S^{2}GAN: Share Aging Factors Across Ages and Share Aging Trends Among Individuals

Zhenliang He[1,2], Meina Kan[1,2], Shiguang Shan[1,2,3], Xilin Chen[1,2]
[1] Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Peng Cheng Laboratory, Shenzhen, 518055, China
zhenliang.he@vipl.ict.ac.cn, {kanmeina,sgshan,xlchen}@ict.ac.cn

Abstract

Generally, we human follow the roughly ***common aging trends****, e.g., the wrinkles only tend to be more, longer or deeper. However, the aging process of each individual is more dominated by his/her* ***personalized factors****, including the invariant factors such as identity and mole, as well as the personalized aging patterns, e.g., one may age by graying hair while another may age by receding hairline. Following this biological principle, in this work, we propose an effective and efficient method to simulate natural aging. Specifically, a* ***personalized aging basis*** *is established for each individual to depict his/her own aging factors. Then different ages share this basis, being derived through* ***age-specific transforms****. The age-specific transforms represent the aging trends which are shared among all individuals. The proposed method can achieve continuous face aging with favorable aging accuracy, identity preservation, and fidelity. Furthermore, befitted from the effective design, a unique model is capable of all ages and the prediction time is significantly saved.*

1. Introduction

Face aging/rejuvenation aims to predict the future/past faces of a given face image as shown in Fig. 1(a), which is not only interesting for entertainment but also valuable for applications such as cross-age face recognition, finding lost children or wanted fugitives. Although face aging has been studied for decades and witnessed various breakthroughs, it is still a challenging task even for human itself because the natural human aging is affected by many factors such as genes, physical damage, disease and living environment, which are indeterministic and quite complicated to be modeled. Besides, it is difficult or even impossible to collect age data for each individual over a long period, and most available datasets are limited to short age span for the same person. Therefore, discontinuous data distribution makes the modeling of long period aging even more challenging.

Generally, traditional face aging approaches can be grouped into two main categories, i.e., the physical model approaches and the prototype approaches. The physical model approaches [38, 28] simulate the aging process by exploiting prior knowledge about the biological structure and aging mechanism. These physical model approaches only consider the overall aging characteristics, but do not specially investigate the personality in the aging process. The prototype approaches [37, 15] divide continuous ages into discrete age groups, and define the average face or low-rank face as the prototype of each age group, then the transition pattern between a pair of age groups is represented by the difference between their corresponding prototypes. Because of averaging, however, the personalized information is eliminated in the prototypes, and therefore the transition pattern is also an average pattern without personality. Shu et al. [32] try to introduce personality by establishing aging coupled dictionaries for each age group, and the personality is represented by the sparse coding coefficients. However, this method produces severe ghost artifacts.

Recently, generative adversarial networks (GANs) [9] and its variants are adopted for face aging with promising results [17, 40, 42]. These methods, powered by deep neural networks and adversarial training, usually train a transformation network to convert an input image to a target age group. Besides, an identity preservation network is often adopted in GAN based methods to keep the identity unchanged during the aging process. Although the identity is kept, as for age, these methods only consider the population-to-population transition between age groups, instead of explicitly considering the personalized transition pattern for each individual between different ages. As a solution, Liu et al. [18] propose a transition pattern discriminator to drive the aging network to capture the personalized transition pattern for each individual.

Without proper consideration of personality may result in missing details. Therefore, in this work, in considera-

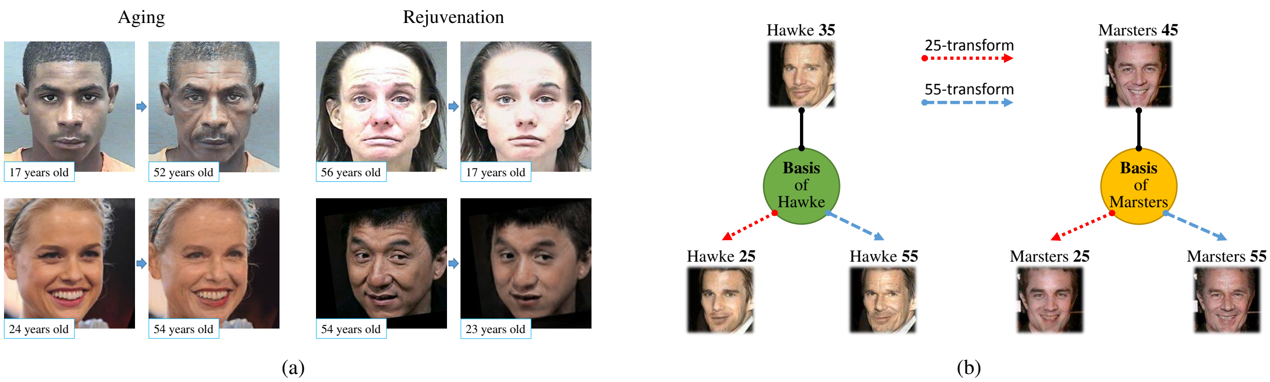


Figure 1. (a) Simulation of face aging and rejuvenation. (b) On the one hand, Hawke and Marsters have their own aging basis/factors respectively, and all their ages are derived from their own basis. On the other hand, both Hawke and Marsters use the same age-specific transform for the same age.

tion of both the *personalized aging factors* and the *common aging trends*, we suggest a novel approach for effective and efficient face aging, named as S^{2}GAN. On the one hand, the personalized aging factors include identity, mole, and even the personalized aging patterns (e.g., one may age by graying hair while another may age by receding hairline). These personalized factors almost remain unchanged through one's whole life, because they are most probably encoded in genes which would not change across ages. Accordingly, to simulate these personalized effects, our approach employs a deep encoder to establish the *personalized aging basis* for each individual depicting his/her own aging factors, while such basis is shared across different ages. On the other hand, given the personalized aging basis, different ages of each individual can be derived from his/her own basis through *age-specific transforms*, which are shared among all individuals to capture those common aging trends (e.g., the wrinkles only tend to be more, longer or deeper if a person ages by wrinkles).

Overall, as shown in Fig. 1(b) the personalized basis is distinct for each individual but **shared** across ages, while the age-specific transform is distinct for different ages but **shared** among all individuals. Such sharing forms a concise S^2-module, embedded in the **GAN** framework, further forms the proposed **S^{2}GAN** approach. Compared to the existing GAN based methods, the main contributions of this work include:

- A new perspective for the natural aging process, i.e., faces at different ages of a specific person are derived from a same personalized aging basis, while the age-specific transforms from the aging basis to the target aged faces are shared among all individuals.
- Lower computational cost. A unique model is capable of all target ages, and thus the prediction time is significantly saved benefited from the sharing mechanism (S^2-module).
- Favorable continuous aging, which can be achieved by interpolating the aging transforms of adjacent age groups, superior to the existing discrete group-wise age synthesis methods [1, 17, 40, 42, 19].
- The S^2-module is orthogonal to the existing methods, therefore can be used as a plug-in module in many recent methods with a transformation network such as [17, 42, 19], to reduce their computational consumption as well as enable the continuous aging, while still keeping their own advantages.

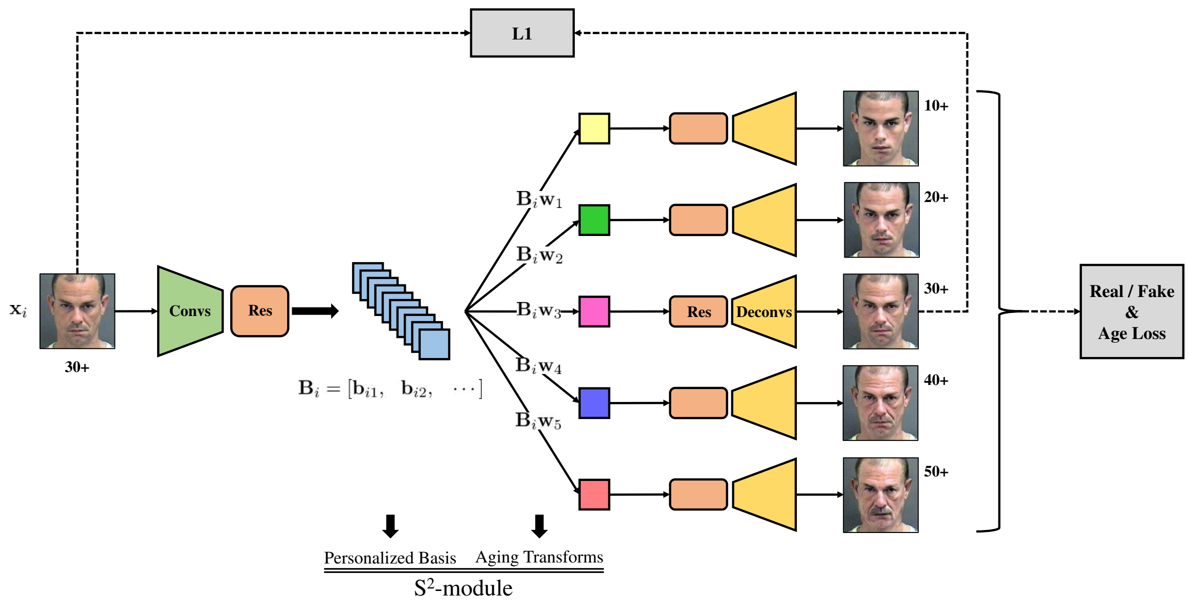
2. Related Work

2.1. Face Aging

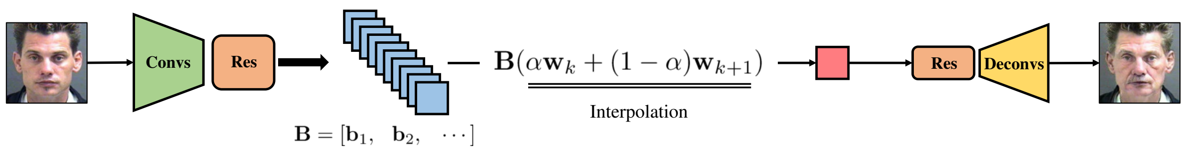
Researchers have made great efforts to face aging with effective and inspirational approaches, and please refer to [29, 8, 7] for a comprehensive survey of face aging. Generally, the face aging approaches can be divided into three groups, i.e., *physical model approaches*, *prototype approaches* and *deep learning approaches*.

Physical model approaches model the aging factors based on biological and physical mechanism, e.g., craniofacial growth [38, 27], skin and wrinkles [41, 3, 2, 28], muscle structure [28, 34] and facial components [35, 36]. Although these mechanical models are dedicatedly designed, they are heavily biased to the imperfect human knowledge of aging mechanism while often computationally expensive.

In the *prototype approaches*, continuous ages are divided into discrete age groups and the average face [4, 31, 37] or a low-rank subspace [15] of each group is defined as its prototype, then the transition pattern between age groups are modeled as the difference between their prototypes. Due to the average, missing personality in the learned transition pattern becomes the main problem of these methods. To remedy the lack of personality, Shu et al. [32] build age coupled dictionaries for each age group, and the sparse coding coefficients of the input image express its personalized transition patterns.



(a) *Training with discrete age groups.* The personalized aging basis is inferred by a deep encoder from an input face, after which different age-specific aging transforms (linear combination coefficients here) are applied on this basis to generate the corresponding age representations. Then these age representations are decoded by the same decoder network to generate corresponding aged faces. The whole model is learned under the age loss, L1 reconstruction loss and adversarial loss.



(b) *Testing with continuous age-specific transforms.* By using the interpolations between transforms (coefficients) of any pair of adjacent age groups, naturally, we achieve favorable continuous face aging.

Figure 2. Overview of the proposed S^2GAN with (a) discrete training, and (b) continuous testing.

Recently, *deep learning approaches* become the state-of-the-arts credited to the powerful non-linearity of the deep networks. Temporal deep models [39, 24, 23] are adopted to model the transition pattern between the adjacent age groups with impressive aging results. Afterwards, the visual fidelity of face aging is largely improved by [1, 45, 18, 17, 40, 42, 19] based on the generative adversarial network (GAN) [9] and its variants. Specifically, Zhang et al. [45] propose a conditional adversarial autoencoder to project the images onto a manifold with an aging axis. However, the aging results tend to be blurry probably because the manifold is constrained as a simple prior distribution (e.g., uniform). Liu et al. [18] propose a transition pattern discriminator to drive the aging network to capture the transition patterns between age groups. Li et al. [17] use three local generators, which are responsible for forehead, eyes, and mouth respectively, cooperating with a global generator to enhance the age generation. Wang et al. [40] adopt identity-preserved conditional GAN achieving convincing identity preservation. Yang et al. [42] propose a pyramid architecture discriminator accepting the high-level age-specific features for finer supervision on the aging details achieving promising results. Commonly, most of these GAN based approaches adopt an identity preservation network [17, 40, 42, 19] to keep the identity during the aging process. As for accurate age generation, [17, 40] apply the age group classification loss, while [42, 19] adopt the adversarial training between a pair of age groups.

2.2. Generative Adversarial Networks

Generative adversarial network (GAN) [9] is a special generative model with adversarial training as its core idea, where a discriminator tries to distinguish the real and fake samples while a generator tries to deceive the discriminator. Theoretically, when the adversarial training attains the Nash equilibrium, the fake distribution is identical to the real one. As extensions, cGAN [21] and AC-GAN [25] accept conditional signals and generate samples satisfying the conditions, e.g., to generate digits specifying a number. GAN and its variants can generate visually realistic images and have shown superiority on many image synthesis tasks [13, 14, 44, 46, 6, 11]. Therefore, most recent face aging methods also adopt GAN for high generation fidelity, as mentioned in Sec. 2.1.

3. S^{2}GAN

The proposed S^{2}GAN mainly consists of three parts, i.e., 1) establishing the personalized aging basis, 2) transforming the basis to the age representations, and 3) decoding the representations to the aged faces, with an overview shown in Fig. 2(a). Firstly, the personalized aging basis is inferred by a deep encoder. Then, age-specific transforms are applied on this basis to obtain the age representations for different age groups. Finally, aged faces within different age groups are obtained by decoding the corresponding age representations. The whole architecture is optimized end-to-end with three objectives, i.e., the age group classification loss for accurate aging, the L1 reconstruction loss for identity preservation and the adversarial loss for fidelity. At the testing phase as shown in Fig. 2(b), for a target age group, an aged face of the input can be obtained by decoding the corresponding age representation which is derived from the aging basis by the corresponding age-specific transform. More favorably, continuous aging can be naturally achieved with interpolations between adjacent age-specific transforms.

We first introduce the key notations before illustrating the details. Let $(\mathbf{x}_i, y_i)$ denote the i^{th} sample where $\mathbf{x}_i$ is the input image and $y_i \in \{1, 2, 3, 4, 5\}$ is the ground truth age group of $\mathbf{x}_i$, e.g., $y_i = 1$ denotes that the age of $\mathbf{x}_i$ falls in [11, 20], while $y_i = 5$ denotes that the age of $\mathbf{x}_i$ falls in 51+. $\mathbf{B}_i$ denotes the personalized basis inferred from $\mathbf{x}_i$, and $\mathbf{w}_k$ denotes the coefficients of the k^{th} age-specific transforms corresponding to the k^{th} age group.

3.1. Formulation

Personalized Aging Basis As mentioned in Sec. 1, the aging process of each individual is dominated by his/her personalized aging factors, which can be depicted by a personalized basis. Here, we use a neural network encoder E to map an input image $\mathbf{x}_i$ to its personalized basis $\mathbf{B}_i = [\mathbf{b}_{i1}, \ \mathbf{b}_{i2}, \ \cdots, \ \mathbf{b}_{im}]$ with m basis vectors. Formally, the personalized basis is obtained as follows,

$$\mathbf{B}_i = E(\mathbf{x}_i). \tag{1}$$

Such personalized basis is unique for each individual, trying to capture those personalized aging factors encoded in genes such as identity and aging by graying hair or receding hairline. These personalized factors generally remain unchanged in one's whole life; therefore one's personalized basis can be shared by all his/her ages.

Age-Specific Transforms Given the aging basis $\mathbf{B}_i$, we can obtain the age-specific representation for an age group by applying the corresponding age-specific transform. This is formulated as a linear combination of the aging basis as follows,

$$\mathbf{r}_i^k = \sum_{j=1}^{m} w_{kj} \mathbf{b}_{ij} = \mathbf{B}_i \mathbf{w}_k, \tag{2}$$

where $\mathbf{w}_k = [w_{k1}, w_{k2}, \cdots, w_{km}]^\top$ is the aging transform corresponding to the k^{th} age group, and thus $\mathbf{r}_i^k$ denotes the age representation for the k^{th} age group of the i^{th} face image. Here, $\mathbf{w}_k$ characterizes the aging trends for the k^{th} age group. Since the aging basis $\mathbf{B}_i$ has already captured the personalized aging factors, the age-specific transform $\mathbf{w}_k$ can be shared among all individuals to simulate the common aging/rejuvenation trends, e.g., $\mathbf{w}_5$ may increase mustache or thins the lips while $\mathbf{w}_1$ may decrease the wrinkles.

As can be seen from Eq. (2), in our method the personalized aging factors are distinct for different individuals but shared across ages, while the common aging trends are shared among individuals but distinct for different ages. This concise design well fits our biological insight and observations about natural aging as analyzed in Sec. 1.

Finally, by decoding the representation for a certain age group via a lightweight decoder G, we can obtain the aged image within that age group, i.e.,

$$\hat{\mathbf{x}}_i^k = G(\mathbf{r}_i^k). \tag{3}$$

3.2. Objective

With expectation, the aged face in Eq. (3), i.e., $\hat{\mathbf{x}}_i^k$, should belong to the target age group, preserve the identity, as well as be with high fidelity. Accordingly, three types of loss are designed to ensure these objectives.

Age Loss for Accurate Aging To ensure that the generated aged face correctly falls into the target age group, a well trained and fixed age group classifier C is used to guide the generation, which follows the AC-GAN [25] spirit for conditional generation. The age loss is formulated as below,

$$l_i^{\text{age}} = -\sum_k \log(C_k(\hat{\mathbf{x}}_i^k)), \tag{4}$$

where $C_k(\cdot)$ denotes the probability that a sample falls into the k^{th} age group, predicted by the classifier C. Therefore, this loss tries to make $\hat{\mathbf{x}}_i^k$ more likely to be in the expected k^{th} age group for each k.

L1 Loss for Identity Preservation Besides accurate aging, identity should not be changed during the aging process. Therefore, L1 reconstruction loss is applied for identity preservation. Specifically, if an input face is in the age group of [31,40], then the generated face within the same age group should be as close as possible to the input. Accordingly, the L1 loss is formulated as follows,

$$l_i^{\text{L1}} = \sum_k \delta(y_i = k) \|\mathbf{x}_i - \hat{\mathbf{x}}_i^k\|_1, \tag{5}$$

where $\delta(y_i = k)$ is 1 if $y_i = k$ is valid else 0. Identity feature reconstruction is another choice for identity preservation [17, 40, 42, 19]; however, it needs an extra deep identity network. Therefore, the L1 reconstruction loss, which is directly applied to the images without effort to tune an extra identity network, is more convenient and favorable. Our elaborate framework makes the L1 reconstruction loss feasible for identity preservation, thus is more flexible than methods such as [17, 40, 42, 19] which can only adopt the identity feature reconstruction.

Adversarial Loss for Fidelity The conditional adversarial training [21] is applied for the fidelity of the aging results. Following [22], the hinge loss is used as the adversarial loss, formulated as

$$l_i^{\text{adv-d}} = \max(1 - D(\mathbf{x}_i, y_i), 0) + \sum_k \max(1 + D(\hat{\mathbf{x}}_i^k, k), 0), \tag{6}$$

$$l_i^{\text{adv-g}} = \sum_k -D(\hat{\mathbf{x}}_i^k, k), \tag{7}$$

where D is the discriminator (real/fake predictor) regularized by the spectral normalization [22]. $l_i^{\text{adv-d}}$ in Eq. (6) denotes the discriminator loss trying to distinguish the real and fake samples, by learning to predict the real pair (x_i, y_i) as ≥ 1 and the fake pair $(\hat{\mathbf{x}}_i^k, k)$ as ≤ -1. As the adversary against the discriminator, the generator loss $l_i^{\text{adv-g}}$ in Eq. (7) tries to push the prediction of the fake pair $(\hat{\mathbf{x}}_i^k, k)$ to be greater than 0 as far as possible, making the fake samples more likely to be realistic as well as in the expected age group.

Overall, the objective of the proposed S[2]GAN is

$$\min_{E,G,\{\mathbf{w}_k\}} \sum_i \lambda_1 l_i^{\text{age}} + \lambda_2 l_i^{\text{L1}} + l_i^{\text{adv-g}}, \tag{8}$$

$$\min_D \sum_i l_i^{\text{adv-d}}, \tag{9}$$

where λ_1 and λ_2 are the hyper-parameters for balancing the losses, and these two objectives are optimized iteratively.

3.3. Discussion

Continuous Aging In the proposed S[2]GAN, multiple linear transforms on the unique aging basis are used to generate age representations for each age group, so naturally they can be interpolated, resulting in continuous aging faces, i.e.,

$$\hat{\mathbf{x}}^{k\alpha} = G(\mathbf{B}(\alpha \mathbf{w}_k + (1-\alpha)\mathbf{w}_{k+1})),\ \alpha \in [0, 1]. \tag{10}$$

Compared to the existing methods which can only generate images with discrete age groups [1, 17, 40, 42, 19], the proposed S[2]GAN with continuous aging is more favorable and practical.

Lower Computational Cost As seen from Fig. 2, the proposed S[2]GAN needs only one model for all target ages. Compared to the existing methods such as [42, 19] which need $n(n-1)$ models for n age groups, the proposed method saves much storage and memory consumption. Besides, since the personalized basis is shared across ages, it takes much less time than most existing models [17, 40, 42, 19] to generate images of all n age groups. The computational consumption of different methods is shown in Table 1. Furthermore, the S[2]-module in Fig. 2(a) is orthogonal to these methods, and therefore it can be inserted into the generators of these methods, to reduce their computational consumption and enable the continuous aging, while preserving their own advantages.

Method	# Models	Prediction Time
Li et al. [17]	1	$nt_e + nt_d$
Wang et al. [40]	1	
Yang et al. [42]	$n(n+1)$	
Liu et al. [19]	$n(n+1)$	
Ours	1	$t_e + nt_d$

Table 1. Computational consumption for n age groups. t_e and t_d denote the prediction time of encoder and decoder respectively.

4. Experiments

Datasets We adopt MORPH [30] and CACD [5] datasets to evaluate the proposed S[2]GAN. MORPH contains 55,349 color images of 13,672 subjects with age annotations ranging from 16 to 77 years old. CACD contains 163,446 color images of 2,000 celebrities with age annotations ranging from 14 to 62 years old. For both MORPH and CACD, we randomly select 80% of the images as the training set and the rest 20% as the testing set. Following [40, 42] which divide ages into groups by every 10 years, for MORPH, we separate the images into 5 groups according to the age: 11-20, 21-30, 31-40, 41-50, and 50+. For CACD, since the images within 11-20 is significantly less than the other age groups, we separate the images into 4 groups: 11-30, 31-40, 41-50 and 50+.

Competitors For fair comparisons, the recent state-of-the-arts including CAAE [45] and IPCGAN [40] are trained by their official codes under the same protocol as the proposed method. We also make the comparison with CONGRE [34], HFA [43], GLCA-GAN [17], Yang et al. [42] and Liu et al. [19] by directly referring to their results in the papers, since these methods have no released codes and are hard to be reproduced with fair accuracy.

Implementation Details CycleGAN [46] architecture is adapted for the age generator. Specifically, with an 256×256 input, an architecture with 2 stride-2 convolutions followed by 6 residual blocks and 1 1×1 convolution are used as the encoder. Feature maps are extracted by the encoder, then they are equally divided into 256 feature blocks along the channel axis, with each representing a basis vector. An architecture with 3 residual blocks followed by 2 stride-$\frac{1}{2}$ convolutions is used as the age representation decoder. For the age group classifier, we train a ResNet-50 [10] and a VGG16 [33], then fix and ensemble them for stronger age supervision. For the discriminators, an archi-

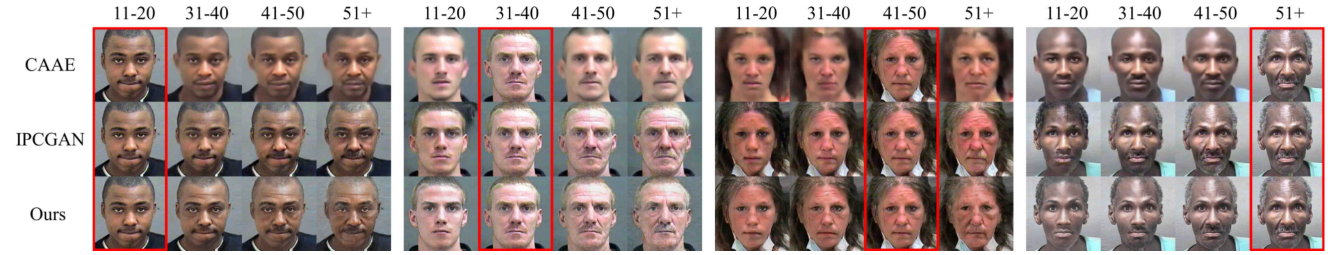


Figure 3. Comparisons on MORPH [30] among CAAE [45], IPCGAN [40] and our S^2GAN. The input images are wrapped in red boxes.

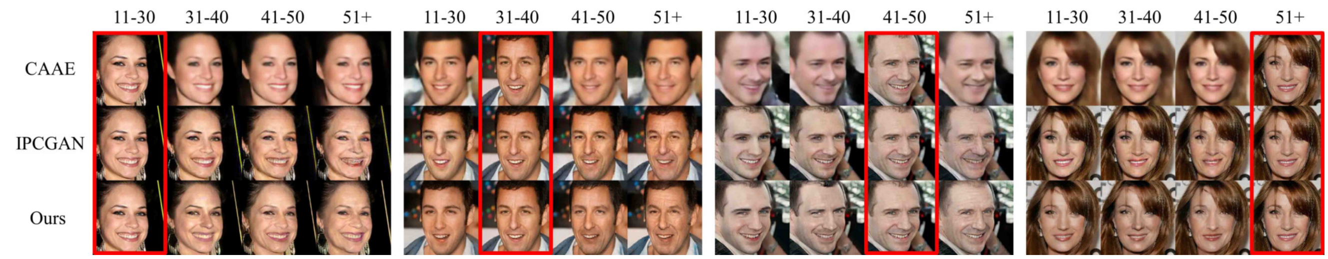


Figure 4. Comparisons on CACD [5] among CAAE [45], IPCGAN [40] and our S^2GAN. The input images are wrapped in red boxes.

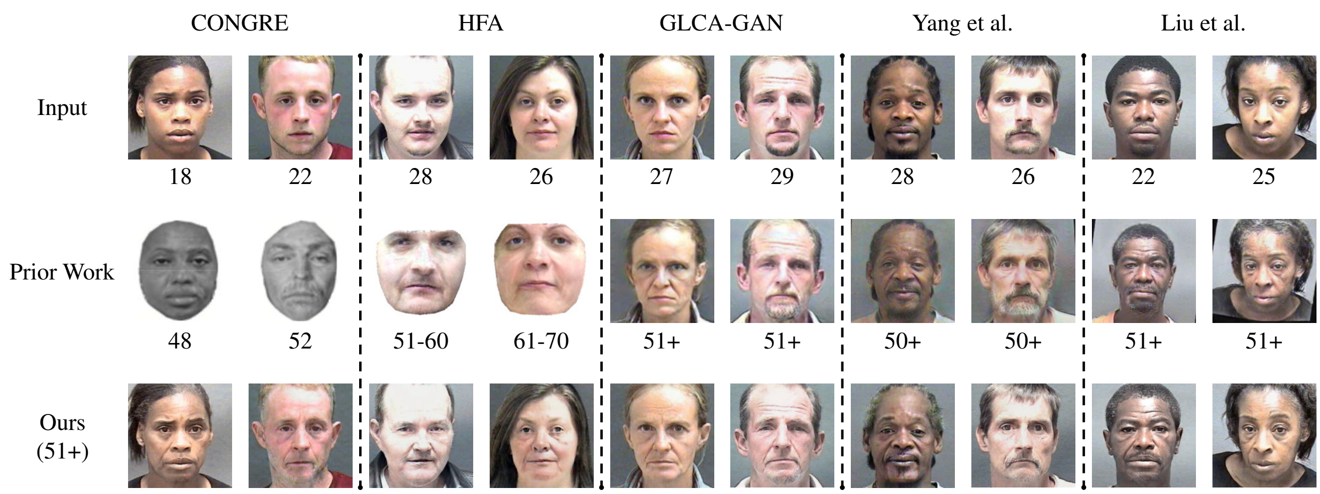


Figure 5. Comparisons with CONGRE [34], HFA [43], GLCA-GAN [17], Yang et al. [42] and Liu et al. [19].

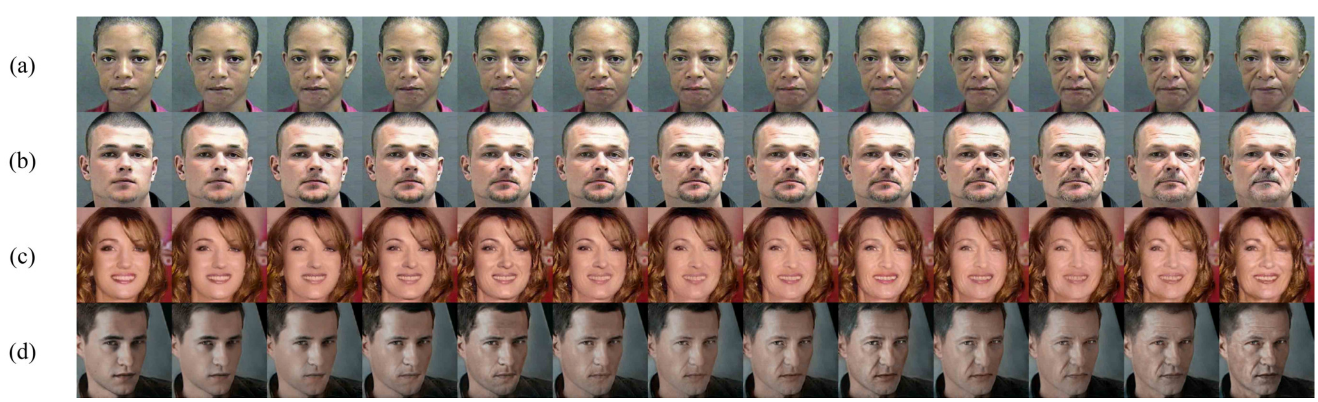


Figure 6. Continuous face aging of the proposed S^2GAN.

Figure 7. Face aging details of the proposed S^{2}GAN.

Method	11-20	21-30	31-40	41-50	51+
CAAE [45]	57.7%/22.0/4.8	53.9%/26.1/4.9	58.7%/30.1/4.0	6.0%/34.6/4.4	5.6%/40.6/5.5
IPCGAN [40]	63.1%/21.4/5.2	48.9%/28.7/5.6	75.7%/35.9/4.7	79.0%/44.8/4.1	56.4%/51.1/4.6
Ours	**95.1%**/18.2/1.4	**93.3%**/25.8/2.7	**92.3%**/35.4/2.8	**95.0%**/45.2/2.5	**89.3%**/53.6/2.7

Table 2. Aging accuracy/mean age of generations/std. on MORPH [30].

Method	11-30	31-40	41-50	51+
CAAE [45]	61.8%/29.6/7.3	43.8%/33.6/7.3	37.9%/37.9/7.3	11.0%/41.9/7.5
IPCGAN [40]	81.9%/27.4/5.1	70.7%/36.2/5.1	74.5%/44.7/4.4	75.6%/52.5/3.9
Ours	**97.2%**/24.0/3.2	**94.9%**/36.0/2.5	**97.2%**/45.7/2.2	**95.2%**/55.3/2.5

Table 3. Aging accuracy/mean age of generations/std. on CACD [5].

tecture with 7 convolutions with strides of 2, 1, 2, 2, 1, 2, 2 respectively followed by 2 fully connected layers is adopted as the global discriminator, while another architecture with 6 convolutions with strides of 2, 1, 2, 2, 1, 2 respectively, is adopted as the local discriminator. Besides, the features from the VGG16 classifier are inserted into the discriminators to enhance the supervision on aging details [42]. Please refer to the supplementary material for more details about the network architectures.

The coefficients in Eq. (8) and (9) are set as $\lambda_1 = (1 \text{ for MORPH and } 10 \text{ for CACD})$ and $\lambda_2 = 5$ at the first 10 epochs, and then λ_2 is reduced to 1.25 at the next 15 epochs. The networks are trained by Adam solver ($\beta_1 = 0.5, \beta_2 = 0.999$) [16] with the batch size of 3 and the learning rate of 0.0002.

4.1. Qualitative Analysis

Face Aging/Rejuvenation The aging results of CAAE [45], IPCGAN [40] and the proposed S^{2}GAN are shown in Fig. 3 and Fig. 4. As seen, CAAE tends to generate blurry images probably because its latent manifold is constrained as a simple distribution (e.g., uniform). IPCGAN achieves a better result on fidelity and identity preservation. However, IPCGAN produces artifacts in several cases such as the rejuvenation to 11-20 of the last object in Fig. 3. The proposed S^{2}GAN generate the aging faces with best visual quality, i.e., correct age, well preserved identity and high fidelity. Moreover in Fig. 5, we compare the proposed S^{2}GAN to some other methods including CONGRE [34], HFA [43], GLCA-GAN [17], Yang et al. [42] and Liu et al. [19]. As can be seen, compared to the traditional approaches CONGRE and HFA, our approach generates better and clearer facial details. Compared to the deep approaches such as Yang et al. [42] and Liu et al. [19], which need $n(n+1)$ models for n age groups, our approaches achieve comparable performance with a unique model.

Continuous Face Aging As mentioned in Sec. 3.3, the proposed S^{2}GAN is naturally applicable for continuous face aging with interpolated aging transforms. The continuous aging results are shown in Fig. 6, from which we can see the proposed S^{2}GAN generates continuously aging images with satisfying visual effect, such as the laugh lines shown in Fig. 6(a), mustache in Fig. 6(b), wrinkles in Fig. 6(c), and hair in Fig. 6(a).

Aging Details The aging details of different facial parts are shown in Fig. 7. As can be seen, the proposed method generates smooth aging changes with high fidelity for different parts such as gradually getting more and deeper forehead wrinkles, longer and deeper laugh lines, or thinner lip, while well keeping the invariant facial details such as scars.

4.2. Quantitative Analysis

Besides the visual results, we also compare all methods in terms of quantitative evaluation for aging accuracy, identity preservation, and fidelity.

Aging Accuracy A well trained continuous age predictor with ResNet-101 [10] architecture by mean-variance loss [26] is adopted to predict the ages of the generated face image. An aging image is considered correct only if its pre-

Method	Average of All Pairs	Hardest Pair	Easiest Pair
CAAE [45]	79.61%	(test,51+): 28.47%	(11-20,21-30): 100%
IPCGAN [40]	98.95%	(11-20,51+): 86.80%	(11-20,21-30): 100%
Ours	**99.69%**	(11-20,51+): 96.08%	(11-20,21-30): 100%

Table 4. Evaluation of identity preservation in terms of face verification rates on MORPH [30].

Method	Average of All Pairs	Hardest Pair	Easiest Pair
CAAE [45]	60.88%	(test,51+): 2.00%	(41-50,51+): 99.97%
IPCGAN [40]	91.40%	(11-30,51+): 62.98%	(41-50,51+): 99.98%
Ours	**98.91%**	(11-30,41-50): 94.08%	(41-50,51+): 99.96%

Table 5. Evaluation of identity preservation in terms of face verification rates on CACD [5].

Method	MORPH	CACD
CAAE [45]	47.7	44.2
IPCGAN [40]	10.4	9.1
Ours	**9.3**	**8.4**

Table 6. Evaluation of fidelity in terms of Fréchet Inception Distance (FID), lower is better.

Method	MORPH	CACD
CAAE	1%	1%
IPCGAN	22%	33%
Ours	**77%**	**66%**

Table 7. The proportion of being chosen as the best in the user study.

dicted age falls into the expected age group, and the aging accuracy is calculated as the percentage of correct aging images. In Table 2 and Table 3, we show the aging accuracy, the mean age and the standard deviation of the generated images for each target age group. As seen, our aging accuracies are much better than the competitors, and the mean age of each group is very close to the group center.

Identity Preservation Face verification is adopted to evaluate the identity preservation of aging results. We conduct the verifications between the test images and the generated faces, i.e., (test,11-20), (test,21-30), ..., (test,51+). We also conduct the verifications between every pair of aging results, i.e., (11-20,21-30), (11-20,31-40), ..., (41-50,51+). Following [42], we adopt an online face analysis tool Face++[1] to obtain the verification scores and the threshold is set as 76.5(@FAR=1e-5). The average, the highest (hardest) and the lowest (easiest) verification rates are shown in Table 4 and 5. As can be seen, our S^{2}GAN achieves the best average face verification rate demonstrating the superior identity preservation of our method. Besides, our method performs very well even on hard pairs, e.g., in Table 4, (11-20,51+) is the hardest pair of both IPCGAN and our S^{2}GAN, while our S^{2}GAN has about 9% improvement over IPCGAN.

Generation Fidelity Fidelity is an important aspect of evaluating any image generation task. We adopt an effective metric - Fréchet Inception Distance (FID) [12, 13, 20] to evaluate the quality of the aging results. The FIDs of the competing methods are reported in Table 6 and the lower FID indicates the better the generation. As can be seen, the proposed S^{2}GAN method achieves a lower FID, therefore, the better fidelity than the competitors.

User Study To evaluate our method under the human perception, we asked 20 volunteers to evaluate the aging results of CAAE [45], IPCGAN [40], and our S^{2}GAN. Specifically, 400 random images (200 from MORPH and 200 from CACD) are chosen as input, and the three methods are respectively used to generate aging images of all age groups for each input. Then, each volunteer is asked to choose the best aging method for each input by considering the aging accuracy, identity preservation, and image quality. Table 7 shows the proportion of each method to be chosen as the best, averaged over all volunteers.

5. Conclusion and Future Works

In this work, we suggest a new perspective of natural aging process, i.e., faces at different ages of a specific person are derived from a same personalized basis, while the age-specific transforms from the aging basis to the target aged faces are shared among all individuals. Based on this perspective, we propose an effective and efficient S^{2}GAN approach for face aging with favorable aging accuracy, identity preservation, fidelity and low computational cost, which is also applicable for continuous face aging. Besides, the S^2-module in our approach can be used as a plug-in module in many GAN-based approaches to reduce their computational consumption and enable the continuous aging.

However, the aging factors are extremely complicated, and then a question is raised: Is a single face image sufficient to infer the whole personalized aging basis/factors? In future works, therefore, we will investigate how to establish a complete personalized aging basis through multiple images for each individual at different ages.

Acknowledgement This work is partially supported by the 973 Program (No. 2015CB351802) and Natural Science Foundation of China (No. 61772496 and No. 61532018).

[1]Face++ Research Toolkit. `http://www.faceplusplus.com`

References

[1] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In *IEEE International Conference on Image Processing*, 2017.

[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*, 2002.

[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*, pages 15–27. 2000.

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

[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*, 2014.

[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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[7] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Tien D Bui. Longitudinal face aging in the wild-recent deep learning approaches. *arXiv:1802.08726*, 2018.

[8] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. In *Advances in Neural Information Processing Systems*, 2014.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[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*, 2019.

[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*, 2017.

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018.

[14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[15] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz. Illumination-aware age progression. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[16] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[17] 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*, 2018.

[18] Si Liu, Yao Sun, Defa Zhu, Renda Bao, Wei Wang, Xiangbo Shu, and Shuicheng Yan. Face aging with contextual generative adversarial nets. In *ACM International Conference on Multimedia*, 2017.

[19] Yunfan Liu, Qi Li, and Zhenan Sun. Attribute enhanced face aging with wavelet-based generative adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[20] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. Are gans created equal? a large-scale study. In *Advances in Neural Information Processing Systems*, 2018.

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

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

[23] Chi Nhan Duong, Kha Gia Quach, Khoa Luu, Ngan Le, and Marios Savvides. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In *IEEE International Conference on Computer Vision*, 2017.

[24] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Tien D Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[25] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International Conference on Machine Learning*, 2017.

[26] Hongyu Pan, Hu Han, Shiguang Shan, and Xilin Chen. Mean-variance loss for deep age estimation from a face. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[27] Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[28] Narayanan Ramanathan and Rama Chellappa. Modeling shape and textural variations in aging faces. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.

[29] Narayanan Ramanathan, Rama Chellappa, and Soma Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages and Computing*, 20(3):131–144, 2009.

[30] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE*

International Conference on Automatic Face and Gesture Recognition, 2006.

[31] Duncan A Rowland and David I Perrett. Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications*, 15(5):70–76, 1995.

[32] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *IEEE International Conference on Computer Vision*, 2015.

[33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.

[34] 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.

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

[36] 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.

[37] 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.

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

[39] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe. Recurrent face aging. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[40] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[41] 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. 1994.

[42] 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*, 2018.

[43] 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.

[44] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *IEEE International Conference on Computer Vision*, 2017.

[45] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[46] 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*, 2017.