Dual Conditional GANs for Face Aging and Rejuvenation

Jingkuan Song[1], Jingqiu Zhang[1], Lianli Gao[1], Xianglong Liu[2], Heng Tao Shen[1*]
[1]Center for Future Media and School of Computer Science and Engineering,
University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Beihang University, Beijing, 100083, China,
{jingkuan.song,jingqiu.zhang}@gmail.com,
lianli.gao@uestc.edu.cn, xlliu@buaa.edu.cn, shenhengtao@hotmail.com

Abstract

Face aging and rejuvenation is to predict the face of a person at different ages. While tremendous progress have been made in this topic, there are two central problems remaining largely unsolved: 1) the majority of prior works requires sequential training data, which is very rare in real scenarios, and 2) how to simultaneously render aging face and preserve personality. To tackle these issues, in this paper, we develop a novel dual conditional GANs (Dual cGANs) mechanism, which enables face aging and rejuvenation to be trained from multiple sets of unlabeled face images with different ages. In our architecture, the primal conditional GAN transforms a face image to other ages based on the age condition, while the dual conditional GAN learns to invert the task. Hence a loss function that accounts for the reconstruction error of images can preserve the personal identity, while the discriminators on the generated images learn the transition patterns (e.g., the shape and texture changes between age groups) and guide the generation of age-specific photo-realistic faces. Experimental results on two publicly dataset demonstrate the appealing performance of the proposed framework by comparing with the state-of-the-art methods.

1 Introduction

Face aging and rejuvenation, also called face age progression and regression, is a meaningful application to predicate what does a person look like at different ages. It has lots of applications, such as finding lost children, cross-age face verification [Park *et al.*, 2010], and entertainment.

In the last decade, a number of prior works have made great progress in face aging, which are roughly divided into two classes, the physical model-based methods [Suo *et al.*, 2010; Tazoe *et al.*, 2012] and the prototype-based methods [Kemelmacher Shlizerman *et al.*, 2014; Tiddeman *et al.*, 2001]. The physical model-based methods propose a parameter model to

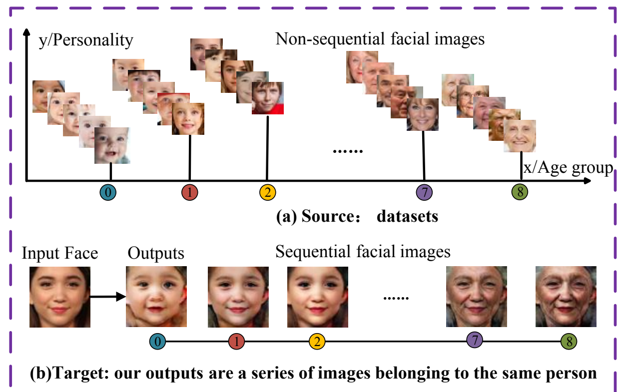

Figure 1: An illustration of our face aging and rejuvenation process. As (a) shows, our training examples are non-sequential and unpaired, and we aim to simultaneously render a series of age-changed facial images of a person and preserve personality, as shown in (b).

specifically describe the changes of faces in different ages. These methods parametrically model shape and texture parameters for different features of each age group, e.g., muscles [Suo *et al.*, 2012], wrinkles [Ramanathan and Chellappa, 2008; Suo *et al.*, 2010] and facial structure [Ramanathan and Chellappa, 2006; Lanitis *et al.*, 2002]. On the other hand, prototype-based methods [Kemelmacher Shlizerman *et al.*, 2014; Tiddeman *et al.*, 2001] divide faces into groups by age, and then construct an average face as its prototype for each age group. After that, these methods can transfer the texture difference between the prototypes to the input facial image.

More recently, the deep learning-based method [Wang *et al.*, 2016; Liu *et al.*, 2017] achieved the state-of-the-art performance. In [Wang *et al.*, 2016], RNN is applied on the coefficients of eigenfaces for age pattern transition. It performs the group-based learning which requires the true age of testing faces to localize the transition state which might not be convenient. In addition, these approaches only provide age progression from younger face to older ones. To achieve flexible bidirectional age changes, it may need to retrain the model inversely. Generative Adversarial Net-

*Corresponding author: Lianli Gao, Heng Tao Shen

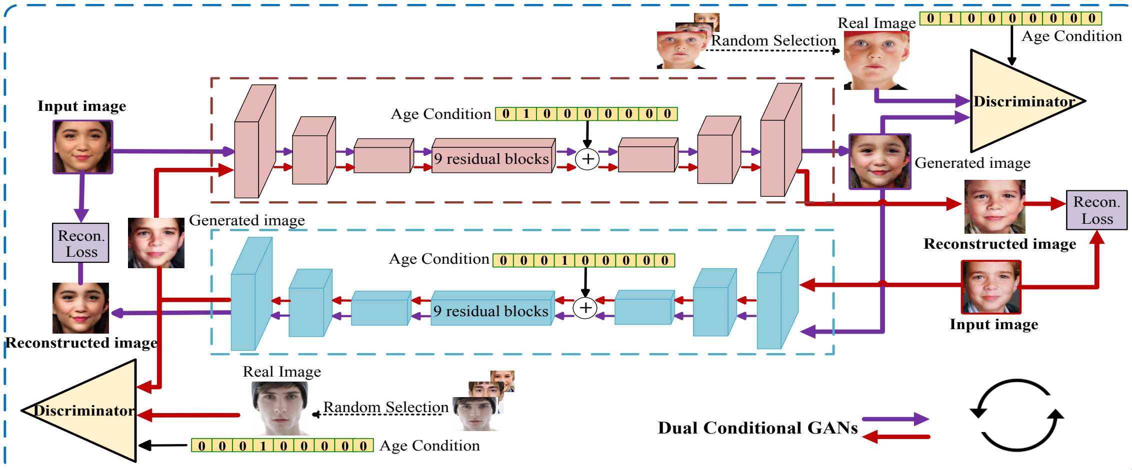

Figure 2: The main structure of our Dual cGANs. The purple flowchart shows the primal cGAN, and the red flowchart shows the dual cGAN. In each of them, an input image is first transformed to other ages based on the age condition, and then it is reconstructed. .

works (GANs) [Goodfellow *et al.*, 2014] have recently made great progress in image-to-image translation [Liu *et al.*, 2017; Radford *et al.*, 2015; Zhu *et al.*, 2017; Antipov *et al.*, 2017; Zhang *et al.*, 2017; Song *et al.*, 2018]. As a variant of the vanilla GANs, conditional GANs (CGANs) [Mirza and Osindero, 2014] are able to control features of generated images with a condition. By taking the age as a condition, [Liu *et al.*, 2017] proposed Contextual Generative Adversarial Nets (C-GANs) to learn the aging procedure. Nevertheless, existing works require the availability of paired samples, i.e., face images of the same person at different ages, and some even require paired samples over a long range of age span, which is difficult to collect.

To avoid the requirement of paired training samples, there are some pioneering works [Zhang *et al.*, 2017; Antipov *et al.*, 2017]. In both of them, the face is first mapped to a latent vector through a convolutional encoder, and then projected to the face manifold conditional on age through a deconvolutional generator. A critical problem is that the effect of condition on the generated face images is not guaranteed. During training, they try to reconstruct the facial image with the same age condition as the input. But during the testing, they obtain the face aging results by combining an input face with different age conditions. In the worst case, if the age has no effect on the reconstructed face images, the training will also converge. Then, it is impossible to achieve face aging/rejuvenation by changing the age condition of the trained network.

To tackle these issues, in this paper, we develop a novel dual conditional GANs mechanism, which enables face aging and rejuvenation to be trained from multiple sets of unlabeled face images with different ages (Figure 1). Specifically, our dual conditional GANs first transform a face image to other ages based on the age condition, and then learns to invert the task. Therefore, a loss function that accounts for the reconstruction error of images can preserve the personal identity in a self-supervised way. Then, the discriminators on the generated images learn the transition patterns (e.g., the shape and texture changes between age groups) and guide the generation of age-specific photo-realistic faces.

2 Dual Conditional GANs

Given multiple sets of training face images $\mathbb{F}_1, \mathbb{F}_2, ..., \mathbb{F}_N$ with N different ages (not necessary to be the same person), our goal is to learn a face aging/rejuvenation model $\mathbf{G}_t$ and a face reconstruction model $\mathbf{G}_s$. Then, given any facial image $\mathbf{x}_i$ with the age of $\mathbf{c}_s$ and the target age condition $\mathbf{c}_t$, this model $\mathbf{G}_t$ can predict the face $\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t)$ of a person at different ages, but still preserve personalized features of the face by $\mathbf{x}'_s = \mathbf{G}_s(\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t), \mathbf{c}_s)$.

We illustrate our proposed Dual cGANs architecture by the scheme in Fig. 2. The scheme shows how a face aging/rejuvenation model is learned in an unsupervised fashion through the interplay between, on the one side, the face aging/rejuvenation process taking the source face image and a target age as input, and, on the other side, the reconstruction process where the images from the generator are compared to the original training images. In the remainder of this section, we describe the network architecture of Dual cGANs, our loss function and learning of parameters.

2.1 System Architecture

As shown in Figure 2, our Dual cGANs consist of the primal cGAN and the dual cGAN. In each of them, there are three components: target generator, source generator and their discriminators. We first introduce the components in the primal cGAN.

Target Generator

The task of our Primal cGAN Target Generator network $\mathbf{G}_t$ is to generate the face $\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t)$ of a person at different ages based on the input image $\mathbf{x}_i$ and target age $\mathbf{c}_t$. Our input and output are both colorful facial images with shape $256 \times 256 \times 3$. Its structure is shown in Fig 3. From this figure, we can see that it is composed of three parts: encoder, age condition and decoder.

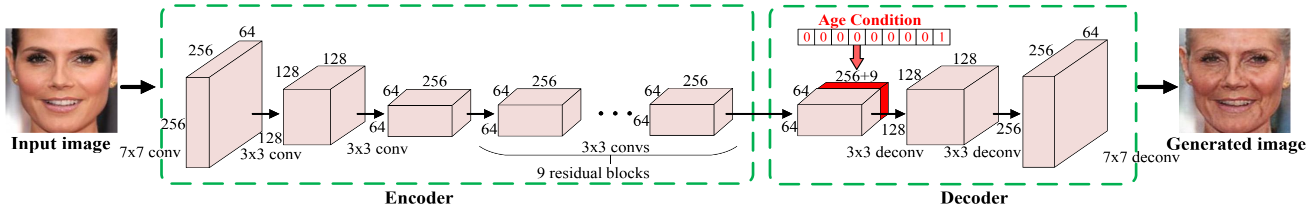

Figure 3: The framework of our generators. It has three major components, i.e., encoder, age condition and decoder.

The "encoder" extracts the features of the input face image. Our structure is similar to [Zhu *et al.*, 2017]. We use 3 groups of convolution layers and 9 residual blocks [He *et al.*, 2016b]. Every convolution group includes a stride-2 convolution layer, a spatial batch-normalization layer and a ReLU nonlinearity layer. The sizes of kernels are $7 \times 7, 3 \times 3, 3 \times 3$ and the numbers of kernels are $64, 128, 256$, respectively. After that we get 256 feature maps. The residual block is composed of 2 convolution layers, each has a zero-padding layer, a stride-1 convolution layer and a batch-normalization layer, and whose convolution has 3×3 kernel size and 256 kernels.

The age condition aims to control the transformation process. Considering the huge range of ages in our dataset, we divide different ages into y groups, and thus our age condition is represented as a y-dim one-hot age vector. Each dim represents a specific age group. To concatenate it with our "decoder", we reshape it as a y-channel tensor, whose dimension is the same as the tensor where we intend to concatenate.

The task of our "decoder" is to generate a face based on the features and target age condition. We also have 3 deconvolution layer groups. All of them share the same basic structure: a deconvolution layer with the size of kernels 3×3, which is followed by a batch-normalization layer and a ReLU as the activation function. The last layer uses a scaled tanh to ensure that the pixels of the output image are in the range [0, 255]. What's more, the first deconvolution layer is concatenated to the reshaped age condition as described in conditional GANs [Mirza and Osindero, 2014]. Thus we can generate different faces of a person at different ages.

Source Generator

The task of our Source Generator network $\mathbf{G}_s$ is to reconstruct the input face $\mathbf{x}'_s = \mathbf{G}_s(\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t), \mathbf{c}_s)$ of a person based on the synthesized image $\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t)$ and source age $\mathbf{c}_s$. Same as $\mathbf{G}_t$, the input and output of $\mathbf{G}_s$ are also colour facial images with dimension $256 \times 256 \times 3$, and it is also composed of an encoder, a decoder and a age condition.

Differently, the input of $\mathbf{G}_s$ is the output of $\mathbf{G}_t$, and we hope the output of $\mathbf{G}_s$ is similar to the input of $\mathbf{x}_i$. The source age condition $\mathbf{c}_s$ drives the reconstruction process to produce face having the same age as the input face. Using $\mathbf{c}_s$ to reconstruct our original face, we have the result $\mathbf{G}_s(\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t), \mathbf{c}_s)$, that is regarded as a reconstruction of the input.

Discriminators

Discriminator aims to distinguish between the generated image (regarded as fake sample) and its ground truth (regarded as real sample). We define two discriminators network, $\mathbf{D}_t$ and $\mathbf{D}_s$. Here we follow the architecture design summarized by [Yi *et al.*, 2017]. For both $\mathbf{D}_t$ and $\mathbf{D}_s$, we set 4 convolution layers with an increasing number of 5×5 filter kernels (64, 128, 256, and 512). We also use batch-normalization layer and Leakly-ReLU after convolution. To judge different images belonging to different age groups, we also add age condition to our discriminator by concatenating reshaped age condition to the first convolution layer. Therefore, with a change of $\mathbf{c}_t$ in $\mathbf{G}_t$, we set the corresponding condition $\mathbf{c}_t$ for $\mathbf{D}_t$. Certainly the above process is similar to $\mathbf{G}_s$, $\mathbf{c}_s$ and $\mathbf{D}_s$.

2.2 The Primal and Dual cGAN Process

As we described above, our primal cGAN process can be formulated as

$$\mathbf{x}_i \rightarrow \mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t) \rightarrow \mathbf{G}_s(\mathbf{G}_t(\mathbf{x}_i, \mathbf{c}_t), \mathbf{c}_s). \tag{1}$$

From the equation we find that $\mathbf{G}_t$ and $\mathbf{G}_s$ have the same type of input and output, and they perform the same function. Therefore, it is necessary to add some constraints to $\mathbf{G}_t$ and $\mathbf{G}_s$. One intuitive way is to force them to share the weights. However, this constraint is too strict. Instead, inspired by dual learning [He *et al.*, 2016a], we add a dual process, which exchange $\mathbf{G}_t$ and $\mathbf{G}_s$. Specifically, given an input $\mathbf{x}_j$ with age $\mathbf{c}_t$, we get the generated result $\mathbf{G}_s(\mathbf{x}_j, \mathbf{c}_s)$ by the generator $\mathbf{G}_s$ and the condition $\mathbf{c}_s$. Then we reconstruct the input using the generator $\mathbf{G}_t$ and the original condition $\mathbf{c}_t$. The dual process can be formulated as

$$\mathbf{x}_j \rightarrow \mathbf{G}_s(\mathbf{x}_j, \mathbf{c}_s) \rightarrow \mathbf{G}_t(\mathbf{G}_s(\mathbf{x}_j, \mathbf{c}_s), \mathbf{c}_t). \tag{2}$$

2.3 Loss Function

The definition of the loss function $\mathcal{L}(\mathbf{G}_s, \mathbf{G}_t, \mathbf{D}_s, \mathbf{D}_t)$ is critical for the performance of our networks. It is composed of three parts: adversarial loss ($\mathcal{L}_{GAN}$) for matching the distribution of outputs to the data distribution, generation loss ($\mathcal{L}_{gene}$) to measure the similarity between specific generated image and the original, and reconstruction loss ($\mathcal{L}_{recons}$) to evaluate what extent the reconstructed images are consistent with the original ones. In the following subsections, we explain our realization of these loss functions.

Adversarial Loss

Adversarial loss tries to classify if the output image is real or fake. GANs can be extended to a conditional model if both the generator $\mathbf{G}$ and discriminator $\mathbf{D}$ are conditioned on some extra information $\mathbf{c}$. The objective of conditional GANs can be expressed as

$$E_{\mathbf{x},\mathbf{c} \sim P_{data}}[\log(1 - \mathbf{D}(\mathbf{G}(\mathbf{x}, \mathbf{c}), \mathbf{c}))] + E_{\mathbf{x},\mathbf{c} \sim P_{data}}[\log \mathbf{D}(\mathbf{x}, \mathbf{c})].$$

In our Dual cGAN, we have a primal cGAN and dual cGAN. An original facial image $\mathbf{x}_i$ is transformed into an age-changed image by the target generator $\mathbf{G}_t$. From the target age group, we randomly choose a real image $\mathbf{x_t}$ as a real sample for the target discrimination $\mathbf{D}_t$. Moreover, both $\mathbf{D}_t$ and $\mathbf{G}_t$ utilize the information of $\mathbf{c}_t$ as condition. By the same token, the source generator $\mathbf{G}_s$ and the source discriminator $\mathbf{D}_s$ can be trained with another input $\mathbf{x}_j$. In this process, $\mathbf{x}_s$, an random image from the source age group, is regarded as a real sample for $\mathbf{D}_s$. We formulate our adversarial loss as:

$$\begin{aligned}\mathcal{L}_{GAN} = & E_{x,c\sim P_{data}}[\log\left(1-\mathbf{D}_t\left(\mathbf{G}_t\left(\mathbf{x}_i,\mathbf{c}_t\right),\mathbf{c}_t\right)\right)] \\ & + E_{x,c\sim P_{data}}[\log \mathbf{D}_t\left(\mathbf{x}_t,\mathbf{c}_t\right)] \\ & + E_{x,c\sim P_{data}}[\log\left(1-\mathbf{D}_s\left(\mathbf{G}_s\left(\mathbf{x}_j,\mathbf{c}_s\right),\mathbf{c}_s\right)\right)] \\ & + E_{x,c\sim P_{data}}[\log \mathbf{D}_s\left(\mathbf{x}_s,\mathbf{c}_s\right)].\end{aligned} \tag{3}$$

Reconstruction Loss

Adversarial loss can help to produce a photo-realistic facial image. However, it alone cannot guarantee that a generated image have the same identify as its original image. Therefore, our Dual cGANs deal with this by defining a reconstruction loss. Using the notations already explained, we model the reconstruction loss as:

$$\begin{aligned}\mathcal{L}_{recons} = & \|\mathbf{G}_s(\mathbf{G}_t(\mathbf{x}_i,\mathbf{c}_t),\mathbf{c}_s)-\mathbf{x}_i\| \\ & + \|\mathbf{G}_t(\mathbf{G}_s(\mathbf{x}_j,\mathbf{c}_s),\mathbf{c}_t)-\mathbf{x}_j\|.\end{aligned} \tag{4}$$

Generation Loss

As we described above, our target age condition $\mathbf{c}_t$ determines the age of the facial image we want to generate. When we happen to choose $\mathbf{c}_t = \mathbf{c}_s$, we expect the generated image $\mathbf{G}_t(\mathbf{x}_i,\mathbf{c}_t)$ to be similar to $\mathbf{x}_i$. Similarly, we expect the generated image $\mathbf{G}_s(\mathbf{x}_j,\mathbf{c}_s)$ to be similar to $\mathbf{x}_j$. This composition is proposed to emphasize on preserving the original person's identity. This can be formulated as:

$$\mathcal{L}_{gene} = \|\mathbf{G}_t(\mathbf{x}_i,\mathbf{c}_t)-\mathbf{x}_i\| + \|\mathbf{G}_s(\mathbf{x}_j,\mathbf{c}_s)-\mathbf{x}_j\|. \tag{5}$$

Finally, the objective function is

$$\begin{aligned}\mathcal{L}(\mathbf{G}_s,\mathbf{G}_t,\mathbf{D}_s,\mathbf{D}_t) = & \mathcal{L}_{GAN}(\mathbf{G}_t,\mathbf{D}_t,\mathbf{G}_s,\mathbf{D}_s) \\ & +\beta\mathcal{L}_{recons}(\mathbf{G}_t,\mathbf{G}_s) + \alpha\mathcal{L}_{gene}(\mathbf{G}_t,\mathbf{G}_s).\end{aligned} \tag{6}$$

α and β are the hyper-parameters to balance difference loss functions. We aim to optimize

$$\arg\min_{\mathbf{G}_s,\mathbf{G}_t}\max_{\mathbf{D}_s,\mathbf{D}_t}\mathcal{L}(\mathbf{G}_s,\mathbf{G}_t,\mathbf{D}_s,\mathbf{D}_t). \tag{7}$$

We use back-propagation (BP) for learning and stochastic gradient descent (SGD) to minimize the loss.

2.4 Simplified Models

To achieve the goal of unsupervised face aging and rejuvenation, there are some simplified models which are variants of Dual cGANs. In this subsection, we introduce two of them.

One is only using a conditional GAN, as illustrated in Fig.3, but restricting $\mathbf{c}_t = \mathbf{c}_s$. We term this method cGAN, and its loss function can be formulated as:

$$\begin{aligned}& E_{x,c\sim P_{data}}[\log\left(1-\mathbf{D}\left(\mathbf{G}\left(\mathbf{x},\mathbf{c}\right),\mathbf{c}\right)\right)] \\ & + E_{x,c\sim P_{data}}[\log \mathbf{D}\left(\mathbf{x},\mathbf{c}\right)] + \lambda|\mathbf{G}(\mathbf{x},\mathbf{c})-\mathbf{x}\|.\end{aligned} \tag{8}$$

The second variant is to train y networks for y age groups. For each network, we treat the facial images with the age $\mathbf{c}_t$ as domain **A**, and the others as another domain **B**. Each network can be realized by a DualGAN [Yi *et al.*, 2017]. In this way, the loss function for each model can be formulated as:

$$\begin{aligned}& E_{x,c\sim P_{data}}[\log(1-\mathbf{D}_B(\mathbf{G}_B(\mathbf{x}_A)))] + E_{\mathbf{x},\mathbf{c}\sim P_{data}}[\log \mathbf{D}_B(\mathbf{x}_B)] \\ & + E_{\mathbf{x},\mathbf{c}\sim P_{data}}[\log(1-\mathbf{D}_A(\mathbf{G}_A(\mathbf{x}_B)))] + E_{\mathbf{x},\mathbf{c}\sim P_{data}}[\log \mathbf{D}_A(\mathbf{x}_A)] \\ & + \gamma|\mathbf{G}_A(\mathbf{G}_B(\mathbf{x}_A))-\mathbf{x}_A\| + \gamma|\mathbf{G}_B(\mathbf{G}_A(\mathbf{x}_B))-\mathbf{x}_B\|.\end{aligned} \tag{9}$$

3 Experiments

We evaluate our Dual cGANs on the task of face aging and rejuvenation. Specifically, the experiments are designed to study the following research questions of our algorithm:

RQ1: How does each component of our algorithm affect the performance?

RQ2: Does the performance of Dual cGANs significantly outperform the state-of-the-art methods?

RQ3: Can our Dual cGANs be generalized to other applications?

3.1 Data Collection

We use UTKFace [Zhang *et al.*, 2017] for training. UTKFace is a large-scale face dataset with long age span (range from 0 to 116 years old). We divide the images into 9 age groups, i.e., 0-3, 4-11, 12-17, 18-29, 30-40, 41-55, 56-65, 66-80, and 81-116, thus we can use a 9-dim one-hot vector to represent the age condition. It is significant that our dataset does not contain multiple images for each person which cover different age groups. It means that our training data is unpaired and non-sequential. To show the generalizability of Dual cGANs, we not only test on the same dataset as the training dataset (i.e., UTKFace), but also test on the face images from CACD [Chen *et al.*, 2014], FGNET [Lanitis *et al.*, 2002], Morph [Ricanek and Tesafaye, 2006] and IMDB-Wiki dataset [Rothe *et al.*, 2015].

3.2 Implementation Details

We build our networks as section 2 describes. Both the input images and the output images have the shape of $256 \times 256 \times 3$, while our condition for concatenating is a one-hot vector. For concatenation, we normalize the pixel values of the input images to [-1, 1]. And the elements of condition vector are also normalized to [-1, 1], while -1 actually means 0. We use the tanh activation function, so the output of our networks is also confined to [-1, 1].

During the training process, we set α=10.0 and β=10.0 for a balance between keeping personality and changing feature. Each input needs to be trained nine times with nine images from different age groups. Our batch size is one, but actually it is nine as we updated the parameters after nine iterations. With four blocks, $\mathbf{G}_s$, $\mathbf{G}_t$, $\mathbf{D}_s$ and $\mathbf{D}_t$, we train the discriminators one step, then two steps on generators as DualGAN [Yi *et al.*, 2017] described. The training time for 80 epochs on NVIDIA TITAN X (12GB) is about 160 hours. All the models are trained by RMSprop optimizer with a weight-decay of 0.9. The learning rate is initialized as 0.00005.

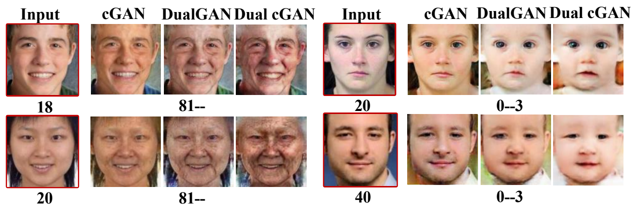

Figure 4: Experimental results of different variants of our method. As described in Sec.2.4, cGAN and DualGAN indicate the two simplified version of our Dual cGANs.

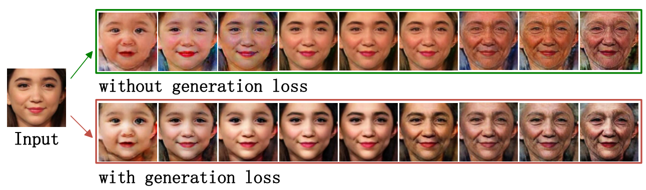

Figure 5: Experiment results for ablating the generation loss

3.3 Ablation Study (RQ1)

Our framework consists of age condition and three loss functions. In this subsection, we study the effect of each component on the performance.

If we ablate the different age conditions, our model degrades to the first simplified model cGAN. Its basic structure is similar to CAAE [Zhang *et al.*, 2017]. Then, we also compare with our second simplified model termed DualGAN. From Figure 4 we can observe that 1) cGAN cannot generate facial images with significant changes according to the ages. One possible reason is that the effect of condition is not guaranteed for the model of cGAN. 2) DualGAN can render images with the aging effect, but some details are not well processed. For example, in the last example, a baby face has mustache. In general, our Dual cGANs can simultaneously render aging face and preserve personality.

Also, we test the performance of our Dual cGANs by excluding the generation loss from our loss function, and the results are shown in Figure 5. Obviously, our generation loss is beneficial for increasing of image reality and preserving personality. Without the generation loss, the generated images are blurred and unrealistic (e.g., the last three facial images in the first line), and the personality is not well preserved (e.g., the first two facial images in the first line).

We tested the performance of Dual cGANs by removing discriminator losses from our ℓ, and the performance is unsatisfactory. Due to the space limit, we do not show these results.

3.4 Compare with the State-of-the-art (RQ2)

To evaluate that the proposed Dual cGANs can generate more photo-realistic results, we qualitatively and quantitatively compare ours with the best results from prior works. We also use the ground truth to compare the personality preserving ability of different methods. There are several prior methods, which can be regarded as a comparative benchmark:

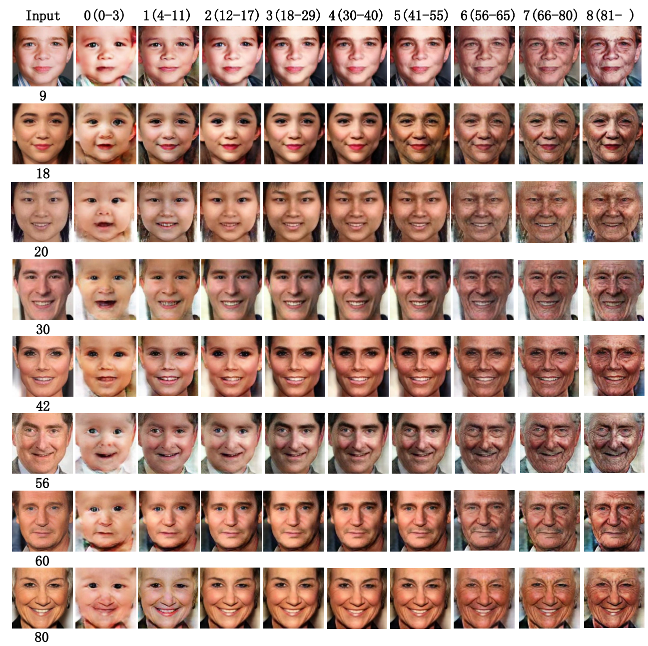

Figure 6: Our results. The leftmost column shows that original images are regarded as inputs, and the remaining nine columns represent the generated images belong to nine age groups $(0-3, 4-11, 12-17\ldots)$

FT demo[1], CDL [Shu *et al.*, 2015], RFA [Wang *et al.*, 2016], CAAE [Zhang *et al.*, 2017] and C-GANs [Liu *et al.*, 2017].

Qualitative Evaluation: We show our results for face aging/rejuvenation in Fig. 6, and the compared results in Fig. 7.

For our method, we have the following observations. First, the generated images are photo-realistic, and the details in skins, muscles, wrinkles, etc. are very clear. It shows the whole process of face aging for a person, i.e., the colours from light to dark, and the skin from smooth to wrinkles. Second, our Dual GAN can realize the transformation of long age span, which is useful to find lost children only with their young faces. For example, in the last line of Figure 6, we can render a vivid young girl's face from an elderly lady. We can also transform between any two age groups, so both face aging and rejuvenation can be achieved. Third, the generated images in each age group have specific subtle features. For example, the child tends to have a round face and no teeth, while the elderly people usually have small eyes and gray hair. These results indicate that our Dual cGANs achieve promising results for face aging and rejuvenation.

Compared with previous works, our Dual GAN can render facial images with clear characteristics belonging to different ages, and well preserve the personality. We can better tell the age of our generated facial images than the compared methods. For example, our generated images with small eyes and dark skin colours are usually 60+. All of our results highly resemble the input face image, but the other methods cannot. For example, the third and sixth results of CAAE are not that similar to the input face.

[1] http://cherry.dcs.aber.ac.uk/transformer

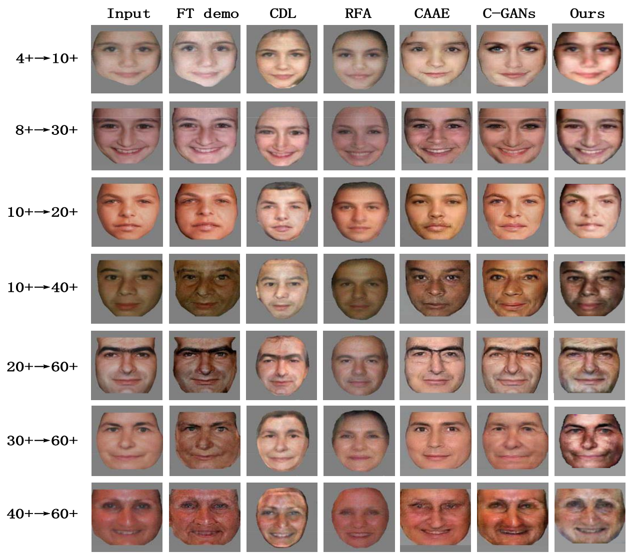

Figure 7: Qualitative results for different methods.

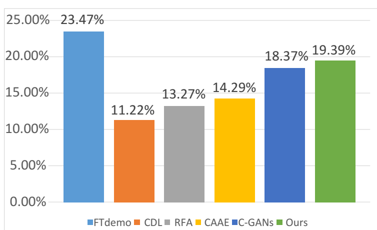

Figure 8: Quantitative evaluation to prior methods.

Quantitative Evaluation: To quantitatively evaluate the performance of the proposed method, we designed a user study. Each volunteer is presented with an original image, a target age and six generated images by Dual cGANs and other prior methods each time. We ask the volunteers to select the best result from the 6 candidates, which will credit 1 point to its method. There are 56 volunteers and we get 392 votes. Figure 8 shows the results. We can see that compared with the unpaired methods CAAE and C-GANs, our method achieves the best performance. FT demo outperforms our Dual cGANs by 4%. However, it requires paired training samples. And it only has 5 age groups for transformation.

Comparison with ground truth: In order to verify whether the personality has been preserved by the proposed Dual cGANs, we compare the generated faces of different methods with the ground truth. Considering the recognition by human eye are accurate, our quantitative evaluation is based on the SeetaFace Identification part of Seetaface Engine[2]. Figure 7 and Table 1 show the generated images and their scores, respectively, for some prior works and our method. Compared to other methods, our method can well preserve the personality. Compared with the best counterpart for personality preserving FT, our Dual cGANs improves it by 7%.

[2]https://github.com/seetaface/SeetaFaceEngine

	Input	FT	CDL	RFA	CAAE	C-GANs	Ours
4+→10+	0.98	0.85	0.48	0.59	0.57	0.47	0.86
8+→30+	0.99	0.86	0.78	0.42	0.51	0.79	0.82
10+→20+	0.97	0.80	0.68	0.44	0.51	0.70	0.81
10+→40+	0.95	0.60	0.65	0.33	0.51	0.49	0.80
20+→60+	0.97	0.77	0.73	0.48	0.65	0.76	0.80
30+→60+	0.97	0.64	0.70	0.41	0.45	0.65	0.72
40+→60+	0.98	0.68	0.59	0.52	0.70	0.75	0.84
Average	0.97	0.74	0.66	0.46	0.56	0.66	0.81

Table 1: Scores of images in Figure 7 compared with their inputs.

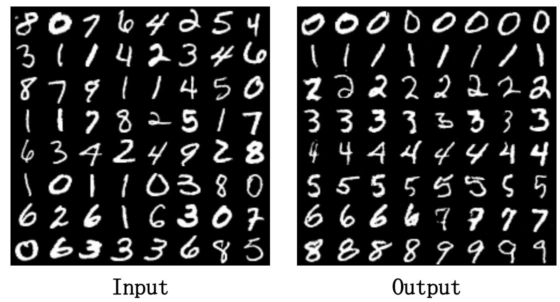

Figure 9: Generated images (right) using an original image from MNIST (left) and a target digital condition (0-9).

3.5 Generalization Ability Study (RQ3)

We also test our networks on MNIST [LeCun *et al.*, 1998], which is a digit dataset containing 60,000 training images and 10,000 test images. Each sample is a 28×28 gray-scale image with the class label from 0 to 9. As Figure 9 shows, we transform test images from MNIST with different class label. Our input is a stitched image for different kinds of digital images. We code different digits as condition, thus we have 10 labels, from 0 to 9, to generate the corresponding digital images.

Clearly, our method is generalized well for digital image generation task on MNIST dataset. The generated digital images have the same label as the target condition, and also have good reality. It indicates that our Dual cGANs can transform images among multiple domains. On the other hand, the generated images resemble to the inputs (e.g., digit's angle, digit's shape and thickness of the digit's stroke). Because the reconstruction loss has strong ability to reconstruct original images, which is helpful for training our networks.

4 Conclusions

In this paper, we develop a novel dual conditional GANs (Dual cGANs) mechanism, which enables face aging and rejuvenation to be trained from multiple sets of unlabeled face images with different ages. In our architecture, the primal conditional GAN transforms a face image to other ages based on the age condition, while the dual conditional GAN learns to invert the task. A loss function of reconstruction error of images can preserve the personal identity, and the discriminators on the generated images learn the transition patterns (e.g., the shape and texture changes between age groups) and guide the generation of age-specific photo-realistic faces. Experimental results on two publicly dataset demonstrate the appealing performance of the proposed framework by comparing with the state-of-the-art methods.

Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. ZYGX2014J063, No. ZYGX2016J085) and the National Natural Science Foundation of China (Grant No. 61772116, No. 61502080, No. 61632007, No. 61602049). This work was partly supported by the 111 Project No. B17008.

References

[Antipov *et al.*, 2017] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. *arXiv preprint arXiv:1702.01983*, 2017.

[Chen *et al.*, 2014] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *ECCV*, pages 768–783. Springer, 2014.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[He *et al.*, 2016a] Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tieyan Liu, and Wei-Ying Ma. Dual learning for machine translation. In *NIPS*, pages 820–828, 2016.

[He *et al.*, 2016b] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[Kemelmacher Shlizerman *et al.*, 2014] Ira Kemelmacher Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz. Illumination-aware age progression. In *CVPR*, pages 3334–3341, 2014.

[Lanitis *et al.*, 2002] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[LeCun *et al.*, 1998] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Liu *et al.*, 2017] Si Liu, Yao Sun, Defa Zhu, Renda Bao, Wei Wang, Xiangbo Shu, and Shuicheng Yan. Face aging with contextural generative adversarial nets. In *ACM Multimedia*, pages 82–90. ACM, 2017.

[Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[Park *et al.*, 2010] Unsang Park, Yiying Tong, and Anil K Jain. Age-invariant face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):947–954, 2010.

[Radford *et al.*, 2015] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[Ramanathan and Chellappa, 2006] Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *CVPR*, volume 1, pages 387–394. IEEE, 2006.

[Ramanathan and Chellappa, 2008] Narayanan Ramanathan and Rama Chellappa. Modeling shape and textural variations in aging faces. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–8. IEEE, 2008.

[Ricanek and Tesafaye, 2006] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *FG*, pages 341–345. IEEE, 2006.

[Rothe *et al.*, 2015] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *ICCV Workshops*, pages 10–15, 2015.

[Shu *et al.*, 2015] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *ICCV*, pages 3970–3978, 2015.

[Song *et al.*, 2018] Jingkuan Song, Tao He, Lianli Gao, Xing Xu, Alan Hanjalic, and Heng Tao Shen. Binary generative adversarial networks for image retrieval. In *AAAI*, 2018.

[Suo *et al.*, 2010] Jinli Suo, Song-Chun Zhu, Shiguang Shan, and Xilin Chen. A compositional and dynamic model for face aging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):385–401, 2010.

[Suo *et al.*, 2012] Jinli Suo, Xilin Chen, Shiguang Shan, Wen Gao, and Qionghai Dai. A concatenational graph evolution aging model. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2083–2096, 2012.

[Tazoe *et al.*, 2012] Yusuke Tazoe, Hiroaki Gohara, Akinobu Maejima, and Shigeo Morishima. Facial aging simulator considering geometry and patch-tiled texture. In *SIGGRAPH*, page 90. ACM, 2012.

[Tiddeman *et al.*, 2001] Bernard Tiddeman, Michael Burt, and David Perrett. Prototyping and transforming facial textures for perception research. *IEEE computer graphics and applications*, 21(5):42–50, 2001.

[Wang *et al.*, 2016] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe. Recurrent face aging. In *CVPR*, pages 2378–2386, 2016.

[Yi *et al.*, 2017] Zili Yi, Hao Zhang, Ping Tan Gong, et al. Dualgan: Unsupervised dual learning for image-to-image translation. *arXiv preprint arXiv:1704.02510*, 2017.

[Zhang *et al.*, 2017] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *CVPR*, 2017.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.