IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 7, 2018, accepted October 16, 2018, date of publication November 9, 2018, date of current version November 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877706

CAAE++: Improved CAAE for Age Progression/Regression

JIANGFENG ZENG[1], XIAO MA[2], AND KE ZHOU[1], (Member, IEEE)
[1]Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China
[2]School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

Corresponding author: Ke Zhou (k.zhou@hust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61802440, Grant 61232004, and Grant 61502189, and in part by the National Key Research and Development Program of China under Grant 2016YFB0800402.

ABSTRACT Face age progression/regression has garnered substantial active research interest due to its tremendous impact on a wide-range of practical applications like searching for missing individuals with photos of childhood, entertainment, and so on. Most existing face aging models have proven to be successful and effective in learning the transformation between age groups with the aid of paired samples, i.e., face images of the same person at different ages. Considering the expensive cost of collecting paired datasets, Conditional Adversarial Autoencoder (CAAE) is designed for face aging task without paired samples and first achieves face age progression and regression in a holistic framework. However, only rough wrinkles are generated because of the insufficient discriminative and generative ability. To tackle this problem, in this paper, we develop a novel generative model based on CAAE, dubbed CAAE++, which defeats the previous CAAE mainly for two enhancements: 1) an auxiliary classifier is added on top of the discriminator, which allows a single discriminator not only distinguishes real images from synthetics but also classifies them into the target age group; and 2) a pre-trained deep face recognition model and a pre-trained age estimation model are exploited to preserve identity and age similarity, respectively. We train CAAE++ on UTKFace dataset and test on FGNET dataset. Experimental results demonstrate the efficacy of our proposed method in terms of fidelity.

INDEX TERMS Age progression/regression, generative adversarial networks, conditional adversarial autoencoder, image generation, generators, face recognition.

I. INTRODUCTION

In this work, we are interested in face age progression/regression, also known as face aging and rejuvenation, whose purpose is to generate face images at different ages and preserve personality in the meantime given one photo of a specific individual. Undoubtedly, many practical applications (e.g., searching for missing children) benefit a lot from the achievements of face age progression/regression. The problem of automatically generating face images with aging effects has drawn a surge of interests in the research community these years, but it is far from being essentially solved.

Over the past decades, most existing studies make contributions to face age progression but ignore face age regression. There are two main paradigms when addressing face aging synthesis: physical model-based [1], [2] and prototype-based [3], [4]. The former simulates the aging effects through complex modeling on facial muscles, wrinkles, etc, but suffers from two shortcomings: requiring substantial face images of the same individuals to cover a long time span and computation-intensive. The latter divides face images into different age groups, learns aging patterns between groups and requires paired samples over a short time span. As an outstanding representative of prototype-based methods, Wang *et al.* [4] devised the Recurrent Face Aging (RFA) model which captured the in-between evolving states between the adjacent age groups and employed a two-layer GRU to model complex dynamic appearance variations. Although some signs of aging are synthesized by these approaches, they heavily rely on the availability of paired samples which are difficult and costly to collect.

The last few years have witnessed remarkable generative capability of deep generative models in image generation [5]–[8], which brings new insights into face progression/regression. Duong *et al.* [9] proposed the Temporal

Deep Restricted Boltzmann Machines based age progression model and captured the non-linear aging changes well. One year later, Duong *et al.* [10] further presented the Temporal Non-Volume Preserving (TNVP) aging approach with a tractable density function for age progression. Although TNVP produced high-quality age-progressed faces, the synthesized face images in a complete aging sequence varied a lot in color, expression and even identity due to modeling without any personality information. Notedly, in spite of the effectiveness of all the aforementioned methods, they still require paired training data to ensure the high-quality of synthesized face images. The extreme challenge in collecting paired samples for training encourages Zhang *et al.* [11] to seek an algorithm that combines the benefits of both Adversarial Autoencoder (AAE) [12] and Generative Adversarial Network (GAN) [5], and first achieves face age progression and regression in a holistic framework without paired input-output examples. Under the framework of Conditional Adversarial Autoencoder (CAAE) [11], not only were the synthesized face images with rough aging effects generated but also the personality information got preserved. However, only rough wrinkles are generated because of the insufficient discriminative and generative ability.

In this paper, aiming to improve the discriminative and generative ability of CAAE, we propose CAAE++ to learn the face manifold for face age progression and regression. We strengthen the discriminator with an auxiliary classifier that allows a single discriminator to judge real/fake images and classify real/fake images into the target age group meanwhile. To further make aging effects of the synthesized face images more clear, we also introduce a pre-trained deep face recognition model and a pre-trained age estimation model to preserve identity and age similarity respectively.

The contributions of our work can be summarized as follows. **First**, we propose a new generative model based on CAAE which enhances the discriminative and generative ability of CAAE, and generates photo-realistic face images for both face aging and rejuvenation. **Second**, the effectiveness of CAAE++ is validated through thorough comparisons with pure CAAE.

The remainder of this paper is organized as follows. In Section II, we discuss the related works on face age progression/regression and overview the Generative Adversarial Networks (GANs) for image generation. Afterwards, we present the proposed CAAE++ detailedly in Section III. Section IV displays extensive experiments conducted to demonstrate the superiority of the proposed architecture over the original CAAE. Finally, we draw a conclusion in Section V.

II. RELATED WORK

A. FACE AGE PROGRESSION AND REGRESSION

The majority of existing approaches aimed to explore face aging problem but ignored face rejuvenation which is of the equal importance to face aging. This area has garnered a surge of active research interest in the computer vision community, but is far from being essentially solved. It is known to all that domain expert knowledge plays a key role in effective hand-crafted feature engineering. For example, Zhang *et al.* [13] propose a hybrid ECG feature extraction method that integrated fiducial- and non-fiducial-based features to extract more comprehensive ECG features and thereby improve the authentication stability. Traditional face aging methods can be roughly classified into two categories: physical model-based methods and prototype-based. The former models the biological facial variance of age, muscle, wrinkle, skin, etc. For example, [14] and [15] aimed to model facial structure as physical aging features, [16] and [17] explored textural variations in aging faces, and [1] captured the muscles. Although some rough aging effects are generated, physical model-based methods extremely rely on paired samples of the same person over a long time span and the biological prior knowledge. The latter aims to learn a transformation between different age groups, thus to some extent relaxing the requirement of paired samples of the same person over a long time span. Wang *et al.* [18] attempted to learn the mapping between corresponding down-sampled and high-resolution faces in a tensor space. Kemelmacher-Shlizerman *et al.* [19] argued that the illumination environment exerted a considerable influence on age progression. Recently, Shu *et al.* [3], [20] encoded the aging patterns by learning two neighboring dictionaries jointly. However, the prototype-based methods are prone to produce ghosting artifacts.

The past decade has witnessed a remarkable success achieved in Computer Vision (CV) [21]–[23], Natural Language Processing (NLP) [24]–[26] and deep hashing [27] by deep neural networks (DNNs) whose key advantage is to automatically learn powerful and effective representations for high dimensional raw data. AI techniques show a promising future for a myriad of issues [28]. For example, Lu *et al.* [29] attempt to address low illumination underwater light field images reconstruction task using deep convolutional neural networks. Some deep learning techniques have been explored for face aging. For instance, in the work of [4] the Recurrent Face Aging(RFA) approach is developed to model complex dynamic appearance variations via a two-layer GRU. The Temporal Deep Restricted Boltzmann Machines method [9] and Temporal Non-Volume Preserving (TNVP) aging method [10] had been proposed one after another for age progression. Although some age-progressed faces of high quality were obtained, the biggest drawback of TNVP was that the synthesized face images in a complete aging sequence varied a lot in color, expression and even identity due to modeling without any personality information. Above all, deep learning techniques relieve us from the burden of complex hand-crafted feature engineering under the guidance of expert knowledge.

More recently, as the most prevalent generative method, Generative Adversarial Networks (GANs) have exhibited an outstanding capability in image generation and have also been

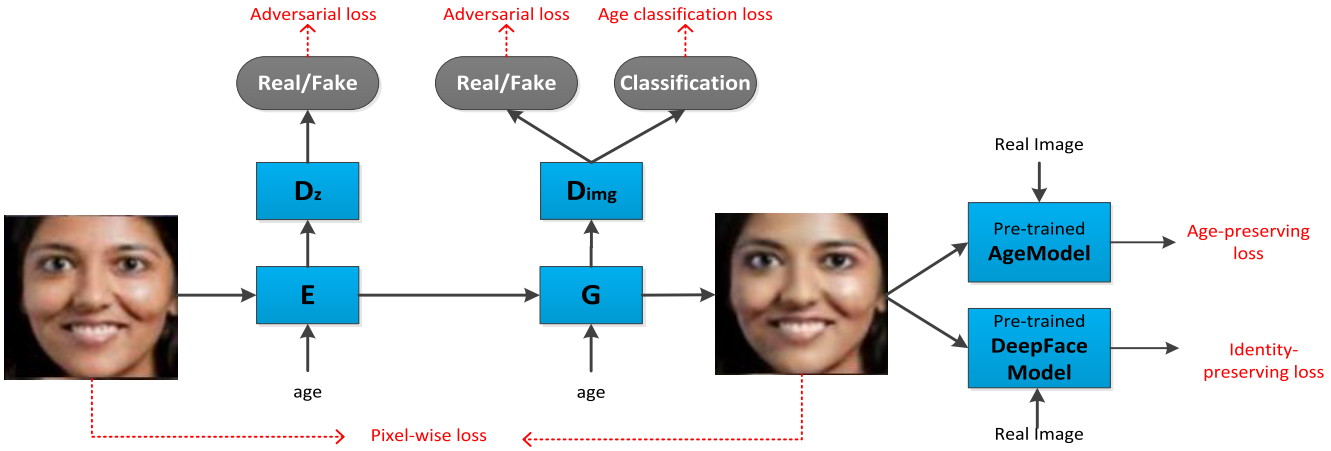


FIGURE 1. Overview of the proposed framework for age progression/regression.

investigated for facial age progression and regression. For example, Liu *et al.* [30] proposed a Contextual Generative Adversarial Net (C-GAN) which explicitly exploits the transition patterns between adjacent age groups. Zhang *et al.* [11] presented a general learning framework coined CAAE which combined an Adversarial Autoencoder (AAE) [12] with a Generative Adversarial Network (GAN) [5]. Both face age progression and regression are tackled in a holistic framework with unpaired samples. However, only rough wrinkles are generated due to the insufficient discriminative and generative ability.

B. GENERATIVE ADVERSARIAL NETWORKS

Recent advances in Generative Adversarial Networks (GANs) have shown impressive performance in a suite of image generation tasks, including image-to-image translation [31]–[34], image super-resolution [35], [36] and text-to-image synthesis [37], [38]. Because of the remarkable generative ability, GAN-based models often serves as data augmentation techniques [39]. A typical GAN [5] contains one generator G which aims to generate realistic-looking synthetics and one discriminator D which is dedicated to distinguishing real training samples from fake samples. The core idea behind GANs is the famous game theory. In other words, the adversarial loss is exloited to train the generator G and the discriminator D alternatively with the purpose of producing a distribution as far as possible to be similar to the true data distribution. Formally, the competition between the generator G and the discriminator D can be formulated using a two-player min-max game as follows:

$$\min_G \max_D \mathbf{E}_{x\sim p_d(x)}\left[\log D(x)\right] + \mathbf{E}_{z\sim p(z)}\left[\log(1-D(G(z)))\right] \quad (1)$$

where x is a training image sampled from the true data distribution $p_d(x)$, and z is a latent vector sampled from a prior distribution $p(z)$ (e.g., standard normal distribution).

By incorporating conditions and constraints into the generation process, conditional GANs (cGANs) [40] further enlarge the application scope, which makes GAN-based models to be the most outstanding generation methods. Odena *et al.* [41] argued that well-conditioned generators always perform better than unconditioned generators since cGANs control the synthesizing process by conditioning on additional valuable information. In order to enhance the discriminator, Augustus [42] added auxiliary classifiers on top of the discriminator and trained the discriminator in a multi-task learning setting. Bao *et al.* [43] proposed CVAE-GAN which combined a variational autoencoder (VAE) with a GAN to address fine-grained image generation. Moreover, because the traditional GANs suffer a lot from the unstable training, some recent work [44]–[46] have made contributions. Li *et al.* [47] improved the Boundary Equilibrium Generative Adversarial Networks (BEGANs) by adding a denoising loss to the discriminator so that the discriminator can learn more useful information about the true data distribution.

III. THE PROPOSED APPROACH

A high-level illustration of our proposed model is shown in Figure 1. We first formulate the problem of face age progressing and regression. Then, we briefly give an overview of Conditional Adversarial Autoencoder (CAAE), followed by the proposed improvements to enhance the discriminative and generative ability of CAAE. At last, the training objective of our model is presented.

A. TASK DEFINITION

A formal problem definition of face age progression and regression task is as follows:

We take for training a dataset X containing a large number of human face images labeled with actual ages, $X = \{(x_1, a_1), (x_2, a_2), \ldots, (x_N, a_N)\}$, where a_i is the age label. It is worth pointing out that we don't need to collect paired samples of the same person at different ages. Considering

the change of human face is a staged procedure, we divide the dataset into M categories according to the age, i.e., $C = \{c_1, c_2, \ldots, c_M\}$. Given the samples, our objective is to learn the manifold for facial image synthesis with age effects. Specifically, for a given test image x and its age category c, face age progression aims to generate the individual's face images after the age category c while face age regression aims to produce artifacts before the age category c.

B. CONDITIONAL ADVERSARIAL AUTOENCODER (CAAE)

CAAE consists of four networks: 1) The encoder network E, which maps the data sample x into a latent variable z, i.e., $E(x) = z$; 2) The discriminator network D_z, which imposes a prior distribution (e.g., uniform distribution) on the output of E; 3) The generator network G, which generates fake image x' given a latent vector z and the age label c, i.e., $G(z, c) = G(E(x), c) = x'$; 4) The discriminator network D_{img}, which aims to distinguish real training samples from synthetics. These four components are seamlessly cascaded together, and the whole pipeline is trained end-to-end.

Formally, in the training stage, CAAE tries to optimize the four networks with four losses. We denote $p_d(x)$ as the true data distribution, and assume the prior distribution to be $z^* \sim p(z)$. First, to ensure that the structure-preserving face images are emitted, the pixel-wise l_2 reconstruction loss is employed between x and x' as follows:

$$L_{pixel} = \left\| x - x' \right\|_2^2 \tag{2}$$

Second, an adversarial loss is added since the encoder E intends to fool the discriminator D_z to believe the generated z is from the prior distribution and can be defined as:

$$\begin{aligned} L_{GD_z} = {} & E_{z^* \sim p(z)} \left[\log D_z \left(z^* \right) \right] \\ & + E_{x \sim p_d(x)} \left[\log \left(1 - D_z \left(E \left(x \right) \right) \right) \right] \end{aligned} \tag{3}$$

Third, another adversarial loss is introduced with the goal of cheating the discriminator D_{img} and can be computed as:

$$\begin{aligned} L_{GD_{img}} = {} & E_{x \sim p_d(x)} \left[\log D_{img} \left(x, c \right) \right] \\ & + E_{x \sim p_d(x)} \left[\log \left(1 - D_{img} \left(G(E \left(x \right), c), c \right) \right) \right] \end{aligned} \tag{4}$$

The last loss brought in is the total variation error, which is effective in removing the ghosting artifacts and denoted as L_{tv}. Here we fail to present the formulation for L_{tv} because the authors didn't formulate it.

To sum up, the final objective function of CAAE becomes

$$L(CAAE) = \lambda_1 L_{pixel} + \lambda_2 L_{tv} + L_{GD_z} + L_{GD_{img}} \tag{5}$$

where λ_1 and λ_2 are the coefficients balancing the smoothness and high resolution, and are assigned to be 100 and 10 respectively in the original implementation.

As shown in Figure 1, besides the four networks depicted above, we further equip CAAE++ with an auxiliary classifier on top of the discriminator D_{img} and two pre-trained deep models, aiming at improving the discriminative and generative ability. In our case we incorporate the age information into the encoder, i.e., $E(x, c) = z$. Correspondingly, three new losses are devised to enhance the visual quality and aging effects of generated face images: age classification loss, age-preserving loss and identity-preserving loss. It's worth pointing out that Wasserstein GAN objective with gradient penalty (WGAN-gp) [46] is applied to stabilizing the training process of the discriminator network D_{img} and the generator G. We elaborate them in the following.

C. IMAGE ADVERSARIAL LOSS

As theoretically analyzed in the work of [48], the traditional GANs suffer from the unstable training process. Arjovsky *et al.* [45] put forward the continuous Earth Mover Distance metric to replace the original Jensen-Shannon (JS) divergence loss function. In order to maintain a Lipschitz constraint, Gulrajani *et al.* [46] devised a gradient penalty to further stabilize the training of GANs. Thus, Equation 4 is replaced as

$$\begin{aligned} L_{GD_{img}} = {} & E_{x \sim p_d(x)} \left[D_{img} \left(x \right) \right] \\ & - E_{x \sim p_d(x)} \left[D_{img} \left(G(E(x), c), c \right) \right] \\ & - \lambda_{gp} E_{\hat{x} \sim p_{\hat{x}}} \left[\left(\left\| \nabla_{\hat{x}} D_{img}(\hat{x}) \right\|_2 - 1 \right)^2 \right] \end{aligned} \tag{6}$$

where $\hat{x} \sim p_{\hat{x}}$ is uniformly sampled along straight lines between pairs of real and generated images, and λ_{gp} is the coefficient to penalize the gradients.

D. AGE CLASSIFICATION LOSS

We add a classifier on top of the discriminator D_{img} and train them in a multi-task learning setting. In detail, the age classification loss from categorizing real images is used to optimize D_{img} and that from categorizing fake images is used to optimize G. We denote them as:

$$L_{cls}^r = E_{x \sim p_d(x)} \left[- \log D_{cls} \left(c | x \right) \right] \tag{7}$$

$$L_{cls}^f = E_{x \sim p_d(x)} \left[- \log D_{cls} \left(c | G(E(x, c), c) \right) \right] \tag{8}$$

By minimizing both of the adversarial and classification errors, the discriminator D_{img} guides the generator G to generate more realistic face images with aging effects than before.

E. AGE-PRESERVING LOSS

When generating the fake image x' given the data sample x and its age label c, one core challenge is keeping age similarity between x' and x. Therefore, we incorporate the age-preserving error to measure the input-output distance in the age feature space. Specifically, the network of deep age estimation [49] is utilized, denoted as Φ_{age}, to obtain age features. We pre-train Φ_{age} on the IMDB-WIKI dataset[1] which is the largest publicly available dataset of face images with gender and age labels. Drawing inspirations from [43], feature matching is utilized to help the generator emit diverse samples and generate structure-preserving samples since it's capable of measuring the feature-level similarity between the true image and its fake. In our case, we extract the output

[1] https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

Algorithm 1 The Training Pipeline of the Proposed CAAE++ Algorithm

Require: b, the batch size. M, class number. Θ_E, initial E network parameters.Â Θ_G, initial G network parameters.Â Θ_D^{img}, initial D_{img} network parameters.Â Θ_D^z, initial D_z network parameters. λ_{adv}=0.001, λ_{pixel}=10, λ_{cls}=1.0, λ_{age}=0.1 and λ_{id}=0.1.

while Θ_G has not converged **do**
  Sample $\{x, c\} \sim p_d$ a batch from the real data;
  $L_{cls}^r \leftarrow -\log D_{cls}(c|x)$
  $z \leftarrow E(x, c)$
  $x' \leftarrow G(z, c)$
  $L_{cls}^f \leftarrow -\log D_{cls}(c|x')$
  Sample $z^* \sim p(z)$ a batch of random noise
  $L_{GD_z} \leftarrow \log D_z(z^*) + \log(1 - D_z(z))$
  $L_{GD_{img}} \leftarrow D_{img}(x) - D_{img}(x', c) - \lambda_{gp}(\|\nabla_{\hat{x}} D_{img}(\hat{x})\|_2 - 1)^2$
  $L_{pixel} \leftarrow \|x - x'\|_2^2$
  $L_{age} \leftarrow \|\Phi_{age}(x) - \Phi_{age}(x')\|_2^2$
  $L_{id} \leftarrow \|\Phi_{face}(x) - \Phi_{face}(x')\|_2^2$
  $L(E\&G) = \lambda_{adv} L_{GD_{img}} + \lambda_{pixel} L_{pixel} + \lambda_{cls} L_{cls}^f + \lambda_{age} L_{age} + \lambda_{id} L_{id}$
  $L(D_{img}) = L_{GD_{img}} + \lambda_{cls} L_{cls}^r$
  $L(D_z) = L_{GD_z}$
  $\Theta_D^z \overset{+}{\leftarrow} -\nabla_{\Theta_D^z}(L(D_z))$
  $\Theta_D^{img} \overset{+}{\leftarrow} -\nabla_{\Theta_D^{img}}(L_{GD_img})$
  $\Theta_E \& \Theta_G \overset{+}{\leftarrow} -\nabla_{\Theta_E \& \Theta_G}(L(E\&G))$
end while

of Φ_{age} before the full connection layer as age features. The squared Euclidean distance is used and the age-preserving loss is then formulated as:

$$L_{age} = \left\|\Phi_{age}(x) - \Phi_{age}(x')\right\|_2^2 \tag{9}$$

As a result, the generator G optimized using L_{age} is forced to generate face images corresponding to the right age labels.

F. IDENTITY-PRESERVING LOSS

In addition to the age-preserving loss, we employ a deep face recognition model [50], denoted as Φ_{face}, to keep the personality well when generating face images in a complete aging sequence. We pre-train Φ_{face} on the CASIA-WebFace dataset.[2] Feature matching technique is exploited as also. In this case, the output of Φ_{face} before the full connection layer is regarded as face features. Similarly, the identity-preserving loss is computed as:

$$L_{id} = \left\|\Phi_{face}(x) - \Phi_{face}(x')\right\|_2^2 \tag{10}$$

Consequently, the synthetics with a long time span preserve personality well via optimizing G with L_{id}.

[2]http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html

G. OBJECTIVE OF CAAE++

Finally, the goal of CAAE++ is to minimize the following objective function:

$$L(CAAE++) = L(E\&G) + L(D_{img}) + L(D_z) \tag{11}$$

where $L(E\&G)$ is used to optimize the encoder E and the generator G together, $L(D_{img})$ is for the discriminator D_{img}, and $L(D_z)$ is for the discriminator D_z. In detail, they are formulated as:

$$L(E\&G) = \lambda_{adv} L_{GD_{img}} + \lambda_{pixel} L_{pixel} + \lambda_{cls} L_{cls}^f + \lambda_{age} L_{age} + \lambda_{id} L_{id} \tag{12}$$

$$L(D_{img}) = L_{GD_{img}} + \lambda_{cls} L_{cls}^r \tag{13}$$

$$L(D_z) = L_{GD_z} \tag{14}$$

where λ_{adv}, λ_{pixel}, λ_{cls}, λ_{age} and λ_{id} are hyper-parameters and are set to be 0.001, 10, 1.0, 0.1, and 0.1, respectively. All these objectives are complementary to each other, and ultimately enable our algorithm to obtain superior results. The whole training pipeline is described in Algorithm 1.

IV. EXPERIMENTS

In this section, we present our experiment settings and conduct experiments show the effectiveness of our proposed approach.

A. DATASETS

We train the proposed method and CAAE on the UTKFace dataset [11] which consists of over 20,000 face images with annotations of age and gender. The images are divided into ten categories according to the age, i.e., $0-5$, $6-10$, $11-15$, $16-20$, $21-30$, $31-40$, $41-50$, $51-60$, $61-70$, and the rest. We crop a 128×128 face region as input images and represent age labels using one-hot vectors. Then quantitative experimental evaluations are made on FGNET [14], the testing dataset, which has 1002 images of 82 individuals aging from 0 to 69.

B. IMPLEMENTATION OF CAAE++

In our experiments, the encoder network E takes the concatenation of the image and the age label as input and is built using 5 stride-2 convolution layers and 1 fully-connected(FC) layer. The convolution layers have 64, 128, 256, 512 and 1024 channels with kernel size of 5×5. The age category information is merged again at the last FC layer. Note that we use the uniform distribution as the prior distribution and its dimension is 50. The G network contains 1 FC layer, followed by 5 deconv layers with 2-by-2 upsampling and 2 stride-1 deconv layer. We concatenate z and the age label at the first FC layer. The 7 deconv layers have 1024, 512, 256, 128, 64, 16 and 3 channels with kernel size of 5×5. Note that the outputs of both E and G are activated by the hyperbolic tangent function. The D_z network consists of 4 FC layers with 64, 32, 16 and 1 channels respectively. We apply the batch normalization to the outputs of the first three FC layers except the output of the last FC layer. The D_{img} network

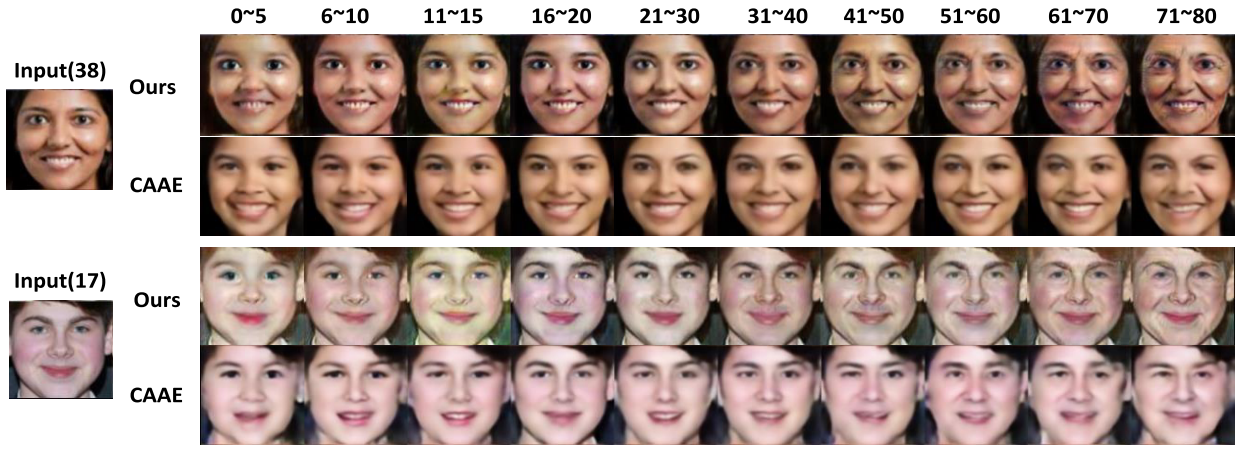


FIGURE 2. Comparison to CAAE for age progression and regression in a complete aging sequence.

FIGURE 3. More results of CAAE++ for age progression/regression. The first column shows the original faces marked with their true ages on the top. The right 10 columns are synthetics of our proposed framework. The generated faces fall into 10 age groups as indicated at the top of each column. The red boxes show the reconstruction results.

is composed of two branches: the adversarial discriminator and the classifier for age classification. The two branches share in common 4 convolution layers with 32, 64, 128 and 256 channels. The batch normalization is also applied after each convolution layer. Then 2 FC layers are used to map features into 1 and 10 units respectively.

In training, we update the four networks with a mini-batch size of 100 through ADAM($\beta_1 = 0.5$). The learning rate begins with 0.0002 and degrades every two epochs.

C. MODEL COMPARISON

Note that we use CAAE[3] implemented by [11]. In order to demonstrate the superiority of our proposed approach over CAAE, we use the same input images to synthesize faces in a complete aging sequence for both age progression and regression as shown in Figure 2. Evidently, compared with CAAE, CAAE++ generates more visually plausible and convincing

[3]https://github.com/ZZUTK/Face-Aging-CAAE

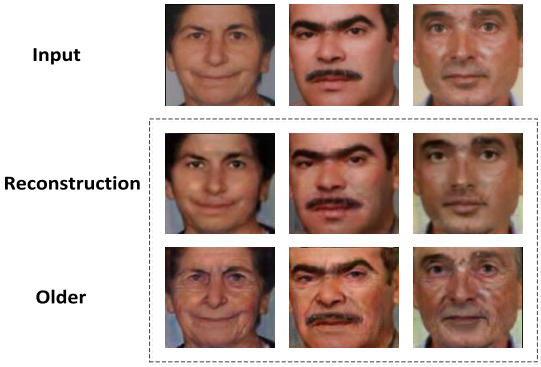


FIGURE 4. Reconstructions and aged synthetics from our model.

synthetics with aging effects and also better preserves the personality along a long age span because the pre-trained deep face recognition model ensures the individual features. Richer textures are captured in older faces, which makes old faces look more photo-realistic and demonstrates our pre-trained deep age estimation model takes effects. More synthesized images of our CAAE++ can be see in Fig 3.

Additionally, our method performs particularly well for image reconstruction and face aging. It can be seen from Figure 4 that CAAE++ perfectly reconstructs the faces and produces wrinkled faces consistent with the aged.

V. CONCLUSION

In this paper, aiming to improve the discriminative and generative ability of CAAE, we first strengthen the discriminator with an auxiliary classifier that allows a single discriminator to differentiate between real and synthesized samples and classify real/fake images into corresponding age groups meanwhile. Second, in order to enhance the generator, a pre-trained deep face recognition model and a pre-trained age estimation model are exploited to ensure the identity and age similarity respectively. Finally, extensive experiments have demonstrated that our proposed CAAE++ can generate photo-realistic face images at different ages for both face aging and rejuvenation.

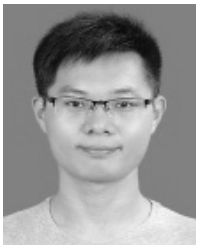
ACKNOWLEDGEMENTS

The authors would like to thank anonymous reviewers for their helpful and thought-provoking suggestions and feedbacks that have helped improve this paper substantially.

REFERENCES

[1] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.
[2] Y. Tazoe, H. Gohara, A. Maejima, and S. Morishima, "Facial aging simulator considering geometry and patch-tiled texture," in *Proc. SIGGRAPH Posters*, 2012, p. 90.
[3] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized age progression with aging dictionary," in *Proc. ICCV*, 2015, pp. 3970–3978.
[4] W. Wang *et al.*, "Recurrent face aging," in *Proc. CVPR*, 2016, pp. 2378–2386.
[5] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.
[6] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 5967–5976.
[7] Z. Zheng, C. Wang, Z. Yu, H. Zheng, and B. Zheng, "Instance map based image synthesis with a denoising generative adversarial network," *IEEE Access*, vol. 6, pp. 33654–33665, 2018.
[8] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. (2017). "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation." [Online]. Available: https://arxiv.org/abs/1711.09020
[9] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, "Longitudinal face modeling via temporal deep restricted Boltzmann machines," in *Proc. CVPR*, 2016, pp. 5772–5780.
[10] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *Proc. ICCV*, 2017, pp. 3755–3763.
[11] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. CVPR*, 2017, pp. 4352–4360.
[12] A. Makhzani, J. Shlens, N. Jaitly, and I. J. Goodfellow. (Nov. 2015). "Adversarial autoencoders." [Online]. Available: https://arxiv.org/abs/1511.05644
[13] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, "PEA: Parallel electrocardiogram-based authentication for smart healthcare systems," *J. Netw. Comput. Appl.*, vol. 117, pp. 10–16, Sep. 2018.
[14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.
[15] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. CVPR*, 2006, pp. 387–394.
[16] N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Amsterdam, The Netherlands, Sep. 2008, pp. 1–8.
[17] A. C. Berg, F. J. P. López, and M. Gonzalez, "A facial aging simulation method using flaccidity deformation criteria," in *Proc. 10th Int. Conf. Inf. Visualisation (IV)*, London, U.K., Jul. 2006, pp. 791–796.
[18] Y. Wang, Z. Zhang, W. Li, and F. Jiang, "Combining tensor space analysis and active appearance models for aging effect simulation on face images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1107–1118, Aug. 2012.
[19] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proc. CVPR*, 2014, pp. 3334–3341.
[20] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, "Personalized age progression with bi-level aging dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 905–917, Apr. 2018.
[21] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2204–2212.
[22] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "DRAW: A recurrent neural network for image generation," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 1462–1471.
[23] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.
[24] D. Bahdanau, K. Cho, and Y. Bengio. (Sep. 2014). "Neural machine translation by jointly learning to align and translate." [Online]. Available: https://arxiv.org/abs/1409.0473
[25] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1412–1421.
[26] T. Rocktäschel, E. Grefenstette, K. M. Hermann, T. Kociský, and P. Blunsom. (Sep. 2015). "Reasoning about entailment with neural attention." [Online]. Available: https://arxiv.org/abs/1509.06664
[27] K. Zhou, J. Zeng, Y. Liu, and F. Zou, "Deep sentiment hashing for text retrieval in social CIoT," *Future Gener. Comp. Syst.*, vol. 86, pp. 362–371, Sep. 2018.
[28] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, 2018.
[29] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comp. Syst.*, vol. 82, pp. 142–148, May 2018.
[30] S. Liu *et al.*, "Face aging with contextual generative adversarial nets," in *Proc. ACM Multimedia Conf. (MM)*, Mountain View, CA, USA, Oct. 2017, pp. 82–90.

[31] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 700–708.
[32] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, Aug. 2017, pp. 1857–1865.
[33] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2242–2251.
[34] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2868–2876.
[35] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 105–114.
[36] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár. (Oct. 2016). "Amortised MAP inference for image super-resolution." [Online]. Available: https://arxiv.org/abs/1610.04490
[37] H. Zhang, T. Xu, and H. Li, "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5908–5916.
[38] T. Xu *et al.* (Nov. 2017). "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks." [Online]. Available: https://arxiv.org/abs/1711.10485
[39] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
[40] M. Mirza and S. Osindero. (Nov. 2014). "Conditional generative adversarial nets." [Online]. Available: https://arxiv.org/abs/1411.1784
[41] A. Odena *et al.* (Jun. 2018). "Is generator conditioning causally related to GAN performance?" [Online]. Available: https://arxiv.org/abs/1802.08768
[42] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. ICML*, 2017, pp. 2642–2651.
[43] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: Fine-grained image generation through asymmetric training," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2764–2773.
[44] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2813–2821.
[45] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, Aug. 2017, pp. 214–223.
[46] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 5769–5779.
[47] Y. Li, N. Xiao, and W. Ouyang, "Improved boundary equilibrium generative adversarial networks," *IEEE Access*, vol. 6, pp. 11342–11348, 2018.
[48] M. Arjovsky and L. Bottou. (Jan. 2017). "Towards principled methods for training generative adversarial networks." [Online]. Available: https://arxiv.org/abs/1701.04862
[49] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, nos. 2–4, pp. 144–157, 2018.
[50] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. CVPR*, 2015, pp. 815–823.



JIANGFENG ZENG is currently pursuing the Ph.D. degree with the Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, China. His current research interests include machine learning, natural language processing, sentiment analysis, deep hashing, and image generation.

XIAO MA received the Ph.D. degree in computer science from the Huazhong University of Science and Technology in 2017. She is currently an Assistant Professor with the School of Information and Safety Engineering, Zhongnan University of Economics and Law, China. Her research interests include recommendation systems, social network analysis, and data mining.

KE ZHOU (M'14) received the B.E., M.E., and Ph.D. degrees in computer science and technology from the Huazhong University of Science and Technology (HUST), China, in 1996, 1999, and 2003, respectively. He is currently a Full Professor with the School of Computer Science and Technology, HUST. He has over 50 publications in journals and international conferences, including performance evaluation, FAST, MSST, ACM MM, SYSTOR, MASCOTS, and ICC. His main research interests include computer architecture, cloud storage, parallel I/O, and storage security.