}\sum_{i=1}^{C_A} \| z_A^{''i} - \hat{z}_A \| \tag{8}$$

$$L_{reg}^{(id)} = \frac{1}{C_I}\sum_{i=1}^{C_I} \| z_I^{'i} - z_I \| + \frac{1}{C_I}\sum_{i=1}^{C_I} \| z_I^{''i} - \hat{z}_I \| \tag{9}$$

where $z_A^{'}$ and $z_I^{'}$ are the inferred representations from the generated images x_r, while $z_A^{''}$ and $z_I^{''}$ are inferred from the generated images x_s.

To alleviate the problem of generating blurry samples in VAEs, the KL distance in Eq. (5) is employed as the adversarial signal to train the inference network E and the generator G adversarially [12]. When training E, the model minimizes the KL-distance of the posterior $q_\phi\left(z_E|x\right)$ from its prior $p_E\left(z_E\right)$ for the real data and maximize it for the generated samples. When training G, the model minimizes this KL-distance for the generated samples. The adversarial training objects for E and G are defined as below:

$$\begin{aligned} L_E^{(adv)} =& L_{kl}^{(ext)}\left(\mu_E, \sigma_E\right) + \alpha\{\left[m - L_{kl}^{(ext)}\left(\mu_E', \sigma_E'\right)\right]^+ \\ &+ \left[m - L_{kl}^{(ext)}\left(\mu_E'', \sigma_E''\right)\right]^+\}, \end{aligned} \tag{10}$$

$$L_G^{(adv)} = L_{kl}^{(ext)}\left(\mu_E', \sigma_E'\right) + L_{kl}^{(ext)}\left(\mu_E'', \sigma_E''\right), \tag{11}$$

where m is a positive margin, α is a weighting coefficient, (μ_E, σ_E), (μ_E', σ_E') and (μ_E'', σ_E'') are computed from the real data x, the reconstruction sample x_r and the new samples x_s, respectively. $[]^+ = max(0, .)$, which has the same meaning in hinge loss.

The total objective function is a weighted sum of the above losses, defined as

$$L_E = L_{rec} + \lambda_1 L_{kl}^{(age)} + \lambda_2 L_{kd}^{(id)} + \lambda_3 L_E^{(adv)}, \tag{12}$$

$$L_G = L_{rec} + \lambda_4 L_{reg}^{(age)} + \lambda_5 L_{reg}^{(id)} + \lambda_6 L_G^{(adv)}, \tag{13}$$

where $\lambda_{1\sim6}$ are the weighted parameters to balance the importance of each loss.

3.4 Inference and Sampling

By regularizing the disentangled representations with the age prior $p_A\left(z_A\right) = \mathbb{N}\left(\mathbf{y}, \mathbf{I}\right)$, identity knowledge prior $p_R\left(z_I\right)$ and extraneous prior $p_E\left(z_E\right) = \mathbb{N}\left(\mathbf{0}, \mathbf{I}\right)$, DAAE is thus a universal framework for face aging, exemplar-based face aging and age estimation.

Face Aging. We concatenate the identity variable z_I, the extraneous variable z_E and a target age variable $\hat{z}_A$ as the input of the generator G, where z_I and z_E are the inferred identity and extraneous information from the input x, while $\hat{z}_A$ is sampled from a distribution $p_A\left(z_A\right) = \mathbb{N}\left(\mathbf{y}, \mathbf{I}\right)$. The face aging result $\hat{x}$ is written as:

$$\hat{x} = G\left(\hat{z}_A, z_I, z_E\right) \tag{14}$$

Exemplar-based Face Aging. We remain the identity and extraneous information unchanged, and transfer the age information from the given exemplar x_e. Specifically, we concatenate the identity variable z_I, the extraneous variable z_E and the age variable z_{A_e} as the input of G, where z_A, z_I and z_{A_e} are from the posterior distribution $q_\phi(z_A|x)$, $q_\phi(z_I|x)$ and $q_\phi(z_{A\_e}|x_e)$, respectively. The exemplar-based face aging result is formulated as:

$$\hat{x} = G\left(z_{A\_e}, z_I, z_E\right) \tag{15}$$

Age Estimation. We calculate the mean value of C-dimension vector μ_A as the age estimation result, defined as:

$$\hat{y} = \tfrac{1}{C} \sum_{i=1}^{C} \mu_A^i \tag{16}$$

where μ_A is one of the output vectors of the inference network E.

4 Experiments

4.1 Datasets and Settings

Datasets We conduct experiments on five popular datasets. **CACD2000** [3] consists of 163,446 color facial images of 2,000 celebrities, where the ages range from 14 to 62 years old. However, there are many dirty data in it, which leads to a challenging model training. **Morph** [26] is the largest publicly available dataset collected in the constrained environment. It contains 55,349 color facial images of 13,672 subjects with ages ranging from 16 to 77 years old. **UTKFace** [38] is a large-scale facial age dataset with a long age span, which ranges from 0 to 116 years old. It contains over 20,000 facial images in the wild. We employ classical 80-20 split on CACD2000, Morph and UTKFace. **FG-NET** [16] contains 1,002 facial images of 82 subjects. We employ it as the testing set to evaluate the generalization of DAAE. **AgeDB** [21] is a manually collected database, which consists of 16,488 images of 568 subjects from 0 to 101 years old.

Experimental Settings Following [17], we employ the multi-task cascaded CNN [37] to detect the faces. All the facial images are cropped and aligned into 224 × 224. Our model is implemented with Pytorch. During training, we choose Adam optimizer [14] with β_1 of 0.9, β_2 of 0.99, a fixed learning rate of 2×10^{-4} and batch size of 16. The trade-off parameters $\lambda_{1\sim6}$ are all set to 1, 100, 1, 100, 100, 1, respectively. Besides, m is set to 200 and α is set to 0.5. More details of the network architectures and training processes are provided in the supplementary materials.

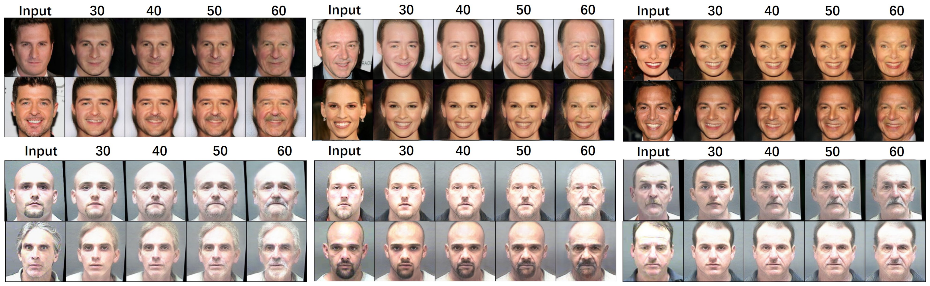

Fig. 3. Face aging results on CACD2000 (the first two rows) and Morph (the last two rows). For each subject, the first column is the input and the rest four columns are the synthesized results in 30, 40, 50 and 60 years old, respectively.

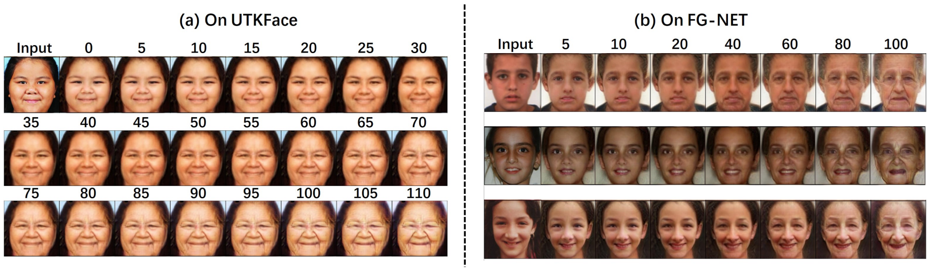

Fig. 4. Face aging results on UTKFace and FG-NET. (a) shows the aging results on UTKFace from 0 to 110 years old. The first image (top left) is the input, the rest are the synthesized results. (b) shows cross-dataset face aging results on FG-NET.

4.2 Qualitative Evaluation of DAAE

Face Aging By manipulating the mean value μ_A and sampling from age distribution, the proposed DAAE can generate facial images with arbitrary ages based on the input. Fig. 3 presents the face aging results on CACD2000 and Morph, respectively. We observe that the synthesized faces are getting older and older with ages growing. Specifically, the face contours become longer, the beards turn white and the nasolabial folds are deepened. Since both CACD2000 and Morph lack of images of children, we conduct face aging on UTKFace. Fig. 4 (a) describes the aging results on UTKFace from 0 to 110 years old. Obviously, from birth to adulthood, the aging effect is mainly shown on craniofacial growth, while the aging effect from adulthood to elder is reflected on the skin aging, which is consistent with human physiology. To evaluate the model generalization, we train our DAAE on UTKFace and test it on FG-NET. The aging results are shown in Fig. 4 (b). The left image of each subject is the input and the rest seven are the generated results from 5 to 100 years old.

The comparison results with previous works, including IAAP [13], RFA [31], RJIVE [28], MA-RCA [22], IPCGANs [32], CAAE [38], Yang et al. [35], GLCA-

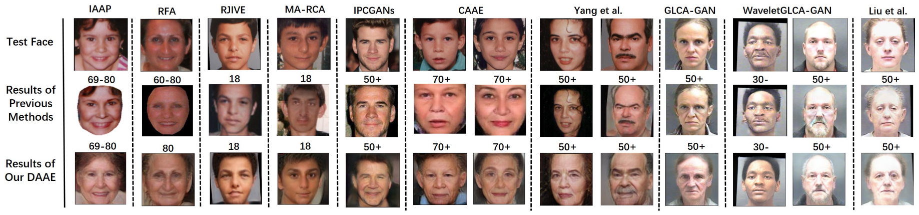

Fig. 5. Comparison with the previous works. The first row is the input face. The second row are the synthesized results of previous methods. The last row are the synthesized results by our DAAE.

GAN [18], waveletGLCA-GAN [17] and Liu et al. [20] are depicted in Fig. 5. We can see that our DAAE generates more obvious or comparable age effects on the input. Besides, previous face aging methods roughly divide the data into four or nine age groups to four or nine times increase the training data for specific age groups, while our DAAE is trained with original age labels.

4.3 Quantitative Evaluation of DAAE

	(a) on Morph				(b) on CACD2000			
Method	Input	AG1	AG2	AG3	Input	AG1	AG2	AG3
CAAE [40]	-	28.13	32.50	36.83	-	31.32	34.94	36.91
Yang et.al [35]	-	42.84	50.78	59.91	-	44.29	48.34	52.02
GLCA-GAN [18]	-	43.00	49.03	54.60	-	37.09	44.92	48.03
Liu et al. [20]	-	38.47	47.55	56.57	-	38.88	47.42	54.05
Ours	-	37.46	49.40	59.67	-	39.21	46.38	51.66
Real Data	28.19	38.89	48.10	58.22	30.73	39.08	47.06	53.68

Table 1. Comparisons of the aging accuracy on Morph and CACD2000.

Aging Accuracy Aging accuracy is an essential quantitative metric for face aging. Following [35, 18], we utilize the online face analysis tool of Face++ [1] to evaluate the ages of the synthesized results on Morph and CACD2000. We divide the testing data of the two datasets into four age groups: 30-(AG0), 31-40(AG1), 41-50(AG2), 51+(AG3). We choose AG0 as the input and synthesize images in AG1, AG2 and AG3. Then we estimate the ages of the synthesized images and calculate the average ages for each group. As shown in Table 1, we compare the DAAE with previous works on Morph and CACD2000. We observe that the generated ages by DAAE are closer to the real data than by CAAE [40] as well as GLCA-GAN [18], and comparable to [35, 20]. Note that [35, 20] need to train a specific model for each age group, while DAAE trains a unified model for arbitrary age synthesis, as well as other tasks.

	(a) on Morph				(b) on CACD2000			
Method	Input	AG1	AG2	AG3	Input	AG1	AG2	AG3
CAAE [40]	-	15.07	12.02	8.22	-	4.66	3.41	2.40
Yang et al. [35]	-	100.00	98.91	93.09	-	99.99	99.81	98.28
GLCA-GAN [18]	-	97.66	96.67	91.85	-	97.72	94.18	92.29
Liu et al. [20]	-	100.00	100.00	98.26	-	99.76	98.74	98.44
Ours	-	99.48	99.36	99.36	-	99.24	99.19	99.19

Table 2. Comparisons of the face verification results(%) on Morph and CACD2000.

Identity Preserving Identity preserving is another important quantitative metric for face aging. We evaluate this performance of DAAE by face verification. We also choose AG0 as the input and synthesize images in AG1, AG2 and AG3. For each testing face in AG0, we evaluate the verification rates between it and its corresponding aging results: [testing face → Age1], [testing face → Age2] and [testing face → Age3]. We adopt Light-CNN [33] as the identity extractor. Following [35, 20], we adopt thresholds=76.5 and FAR=1e-5 in our face verification experiments. The comparison results on Morph and CACD2000 are reported in Table 5. It is worth noting that it is unfair to directly compare DAAE with [20]. Because [20] utilizes extra attribute labels, including gender and race, to improve aging performance. Besides, [35, 20] need to train a specific model for each age group.

Age-Invariant Face Verification Following the testing protocol in [28], we evaluate our method on AgeDB. As shown in Table 3, DAAE achieves promising performance of face verification on AgeDB. The qualitative results are reported in the supplementary materials.

		5 years	10 years	20 years	30 years
RJIVE [28]	AUC	0.686	0.654	0.633	0.584
	Accuracy	0.637	0.621	0.598	0.552
Ours	AUC	0.989	0.988	0.986	0.981
	Accuracy	0.969	0.969	0.956	0.953

Table 3. Comparisons of Mean AUC and Accuracy on AgeDB.

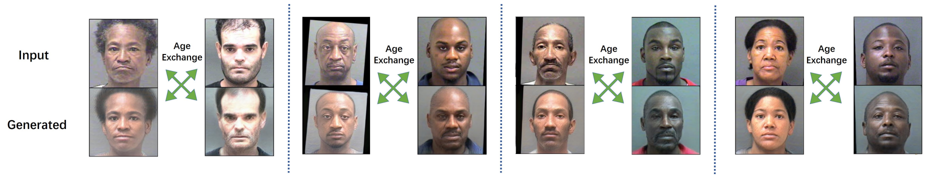

Fig. 6. Exemplar-based face aging results on Morph. For each image group, the first row are the input and the second row are the aging results with age information z_A exchanged in the group.

4.4 More Facial Age Analysis by DAAE

The previous face aging methods [40, 32, 18, 17] directly concatenate an age label to control the aging process, which are limited in handling various age analysis. Benefiting from the disentangling and modeling of age, identity and extraneous representations in the latent space, the proposed DAAE is able to realize more age-related tasks, such as exemplar-based face aging and age estimation.

Exemplar-based Face Aging Given an exemplar image x_e, the DAAE first extracts its age information z_{A_e} and then transfers it to the input x. Fig. 6 presents some results under this situation on Morph. We observe that the identity and extraneous information are preserved across rows, and the age information, such as wrinkles and beards, is changed according to the given exemplar. This demonstrates that our DAAE effectively disentangles age and age-irrelevant representations in the latent space.

Age Estimation To further demonstrate the disentangling ability of DAAE, we conduct age estimation on Morph. We detail the evaluation metrics in the supplementary materials. Following [24], we report the mean absolute error (MAE). As shown in Table 4, the age estimation result of DAAE on Morph is nearly as good as the state-of-the-arts, which demonstrates that the age representation is well learned from the given image.

4.5 Ablation Study

We report face verification results of DAAE and its five variants for a comprehensive comparison as the ablation study. Table 5 presents the comparison results. For the setting I, age with extraneous information is passed into the first few layers of the generator, while identity with another extraneous information is passed into the last few layers. For the setting II, we concatenate the age, identity and extraneous vectors and send it to the input layer of the generator. We observe that the face verification accuracy will decrease when one of the three losses is removed or the generator's architecture is changed. These phenomena indicate that each component in our method is essential for face aging.

Methods	Pre-trained	Morph
OR-CNN[23]	-	3.34
DEX[27]	IMDB-WIKI	2.68
Ranking [4]	Audience	2.96
Posterior[39]	-	2.87
SSR-Net[36]	IMDB-WIKI	2.52
M-V Loss[24]	-	2.51
ThinAgeNet [7]	MS-Celeb-1M	2.35
Ours	-	2.23

Table 4. Comparisons with state-of-the-art methods on Morph. Lower MAE is better.

Testing Face	AG1	AG2	AG3
w/oL_{adv}	95.74	95.61	95.60
$w/oL^{(age)}_{reg}$	97.84	97.89	97.89
$w/oL^{(id)}_{reg}$	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 5. Face verification results(%) of the ablation study on Morph.

5 Conclusion

This paper proposes a Disentangled Adversarial Autoencoder (DAAE) for facial age analysis. Specifically, we assign two different priors as well as identity distillation to assist disentangling the facial images into three independent factors: age, identity and extraneous information. A hierarchical conditional generator is introduced to produce photo-realistic images by transferring the identity and age information layer-by-layer. Finally, we propose a disentangled adversarial learning mechanism by training encoder and generator with age preserving regularization and identity knowledge distillation in an introspective adversarial manner. To the best of our knowledge, DAAE is the first attempt to achieve facial age analysis, including face aging, exemplar-based face aging and age estimation in a universal framework. This indicates that DAAE can efficiently formulate the facial age prior, which contributes to interpretable facial age manipulation. The qualitative and quantitative experiments demonstrate the superiority of the proposed DAAE on five popular datasets.

Acknowledgement

This work is partially funded by National Natural Science Foundation of China (Grant No. U1836217) and Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) (NO.2019JZZY010119).

References

1. Face++ research toolkit. megvii inc. http://www. faceplusplus.com ([Online])
2. Burgess, C.P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., Lerchner, A.: Understanding disentangling in beta-vae. arXiv preprint arXiv:1804.03599 (2018)
3. Chen, B.C., Chen, C.S., Hsu, W.H.: Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. IEEE Transactions on Multimedia **17**(6), 804–815 (2015)
4. Chen, S., Zhang, C., Dong, M., Le, J., Rao, M.: Using ranking-cnn for age estimation. In: CVPR (2017)
5. De Vries, H., Strub, F., Mary, J., Larochelle, H., Pietquin, O., Courville, A.C.: Modulating early visual processing by language. In: NeurIPS (2017)
6. Gao, B.B., Zhou, H.Y., Wu, J., Geng, X.: Age estimation using expectation of label distribution learning. In: IJCAI (2018)
7. Gao, B.B., Zhou, H.Y., Wu, J., Geng, X.: Age estimation using expectation of label distribution learning. In: IJCAI (2018)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
9. Gulrajani, I., Kumar, K., Ahmed, F., Taiga, A.A., Visin, F., Vazquez, D., Courville, A.: Pixelvae: A latent variable model for natural images. arXiv preprint arXiv:1611.05013 (2016)
10. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
11. Hou, X., Shen, L., Sun, K., Qiu, G.: Deep feature consistent variational autoencoder. In: WACV (2017)
12. Huang, H., He, R., Sun, Z., Tan, T., et al.: Introvae: Introspective variational autoencoders for photographic image synthesis. In: NeurIPS (2018)
13. Kemelmacher-Shlizerman, I., Suwajanakorn, S., Seitz, S.M.: Illumination-aware age progression. In: CVPR (2014)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: ICLR (2014)
16. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(4), 442–455 (2002)
17. Li, P., Hu, Y., He, R., Sun, Z.: Global and local consistent wavelet-domain age synthesis. IEEE Transactions on Information Forensics and Security (2018)
18. Li, P., Hu, Y., Li, Q., He, R., Sun, Z.: Global and local consistent age generative adversarial networks. ICPR (2018)
19. Li, P., Hu, Y., Wu, X., He, R., Sun, Z.: Deep label refinement for age estimation. Pattern Recognition **100**, 107–178 (2020)
20. Liu, Y., Li, Q., Sun, Z.: Attribute-aware face aging with wavelet-based generative adversarial networks. In: CVPR (2019)
21. Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S.: Agedb: the first manually collected, in-the-wild age database. In: CVPRW (2017)
22. Moschoglou, S., Ververas, E., Panagakis, Y., Nicolaou, M.A., Zafeiriou, S.: Multi-attribute robust component analysis for facial uv maps. IEEE Journal of Selected Topics in Signal Processing **12**(6), 1324–1337 (2018)

23. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output cnn for age estimation. In: CVPR (2016)
24. Pan, H., Han, H., Shan, S., Chen, X.: Mean-variance loss for deep age estimation from a face. In: CVPR (2018)
25. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: ICML (2014)
26. Ricanek, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: FGR (2006)
27. Rothe, R., Timofte, R., Van Gool, L.: Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision **126**(2-4), 144–157 (2018)
28. Sagonas, C., Ververas, E., Panagakis, Y., Zafeiriou, S.: Recovering joint and individual components in facial data. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(11), 2668–2681 (2017)
29. Semeniuta, S., Severyn, A., Barth, E.: A hybrid convolutional variational autoencoder for text generation. arXiv preprint arXiv:1702.02390 (2017)
30. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: NeurIPS (2015)
31. Wang, W., Cui, Z., Yan, Y., Feng, J., Yan, S., Shu, X., Sebe, N.: Recurrent face aging. In: CVPR (2016)
32. Wang, Z., Tang, X., Luo, W., Gao, S.: Face aging with identity-preserved conditional generative adversarial networks. In: CVPR (2018)
33. Wu, X., He, R., Sun, Z., Tan, T.: A light cnn for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security **13**(11), 2884–2896 (2018)
34. Wu, X., Huang, H., Patel, V.M., He, R., Sun, Z.: Disentangled variational representation for heterogeneous face recognition. In: AAAI (2019)
35. Yang, H., Huang, D., Wang, Y., Jain, A.K.: Learning face age progression: A pyramid architecture of gans. In: CVPR (2018)
36. Yang, T.Y., Huang, Y.H., Lin, Y.Y., Hsiu, P.C., Chuang, Y.Y.: Ssr-net: A compact soft stagewise regression network for age estimation. In: IJCAI (2018)
37. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016)
38. Zhang, Zhifei, S.Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: CVPR (2017)
39. Zhang, Y., Liu, L., Li, C., et al.: Quantifying facial age by posterior of age comparisons. arXiv preprint arXiv:1708.09687 (2017)
40. Zhang, Z., Song, Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: CVPR (2017)