Continuous Face Aging via Self-estimated Residual Age Embedding

Zeqi Li*
ModiFace
lizeqi@cs.toronto.edu

Ruowei Jiang
ModiFace
irene@modiface.com

Parham Aarabi
ModiFace
parham@modiface.com

Abstract

Face synthesis, including face aging, in particular, has been one of the major topics that witnessed a substantial improvement in image fidelity by using generative adversarial networks (GANs). Most existing face aging approaches divide the dataset into several age groups and leverage group-based training strategies, which lacks the ability to provide fine-controlled continuous aging synthesis in nature. In this work, we propose a unified network structure that embeds a linear age estimator into a GAN-based model, where the embedded age estimator is trained jointly with the encoder and decoder to estimate the age of a face image and provide a personalized target age embedding for age progression/regression. The personalized target age embedding is synthesized by incorporating both personalized residual age embedding of the current age and exemplar-face aging basis of the target age, where all preceding aging bases are derived from the learned weights of the linear age estimator. This formulation brings the unified perspective of estimating the age and generating personalized aged face, where self-estimated age embeddings can be learned for every single age. The qualitative and quantitative evaluations on different datasets further demonstrate the significant improvement in the continuous face aging aspect over the state-of-the-art.

1. Introduction

Face aging, also known as age progression, aims to aesthetically render input face images with natural aging and rejuvenating effects while preserving identity information of the individual. With recent advances in deep learning, face synthesis has also shown substantial improvement on image fidelity and the age precision in the simulated face images [10, 44, 25]. A major challenge to solve a variety of remaining problems (e.g. continuous aging) is the lack of data. For example, many research works of face aging [20, 44, 46, 10] need to group images into 4-5 age groups (such as <30, 30-40, 40-50, 50+) and can only generate images within a target age group, due to the limited amount of data at each age. Another important problem is how to maintain personal traits in age progression, as aging patterns may differ for each individual.

Traditional face aging contains mainly two approaches: physical model-based [3, 45] and prototype-based [39, 16]. The physical model-based methods often consist of complex physical modeling, considering skin wrinkles, face shape, muscle changes, and hair color, etc. This type of method typically requires a tremendous amount of data and is very expensive computationally. Prototype-based methods firstly explore group-based designs by computing an average face within the pre-defined age groups, which fails to retain personalized aging information. Further, all those methods are not applicable to continuous face aging.

Following the success of recent generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs) [9], on the image translation tasks, researchers have dedicated efforts in adapting those methods to face synthesis. IPCGAN [44] has shown significant progress in generating face images with evident aging effects by enforcing an age estimation loss. Later variation [46] creates a pyramid structure for the discriminator to improve face aging understanding at multiple scales. Continuous aging was not explored among these methods. He et al. [10] introduced a multi-branch generator for the group-based training and proposed the idea to approximate continuous aging via linear interpolation of latent representations between two adjacent age groups. The authors of [25] also tackle the problem using a similar linear interpolation approach, which is performed on the learned age latent code between two neighboring groups instead. These types of methods make an assumption that the age progression is linear between the two adjacent groups and the learned group embedding can be used directly as the median age embedding. Consequently, this may result in a shift of target age in the generated images. Intuitively, this nonlinearity can be interpreted as: people do not age at the same speed for different stages. Moreover, such interpolation-based methods may alter personal traits when disentanglement is imperfect.

---

*This work is done during Zeqi Li's full-time employment at ModiFace.

To address the aforementioned problems, we propose a novel approach to achieve continuous aging by a unified network where a simple age estimator is embedded into a regular encoder-decoder architecture. This allows the network to learn self-estimated age embeddings of all ages, thus representing the continuous aging information without manual efforts in selecting proper anchor age groups. Given a target age, we derive a personalized age embedding which considers two aspects of face aging: 1) a personalized residual age embedding at the current age, which preserves the individual's aging information; 2) exemplar-face aging basis at the target age, which encodes the shared aging patterns among the entire population. We describe the detailed calculation and training mechanism in **Method**. The calculated target age embedding is then used for final image generation. We experiment extensively on FFHQ [15] and CACD2000 [5] datasets. Our results, both qualitatively and quantitatively, show significant improvement over the state-of-the-art in various aspects. Our main contributions are:

- We propose a novel method to self-estimate continuous age embeddings and derive personalized age embeddings for face aging task by jointly training an age estimator with the generator. We quantitatively and qualitatively demonstrate that the generated images better preserve the personalized information, achieve more accurate aging control, and present more fine-grained aging details.

- We show that our continuous aging approach generates images with more well-aligned target ages, and better preserves detailed personal traits, without manual efforts to define proper age groups.

- Our proposed idea to self-estimate personalized age embedding from a related discriminative model can be easily applied to other conditional image-to-image translation tasks, without introducing extra complexity. In particular, tasks involving a continuous condition and modeling (e.g. non-smile to smile), can benefit from this setup.

2. Related Work

2.1. Face Aging Model

Traditional methods can be categorized as physical model-based approaches [3, 45, 36] and prototype-based approaches [32, 39, 16, 17]. The physical model-based methods focuses on creating models to address specific sub-effects of aging, such as skin wrinkles [45, 2, 3], cranio-facial growth [40, 28], muscle structure [36, 29], and face components [37, 38]. These methods are often very complicated, which typically require a sequence of face images of the same person at different ages and expert knowledge

of the aging mechanism. The prototype-based approaches [32, 39, 4] explore face progression problem using group-based learning where an average face is estimated within each age group. However, personalized aging patterns and identity information are not well-preserved in such strategies. In [43, 47, 35], sparse representation of the input image have been utilized to express personalized face transformation patterns. Though the personalized aging patterns are preserved to some extent by such approaches, the synthesized images suffer from quality issues.

Recently, deep learning approaches have been adopted to model personalized aging transformations. Wang et al. [42] proposed a recurrent neural network model, leveraging a series of recurrent forward passes for a more smooth transition from young to old. Later GAN-based works [18, 44, 46] have shown superior breakthroughs on the fidelity of images. Li et al. [18] designed three subnets for local patches and fused local and global features to obtain a smooth synthesized image. IPCGAN [44] enforces an age estimation loss on the generated image and an identity loss to achieve good face aging effects. More efforts have also been made to address age accuracy and identity permanence. Yang et al.[46] and Liu et al. [20] introduce a modification of discriminator losses to guide a more accurate age of the output images. Authors of [21] improved image quality of synthesized images by using a wavelet packet transformation and multiple facial attribute encoding. However, these methods [44, 46, 20] condition the output image by concatenating one-hot vector representing the target age groups. To obtain a continuous aging condition, the vector will be extended to a much larger dimension, which makes training unstable and more complicated. Furthermore, it requires a tremendous amount of training images.

Though some works [49, 1, 34], which aim to interpolate features in the latent space, provided a direction to support continuous aging, they have limited ability to produce high-quality images while preserving the identity. In [10], the authors proposed to linear interpolate feature vectors from adjacent age groups upon group-based training to achieve continuous aging progression. Similarly, [25] linearly interpolates between two adjacent anchor age embeddings. These methods follow the assumption that the embeddings are aligned linearly between anchors, which makes the decision of anchor ages crucial. In this work, we present continuous self-estimated age embeddings free of manual efforts while achieving better continuous age modeling.

2.2. Generative Adversarial Networks

Generative adversarial networks [9] have been a popular choice on image-to-image translations tasks. CycleGAN [50] and Pix2Pix [14] explored image translations between two domains using unpaired and paired training samples respectively. More recent works [6, 19] proposed training
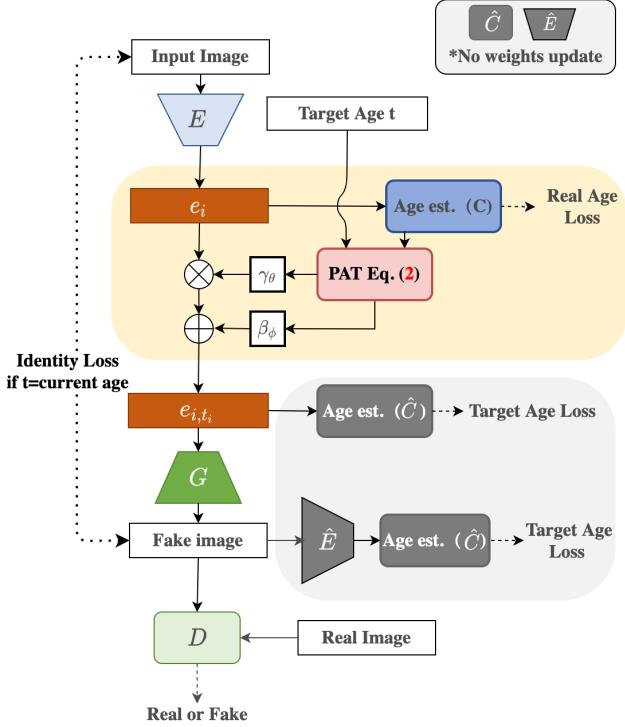
Figure 1. Model architecture: An age estimator is jointly trained with an image generator, where **E** is the shared encoder and **C** is branched off for the age estimation task. The personalized age embedding transformation (**PAT**, **Eq. (2)**) is based on two components: 1) residual aging basis at the current age; 2) exemplar-face aging basis at the target age. Then the transformed identity encoding is decoded by **G**. The whole model is learned with the age losses, identity loss, and the adversarial loss.

2.3. Face Age Estimation

The task to predict apparent age refers to the regression problem that estimates a continuous numerical value for each given face image. Deep Expectation of Apparent Age (DEX) [31] proposed a method to achieve a MAE of 3.25 on MORPH II [30], by combining classification loss and regression loss. Pan et al. [26] proposed to use mean-variance loss on the probability distribution to further improve the MAE to 2.16 on MORPH II.

techniques to enable multi-domain translation. In [22, 24], authors firstly explored conditional image generation as extensions to basic GANs. Later works [7, 27] have further shown superiority on many conditional image translation tasks, by transforming and injecting the condition into the model in a more effective manner.

3. Method

3.1. Formulation

As shown in **Fig. 1**, our model consists of four components: 1) identity encoding module **E**; 2) age estimation module **C**; 3) personalized age embedding transformation module **PAT**; 4) aged face generation module **G**. During inference, we apply an encoder network **E** to extract the identity information from the given image x_i, where the encoding is denoted as $e_i = \mathbf{E}(x_i)$. Then an embedded age estimator **C** is used to obtain the age probability distribution of the identity encoding. Based on the self-estimated age distribution and the target age t, we apply a personalized age embedding transformation **PAT** on the identity encoding e_i. Lastly, the synthesized face is decoded from the transformed identity encoding $\mathbf{PAT}(e_i, t)$ by the generator **G**. All modules are optimized jointly end-to-end under three objectives including the mean-variance age loss [26] for accurate aging, the $L1$ reconstruction loss for identity preservation, and the adversarial loss for image realism. Unlike many prior face aging works [44, 10] in which require a pre-trained age classifier to guide the face aging training, our model directly obtains a self-estimated age embedding by utilizing a unified framework for achieving face aging and age estimation at the same time. More favorably, the embedded age estimator not only enables personalized continuous age transformation in a more accurate manner, compared to the interpolation-based approach, but also provides the guidance for face image generation.

Identity Age Estimation Module (C) In prior works [44, 10], face aging and face age estimation are treated as two independent tasks where an age estimation model, usually a classifier, is pre-trained separately and then used to guide the generator to realize natural aging effects. As the two mentioned tasks are intrinsically related, both goals can be achieved with one unified structure by sharing an encoder **E**. The age estimator **C**, in our case containing a global average pooling layer and a fully-connected layer, is branched off from **E**. Finally, the age probability distribution $p_i \in R^K$ can be obtained by performing the softmax function, where K denotes the number of age classes. Without introducing too much extra complexity, such unified design also provides three advantages. Firstly, it eliminates the need to acquire a well-trained age estimator model beforehand. Secondly, age estimation on the identity encoding helps the model to establish a more age-specific identity representation. Thirdly, the weight W_C in the fully-connected layer is also used as the age embedding bases (bias terms are set to zero) which encodes the exemplar-face information from a metric learning perspective. In notation:

$$a_j = W_C[j], \tag{1}$$

where $W_C \in \mathbb{R}^{K \times D}, a_j \in \mathbb{R}^D$ and D equals to the channel

dimension of the identity encoding.

Personalized Age embedding Transformation (PAT)
Face aging is a challenging and ambiguous task in nature as different facial signs/symptoms ages differently for different people at different stages. Thus, personalization is desired in performing face aging. In our design, we characterize this personalization by a residual age embedding calculated from the age probability distribution $p_i \in \mathbb{R}^K$ and the exemplar-face aging basis $a_j \in \mathbb{R}^D$ where i denotes the sample i and $j \in 1, 2, \ldots, K$ denotes the age. $p_{i,j} \in R$ is the probability at age j for sample i. To obtain the personalized aging basis for any target age t_i, we formulate the process as the following operation:

$$\tilde{a}_{i,t_i} = (\sum_{j=1}^{K} p_{i,j} a_j - a_{j=[m_i]}) + a_{j=t_i}, \quad (2)$$

The $\sum_{j=1}^{K} p_{i,j} a_j$ term represents the personalized aging basis of the identity by taking the expected value of the aging basis based on the age probability distribution. Then we can obtain the residual age embedding by subtracting the exemplar-face aging basis at the current (self-estimated) age $a_{j=[m_i]}$ from the personalized aging basis. The residual age embedding preserves the identity's personalized factors while removing the prevailing aging factors at the self-estimated age. The final personalized target age embedding $\tilde{a}_{i,t_i}$ is obtained by adding the exemplar-face aging basis at the target aging basis $a_{j=t_i}$, which encodes the shared aging factors at the target age among the entire population. With the personalized target age embedding $\tilde{a}_{i,t_i}$, we then apply an affine projection transformation to derive the scale and shift coefficients for the original identity encoding $E(x_i) = e_i$, similar to Conditional BN [8] and AdaIN [13]:

$$\textbf{PAT}(e_i, t_i) = e_{i,t_i} = \gamma_\theta(\tilde{a}_{i,t_i}) e_i + \beta_\phi(\tilde{a}_{i,t_i}), \quad (3)$$

In our experiments, we do not observe significant performance difference w/wo $\beta_\phi(\tilde{a}_{i,t_i})$.

Continuous Aging As the aging bases from the fully-connected layer encode every single age, any integer target age is naturally supported. While some previous group-based approaches only model a few anchor age groups and achieving continuous aging via linear interpolation in the latent space. Our proposed method, however, explicitly models a fine-controlled age progression for each age and also supports float target age via a weighted sum of 2 neighboring integer age embedding bases.

3.2. Objective

The design of the objectives ensures the synthesized face image reflects accurate age progression/regression, preserves the identity, and looks realistic.

Mean-Variance Age Loss The age loss plays two roles in our network: 1) it helps the estimator learn good aging bases for all ages; 2) it guides the generator by estimating the age of the generated fake images. To achieve both goals, we adopt the mean-variance age loss proposed by [26]. Given an input image x_i and an age label y_i, the mean-variance loss is defined as below:

$$\begin{aligned} \mathbf{L}_{mv} &= L_s + \lambda_{mv1} L_m + \lambda_{mv2} L_v \\ &= \frac{1}{N} \sum_{i=1}^{N} -log p_{i,y_i} + \frac{\lambda_1}{2}(m_i - y_i)^2 + \lambda_2 v_i, \end{aligned} \quad (4)$$

where $m_i = \sum_{j=1}^{K} j p_{i,j}$ is the mean of the distribution and $v_i = \sum_{j=1}^{K} p_{i,j} * (j - m_i)^2$ is the variance of the distribution.

In addition to being more effective than other losses on the age estimation task, mean-variance loss also satisfies our needs to learn a relatively concentrated age distribution while capturing the age continuity for the adjacent aging bases. The supervised age loss is formulated as below:

$$\mathbf{L}_{real} = L_{mv}(C(E(x)), y), \quad (5)$$

For guiding face aging, we apply the embedded age estimator at both the transformed identity encoding level and the generated image level (as shown in **Fig.** 1).

$$\begin{aligned} \mathbf{L}_{fake} &= \lambda_{fake1} L_{mv}(\hat{C}(PAT(E(x), t)), t) + \\ &\quad \lambda_{fake2} L_{mv}(\hat{C}(\hat{E}(G(PAT(E(x), t)))), t), \end{aligned} \quad (6)$$

When the age estimator $\hat{\mathbf{C}}$ and encoder $\hat{\mathbf{E}}$ are used on the transformed identity encodings and fake images, their weights are not updated during backpropagation.

L1 Reconstruction Loss Another important aspect is to preserve the identity of the individual. We apply $L1$ pixel-wise reconstruction loss on the synthesized face by setting the target age to its self-estimated age. Specifically, it is formulated as below:

$$\mathbf{L}_{idt} = \frac{1}{N} \sum_{i}^{N} ||G(PAT(E(x_i), m_i)) - x_i||_1, \quad (7)$$

We have also experimented with a cycle-consistency loss as proposed in StarGAN [6] to enforce the identity criteria but found that the pixel-wise $L1$ reconstruction loss is sufficient to achieve the goal without extensive efforts in tuning the hyper-parameters.

Adversarial Loss To produce high fidelity images, we apply GAN loss in the unconditional adversarial training
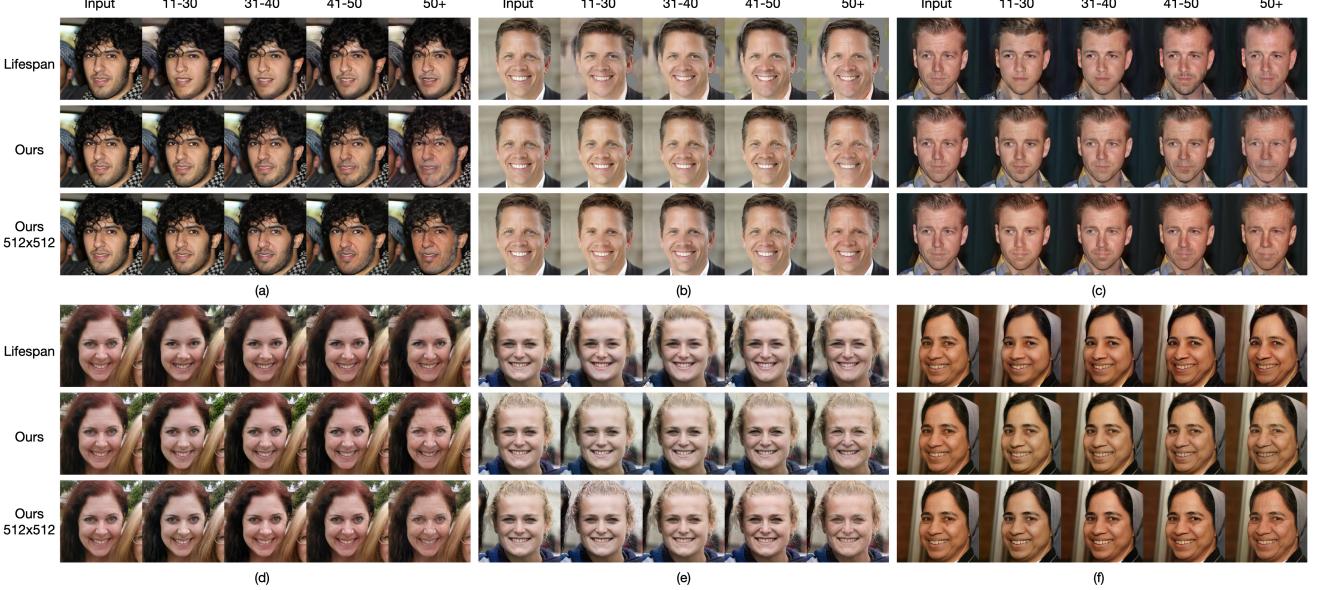
Figure 2. Comparisons on FFHQ [15] among Lifespan[25], ours and ours (512x512). Lifespan does not have an explicit age group at 11-30 and 41-50 so images for these 2 groups are generated using linear interpolation between 2 neighboring anchor classes. As shown, our generated images provide more aging details, such as skin wrinkles and color of the beard, on different parts of the face. In the example (f) in particular, both of our models well preserve her personal traits (a mole), comparing to the Lifespan model.

Figure 3. Comparisons on CACD2000 [5] among CAAE[49] IPCGAN [44], S^{2}GAN [10] and ours. The input images are wrapped in red boxes.
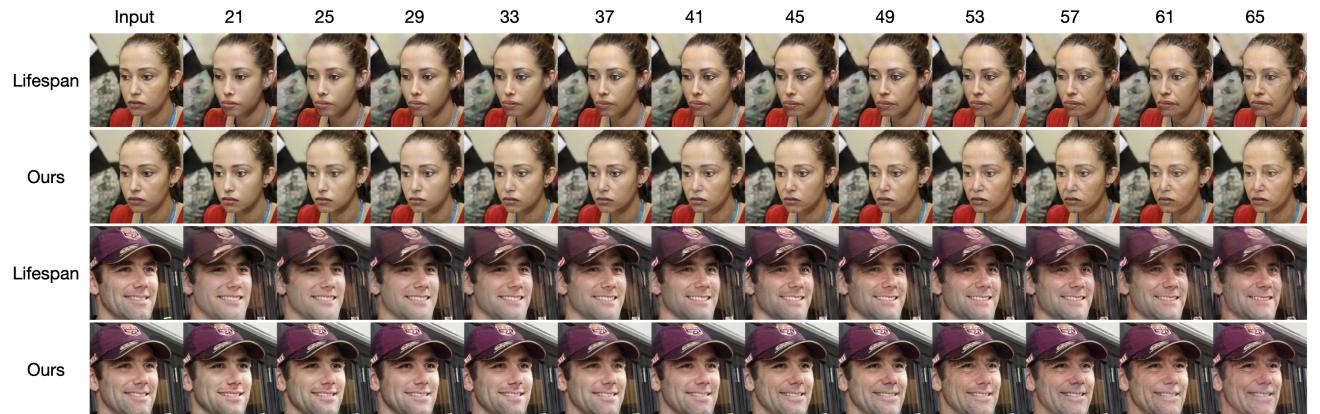

Figure 4. Continuous aging from 21 to 65. The age gap is chosen as 4 due to the limited space. As shown, the linear interpolation-based method used by Lifespan [25], some personal traits are altered (such as mouth shape, beard, hats). Further, our method generates more realistic aging effects with minimal artifacts. Continuous aging of 1-year incremental change in Supplementary.

manner. More specifically, we adopt PatchGAN [14] discriminator and optimize on the hinge loss, formulated as the following:

$$\mathbf{L}_{adv-D} = E_{z \sim p_{data}(z)}[max(1 - D(z), 0)] + \\ E_{(x,t) \sim p_{data}(x)}[max(1 + D(G(PAT(E(x), t))), 0)], \quad (8)$$

where we denote the data distribution as $x \sim p_{data}(x)$ and $z \sim p_{data}(z)$.

$$\mathbf{L}_{adv-G} = E_{(x,t) \sim p_{data}(x)}[-D(G(PAT(E(x), t)))], \quad (9)$$

In the experiment, we observe that sampling real examples of the age equal or close to the target age t_i for training the discriminator helps to stabilize the learning process.

All objectives are optimized jointly with different balancing coefficients as the following:

$$\min_{E,C,PAT,G} \lambda_{age}(L_{real} + L_{fake}) + \lambda_{idt} L_{idt} + \lambda_{adv} L_{adv-G}, \quad (10)$$

$$\min_{D} L_{adv-D}, \quad (11)$$

4. Experiments

Datasets We evaluated our model on FFHQ [15] and CACD2000 [5]. FFHQ includes 70000 images with 1024x1024 resolution. Following the data preprocessing procedures as [25], we take images with id 0-68999 as the training set and 69000-69999 for testing and filter out images with low confidence in differentiating the gender, low confidence in estimating the age, wearing dark glasses, extreme pose, and angle based on the facial attributes annotated by Face++[1]. As the annotation from [25] only includes the age group label, we acquire the age label information from [48]. To reconcile both age group labels and age labels, we further filter out images in which the age label disagrees with the age group label. This results in 12488 male and 13563 female images for training, and 279 male and 379 female images for testing. CACD2000 consists of 163446 images where age ranges from 14 to 62 years old. We randomly take 10% of data for evaluation. We use Face++ to separate the images into male and female and extract the facial landmarks using Dlib[2].

Implementation Since aging patterns are different between males and females, we train two separate models

on the FFHQ dataset for both 256x256 and 512x512 resolutions. Model architecture is modified based on Cycle-GAN [50]. Please refer to the Supplementary for the detailed model architecture and optimization settings. λ_{mv1} and λ_{mv2} are set to 0.05 and 0.005 in **Eq.** (4). λ_{fake1} and λ_{fake2} are set to 0.4 and 1 in **Eq.** (6). In **Eq.** (10), λ_{age}, λ_{idt}, and λ_{adv} are set to 0.05, 1, and 1 respectively.

4.1. Qualitative Evaluation

Face Aging We present our test results on FFHQ, comparing with results from [25]. Images for [25] are generated using their provided code[3]. To illustrate the model performance across different ages, we show 6 input examples from 4 representative age groups (<30, 30-40, 40-50, 50+) and generate the results for each group. The target ages for our model are chosen as 25, 35, 45, and 55 respectively. As can be seen in **Fig.** 2, the images generated by our model result in fewer artifacts and exhibit more clear aging details, such as beard color change (example a,c) and wrinkles on different parts of the face (see example b,c,d,e). A convincing detail in example (f) shows that the personal traits (a mole) are well preserved using our models.

We also directly generates images on CACD2000 using the models trained on FFHQ in the resolution of 256x256 to compare with CAAE[49], IPCGAN [44], and S^{2}GAN [10] in **Fig.** 3. The demonstrated images are the presented examples in [11], which is the state-of-the-art work on CACD2000. For all age groups, our model presents more evident and fine-grained aging effects comparing with all previous works.

Continuous Aging In **Fig.** 4, we illustrate some examples of continuous aging results comparing with [25]. We choose an age step of 4 to present due to the limited space. A gradual and smooth natural aging process (e.g. wrinkle depth change, beard, pigmentation on face) can be observed from our images while retaining personal traits. The interpolation-based method in Lifespan, however, lacks the ability to generate images of well-aligned target ages and does not preserve certain personalized information.

Aging Details Here, we show that the generated images express a significant level of aging details on different parts of the face. In **Fig.** 5, we demonstrate three enlarged face crops from the generated images, which give a clear and detailed view of enhanced wrinkles, skin smoothness, color change of beard and eyebrow.

4.2. Quantitative Evaluation

Identity Preservation To evaluate identity preservation, we adopt the face verification rate metric. Specifically, we followed the evaluation protocol of [10] on an age group basis for a fair comparison with prior works. We calculate

---

[1]Face++ facial attribute annotation API: https://www.faceplusplus.com/

[2]Dlib toolkit: http://dlib.net/

[3]Lifespan official code: https://github.com/royorel/Lifespan_Age_Transformation_Synthesis
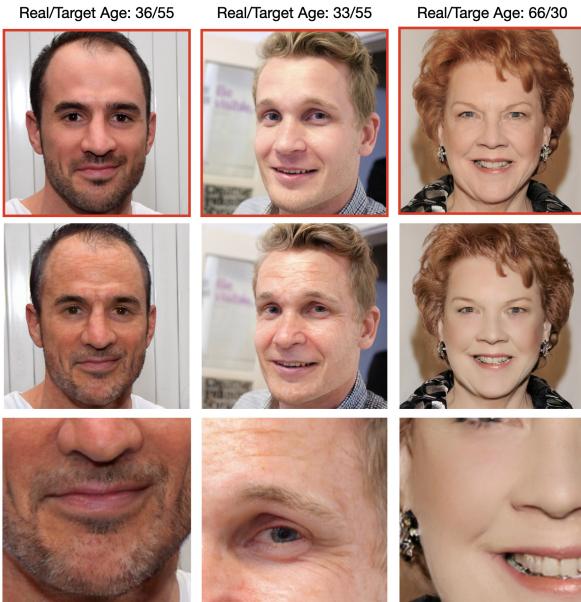
Figure 5. Aging details: enlarged face crops to show details for beard, wrinkle, and skin smoothness. Input images are in red boxes.

the face verification rate between all combination of image pairs, i.e. (test, 10-29), (test, 30-39), ...,(30-39, 40-49), (40-49, 50-59). Face verification score is obtained from Face++ and the threshold is set as 76.5 (@FAR=1e-5). The complete results are presented in **Table** 1 and 2 for CACD2000 and FFHQ respectively. As the results suggest, our model achieves the highest face verification rate for both datasets among all candidates, which indicates it best meets the identity preservation requirement of the task.

	Average of All Pairs
CAAE [49]	60.88%
IPCGAN [44]	91.40%
S^{2}GAN [10]	98.91%
Lifespan [25]	93.25%
Ours	**99.97%**

Table 1. Evaluation of identity preservation in terms of face verification rates on CACD2000 [5]. Pair-wise results are presented in Supplementary.

	Average of All Pairs
Lifespan [25]	87.11%
Ours	**99.98%**

Table 2. Evaluation of identity preservation in terms of face verification rates on FFHQ [15]. Pair-wise results are presented in Supplementary.

Aging Accuracy In terms of assessing aging accuracy, we use an unbiased age estimator to infer the age of the generated images. To be able to compare with prior group-

	10-29	30-39	40-49	50+
CAAE [49]	29.6	33.6	37.9	41.9
S^{2}GAN [10]	24.0	36.0	45.7	55.3
IPCGAN [44]	27.4	36.2	44.7	52.5
IPCGAN [44] (Face++)	42.4	47.1	51.9	56.0
Lifespan [25] (Face++)	-	40.2	-	64.3
Ours (Face++)	30.5	38.7	46.9	60.0

Table 3. Comparison of the mean age of generated images in each age group evaluated using Face++ on CACD2000 [5].

	10-29	30-39	40-49	50+
Lifespan [25]	-	38.4	-	63.8
Ours	30.7	38.4	47.7	62.1

Table 4. Comparison of the mean age of generated images in each age group evaluated using Face++ on FFHQ [15].

based methods on CACD2000, we generate our images aligning with their age group settings in which we adaptively increment/decrement by a factor of 10 (age group size) from input image's real age as the target age for generation, i.e. target age 33 is used for generating an image of age group 30-40 given current age of 23. As we neither have the access to [10]'s evaluation age estimator nor their pretrained model for assessing our model and doing a direct comparison, we instead use Face++'s age estimation results on our model and one of accessible prior work IPCGAN [44], which is also evaluated in [10] to show relative comparison. Evaluation of FFHQ follows the same procedure as CACD2000. The evaluation results are shown in **Table** 3 and 4 for CACD2000 and FFHQ respectively. As the results suggest, our model evaluated using Face++ has a more reasonable mean age at each age group than IPCGAN [44] and Lifespan [25] on CACD2000 and has a similar performance as Lifespan on FFHQ.

Image Fidelity Considering the image fidelity, we adopt the Fréchet Inception Distance (FID) [12] metric to evaluate our model. Similar to the image generation settings as before, we calculated the FID on the generated images corresponding to the same age group as theirs on CACD2000. For comparing with [25] on FFHQ, we calculate the FID on the generated images, that share the same age group range. The results are shown in the **Table** 5. On both datasets, our model achieves the lowest FID, which quantitatively demonstrates superiority in the image quality aspect.

4.3. Model Interpretability and Ablation Study

Continuous Aging To evaluate how well our model generates synthesized images in a continuous setting, we use an age estimator to predict age on the generated fake images from 25 to 65 of our approach and the linear interpolation approach performed between anchor aging bases. The an-

Figure 6. Linear interpolation between transformed identity encodings. Real images are in red boxes. From left to right, we linearly interpolate between two images' transformed identity encodings at the same target age 65. Personal traits, such as eye color and teeth shape, smoothly change from one person to the other.

	CACD2000	FFHQ
CAAE [49]	44.2	-
IPCGAN [44]	9.1	-
S^{2}GAN [10]	8.4	-
Lifespan [25]	11.7	26.2
Ours	**6.7**	**18.5**

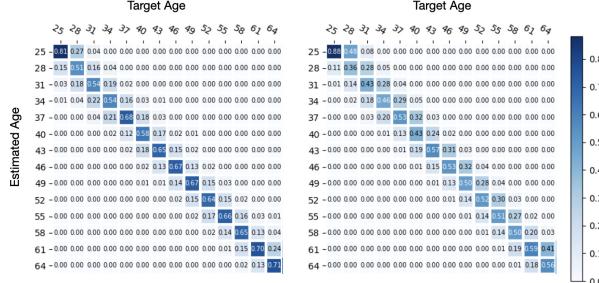Table 5. FID evaluation: lower is better.

Figure 7. Confusion matrices of continuous face aging. Left: age estimation on the self-estimated aging embeddings (proposed). Right: age estimation on the linear interpolated aging embeddings. The age step is chosen as 3 based on the MAE of the estimator.

chor basis is generated by taking the mean of every aging bases within an age group. We calculate a confusion matrix in terms of aging accuracy for each approach using the age estimator jointly trained on the FFHQ dataset. **Fig.** 7 indicates that our generated fake images express a more evident continuous aging trend with much higher aging accuracy than the linear interpolation approach.

Interpolation between Two Identities in Latent Space In **Fig.** 6, we further illustrate that our proposed model also learns a disentangled representation of age and identity in latent space. We linearly interpolate between the two transformed identity encodings of the same age and different identities and then generate images for the interpolated encodings. As shown in the figure, the identity changes gradually while maintaining the respective age.

Use of the Residual Embedding One of the key innovative design of our model architecture is the formulation of the personalized age embedding, which incorporates both

personalized aging features of the individual and shared aging effects among the entire population. To better illustrate and understand the effectiveness of the design, we train a model without adding the residual embedding (i.e. directly applying the target age's exemplar-face aging basis $a_{i,j=t_i}$), and compare with the proposed method.

In **Fig.** 8, we display a few examples with highlighted/enlarged regions comparing results w/wo residual embeddings. Noticeably, more unnatural artifacts and a tendency to examplar-face modification are observed in the images generated without residual embeddings.
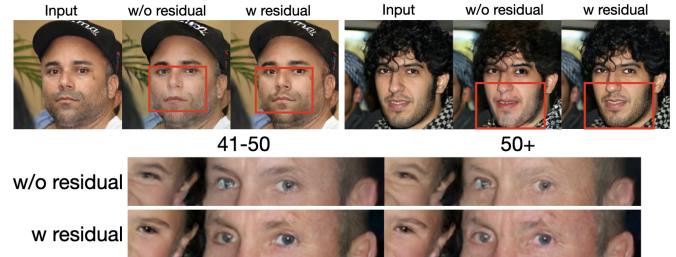

Figure 8. Enlarged ablation examples. Top-row target age: 11-30. The nose shape and beard were not preserved in the w/o residual example (top). Artifacts in eyes are commonly seen in older age groups w/o residual (bottom).

5. Conclusions

In this work, we introduce a novel approach to the task of face aging with a specific focus on the continuous aging aspect. We propose a unified framework to learn continuous aging bases via introducing an age estimation module to a GAN-based generator. The designed **PAT** module further enhances the personalization of the exemplar-face aging bases, which results in more natural and realistic generated face images overall. The experiments qualitatively and quantitatively show superior performance on the aging accuracy, identity preservation, and image fidelity on two datasets compared to prior works. Furthermore, the proposed network structure can also be applied to other multi-class domain transfer tasks to avoid group-based training and achieve a more accurate continuous modeling.

8

References

[1] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In *2017 IEEE international conference on image processing (ICIP)*, pages 2089–2093. IEEE, 2017. 2

[2] Yosuke Bando, Takaaki Kuratate, and Tomoyuki Nishita. A simple method for modeling wrinkles on human skin. In *Pacific Conference on Computer Graphics and Applications*, pages 166–175. Citeseer, 2002. 2

[3] Laurence Boissieux, Gergo Kiss, Nadia Magnenat Thalmann, and Prem Kalra. Simulation of skin aging and wrinkles with cosmetics insight. In *Computer Animation and Simulation 2000*, pages 15–27. Springer, 2000. 1, 2

[4] D Michael Burt and David I Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259(1355):137–143, 1995. 2

[5] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European conference on computer vision*, pages 768–783. Springer, 2014. 2, 5, 6, 7, 11, 14

[6] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018. 2, 4

[7] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020. 3

[8] Harm de Vries, Florian Strub, Jeremie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C Courville. Modulating early visual processing by language. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6594–6604. Curran Associates, Inc., 2017. 4

[9] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Advances in neural information processing systems*, 3(06), 2014. 1, 2

[10] Zhenliang He, Meina Kan, Shiguang Shan, and Xilin Chen. S2gan: Share aging factors across ages and share aging trends among individuals. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9440–9449, 2019. 1, 2, 3, 5, 6, 7, 8, 14

[11] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing*, 28(11):5464–5478, 2019. 6

[12] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems*, pages 6626–6637, 2017. 7

[13] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017. 4

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017. 2, 6

[15] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 2, 5, 6, 7, 11, 14

[16] 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*, pages 3334–3341, 2014. 1, 2

[17] 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. 2

[18] Peipei Li, Yibo Hu, Qi Li, Ran He, and Zhenan Sun. Global and local consistent age generative adversarial networks. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1073–1078. IEEE, 2018. 2

[19] Ming Liu, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, Wangmeng Zuo, and Shilei Wen. Stgan: A unified selective transfer network for arbitrary image attribute editing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3673–3682, 2019. 2

[20] Si Liu, Yao Sun, Defa Zhu, Renda Bao, Wei Wang, Xiangbo Shu, and Shuicheng Yan. Face aging with contextual generative adversarial nets. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 82–90, 2017. 1, 2

[21] Yunfan Liu, Qi Li, and Zhenan Sun. Attribute-aware face aging with wavelet-based generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11877–11886, 2019. 2

[22] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3

[23] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. 11

[24] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651, 2017. 3

[25] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 5, 6, 7, 8, 14

[26] Hongyu Pan, Hu Han, Shiguang Shan, and Xilin Chen. Mean-variance loss for deep age estimation from a face. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5285–5294, 2018. 3, 4

[27] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Gaugan: semantic image synthesis with spatially adaptive normalization. In *ACM SIGGRAPH 2019 Real-Time Live!* 2019. 3

[28] Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 387–394. IEEE, 2006. 2

[29] Narayanan Ramanathan and Rama Chellappa. Modeling shape and textural variations in aging faces. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8. IEEE, 2008. 2

[30] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE, 2006. 3

[31] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 10–15, 2015. 3

[32] Duncan A Rowland and David I Perrett. Manipulating facial appearance through shape and color. *IEEE computer graphics and applications*, 15(5):70–76, 1995. 2

[33] Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in neural information processing systems*, pages 901–909, 2016. 11

[34] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9243–9252, 2020. 2

[35] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *Proceedings of the IEEE international conference on computer vision*, pages 3970–3978, 2015. 2

[36] 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. 2

[37] Jinli Suo, Feng Min, Songchun Zhu, Shiguang Shan, and Xilin Chen. A multi-resolution dynamic model for face aging simulation. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. 2

[38] 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, 2009. 2

[39] 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. 1, 2

[40] James T Todd, Leonard S Mark, Robert E Shaw, and John B Pittenger. The perception of human growth. *Scientific american*, 242(2):132–145, 1980. 2

[41] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 11

[42] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe. Recurrent face aging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2378–2386, 2016. 2

[43] Wei Wang, Yan Yan, Stefan Winkler, and Nicu Sebe. Category specific dictionary learning for attribute specific feature selection. *IEEE Transactions on Image Processing*, 25(3):1465–1478, 2016. 2

[44] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7939–7947, 2018. 1, 2, 3, 5, 6, 7, 8, 14

[45] Yin Wu, Nadia Magnenat Thalmann, and Daniel Thalmann. A plastic-visco-elastic model for wrinkles in facial animation and skin aging. In *Fundamentals of Computer Graphics*, pages 201–213. World Scientific, 1994. 1, 2

[46] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. Learning face age progression: A pyramid architecture of gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 31–39, 2018. 1, 2

[47] 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*, 25(6):2493–2507, 2016. 2

[48] Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. High resolution face age editing. *arXiv preprint arXiv:2005.04410*, 2020. 6

[49] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818, 2017. 2, 5, 6, 7, 8, 14

[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2, 6

6. Supplementary

6.1. Network Architecture and Optimization Settings

During training, we use Adam optimizer with the learning rate of 0.0002 and batch size of 20 and 5 for 256 and 512 model respectively. The model is trained for 200 epochs and learning rate is linearly decayed over last 100 epochs.

Layer	Stride	Act.	Norm	Output Shape
Input	-	-	-	256x256x3
Conv. 7 x 7	1	ReLU	Spectral	256x256x64
Conv. 3 x 3	2	ReLU	Spectral	128x128x128
Conv. 3 x 3	2	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256
Res. Block	1	ReLU	Spectral	64x64x256

Table 6. Identity Encoder **E** specification. Spectral means spectral normalization [23] is applied after each convolutional layer.

Layer	Stride	Act.	Norm	Output Shape
Encoding	-	-	-	64x64x256
Res. Block	1	ReLU	Instance	64x64x256
Res. Block	1	ReLU	Instance	64x64x256
Res. Block	1	ReLU	Instance	64x64x256
Deconv. 3 x 3	2	ReLU	Instance	128x128x128
Deconv. 3 x 3	2	ReLU	Instance	256x256x64
Conv. 7 x 7	1	Tanh	-	256x256x3

Table 7. Generator **G** specification. Instance means instance normalization [41] is applied after each convolutional layer.

Layer	Norm	Output Shape
Encoding	-	64x64x256
GAP	-	1x1x256
Flatten	-	256
Linear	Weight	100

Table 8. Age estimator **C** specification. GAP means global average pooling. Weight means weight normalization [33] is applied to linear layer (bias term are set to zero).

Detailed network architectures for **E**, **G** and **C** are presented in **Table** 6, 7 and 8 respectively.

6.2. Pair-wise Identity Preservation Results

Here, we provide the complete pair-wise identity preservation comparison using Face++ in **Table** 9 and 10 for CACD2000 [5] and FFHQ [15], respectively. As can be seen, our model achieves the highest verification rate in every aspects compared to prior works.

6.3. More Aging Results

Continuous Aging. We generate the complete continuous aging results of a person from age 20 to age 69 and the results are displayed in **Fig.** 11. As shown, aging proceeds in a natural and gradual manner.

Enlarged Comparison of group 50+. In **Fig.** 9, we show the enlarged generated images of age group 50+. Our model is able to generate fine aging details aligned with the target age group.

6.4. Limitations

While our work can generate natural face aging, we also observe some failure cases when generating outputs for input image with hats and glasses or faces with heavy make-ups (in **Fig.** 10). The model also does not work well for extreme target age like 95-year-old, where the corresponding exemplar-face aging basis is hardly trained due to lack of data for those minority classes.
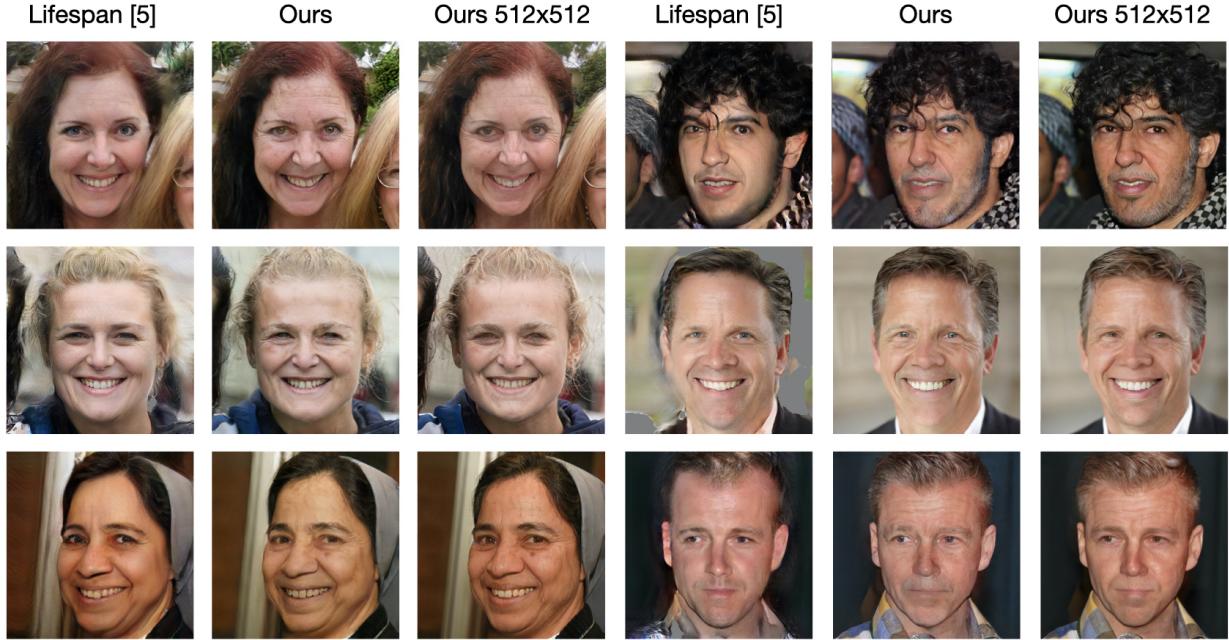
Figure 9. Enlarged examples of generated age 50+. Our model demonstrate better details in aging effects such as wrinkles and beard change.
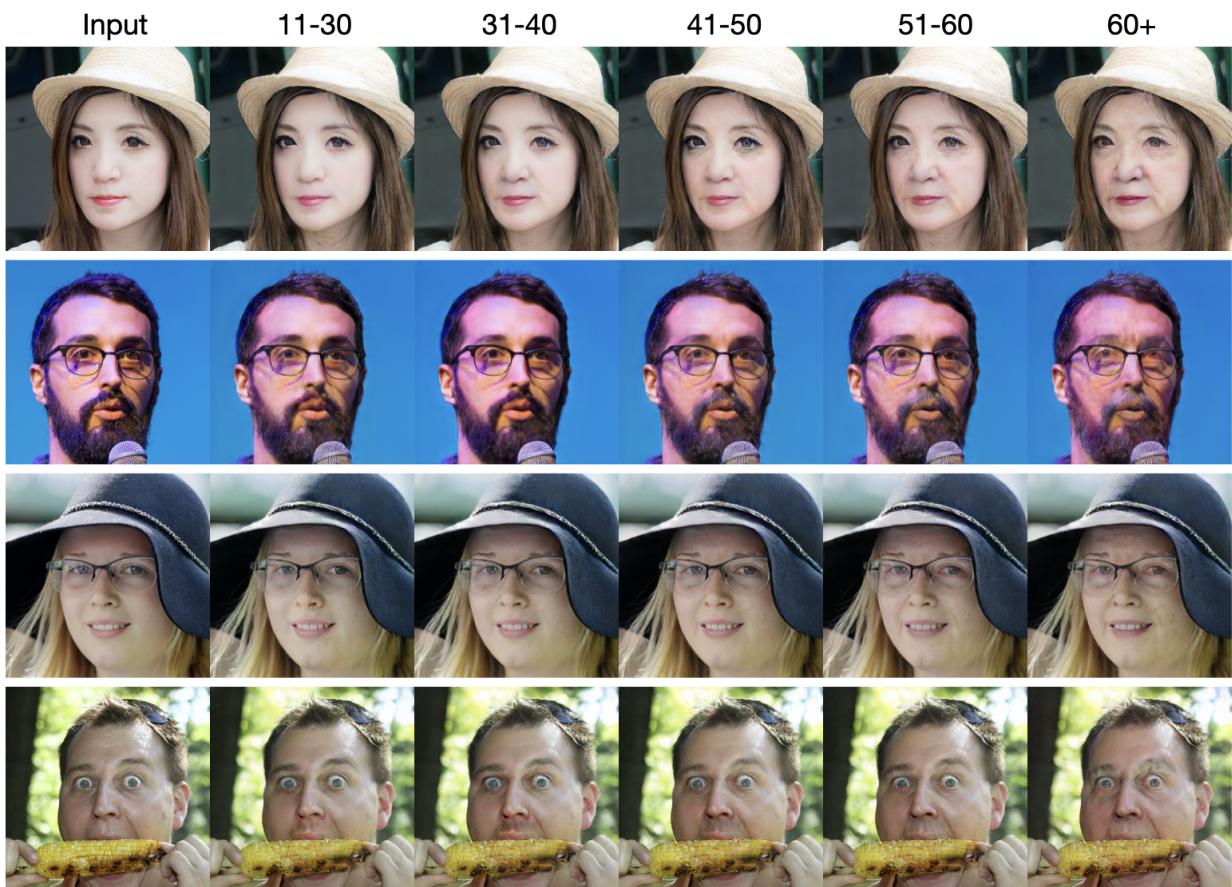

Figure 10. Failure cases: heavy makeup with hat, glasses with bad lighting. Our model could not best capture the personalized information such as skin texture in these cases.
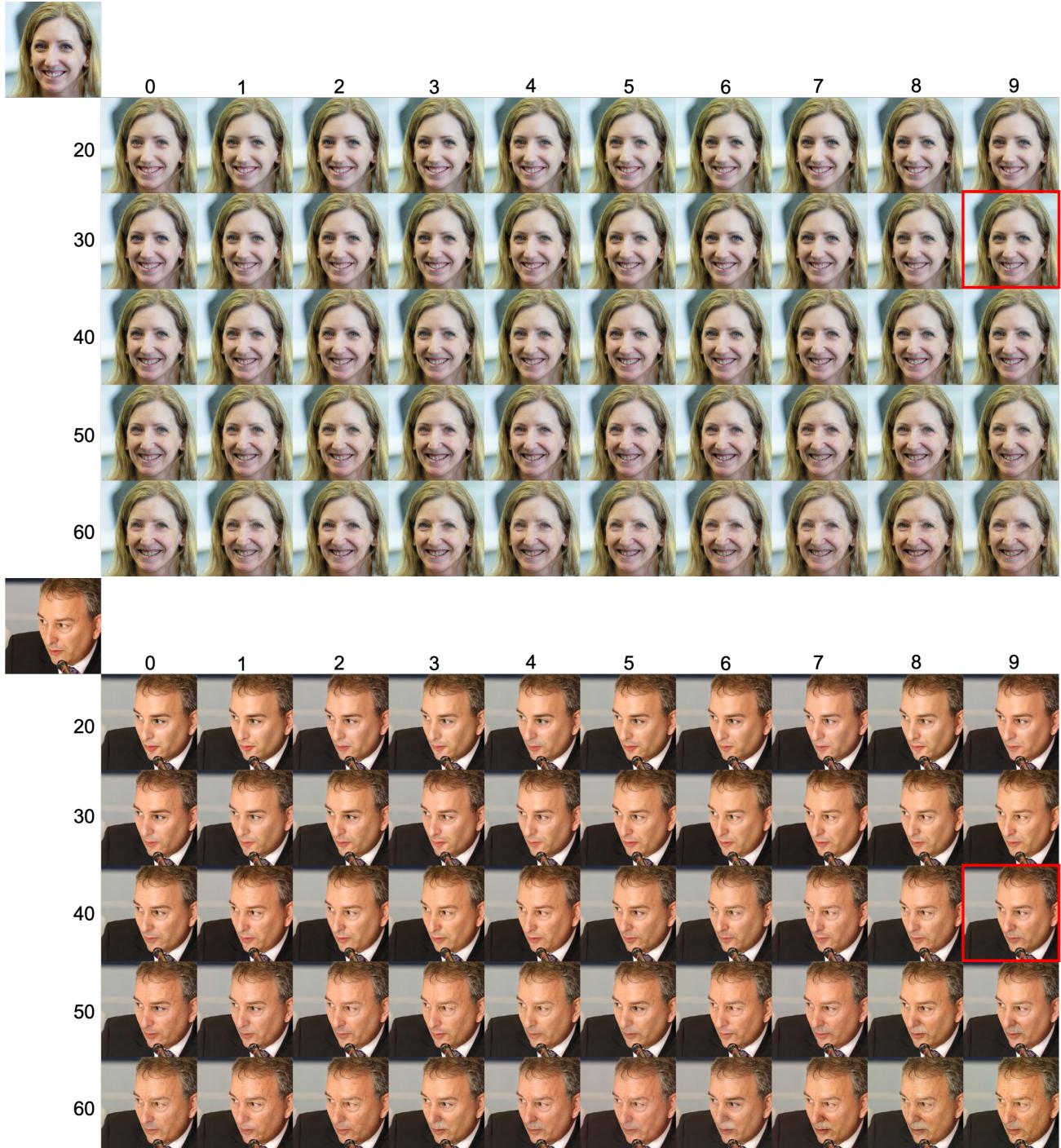
Figure 11. Complete continuous aging results from age 20 to 69. Input is at the top left corner of each image grid. Generated image of real age is in the red box.

	Average of All Pairs	Hardest Pair	Easiest Pair
CAAE [49]	60.88%	(test, 50+): 2.0%	(40-49, 50+): 99.97%
IPCGAN [44]	91.40%	(10-29, 50+): 62.98%	(40-49, 50+): 99.98%
S^{2}GAN [10]	98.91%	(10-29, 40-49): 94.08%	(40-49, 50+): 99.96%
Lifespan [25]	93.25%	(test, 50-69): 80.94%	(30-39, 50-69): 99.75%
Ours	**99.97**%	(test, 40-49): 99.96%	(test, 30-39): 100.00%

Table 9. Complete evaluation of identity preservation in terms of face verification rates on CACD2000 [5].

	Average of All Pairs	Hardest Pair	Easiest Pair
Lifespan [25]	87.11%	(test, 50-69): 72.32%	(30-39, 50-69): 98.85%
Ours	**99.98**%	(test, 60+): 99.96%	(test, 30-39): 100.00%

Table 10. Complete evaluation of identity preservation in terms of face verification rates on FFHQ [15].