Global and Local Consistent Wavelet-domain Age Synthesis
Peipei Li†, Yibo Hu†, Ran He Member, IEEE and Zhenan Sun Member, IEEE

Abstract—Age synthesis is a challenging task due to the complicated and non-linear transformation in human aging process. Aging information is usually reflected in local facial parts, such as wrinkles at the eye corners. However, these local facial parts contribute less in previous GAN based methods for age synthesis. To address this issue, we propose a Wavelet-domain Global and Local Consistent Age Generative Adversarial Network (WaveletGLCA-GAN), in which one global specific network and three local specific networks are integrated together to capture both global topology information and local texture details of human faces. Different from the most existing methods that modeling age synthesis in image-domain, we adopt wavelet transform to depict the textual information in frequency-domain. Moreover, five types of losses are adopted: 1) adversarial loss aims to generate realistic wavelets; 2) identity preserving loss aims to better preserve identity information; 3) age preserving loss aims to enhance the accuracy of age synthesis; 4) pixel-wise loss aims to preserve the background information of the input face; 5) the total variation regularization aims to remove ghosting artifacts. Our method is evaluated on three face aging datasets, including CACD2000, Morph and FG-NET. Qualitative and quantitative experiments show the superiority of the proposed method over other state-of-the-arts.

Index Terms—Age synthesis, wavelet transform, generative adversarial network, global and local features.

I. INTRODUCTION

H

UMAN age synthesis, also known as age progression/regression, has captured much attention in many real-world applications such as finding missing children, age estimation, cross-age face verification, social entertainment, etc. It aims to aesthetically rerender a given face with natural aging effects but still preserve personality. Impressive progress has been made on age synthesis in recent years and many methods [1], [2], [3], [4], [5] have been proposed. However, large-scale age synthesis is still a very challenging problem for several reasons. 1) Rigid requirement for the training and testing datasets, i.e., most existing works require face images of the same subject at different ages. 2) Faces captured in the wild are usually influenced by the large variations of illumination, pose, resolution and expression. 3) Different subjects have different aging processes. There are many internal and external influential factors, e.g., gene, gender, living style.

†These authors contributed to the work equally.
P. Li, Y. Hu, R. He and Z. Sun are with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, 100190, Beijing, China. E-mail: peipei.li@cripac.ia.ac.cn; yibo.hu@cripac.ia.ac.cn; rhe@nlpr.ia.ac.cn; znsun@nlpr.ia.ac.cn.
P. Li, R. He and Z. Sun are also with the University of Chinese Academy of Sciences, Beijing, China.

In prior works, physical modeling based methods and prototype-based methods are mainly used for age synthesis. Physical modeling based methods build computational models for age progression/regression to simulate human biological and aging mechanisms of the cranium, muscles or facial skin etc. Although rich and obvious textures can be generated, these approaches depend on many face sequences of the same subject covering a long age span, which is hard and expensive to obtain. Besides, in prototype-based methods, only texture synthesis is considered. Training data is divided into different age groups. Average face of each age group is regarded as its prototype, thus the aging pattern is the difference between prototypes of target and source age groups. However, prototype-based approaches may ignore identity and different subjects share the same aging pattern. Recently, deep learning based methods have made a breakthrough in computer vision. Training with large-scale data, deep neural networks substantially show better performance than previous methods in image generation. Generative Adversarial Networks (GANs) [6] as a specific type of deep neural network, has proven capable of generating photo-realistic images and is therefore introduced to deal with age synthesis [7], [8], [1], [2], [3], [4], [5]. However, the generated results of existing GAN based methods are often over-smoothed and lack of detailed texture information. Fine-grained age synthesis remains an open and challenging task.

Fig. 1. Demonstration of our age regression results. For each subject, the first column is the input, while the rest three columns are the rejuvenating results from older to younger. (Zoom in for a better view.)
We tackle this fine-grained age synthesis from two aspects: modeling aging patterns of age-sensitive areas with local networks and introducing wavelet transform to age synthesis. In the first place, aging texture information is usually reflected in the local facial parts, such as wrinkles at the eye corners or on the forehead. However, existing GAN based models cannot fully capture these rich textures in the local facial parts. Secondly, most current age synthesis methods only depend on modeling aging process in image-domain. Thus the outputs of these methods tend to be over-smoothed and lack of textural details. Since the subtle texture information is more salient and robust in frequency-domain, we introduce wavelet transform to age synthesis.

In this paper, a novel Wavelet-domain Global and Local Consistent Age Generative Adversarial Network (WaveletGLCA-GAN) is proposed for fine-grained age synthesis. It consists of four parts: wavelet-domain global and local consistent age generator, discriminator, age evaluation network, identity evaluation network. Fig. 1 and Fig. 2 depict the rejuvenating results and framework of WaveletGLCA-GAN, respectively.

The wavelet-domain global and local consistent age generator (WaveletGLCA-G) takes a random age face as an input and predicts the corresponding wavelet coefficients before reconstructing the target age output. Apart from one global specific network like previous works, three local specific networks are imposed to capture the aging texture information in forehead, eyes and mouth parts, respectively. Besides, the wavelet coefficient prediction network is employed to predict wavelet coefficients of the target age faces. Instead of manipulating a whole face, our WaveletGLCA-G learns the residual face, which can accelerate the convergence while preserving the identity and background information. The discriminator is proposed to harness target age information and try to distinguish the synthesized faces from the real target age faces. Through adversarial training, the WaveletGLCA-G is forced to produce photo-realistic target age images.

Aging accuracy is one of the two critical requirements in age synthesis. Thus, we adopt the age evaluation network to evaluate and reinforce the aging accuracy in age synthesis. The age evaluation network takes the generated face as an input and estimates its age. Meanwhile, identity preserving performance is the other critical requirement in age synthesis. An identity evaluation network is introduced to evaluate and enhance the identity preserving performance in age synthesis. It takes the generated and participant faces as inputs and computes their similarity scores to evaluate how much identity information can be retained.

The main contributions of this work are as follows:

1) A global and local consistent age generator is proposed for deep facial age synthesis, in which three local specific networks and one global specific network are integrated together to capture both global topology and local texture information of human faces.

2) Different from the most existing studies that model age synthesis in image-domain, we transform age synthesis to a wavelet coefficient prediction task in frequency-domain. To the best of our knowledge, this is the first attempt to transform age synthesis to wavelet coefficient prediction task in GAN-based framework for age synthesis.

3) The proposed WaveletGLCA-GAN simultaneously achieves age progression and regression in the same framework with the given age labels. Experimental results show that the proposed approach outperforms state-of-the-art methods in terms of qualitative and quantitative metrics.

This paper is an extension of our previous method GLCA-GAN [9]. Apart from providing more in-depth analysis and more extensive experiments, the major difference between this paper and its previous version lies in three-folds: 1) Age synthesis is extended from image-domain generative adversarial network (GLCA-GAN) to wavelet-domain generative adversarial network (WaveletGLCA-GAN) with the impose of frequency-domain information which is more sensitive to texture details. 2) Age synthesis is transformed into wavelet coefficients prediction task. Wavelet coefficient prediction network as well as wavelet reconstruction networks are used to capture and model the delicate texture information in wavelet-domain. Besides, higher-resolution age synthesis is achieved with pixel-size extended from 128 × 128 to 224 × 224. 3) Age estimation and identity verification are proposed to evaluate the aging accuracy and identity preserving performance of WaveletGLCA-GAN.

II. RELATED WORK

A. Image-to-image Translation

Age synthesis is a subproblem of image-to-image translation [10], [11], [12], [13] which transforms the input image from the source domain to the target domain with the original structure or semantics preserving. In recent years, image translation has been widely investigated. A host of recent works have explored the ability of Generative Adversarial Networks (GANs) [6] for image-to-image translation. With a generator and a discriminator, GANs play a min-max game to learn the target image prior. pix2pix [14] learns image-to-image translation in a supervised manner using cGANs [15], which requires paired data samples. StarGAN [16] adds a mask vector to the domain label, which enables joint training between multiple domains. To alleviate the requirement to paired data, unpaired image-to-image translation has been proposed. Cycle-consistent Adversarial Networks (CycleGANs) [17] is
Fig. 3. General framework of WaveletGLCA-G, including global specific network, local specific networks, feature fusion network, wavelet coefficient prediction network as well as wavelet reconstruction network.

successfully applied for unpair image-to-image translation. Recently a Couple-Agent Pose-Guided Generative Adversarial Network (CAPG-GAN) [18] is proposed for arbitrary view face synthesis which combines prior domain knowledge of pose and local structure of face to reinforce the realism of results. [19] introduces the idea of asymmetric style transfer and proposes PairedCycleGAN, which involves two asymmetric networks for makeup transfer and removal.

B. Age Synthesis

The facial age synthesis methods can be mainly classified into three categories: physical modeling based, prototype based and deep learning based methods.

In the early time, physical modeling based methods and prototype based methods are mainly adopted in age synthesis. The physical modeling based methods [20], [21], [22], [23], [24] simulate age progression/regression by modeling human biological and physical mechanisms, e.g., muscles, facial contours. However, this kind of approaches need abundant face sequences of the same subject covering a long age span, which is expensive and hard to obtain. Besides, prototype-based approaches [25], [26] regard the average face of each age group as its prototype, thus the aging pattern is the difference between the prototypes of target and source age groups. However, prototype-based approaches ignore identity information. To handle this problem, Shu et al. [27] propose an age synthesis method based on dictionary learning to reconstruct the aging face.

Recently, deep learning based approaches [7], [8], [1], [2], [3], [4], [5] have shown considerable ability in age synthesis. Wang et al. [3] propose a recurrent neural network to model a smoother face aging process. Zhang et al. [1] apply Conditional Adversarial Autoencoder (CAAE) to synthesize target age faces with target age labels. In addition, Zhou et al. [4] argue that occupation information influences the individual aging process and propose an occupational-aware adversarial face aging network. To make the most of the discriminative ability of GAN, Yang et al. [2] put forward a multi-pathway discriminator to refine the aging/rejuvenating results. Duong et al. [3] present a generative probabilistic model to simulate the aging mechanism of each age stage. Wang et al. [5] impose an identity-preserved term and an age classification term into age synthesis and propose the Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs). However, the age synthesis results of existing methods are still over-smoothed and lack of detailed texture information, which hurts the aging accuracy and identity preserving performance.

III. APPROACH

Age synthesis aims to synthesize a target age face image $\hat{I}_t$ from a given age face image $I_g$. Our goal is to learn such a synthesizer that can infer the corresponding age images. The overall framework of our proposed Wavelet-domain Global and Local Consistent Age Generative Adversarial Network (WaveletGLCA-GAN) is depicted in Fig. 2, which mainly consists of four parts: wavelet-domain global and local consistent age generator, discriminator, age evaluation network, identity evaluation network.

In WaveletGLCA-GAN, age information is encoded into a one-hot vector $l$ with the target age being 1. In this paper we divide the ages into four age groups, thus the age vector $l \in R^{1 \times 4}$. We stack $l$ to $L \in R^{H \times W \times 4}$ as the input age label, where $H$ and $W$ are the height and width of the input image, respectively. Given an input age face $I_g$, forehead, eyes and mouth local patches $p_1$, $p_2$, $p_3$ are first cropped according to the five facial landmarks. Then the age label $L$ is concatenated to both $I_g$ and three local facial patches $p_1$, $p_2$, $p_3$. The new global tensor $[I_g, L]$ and three local tensor $[p_1, L]$, $[p_2, L]$, $[p_3, L]$ are fed into WaveletGLCA-G to synthesize the target age image $\hat{I}_t$. Five types of losses are adopted to supervise the WaveletGLCA-G to achieve accurate age generation under the premise of preserving the background and identity information.

This section firstly introduces the wavelet transform and the architecture of the WaveletGLCA-G, discriminator, age evaluation network and identity evaluation network. Then the
details of five supervised loss functions, including conditional adversarial loss, identity preserving loss, age preserving loss, pixel-wise loss and total variation regularization, are presented later.

A. Wavelet Transform

Most current age synthesis methods only model aging process in image-domain, whose synthesized results usually tend to be over-smoothed and lack of textural details. To alleviate this limitation, we employ wavelet packet transform, which decomposes the image into the approximation (low-frequency component) and the details (high-frequency components). In this paper, the low-frequency component is related to the general face information, while the high-frequency components are related to the detailed aging texture information.

Fig. 5 depicts the process of wavelet decomposition and reconstruction. As shown in Fig. 5(a), the original image is firstly decomposed into an approximation coefficient and three detailed coefficients. Then according to the wavelet components, the image is reconstructed as shown in Fig. 5(b). Inspired by this reversible process, we propose this wavelet-domain age synthesis method.

Note that in the proposed method, the wavelet coefficients are not explicitly decomposed from the original image by wavelet decomposition, but are directly learned from the global and local fused features. Specifically, we initialize the wavelet reconstruction network with a pre-computed wavelet decomposition, but are directly learned from the corresponding target age faces. As shown in Fig. 3, we concatenate the given image as the input of WaveletGLCA-G and synthesize the target age image $\hat{I}^t$.

$$\hat{I}^t = G_{\theta_G}(I^g, L)$$

where $\theta_G$ is the parameter of WaveletGLCA-G. In this paper, we choose the image size of 224 and divide the ages of subjects into four age groups, thus $L \in \mathbb{R}^{224 \times 224 \times 4}$. The detailed architecture of WaveletGLCA-G is shown in Table I.

Specifically, the global specific network $G_g$ takes the whole image $I^g \in \mathbb{R}^{224 \times 224 \times 3}$ concatenated with the target age label $L$ as the input $(I^g, L) \in \mathbb{R}^{224 \times 224 \times 7}$. An encoder-decoder architecture with four strided convolutional layers, two residual blocks and two deconvolutional layers are exploited to extract and transform the global feature maps $G_g(I^g, L) \in \mathbb{R}^{112 \times 112 \times 256}$. In addition, the three local specific networks $G_{li}, i \in \{1,2,3\}$ take forehead, eyes and mouth local patches $p_i$ concatenated with the target age label $L$ as the inputs and learn a separate set of filters to extract and transform the corresponding local feature maps $G_{li}(p_i, L) \in \mathbb{R}^{112 \times 112 \times 256}$. All the local specific networks adopt the encoder-decoder architecture with three strided convolutional layers, two residual blocks and one deconvolutional layer.

The feature fusion network fuses the four feature embeddings of global and local networks by two residual blocks. The wavelet coefficient prediction network has four parallel independent subnetworks. Each of the subnetworks takes the whole output of the fusion network as its input and predicts the corresponding wavelet coefficients, including the approxi-
TABLE I
THE ARCHITECTURE OF WAVELETLGCA-G $G_{θ_G}$.

| Layer       | Input | Filter/Stride | Output Size |
|-------------|-------|---------------|-------------|
| G_conv1     | 3 x 3 | 3 x 3/1       | 224 x 224 x 64 |
| G_conv2     | 3 x 3 | 3 x 3/2       | 112 x 112 x 128 |
| G_conv3     | 3 x 3 | 3 x 2/2       | 56 x 56 x 256 |
| G_conv4     | 3 x 3 | 3 x 3/2       | 28 x 28 x 512 |
| G_conv5     | 3 x 3 | 3 x 3/1       | 28 x 28 x 512 |
| G_deconv1   | 3 x 3 | 3 x 3/2       | 56 x 56 x 256 |
| G_deconv2   | 3 x 3 | 3 x 3/2       | 112 x 112 x 256 |

TABLE II
THE ARCHITECTURE OF DISCRIMINATOR $D_{θ_D}$.

| Layer       | Input | Filter/Stride | Output Size |
|-------------|-------|---------------|-------------|
| D_conv1     | 3 x 3 | 3 x 3/1       | 112 x 112 x 32 |
| D_res1      | 3 x 3 | 3 x 3/1       | 112 x 112 x 64 |
| D_res2      | 3 x 3 | 3 x 3/1       | 56 x 56 x 64  |
| D_res3      | 3 x 3 | 3 x 3/1       | 28 x 28 x 128 |
| D_res4      | 3 x 3 | 3 x 3/1       | 28 x 28 x 128 |
| D_res5      | 3 x 3 | 3 x 3/1       | 7 x 7 x 256   |
| D_res6      | 3 x 3 | 3 x 3/1       | 7 x 7 x 256   |

is preserved. We choose a pretrained Light CNN [28] as the identity evaluation network and fix the parameters during training procedure.

4) Architecture of Age Evaluation Network: An age evaluation network is introduced in WaveletGLCA-GAN to reinforce the aging accuracy in age synthesis. It takes the generated face $I'$ as the input and estimate the corresponding age to evaluate how much age-related information has been learned. We choose a pretrained VGG16 [29] as the age evaluation network and fix the parameters during training procedure.

C. Training Losses

A weighted sum of five losses is imposed to supervise the proposed WaveletGLCA-GAN in an end-to-end manner, including conditional adversarial loss, identity preserving loss, age preserving loss, pixel-wise loss and total variation regularization.

1) Conditional Adversarial Loss: To make the synthesized images indistinguishable from the real data and incorporate prior domain knowledge (age information from target age group), the adversarial loss conditioned on the target age label is introduced to our WaveletGLCA-GAN. In addition, for avoiding producing artifacts as in [30], our discriminator $D_{θ_D}$ distinguishes the $6 \times 6$ local image patches separately. The adversarial loss is formulated as:

$$
L_{adv} = \min_{θ_G} \max_{θ_D} E_{I, L \sim p(I', L)} \left[ \log D_{θ_D}(I', L) \right] + E_{I, L \sim p(I, L)} \left[ \log (1 - D_{θ_D}(I, L)) \right] + E_{I, L \sim p(I, L)} \left[ \log (1 - D_{θ_D}(G_{θ_G}(I, L), L)) \right]
$$

(2)

where $I'$ denotes the real face of age group $L$. Parameters of the generator and the discriminator are trained alternately to optimize the min-max problem.

2) Identity Preserving Loss: It is crucial to keep identity information in the age synthesis. However, the synthetic faces based on GAN are usually close to the real data only in pixel space, not in semantic space. Hence, we introduce an identity preserving loss to our model. Following [31], the identity preserving loss is formulated as:

$$
L_{id} = ||φ_I(I') - φ_I(G_{θ_G}(I, L))||_2^2 + ||φ_P(I') - φ_P(G_{θ_D}(I', L))||_F
$$

(3)

mation coefficient, the detail horizontal, vertical and diagonal coefficients. All the predicted coefficients are half the size of the input, i.e., $112 \times 112 \times 3$.

Finally, the reconstruction network takes the four predicted wavelet coefficients as its input and reconstructs the target age face $\hat{I}^t \in R^{224 \times 224 \times 3}$. The reconstruction network is realized by a deconvolutional layer, which is initialized by a constant wavelet reconstruction matrix. The parameters of the reconstruction network are pre-computed and can be downloaded from the released code. Note that, during training, the parameters of reconstruction network are fixed. In this way, the gradients from the age evaluation and identity evaluation networks can be propagated backwards, which leads to an end-to-end training of the proposed WaveletGLCA-G.

2) Architecture of Discriminator: To exploit the prior domain knowledge of ages, we impose age labels into the discriminator $D_{θ_D}$ to further force the generation of age-specific faces. $θ_D$ is the parameter of the proposed discriminator. Particularly, both the input face $I'$ and the synthetic face $\hat{I}^t$ with the target age label $L$ are treated as negative samples, while the real face $I^t$ with $L$ is the positive sample. The discriminator takes them as the inputs and outputs a scalar, representing the probability that the inputs come from the real data. Based on the GAN principle, our discriminator $D_{θ_D}$ forces the WaveletGLCA-G to synthesize realistic and plausible target age faces, which are indistinguishable with the real faces. The detailed architecture of our discriminator is shown in Table II.

3) Architecture of Identity Evaluation Network: An identity evaluation network is employed to preserve the identity information in age synthesis. It takes the generated face $\hat{I}^t$ and the input face $I^t$ as the inputs and computes their similarity scores to evaluate how much identity information
where $\varphi_f$ and $\varphi_p$ denote the feature extractors of the fully-connected layer and the last pooling layer of the pre-trained light CNN-29 network \cite{28} respectively.

3) **Age Preserving Loss**: Age accuracy is another key issue of age progression/regression. In our model, an age preserving loss based on the pre-trained VGG16 structure is utilized to enhance the age accuracy of the synthetic face, which can be formulated as:

$$L_{age} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{3} (1 \{ y^i = j \} \log(s))$$

(4)

where $s$ denotes the output of the final softmax layer of VGG16, $y^i$ denotes the age class of the i-th synthetic face. When $y^i$ is equal to the $j$, $1 \{ y^i = j \} = 1$, otherwise $1 \{ y^i = j \} = 0$.

4) **Pixel-wise Loss**: By minimizing the adversarial, identity preserving and age preserving losses, $G_{\theta_G}$ can generate photo-realistic faces with accurate identity and age information. However, minimizing these losses can’t guarantee the background well preserved. Since there is no ground-truth target face in age synthesis, a pixel-wise $L2$ loss between the synthetic and the input faces is adopted to preserve the background.

$$L_{pixel} = \frac{1}{C \times H \times W} \| I^p - G_{\theta_G}(I^g, L) \|_2^2$$

(5)

where $C$, $H$, $W$ are the channel, height and width of the image respectively. It is worth noting that Eq. (5) forces the synthetic face image to be similar with the input face image. Thus we update this pixel-wise loss at every 10 iterations to balance the age synthesis and the background preservation.

5) **Total Variation Regularization**: The images generated by GAN usually have some ghosting artifacts \cite{32}, which deteriorates visualization as well as recognition performance. A total variation regularization term is thus imposed to remove the unfavorable artifacts.

$$L_{tv} = \sum_{c=1}^{C} \sum_{w,h=1}^{W,H} | \hat{I}_{w+1,h,c}^t - \hat{I}_{w,h,c}^t | + | \hat{I}_{w,h+1,c}^t - \hat{I}_{w,h,c}^t |$$

(6)

6) **Overall Loss**: Finally, the total supervised loss is a weighted sum of the above five losses. $G_{\theta_G}$ and $D_{\theta_D}$ are trained alternately to optimize the mini-max problem. Taking all loss functions together, the overall loss can be written as:

$$L = \lambda_1 L_{adv} + \lambda_2 L_{ip} + \lambda_3 L_{age} + \lambda_4 L_{pixel} + \lambda_5 L_{tv}$$

(7)

where $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$ are trade-off parameters that control the relative importance among conditional adversarial loss, identity preserving loss, age preserving loss, pixel-wise loss and total variation regularization.

IV. EXPERIMENTS

WaveletGLCA-GAN provides a flexible way to render the input face to any age group controlled by the age label. It produces photo-realistic faces with accurate identity and age information. In the following subsections, we begin with an introduction of datasets and settings. Then we demonstrate the superiority of our WaveletGLCA-GAN on both qualitative visualization and quantitative face verification and age estimation. Lastly, we conduct ablation study to demonstrate the benefits gained from local specific networks and wavelet coefficient prediction network.

A. Datasets and Settings

1) **Datasets**: We employ three datasets, including CACD2000 \cite{33}, Morph \cite{34} and FG-NET \cite{35}, to evaluate the proposed WaveletGLCA-GAN.

- **CACD2000** \cite{33}: The CACD2000 dataset contains 163,446 color images of 2,000 celebrities with age range from 14 to 62 years old. It has more than 80 images for per subject on average. The maximum age gap between images for the same subject is 10 years. With variations of age in the wild, it has been widely used to evaluate age synthesis and age invariant verification performance under unconstrained environments.

- **Morph** \cite{34}: The Morph dataset is the largest public available dataset for evaluating age synthesis, age estimation and age invariant verification, which contains 55,349 color images of 13,672 subjects with age and gender information. The subject ages of Morph range from 16 to 77 years old. It is about four images of each subject on average in the Morph dataset.

- **FG-NET** \cite{35}: The FG-NET dataset is a popular benchmark for age synthesis, age estimation and age invariant verification, which only contains 1,002 images of 82 subjects. The ages of FG-NET range from 0 to 69, while the maximum age gap between images for the same subject is 54 years. We adopt FG-NET as testing set to make fair comparisons with prior works.

Five-fold cross-validation is conducted on CACD2000 and Morph. In particular, we divide the dataset into five folds with one fold for testing and the other four folds for training. On CACD2000, each fold contains 400 subjects and the face images in each fold are divided into four age groups: 14-30, 31-40, 41-50, 51-62 with about 9,536, 8,507, 7,890 and 5,959 face images, respectively. On Morph, the face images in each fold are divided into four age groups: 16-30, 31-40, 41-50, 51-77 with about 4,946, 3,207, 2,266 and 616 face images. Note that, for the settings of the two datasets, there are no overlap subjects between the training and testing sets.

2) **Experimental Settings**: The identity evaluation network is pretrained on MSCeleb-1M \cite{36}. The age evaluation network is pretrained on IMDB-WIKI \cite{37}, while fine-tuned on the corresponding training sets. During training procedure, we fix the parameters of the identity evaluation network and age evaluation network. We utilize multi-task cascaded CNN \cite{38} to detect the bounding box and five facial landmarks of the face images and align the face images by affine transformation to make the two eyes horizontal. According to the detected landmarks, we crop three subregions of forehead, eyes and mouth. All the images are resized to 224 $\times$ 224. Then we convert all the images in the range [0, 255] to the tensors in
the range [0.0, 1.0] and send them into the WaveletGLCA-G. Finally, the output of the WaveletGLCA-G is a tensor in the range [0.0, 1.0]. Before storing as images, the output tensors are multiplied by 255 and taken integers downwardly.

Our model is implemented with Pytorch. During training, we choose Adam optimizer with $\beta_1$ of 0.5, $\beta_2$ of 0.99, learning rate of $2 \times 10^{-4}$, and the batch size is set to 8. For CACD2000, the trade-off parameters $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ and $\lambda_5$ are set to 1.00, 0.01, 120.00, 45.00 and 0.00003, respectively. While for Morph, they are set to 1.00, 0.01, 80.00, 45.00 and 0.0001, respectively. In addition, we update pixel-wise loss at every 10 iteration, and update the discriminator for every 2 generator iterations.

On one GeForce GTX TITAN X and 32 Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz, it takes 72 hours for Morph and 108 hours for CACD2000.

Specifically, the global specific network $G_g$ takes the whole image $I^g \in R^{224 \times 224 \times 3}$ concatenated with the target age label $L \in R^{224 \times 224 \times 4}$ as its input $(I^g, L) \in R^{224 \times 224 \times 7}$ and its output is $G_g(I^g, L) \in R^{112 \times 112 \times 256}$. Simultaneously, the three local specific networks $G_{il}, i \in \{1, 2, 3\}$ take the forehead, eyes and mouth local patches $p_i \in R^{224 \times 224 \times 3}$ concatenated with the target age label $L \in R^{224 \times 224 \times 4}$ as the inputs and learn a separate set of filters to extract and transform the corresponding local feature maps $G_{il}(p_i, L) \in R^{112 \times 112 \times 256}$. The feature fusion network concatenates the $G_g(I^g, L)$ and three $G_{il}(p_i, L)$ as its input and outputs a fused feature $f \in R^{112 \times 112 \times 512}$ for wavelet coefficient prediction network. Then the wavelet coefficient prediction network takes $f$ as its input and predicts four specific wavelet coefficients $w_{ij}, j \in \{1, 2, 3, 4\}$ by four individual subnetworks, where all the $w_{ij} \in R^{112 \times 112 \times 3}$. Finally, the wavelet reconstruction...
network concatenates the four \(w_j, j \in \{1, 2, 3, 4\}\) as its input and reconstructs the target age face \(\hat{I}^t \in \mathbb{R}^{224 \times 224 \times 3}\).

The age evaluation network takes the synthesized image \(\hat{I}^t \in \mathbb{R}^{224 \times 224 \times 3}\) as its input. It predicts the age \(\hat{y}\) of \(\hat{I}^t\). The identity evaluation network takes the input image and the synthesized image as its inputs and outputs the face features. Besides, the discriminator takes the input image, the synthesized image and real image in the target age as its inputs, and outputs a scalar, representing the probability that the inputs come from the real data.

B. Qualitative Evaluation of WaveletGLCA-GAN

In this subsection, we present and compare the visualization results against state-of-the-art age synthesis methods.

1) Results of Age Progression and Regression: WaveletGLCA-GAN can synthesize any age faces controlled by the age labels. Hence, it is able to achieve both age progression and regression. We firstly show the age progression results on CACD2000 and Morph in Fig. 6 and Fig. 8 respectively. The first column of each subject is the input face under 30 years old. The second, third and fourth columns are the synthetic results of WaveletGLCA-GAN for 31-40, 41-50 and 51-62(77) in turn. We can see that as the age labels increasing, the synthetic results become older and older. Specifically, nasolabial folds and the lines at the sides of the eyes are deepened, white hair and beards emerge and gradually increase. Meanwhile, the age regression results on CACD2000 and Morph are shown in Fig. 7 and Fig. 9 respectively. The first column of each subject is the input beyond 51 years old. The second, third and fourth columns are the rejuvenating results of WaveletGLCA-GAN for 41-50, 31-40 and 14(16)-30 in turn. Obviously, with the age label decreasing, the white
beards and hair are gradually turned into black, dry skin begin to restore elasticity and wrinkles are reduced or even disappear.

In addition, the changes of the synthetic results are consecutive and consistent with the age label changes. Fig. 10 details the changes of synthetic results in the age rejuvenating process. In Fig. 10(a), the hairline gradually becomes lower, the hair darkens and thickens. Fig. 10(b) shows that with the age labels decreasing, the lines at the sides of the eyes get smoother, the eyes get bigger and brighter. In Fig. 10(c), the beards become less and turn black from white. Fig. 10(d) demonstrates the rejuvenating consistency of the WaveletGLCA-GAN. Besides, as is shown in Fig. 11, the WaveletGLCA-GAN is robust to pose, expression and illumination variations. For example, the occlusion, i.e., microphone, glasses, sunglasses, hat and makeup, is also well preserved in the age rejuvenating process. The proposed WaveletGLCA-GAN also has pretty generalization performance. Fig. 12 presents the face rejuvenating results with the sketch face images as the inputs.

2) Wavelet Packet Decomposition in Age Synthesis: Fig. 13 depicts the age progression results on Morph and its corresponding visual results of wavelet packet decomposition. The low-frequency component is related to the general face information, while the three high-frequency components are related to the detailed aging texture information. We observe that as the age increasing, the texture in both low-frequency and high-frequency components are more abundant, and the hair in the low-frequency component gradually becomes white from black.

3) Comparison with Prior Works: Different from the most previous works, which only focus on the age synthesis of cropped faces, our proposed WaveletGLCA-GAN can produce the entire face images, including complete foreground and background. We compare the aging results of WaveletGLCA-
Fig. 9. Age regression results on the Morph dataset for 27 different subjects. For each subject, the leftmost column shows the input face, while the rest three columns are synthetic faces from older to younger. (Zoom in for a better view.)

(a) Hair Rejuvenating  
(b) Eyes Rejuvenating  
(c) Mouth Rejuvenating  
(d) Half face Rejuvenating

Fig. 10. Illustration of visual fidelity. For each subject, the first column is the input, while the rest three columns are the rejuvenating results. (Zoom in for a better view.)

GAN with different methods, including HFA: hidden factor analysis [39], FT demo: Face Transformer demo [40], CDL: coupled dictionary learning [27], RFA: recurrent face aging [8], CAAE: conditional adversarial autoencoder [1], C-GAN: contextual generative adversarial nets [41], GLCA-GAN: global and local consistent age generative adversarial networks [9], Yang et.al [2] as well as a popular mobile aging application, i.e. AgingBooth [42]. Fig. 14 depicts the comparisons. Note that our WaveletGLCA-GAN is trained only on CACD2000 and tested on FG-NET. For fair comparison, we choose the same faces as their papers, and directly cite their synthetic results. As respected, WaveletGLCA-GAN obtains the best visual results with both backgrounds and foregrounds well preserved, while [39], [1], [40], [27], [8], [41] only generate cropped faces. Compared with previous best method [2], the synthetic images of our WaveletGLCA-GAN are clearer and have smaller color deviation, which...
C. Quantitative Evaluation of WaveletGLCA-GAN

Age accuracy and identity preserving are two underlying requirements for age synthesis. We conduct age estimation and face verification to quantitatively evaluate how much age information and identity information are predicted or preserved. Specifically, if the generated images can result in better aging accuracy, more precise age information is predicted. If the generated images result in better verification accuracy, more identity information is preserved during aging/rejuvenating process.

1) Age Estimation: Different subjects have different aging processes, but the overall trends of the synthetic faces should be robust and consistent with the real data. Following [2], we apply the online face analysis tool of Face++ [43] to estimate the ages of synthetic faces. Table III shows the age estimation results for different age groups on CACD2000. We choose faces under 30 years old (age group0, AG0) as the input test images, synthesize faces in 31-40 (AG1) 41-50 (AG2) and 51-62 (AG3), and calculate the average age estimated by [43]. As shown in Table III, the mean values of the natural faces for 31-40, 41-50 and 51-62 are 39.15, 47.14 and 53.87, respectively, while the mean values of the generated results of WaveletGLCA-GAN are 37.56, 48.13 and 54.17. In addition, we compare WaveletGLCA-GAN with CAAE [1], Yang et al. [2] and GLCA-GAN [9] on CACD2000, and observe that the results of WaveletGLCA-GAN are closer to the natural faces.

![Fig. 11. Robustness to occlusion, glasses, hat, makeup and pose. For each subject, the first column is the input, while the rest three columns are the rejuvenating results. (Zoom in for a better view.)](image)

![Fig. 12. Generalization to sketch images. For each subject, the first column is the input, while the rest three columns are the rejuvenating results. (Zoom in for a better view.)](image)

| Method         | AG1   | AG2   | AG3   |
|----------------|-------|-------|-------|
| CAAE [1]       | 31.32 | 34.94 | 36.91 |
| Yang et al. [2]| 44.29 | 48.34 | 52.02 |
| GLCA-GAN [9]   | 37.09 | 44.92 | 48.03 |
| WaveletGAN     | 39.12 | 44.66 | 47.18 |
| WaveletGLCA-GAN| 37.56 | 48.13 | 54.17 |
| Real Data      | 39.15 | 47.14 | 53.87 |

Meanwhile, Table IV presents the age estimation results of different methods on Morph. We choose faces under 30 years old (age group0, AG0) as the input test images, synthesize faces in 31-40 (AG1) 41-50 (AG2) and 51-77 (AG3), and calculate the average age estimated by [40]. The mean values of the natural faces for 31-40, 41-50 and 51-67 are 38.59, 48.24 and 58.28 respectively, while the mean values of the synthetic results of WaveletGLCA-GAN are 38.36, 46.90 and 59.14. In addition, we compare WaveletGLCA-GAN with CAAE [1], Yang et al. [2] and GLCA-GAN [9] on Morph, and observe that the age accuracy of WaveletGLCA-GAN outperforms other methods.

![Fig. 11. Robustness to occlusion, glasses, hat, makeup and pose. For each subject, the first column is the input, while the rest three columns are the rejuvenating results. (Zoom in for a better view.)](image)

| Method         | AG1   | AG2   | AG3   |
|----------------|-------|-------|-------|
| CAAE [1]       | 28.13 | 32.50 | 36.83 |
| Yang et al. [2]| 42.84 | 50.78 | 59.91 |
| GLCA-GAN [9]   | 43.00 | 49.03 | 54.60 |
| WaveletGAN     | 44.62 | 49.92 | 55.17 |
| WaveletGLCA-GAN| 38.36 | 46.90 | 59.14 |
| Real Data      | 38.59 | 48.24 | 58.28 |

2) Face Verification: We evaluate the identity preserving performance of WaveletGLCA-GAN by face verification for both age progression and regression. In the age progression setting, for each testing face, we evaluate the verification rate between the input image and its corresponding age synthetic result: [test face→AG1], [test face→AG2] and [test face→AG3]. Following [2], we also conduct face verification among the synthetic faces: [AG1→AG2], [AG2→AG3] and [AG3→AG1].
Fig. 13. Illustration of age progression and its corresponding wavelet packet decomposition. Top row: age progression results on the Morph dataset. For each subject, the leftmost column depicts the input face, while the rest three columns are synthetic results from younger to older. The second and bottom rows: the 1-level full wavelet packet decomposition and normalized 1-level full wavelet packet decomposition of the corresponding images in the same columns. (Zoom in for a better view.)

Fig. 14. Comparison with prior works FT demo [40], CDL [27], RFA [8], CAAE [1], C-GAN [41], AgingBooth App [42], GLCA-GAN [9], and Yang et.al [2]. (Zoom in for a better view.)

TABLE V
FACE VERIFICATION RESULTS(%) ON CACD2000 AND MORPH.

| Progression | CACD2000 | Morph |
|-------------|----------|-------|
| Test Face→AG1 | 97.71 | 99.94 |
| Test Face→AG2 | 96.07 | 99.92 |
| Test Face→AG3 | 95.25 | 99.12 |
| AG1→AG2 | 99.22 | 99.94 |
| AG1→AG3 | 96.01 | 99.68 |
| AG2→AG3 | 97.59 | 99.86 |

| Regression | CACD2000 | Morph |
|------------|----------|-------|
| Test Face→AG2 | 98.85 | 100 |
| Test Face→AG1 | 98.91 | 100 |
| Test Face→AG0 | 98.53 | 99.85 |
| AG2→AG1 | 99.62 | 100 |
| AG2→AG0 | 98.06 | 99.70 |
| AG1→AG0 | 99.00 | 100 |

[AG1→AG3] and [AG2→AG3]. Similar with age progression, in the age regression, we evaluate on [test face→AG2], [test face→AG1], [test face→AG0], [AG2→AG1], [AG2→AG0] and [AG1→AG0]. For CACD2000, we randomly choose negative pairs, and make positive pairs: negative pairs = 1 : 100. For Morph, we randomly choose negative pairs, and make positive pairs: negative pairs = 1 : 200, due to the lack of old age images (about 616 images of 51-77 in each fold). Light CNN [28] is employed as the feature extractor in our experiments. We report TAR@FAR=10^{-5} on CACD2000. While due to the lack of old age images in Morph, we only report TAR@FAR=10^{-5} for age progression and TAR@FAR=10^{-4} for age regression. Table V presents the face verification results.

D. Ablation Study

1) Contribution of Local Specific Networks: To prove the effectiveness of the proposed three local specific networks, we compare the age accuracy of WaveletGLCA-GAN with Wavelet-GAN (The model only use global specific network to synthesize target age faces and keep others consistent with WaveletGLCA-GAN). As Tables III and IV shown, WaveletGLCA-GAN is superior to Wavelet-GAN in age accuracy.

2) Contribution of Wavelet Coefficient Prediction Net: One assumption of WaveletGLCA-GAN is that the introduced of wavelet transform advances the generation of more subtle age-related texture information. Fig. 15 presents the visual comparisons between WaveletGLCA-GAN and GLCA-GAN.
have white-blonde hair, and there are also many facial images Besides in CACD2000, many facial images under 20 years old ties, whose faces usually look younger than their actual ages. only composed of images of European and American celebri-
cation/regression results, there is room for improvement. The 
E. Limitation and Discussion

Although our method achieves photo-realistic age progres-
sion/regression results, there is room for improvement. The 
hair color can be modified on the Morph database, while 
cannot on the CACD2000 database. Because the CACD2000 is 
only composed of images of European and American celebrit-
ies, whose faces usually look younger than their actual ages. 
Besides in CACD2000, many facial images under 20 years old 
have white-blonde hair, and there are also many facial images 
beyond 50 years old have colorful instead of white hair, which 
leads to an extreme difficulty for the model learning to change 
the hair color. Since the data in Morph is clean and our model 
has the ability to simulate the color changes of hair and beards. 
However, if the input face has thick black long hair as shown 
in Fig. [7] there sometimes appears blurry and white irregular 
areas in the age progression/regression results. The change of 
hair color is important to face age progression/regression 
process, thus we will continue to explore this in the further work.

Besides, most of the current age synthesis models mainly 
focus on the aging or rejuvenating between adult faces. How-
ever, it is equally important and more challenging to explore 
the aging or rejuvenating between adult and baby faces. In the 
future, we intend to model the continuous aging mechanism 
among all the ages.

VI. ACKNOWLEDGMENT

This paper is funded by State Key Development Pro-
gram (Grant No. 2017YFC0821602, 2016YFB1001001, 
2016YFB1001000), the National Natural Science Foundation 
of China (Grants No. 61622310, 61427811, 61573360), and 
Beijing Natural Science Foundation (Grants No. JQ18017).

REFERENCES

[1] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by con-
ditional adversarial autoencoder," in Proc. IEEE Conf. Comput. Vis. 
Pattern Recognit, Jun. 2017, pp. 4352–4360.
[2] H. Yang, D. Huang, Y. Wang, and A. K. Jain, “Learning face age 
progression: A pyramid architecture of gans,” Proc. IEEE Conf. Comput. 
Vis. Pattern Recognit, Jun. 2018.
[3] 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. IEEE Int. Conf. Comput. 
Vis., Oct. 2017, pp. 3755–3763.
[4] S. Zhou, W. Zhao, J. Feng, H. Lai, Y. Pan, J. Yin, and S. Yan, 
“Personalized and occupational-aware age progression by generative 
adversarial networks,” CoRR, vol. abs/1711.09368, 2017.
[5] Z. Wang, X. Tang, W. Luo, and S. Gao, “Face aging with identity-
preserved conditional generative adversarial networks,” in Proc. IEEE 
Conf. Comput. Vis. Pattern Recognit, Jun. 2018, pp. 7939–7947.
[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, 
S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in 
Proc. Adv. Neural Inf. Process. Syst., Dec. 2014, pp. 2672–2680.
Fig. 16. Aging results on Morph with different trade-off parameters of TV loss, including 0, 0.0001 and 0.001. The first column of each subject is the input image, while the rest columns are the age progression results with different trade-off parameters. (Zoom in for a better view.)

Fig. 17. Typical failure case of our method. For each image pair, the left shows the input face under 30 years old, and the right is the synthetic result beyond 50 years old. (Zoom in for a better view.)

[7] C. Nhan Duong, K. Luu, K. Gia Quach, and T. D. Bui, “Longitudinal face modeling via temporal deep restricted boltzmann machines,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2016, pp. 5772–5780.

[8] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, “Recurrent face aging,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2016, pp. 2378–2386.

[9] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, “Global and local consistent age generative adversarial networks,” Proc. IEEE Int. Conf. Pattern Recognit, Aug. 2018.

[10] H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet-srnet: A wavelet-based cnn for multi-scale face super-resolution,” in booktitle=Proc. IEEE Int. Conf. Comput. Vis., pages=1068–1079, Oct. 2017.

[11] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in Proc. Adv. Neural Inf. Process. Syst., Dec. 2017, pp. 700–708.

[12] M.-Y. Liu and O. Tuzel, “Coupled generative adversarial networks,” in Proc. Adv. Neural Inf. Process. Syst. Dec. 2016, pp. 469–477.

[13] H. Huang, R. He, Z. Sun, T. Tan et al., “Introvae: Introspective variational autoencoders for photographic image synthesis,” in Proc. Adv. Neural Inf. Process. Syst. Dec. 2018, pp. 52–63.

[14] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” pp. 5967–5976, Jun. 2016.

[15] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” Comput. Sci., pp. 2672–2680, 2014.

[16] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” Jun. 2018.

[17] 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., Oct. 2017, pp. 2242–2251.

[18] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun, “Pose-guided photorealistic face rotation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2018.

[19] H. Chang, J. Lu, F. Yu, and A. Finkelstein, “Pairdecyclegan: Asymmetric style transfer for applying and removing makeup,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2018.

[20] M.-H. Tsai, Y.-K. Liao, and I.-C. Lin, “Human face aging with guided prediction and detail synthesis,” Multimed. Tools Appl., vol. 72, no. 1, pp. 801–824, Sep. 2014.

[21] 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, Feb. 2009.

[22] J. T. Todd, L. S. Mark, R. E. Shaw, J. B. Pittenger et al., “The perception of human growth,” Sci. Am., vol. 242, no. 2, pp. 132–144, Feb. 1980.

[23] J. Suo, X. Chen, S. Shan, W. Gao, and Q. Dai, “A concatenational graph evolution aging model,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2083–2096, Jan. 2012.

[24] N. Ramanathan and R. Chellappa, “Modeling age progression in young faces,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, vol. 1, Jun. 2006, pp. 387–394.

[25] B. Tidemann, M. Burt, and D. Perrett, “Prototyping and transforming facial textures for perception research,” (IEEE Comput. Graph., vol. 21, no. 5, pp. 42–50, Oct. 2001.

[26] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, “Illumination-aware age progression,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2014, pp. 3334–3341.

[27] X. Shi, J. Tang, H. Lai, L. Liu, and S. Yan, “Personalized age progression with aging dictionary,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 3970–3978.

[28] X. Wu, R. He, Z. Sun, and T. Tan, “A light cnn for deep face representation with noisy labels,” IEEE Trans. Inf. Forensics Security, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Comput. Sci., 2014.

[30] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Jun. 2017, pp. 2242–2251.

[31] Z. Li, Y. Hu, M. Zhang, M. Xu, and R. He, “Protecting your faces: Meshfaces generation and removal via high-order relation-preserving cyclegan,” in Proc. Int. Conf. Biometrics, Feb. 2018, pp. 61–68.

[32] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” Comput. Sci., 2015.

[33] B.-C. Chen, C.-S. Chen, and W. H. Hsu, “Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset,” IEEE Trans. Multimedia, vol. 17, no. 6, pp. 804–815, Apr. 2015.

[34] K. Ricanek and T. Tesafaye, “Morph: A longitudinal image database of normal adult age-progression,” in Proc. IEEE Int. Conf. Automat. Face Gesture Recognit. IEEE, Apr. 2006, pp. 341–345.

[35] A. Lantinis, 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, Aug. 2002.

[36] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition,” in Proc. Eur. Conf. Comput. Vis., Oct. 2016, pp. 87–102.

[37] R. Rothe, R. Timofte, and L. Van Gool, “Deep expectation of real faces,” in Proc. Eur. Conf. Comput. Vis., vol. 24, no. 4, pp. 449–464, Oct. 2018.

[38] K. Zhang, Z. Zem, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal, Proc. Let., vol. 23, no. 10, pp. 1499–1503, Aug. 2016.

[39] H. Yang, D. Huang, Y. Wang, H. Wang, and Y. Tang, “Face aging effect of human growth,” Int J Comput Vis., vol. 126, no. 2-4, pp. 144–157, Apr. 2018.

[40] K. Zhang, Z. Zem, Z. Li, and Y. Qiao, “Joint face detection and alignment using hidden factor analysis joint sparse representation,” IEEE Trans. Image Process, vol. 25, no. 6, pp. 2493–2507, Mar. 2016.

[41] “Face transformer (ft) demo,” http://cherry.dcs.aber.ac.uk/transformer/, [Online].

[42] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan, “Face aging with contextual generative adversarial nets,” in Proc. ACM Int. Conf. Multimedia, Oct. 2017, pp. 82–90.

[43] “Agingbooth. pivi and co.” https://itunes.apple.com/us/app/ aging-booth/id357467791?mt=8/, [Online].

[44] “Face++ research toolkit. megvii inc.” http://www.faceplusplus.com [Online].