Age Progression and Regression with Spatial Attention Modules

Qi Li\textsuperscript{1,2*}, Yunfan Liu\textsuperscript{1*}, Zhenan Sun\textsuperscript{1,2†}

\textsuperscript{1}Center for Research on Intelligent Perception and Computing, CASIA
\textsuperscript{2}National Laboratory of Pattern Recognition, CASIA
\{qli, znsun\}@nlpr.ia.ac.cn, yunfan.liu@cripac.ia.ac.cn

Abstract

Age progression and regression refers to aesthetically rendering a given face image to present effects of face aging and rejuvenation, respectively. Although numerous studies have been conducted in this topic, there are still two major problems: 1) multiple models are usually trained to simulate different age mappings, and 2) the photo-realism of generated face images is heavily influenced by the variation of training images in terms of pose, illumination, and background. To address these issues, in this paper, we propose a framework based on conditional Generative Adversarial Networks (cGANs) to achieve age progression and regression simultaneously. Particularly, since face aging and rejuvenation are largely different in terms of image translation patterns, we model these two processes using two separate generators, each dedicated to one age changing process. In addition, we exploit the spatial attention mechanism to limit image modifications to regions closely related to age changes, so that images with high visual fidelity could be synthesized for in-the-wild cases. Experiments on multiple datasets demonstrate the ability of our model in synthesizing lifelike face images at desired ages with personalized features well preserved.

1 Introduction

Age progression and regression, also known as face aging and rejuvenation, aims at predicting the appearance of a given face at different ages. Its applications range from social security to digital entertainment, including face age editing and cross-age identification. Despite the appealing practical value, the lack of labeled age data of the same subject covering a large time span and the great change in appearance over a long time interval collectively make age progression and regression a difficult problem.

In the last two decades, many approaches have been proposed to tackle this issue, and they could be roughly divided into three categories: physical model-based methods, prototype-based methods, and deep learning-based methods. Physical model-based methods simulate the change of facial appearance over time by manipulating the parametric anatomical model of human faces [Todd et al., 1980; Lanitis et al., 2002; Tazoe et al., 2012]. However, these methods are usually computationally expensive and the complex aging rules do not generalize well. As for the prototype-based methods [Suo et al., 2010; Kemelmacher-Shlizerman et al., 2014], face images are firstly divided into several age groups and an average face is computed as the prototype for each group. After that, transition patterns between prototypes are learned and then applied to input faces to render effects of age changing. The main problem of prototype-based methods is that the identity information is not well preserved in the generation results, as personalized facial features are largely lost when computing the average faces.

Deep learning-based face aging methods have achieved state-of-the-art performance in recent years. With the success of Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] in generating visually appealing images, many efforts have been made to solve age progression and regression using GAN-based frameworks. In [Li et al., 2018] and [Yang et al., 2018], deep convolutional GANs are proposed to synthesize face images at given ages. However, in
these two works, models have to be trained repeatedly for different source or target ages, which heavily increases the computational cost. In both [Zhang et al., 2017] and [An-
tipov et al., 2017], different age mappings are modeled by a
single framework with the target age as a prior condition, but
no constraint is enforced to guarantee the target age fulfill-
ment. To solve this problem, [Song et al., 2018] integrates
the target age condition into discriminators to supervise the
apparent age of generated faces. Although face images with
obvious signs of age changing could be obtained, image con-
tents irrelevant to age changes (e.g. image background) are
not well preserved in the output, resulting in severe ghosting
artifacts.

To tackle the above-mentioned issues, in this paper, we
develop a conditional GAN based framework to solve age
progression and regression simultaneously. According to ex-
amples shown in Figure 1, the age progression process and
regression process are largely different from each other in
terms of image translation patterns. Therefore, unlike pre-
vious works, we propose to model these two processes using
two separate generators, each dedicated to one age changing
process. In addition, aging could be considered as adding
representative signs (e.g. wrinkles, eye bags, and laugh lines)
to the original input, while rejuvenation is to do the opposite.
That is to say, image modifications are supposed to be lim-
ited to those regions highly relevant to age changes. To this
end, the spatial attention mechanism is naturally adopted to
constrain image translations, and help to improve the quality
of generation results by minimizing the chance of introducing
distortions and ghosting artifacts. In brief, the main contribu-
tions of our work could be summarized as follows,

- We propose to solve age progression and regression us-
ing a unified conditional GAN based framework. Par-
nicularly, we employ a pair of generators to perform two
opposite tasks, face aging and rejuvenation, which take
face images and target age conditions as input and syn-
thesize photo-realistic age translated faces.

- The spatial attention mechanism is introduced to our
model to limit modifications to those regions that are rel-
vant to convey aging and rejuvenation, so that ghosting
artifacts could be suppressed and the quality of synthe-
sized images could be improved.

- Extensive experiments on three age databases are
conducted to comprehensively evaluate the proposed
method. Both qualitative and quantitative results
demonstrate the effectiveness of our model in accurately
synthesizing face images at desired ages with identity
information being well-preserved.

2 Method

2.1 Problem Formulation

Given a young face image $I_y$ at age $\alpha_y$, we aim to learn an
age progressor $G_p$ to realistically translate $I_y$ into an older
face image $I_o$ at age $\alpha_o$ ($\alpha_o > \alpha_y$), and an age regressor
$G_r$ to do the reverse. To be specific, $G_p$ takes an face im-
age $I_y$ and the target age condition $\alpha_o$ as input, and generates
the aged face image $I_o = G_p(I_y, \alpha_o)$. However, due to the
usage of unpaired aging data, the mapping $I_y \rightarrow I_o$ is highly
under-constrained and translation patterns other than age pro-
gression might be learned [Zhu et al., 2017].

To deal with this problem, an inverse mapping $I_o \rightarrow I_y'$ is
usually adopted to reconstruct the input, and the constraint
$I_o \approx I_y$ is enforced to regulate the mappings. It is worth not-
ning that, the inverse mapping is essentially an age regression
process, thus is supposed to be naturally accomplished by the
age regressor $G_r$, i.e., $I_y' = G_r(I_o, \alpha_o)$. Similarly, for face
rejuvenation, $G_r$ simulates the age regression process and $G_p$
serves as the inverse mapping. In this way, we integrate $G_r$
and $G_r$ into a single framework, which is a unified solution for
both age progression and regression.

The framework of the proposed model is illustrated in Fig-
ure 3. The training process contains two data flow cycles: an
age progression cycle and an age regression cycle. For the age
progression cycle, discriminator $D_p$ is employed to encour-
age the synthesized older face $G_p(I_y, \alpha_o)$ to be as realistic as
the real aged face $I_o$, and the estimated age of $G_p(I_y, \alpha_o)$ to
be close to the target age condition $\alpha_o$. Similar for $D_r$ in the
age regression cycle.

2.2 Network Architecture

In this section, we describe the architecture of the generator
and discriminator in detail. For brevity, in the following dis-
cussion, we collectively refer to $G_p$ and $G_r$, as $G$ if there is no
need to distinguish the translation directions, and similar for
discriminators $D_p$ and $D_r$, as $D$.

Spatial Attention based Generator: Since the age pro-
Figure 3: The framework of the proposed model. (a) $G_p$ and $G_r$ perform age progression and regression given the conditional age vector $\alpha_o$ and $\alpha_p$, respectively. Reconstruction loss is used to ensure that personalized features in the input image is preserved in the output. (b) $D_p$ and $D_r$ are discriminators designed to distinguish real images from synthetic ones and estimate the age of the input face image, and they are involved in the age progression cycle and regression cycle, respectively.

gressor $G_p$ and age regressor $G_r$ serve equivalent functions, they share the same network architecture. Therefore, we take $G_p$, for example to describe the detailed architecture, and $G_r$ is different only in terms of input and output. The structure of $G_p$ is shown in Figure 2.

Most of existing works on face age editing use generators with single pathway to predict the whole output image [Zhang et al., 2017], [Yang et al., 2018], [Song et al., 2018], where the divergence between data distributions of the entire image in the source and target age domain is minimized. Consequently, unintended correspondences between image contents other than age translation (e.g. background textures) would be inevitably established, which increases the chance of introducing age-irrelevant changes and ghosting artifacts (as shown in Figure 2(b)).

To solve this problem, we introduce the spatial attention mechanism into our framework via adding a branch to $G_p$, that estimates the attention mask describing the contribution of each pixel to the final output. To be specific, as shown in Figure 2(a), a fully convolutional network (FCN) $G^A_p$ is used to regress the attention mask, which is fused with the output of another FCN $G^R_p$ to produce the final output. Mathematically, this process could be formulated as:

$$ I_o = G^A_p(I_y, \alpha_o) \cdot I_y + (1 - G^A_p(I_y, \alpha_o)) \cdot G^R_p(I_y, \alpha_o) \; (1) $$

where $\alpha_o$ is the one-hot condition vector indicating the target age group, $G^A_p(I_y, \alpha_o) \in [0, 1]^{H \times W}$ is the attention mask and $G^R_p(I_y, \alpha_o) \in \mathbb{R}^{H \times W \times 3}$ models detailed translations within the attended regions. The greatest advantage of adopting attention mechanism is that the generator could focus only on rendering effects of age changes within specific regions, and irrelevant pixels could be directly retained from the original input, resulting in less distortion and finer image details.

**Discriminator:** Discriminator $D$ is trained to distinguish synthetic face images from real ones and check whether a generated face image belongs to the desired age group. The architecture of $D$ we use is similar to that in PatchGAN [Isola et al., 2017] which has achieved success in a number of image translation tasks [Zhu et al., 2017], [Pumarola et al., 2018]. Concretely, we use a series of six convolutional layers with increasing number of $4 \times 4$ filters, and each layer is followed by a Leaky ReLU unit. In addition, similar to [Odena et al., 2017], to check whether a synthetic image belongs to the age group represented by the corresponding target age condition, we append an auxiliary fully connected network to the top of $D$ to predict the age of the face image. Given an input image $I$, we denote the output of the convolutional layers by $D^1(I)$ and the result of age estimation by $D^{a}(I)$.

**2.3 Loss Function**

The loss function of the proposed model contains four parts: an adversarial loss to encourage the distribution of generated images to be indistinguishable from that of real images, a reconstruction loss to preserve personalized features, an attention activation loss to prevent saturation, and an age regression loss to measure the target age fulfillment.

**Adversarial Loss:** Adversarial loss describes the objective of a minimax two-player game between the generator $G$ and the discriminator $D$, where $D$ aims to classify real images from fake ones and $G$ attempts to fool $D$ with lifelike synthetic images. Unlike regular GANs [Goodfellow et al., 2014], least square adversarial loss [Mao et al., 2017] is employed in our model to improve the quality of generated images and stabilize the training process. Mathematically, the objective of adversarial loss could be formulated as follows,

$$ \mathcal{L}_{GAN} = \mathbb{E}_{I_y}[(D^f_p(G_p(I_y, \alpha_o)) - 1)^2] + \mathbb{E}_{I_y}[(D^f_p(I_y) - 1)^2] + \mathbb{E}_{I_y}[D^f_p(G_p(I_y, \alpha_o))^2] $$

$$ + \mathbb{E}_{I_y}[D^f_p(G_r(I_y, \alpha_y))^2] $$

$$ + \mathbb{E}_{I_y}[(D^f_r(G_r(I_y, \alpha_y) - 1)^2] + \mathbb{E}_{I_y}[D^f_r(G_r(I_y, \alpha_y))^2] $$

(2)

**Reconstruction Loss:** With the adversarial loss, $G$ learns to generate lifelike face images at the target age. However,
these is no guarantee that personalized features in the input image are preserved in the output since no ground-truth supervision is available. Therefore, a reconstruction loss is employed to penalize the difference between the input image and its reconstruction, which could be formulated as

\[
\mathcal{L}_{\text{recon}} = \mathbb{E}_{I_o} \left[ ||G_r(I_o, \alpha_o)) - I_o||_1 \right] + \mathbb{E}_{I_y} \left[ ||G_p(I_y, \alpha_y)) - I_y||_1 \right]
\] (3)

Here we use the L1-norm to encourage less blurred results.

**Attention Activation Loss:** In equation (2) and (3), simply setting \(G_p\) and \(G_r\) to identity mapping would minimize these loss terms, which is definitely not what we expected. In this case, as shown in equation (1), all elements in \(G_y(I_y, \alpha_y)\) saturate to 1 and thus \(I_y = I_o\). To prevent this from happening, an attention activation loss is used to constrain the total activation of the attention mask, which could be written as

\[
\mathcal{L}_{\text{actv}} = \mathbb{E}_{I_y} \left[ ||G_y(I_y, \alpha_y)||_2 \right] + \mathbb{E}_{I_o} \left[ ||G_y(I_o, \alpha_o)||_2 \right]
\] (4)

**Age Regression Loss:** Apart from being photo-realistic, synthesized face images are also expected to satisfy the target age condition. Therefore, an age regression loss is adopted to force generators to reduce the error between estimated ages and target ages, which could be expressed as

\[
\mathcal{L}_{\text{reg}} = \mathbb{E}_{I_y} \left[ ||D^p_g(G_p(I_y, \alpha_y)) - \alpha_o||_2 \right] + \mathbb{E}_{I_o} \left[ ||D^p_g(G_p(I_o, \alpha_o)) - \alpha_y||_2 \right] + \mathbb{E}_{I_y} \left[ ||D^r_g(G_r(I_y, \alpha_y)) - \alpha_y||_2 \right] + \mathbb{E}_{I_o} \left[ ||D^r_g(G_r(I_o, \alpha_o)) - \alpha_o||_2 \right]
\] (5)

By optimizing equation (5), the auxiliary regression network \(D^y\) gains the age estimation ability, and the generator \(G\) is encouraged to accurately render fake faces at desired ages.

**Overall Loss:** The final full loss function of the proposed model could be formulated as the linear combination of all previously defined losses:

\[
\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{\text{recon}} \mathcal{L}_{\text{recon}} + \lambda_{\text{actv}} \mathcal{L}_{\text{actv}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}}
\] (6)

where \(\lambda_{\text{recon}}\), \(\lambda_{\text{actv}}\), and \(\lambda_{\text{reg}}\) are coefficients balancing the relative importance of each loss term. Finally, \(G_p\), \(G_r\), \(D_p\), and \(D_r\) are solved by optimizing:

\[
\min_{G_p, G_r, D_p, D_r} \max_{\lambda_{\text{recon}}, \lambda_{\text{actv}}, \lambda_{\text{reg}}} \mathcal{L}
\] (7)

### 3 Experiments

#### 3.1 Datasets

Three publicly available face aging datasets, Morph [Ricanek and Tesafaye, 2006], CACD [Chen et al., 2015], and UTKFace [Zhang et al., 2017] are used in our experiments. Following [Yang et al., 2018] and [Li et al., 2018], we divide
images in Morph and CACD into 4 age groups (30+, 31-40, 41-50, 51+), and UTKFace into 9 age groups (0-3, 4-11, 12-17, 18-29, 30-40, 41-55, 56-65, 66-80, and 81-116) according to Song et al. For each dataset, we randomly select 80% images for training and the rest for testing, and ensure that these two sets do not share images of the same subject.

### 3.2 Implementation Details

All face images are aligned according to both eyes and then resized to 256 × 256. We train the model for 30 epochs with batchsize of 24, using the Adam optimizer with learning rate set to 1e-4. Optimization over generators is performed every 5 iterations of discriminators. As for the balancing hyper-parameters $\lambda_{reg}$, $\lambda_{actv}$, and $\lambda_{actv}$, we first initialize them to make all losses to be of the same order of magnitude as the $L_{GAN}$, then divide them by 10 except for $\lambda_{reg}$ to emphasize the importance of accurate age simulation.

### 3.3 Qualitative Results

Sample results of age progression and regression are shown in Figure 4. Although input faces cover a wide range of populations in terms of age, race, gender, pose and expression, the model successfully renders photo-realistic and diverse age changing effects. In addition, it could be observed that identity permanence is well-achieved in all generated face images.

Figure 5 displays the attention masks for sample generation results. Note that the network has learned to focus its attention onto face regions most relevant to representative signs of age changes (e.g., wrinkles, laugh lines, mustache) in an unsupervised manner, and leave other parts of the image unattended. Figure 5(b) shows how attention maps help to deal with occlusions and complex background, that is, by assigning lower attention scores to pixels in these regions. This allows pixels in unattended areas to be directly copied from the original input image, which improves the visual quality of generated face images, especially for in-the-wild cases. The proposed method is highly scalable as it could be naturally extended to different age span and age group divisions, as shown in Figure 5(c).

To demonstrate the effectiveness of our model, we compare the proposed method with several benchmark approaches: CONGRE [Suo et al., 2012], HFA [Yang et al., 2016], GLCA-GAN [Li et al., 2018], Pyramid-architected GAN (referred to as PAG-GAN) [Yang et al., 2018], IPCGAN [Wang et al., 2018], CAAE [Zhang et al., 2017], and Dual cGANs [Song et al., 2018]. Results are shown in Figure 6. It is clear that traditional face aging methods, CONGRE and HFA, only render subtle aging effects within tight facial area, while our method could simulate the aging process on the entire face image. As for GAN-based methods, GLCA-GAN, PAG-GAN, and IPCGAN, our model is better at suppressing ghosting artifacts and color distortion as well as rendering enhanced aging details. This is because the attention module enables the model to retain pixels in areas irrelevant to age changes instead of re-estimating them, which avoids introducing additional noise and distortions. This is also confirmed by the comparison between our method and Dual cGANs, as background, hair regions, and face boundaries are better maintained in the results of our model.

### 3.4 Quantitative Evaluations

In this subsection, we report quantitative evaluation results on age translation accuracy and identity preservation. For age translation accuracy, we calculate the error between estimated ages of real and fake face images, and for identity preservation, face verification rates are reported along with verification scores. The threshold is set to 76.5@FAR=1e-5 for all identity preservation experiments according to the protocol of Face++ API. For the sake of fairness, we compare our method with state-of-the-art approaches CAAE, IPCGAN, and Dual cGANs, which all attempt to solve age progression and regression via a single unified framework. To be objective, all metrics are estimated by the publicly available online face analysis tools of Face++ so that results are more objective and reproducible compared to those obtained by user study.

According to results shown in Table 1, our method achieves the best performance in both age translation accuracy and identity preservation on all three datasets, and outperforms other methods by a clear margin. CAAE produces over-identity preservation on all three datasets, and outperforms other methods by a clear margin. CAAE produces over-
especially for in-the-wild images in CACD and UTKFace. In addition, compared to IPCGAN, our method generates images with higher resolution (256 × 256 vs. 128 × 128) and do not require pre-trained age classifier and identity feature extractor, which significantly simplifies the training process. The major difference between our method and Dual cGANs is the adoption of spatial attention modules, and this explains why our method outperforms Dual cGANs under both metrics, which again demonstrates the effectiveness of attention mechanism in improving the quality of generated images.

3.5 Ablation Study

Experiments are conducted to fully explore the contributions of adopting attention modules (ATT) and using ordered input (OI) in simulating age translations. According to results shown in Table 1, either removing attention modules or using unordered training pairs would cause performance drops under both metrics, which demonstrate that both ATT and OI are critical to reach the optimal performance.

In addition, it could be observed that the advantage brought by ATT and OI increases as it goes from controlled training samples (Morph) to in-the-wild face images (CACD and UTKFace), and from relatively concentrated face attributes (white and black people of 20 to 60 years old in Morph) to more diverse data distributions (face images of up to 5 races covering an larger age span in CACD and UTKFace). This is mainly because that attention modules save the network from being interfered by variations irrelevant to age changes, which brings more advantages to in-the-wild cases than controlled ones. In addition, arranging input training samples by age enables the two generators to focus only on one direction of age change, which facilitates the convergence of the overall model and improve the quality of generation results.

3.6 Generalization Ability Study

We evaluate the generalization ability of our method by applying the model trained on CACD to images in FG-NET [Lanitis et al., 2002], CelebA [Liu et al., 2015], and IMDB-WIKI [Rothe et al., 2016]. For those images without age labels, apparent ages are detected using the Face++ API. Sample results shown in Figure 7 demonstrate that our model generalizes well to face images with different data distributions, and accurate age labels are not strictly required.

| Morph | CACD | UTKFace |
|-------|------|---------|
| Age Est. Error | Veri. Rate (%) | Age Est. Error | Veri. Rate (%) | Age Est. Error | Veri. Rate (%) |
| CAAE | 10.34 ± 5.63 | 34.83 (71.75) | 5.16 ± 7.08 | 3.59 (59.90) | 11.64 ± 10.41 | 8.07 (59.40) |
| IPCGAN | 1.74 ± 7.44 | 99.86 (94.04) | 8.11 ± 9.69 | 99.19 (91.60) | 7.51 ± 11.66 | 97.32 (92.63) |
| Dual cGANs | 2.44 ± 6.03 | 99.99 (93.15) | 3.28 ± 8.01 | 99.88 (93.85) | 7.01 ± 11.29 | 92.29 (87.73) |
| w/o OI | 1.58 ± 6.50 | 100.00 (95.48) | 1.92 ± 9.36 | 99.91 (96.30) | 5.49 ± 11.48 | 97.08 (92.38) |
| w/o ATT | 1.53 ± 6.52 | 100.00 (95.50) | 1.89 ± 9.07 | 99.92 (96.52) | 6.81 ± 12.29 | 97.24 (94.47) |
| Ours | 1.53 ± 6.50 | 100.00 (95.67) | 1.78 ± 7.53 | 99.92 (96.13) | 4.77 ± 10.59 | 98.10 (92.74) |

Table 1: Comparison of quantitative measurements, age estimation error (Age Est. Error) and face verification rate (Veri. Rate) on three datasets. Due to the limited space, we only report the mean value and standard deviation of age estimation error computed over all age groups. Verification scores are shown in brackets to provide more information for comprehensive comparisons between different methods.
cessful identity preservation, especially for in-the-wild cases.

References

[Antipov et al., 2017] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. *IEEE International Conference on Image Processing (ICIP)*, pages 2089–2093, 2017.

[Chen et al., 2015] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia (TMM)*, 17(6):804–815, 2015.

[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 *Advances in neural information processing systems (NIPS)*, pages 2672–2680, 2014.

[Isola et al., 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, 2017.

[Kemelmacher-Shlizerman et al., 2014] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz. Illumination-aware age progression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (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 (TPAMI)*, 24(4):442–455, 2002.

[Li et al., 2018] Peipei Li, Yibo Hu, Qi Li, Ran He, and Zhenan Sun. Global and local consistent age generative adversarial networks. In *International Conference on Pattern Recognition (ICPR)*, 2018.

[Liu et al., 2015] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2672–2680, 2015.

[Mao et al., 2017] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[Odena et al., 2017] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 2642–2651, 2017.

[Pumarola et al., 2018] A. Pumarola, A. Agudo, A.M. Martinez, A. Sanfeliu, and F. Moreno-Noguer. Ganimation: Anatomically-aware facial animation from a single image. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[Ricanek and Tesafaye, 2006] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *International Conference on Automatic Face and Gesture Recognition (FG)*, pages 341–345, 2006.

[Rothe et al., 2016] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*, 2016.

[Song et al., 2018] Jingkuan Song, Jingqiu Zhang, Lianli Gao, Xianglong Liu, and Heng Tao Shen. Dual conditional gans for face aging and rejuvenation. In *Proceedings of the Twenty-Seven International Joint Conference on Artificial Intelligence*, pages 899–905, 2018.

[Suo et al., 2010] J. Suo, S. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(3):385–401, 2010.

[Suo et al., 2012] Jini Suo, Xilin Chen, Shiguang Shan, Wen Gao, and Qionghai Dai. A concatenational graph evolution aging model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 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 *ACM SIGGRAPH 2012 Posters*, page 90, 2012.

[Todd et al., 1980] James T Todd, Leonard S Mark, Robert E Shaw, and John B Pittenger. The perception of human growth. *Scientific American*, 242(2):132–144, 1980.

[Wang et al., 2018] Z. Wang, W. Luo X. Tang, and S. Gao. Face aging with identity-preserved conditional generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[Yang et al., 2016] Hongyu Yang, Di Huang, Yunhong Wang, Heng Wang, and Yuanyan Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Transactions on Image Processing (TIP)*, 25(6):2493–2507, 2016.

[Yang et al., 2018] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. Learning face age progression: A pyramid architecture of gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 31–39, 2018.

[Zhang et al., 2017] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5810–5818, 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. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.