

# Re-Aging GAN: Toward Personalized Face Age Transformation

Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park

Dept. of Information and Communication Engineering, Inha University, Incheon 22212, Korea

{farhodfm, csehong, pik}@inha.ac.kr

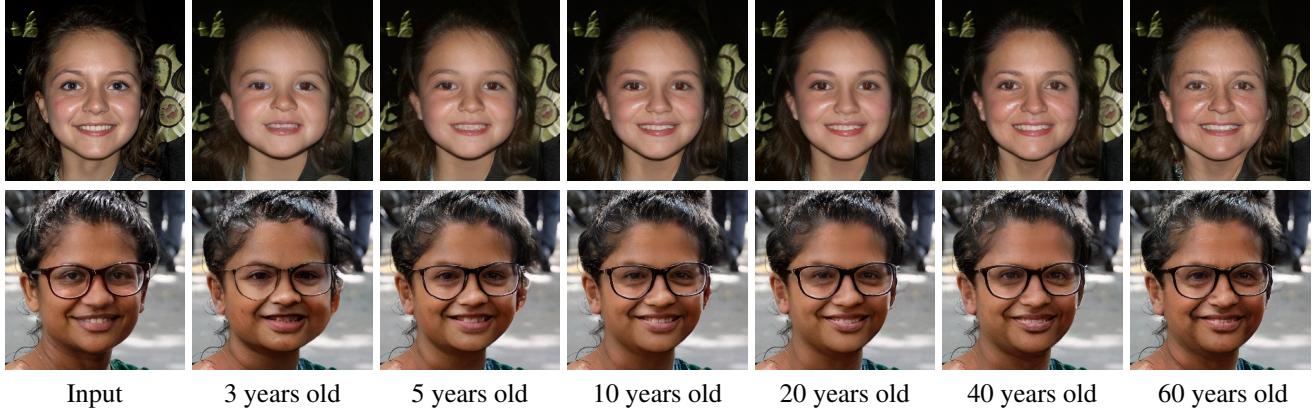


Figure 1: Continuous age transformation results by our method on the FFHQ dataset and synthesized image of StyleGAN2. The first column shows the input images, whereas the remaining ones are generated images. Note that our model is trained only on the FFHQ; thus, the results in the first row show excellent generalization capability of the proposed method.

## Abstract

*Face age transformation aims to synthesize past or future face images by reflecting the age factor on given faces. Ideally, this task should synthesize natural-looking faces across various age groups while maintaining identity. However, most of the existing work has focused on only one of these or is difficult to train while unnatural artifacts still appear. In this work, we propose Re-Aging GAN (RAGAN), a novel single framework considering all the critical factors in age transformation. Our framework achieves state-of-the-art personalized face age transformation by compelling the input identity to perform the self-guidance of the generation process. Specifically, RAGAN can learn the personalized age features by using high-order interactions between given identity and target age. Learned personalized age features are identity information that is recalibrated according to the target age. Hence, such features encompass identity and target age information that provides important clues on how an input identity should be at a certain age. Experimental result shows the lowest FID and KID scores and the highest age recognition accuracy compared to previous methods. The proposed method also demonstrates the visual superiority with fewer artifacts, identity preservation, and natural transformation across various age groups.*

## 1. Introduction

A face age transformation task is dedicated to learning age progression or regression of a given face according to the target age. Here, target age implies an explicit conditioning factor that guides the transformation to produce facial images with a certain age. That is, we can set any target age for an input face image, and expect to have an output face depicting the target age characteristics, as shown in Figure 1. Ideally, age transformation models should satisfy the following properties. *First*, a model should take into account the identity of the person while progressing/regressing the face age and sustain it mostly unaltered, *i.e.*, identity preservation. *Second*, a model should be able to generate natural-looking faces corresponding to the target age across various age groups.

In this regard, a number of works on face age transformation have been introduced [1, 48, 46, 39, 12, 42, 33, 47, 24, 45]. These methods, on the basis of powerful generative adversarial networks (GANs) [10, 30], train deep neural networks to perform a robust age transformation of the input face. Aside from this, several methods [43, 46] adopted additional mechanisms (*i.e.*, networks and constraints) for identity preservation to ensure that face identity is unaltered during the age transformation process. However, even with

improved approaches, existing methods tend to generate images with visible artifacts and/or unnatural-looking faces which surely lowers down the image quality and its perception. Another important aspect that should be considered is a wide-range age transformation, specifically, an age regression process for rejuvenating input face. Most existing works scarcely address this process and more focus on progression. Although a few methods can operate on face age regression and provide good performance, their results still suffer from artifacts near and/or on face regions and contain no background due to the tight face cropping. Overall, even with such considerable efforts, few models cover all critical factors such as identity-preserved age transformation across a wide range of age groups.

In this paper, we consider all these important factors for face age transformation and propose a simple yet effective framework called Re-Aging GAN (RAGAN). Unlike recent methods, we endeavor to design RAGAN in a single framework for making model training easier and more scalable. Our generator comprises three sub-networks, namely encoder, age modulator, and decoder, each of which is designed with a specific role for face age transformation. As it is necessary to meet the identity preservation property as well as background for generating accurate and visually plausible transformed images, we make an encoder to detach the face region from the background such that it solely focuses on extracting identity features. We are aware that identity might be disrupted in the learning process without explicit guidance. Thus, we consider modulating identity features in agreement with the given target age by means of an age modulator plugged in-between encoder and decoder. Straightforward incorporation of this network aims to provide personalized age-aware features and so self-guide the decoding process. By leveraging such features, the decoder learns to smoothly add target age characteristics on identity features in an optimal manner. The decoder further improves the visual perception of images by mapping back the background. As the whole process focuses on the face region, it allows the generator to pay more attention to transforming the facial region and helps to avoid affecting the background (*i.e.*, color change, artifacts). For the discriminator part, we follow the advanced approaches [7, 29, 26] consisting of conventional age-related loss.

The main contributions of this work are as follows:

- We introduce a personalized self-guidance scheme that enables transforming the input face across various target age groups while preserving identity.
- We successfully perform face age transformation using only a single generative and discriminative model trained effectively in an unpaired manner.
- We qualitatively and quantitatively demonstrate the superiority of RAGAN over state-of-the-art methods through extensive experiments and evaluations.

## 2. Related Works

**Face aging with GANs** Many recent works extensively applied GAN approach in facial aging. A concept of conditional GANs (cGANs) [30] allowing to augment age characteristics in generated images is utilized in [1, 48, 43, 39, 24]. However, models lead to produce blurry images, and damage identity information [12, 47]. In turn, several methods [43, 46, 23, 22] provided improvements by adopting mechanisms such as L2 constraint, pre-trained identity classifier, *etc*. In addition to the identity preservation network, [22] also introduced wavelet transform to capture texture details for age synthesis. Similarly, wavelet information is used in [28, 27], but in a multi-scale manner in the discriminator. However, their results still suffer from noticeable artifacts [47]. The work of [12] proposed S<sup>2</sup>GAN which applies age-specific transforms on the encoded personalized aging basis to obtain and decode age-specific representation of the input face. Despite the plausible results, S<sup>2</sup>GAN has less ability to depict age progression [33].

To generate high-resolution images, [47] proposed a face age editing model operating in high-resolution (HRFAE). The method relies on a straightforward approach that re-weights the encoded features by means of the output of a single fully connected layer. Similarly, [9] proposed incorporating aging maps by using SPADE blocks [34] in the decoding process to obtain high-resolution face images. Although these methods maintain good identity preservation and age information, it is limited to perform mostly age progression than regression. LATS [33] presented a method for synthesizing lifespan age transformations. Their decoder operates in a similar manner as StyleGAN2 [17] and performs modulated convolutions on identity features while injecting the target age latent vector learned from the mapping network. Despite the impressive performance, LATS primarily focuses on face synthesis without using background information, which leads to unexpected artifacts.

**Image-to-Image Translation** In addition to the facial age-specific approaches, image-to-image translation [5, 6, 20, 35, 36, 44] could be applied for the face age transformation task. There have been several attempts to manipulate the facial age attributes in a multi-domain transfer fashion. To perform face age manipulation (*i.e.*, young to old, and vice-versa), StarGAN [6] and STGAN [25] assume that the age domain is distinct from other face-related domains. However, the appearance of the face due to the specific age is highly correlated with other facial attributes. For instance, when translating between different age groups using multi-domain transfer, styles that are not related to age, such as gender, accessories, and ethnicity can also be transferred. Besides, multi-domain transfer-based approaches struggle for modeling the shape (*i.e.*, head) deformation associated with age transformation. We refer to the work [33] providing extensive analysis on multi-domain transfer approaches.

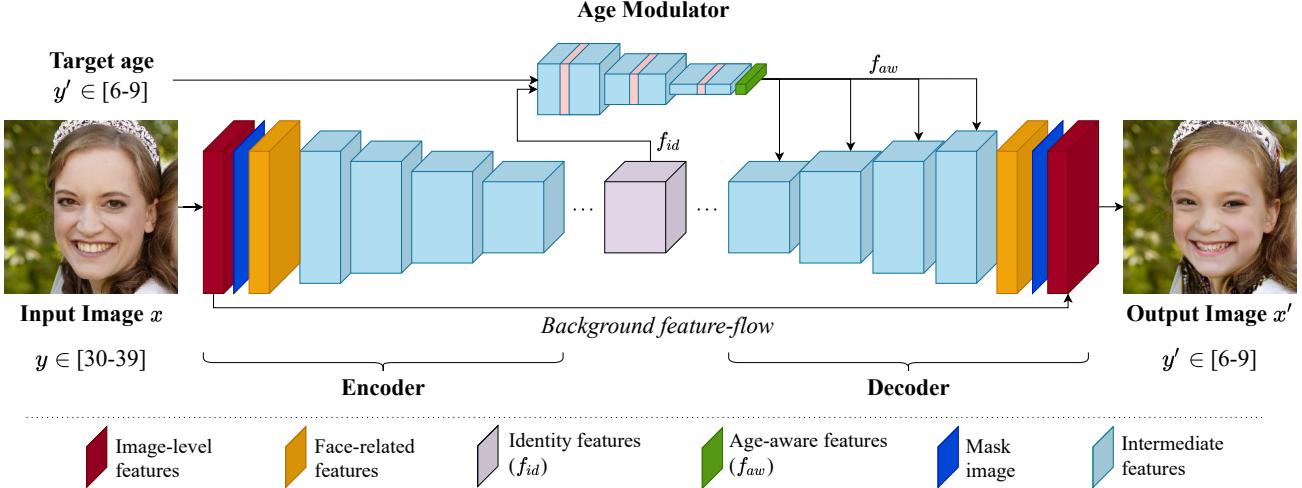


Figure 2: Overview of our proposed RAGAN framework. Given an input image, our framework transforms it to target age  $y'$  by means of self-guiding features that convey personalized age-aware information.

### 3. Proposed Method

#### 3.1. Overview

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be the sets of images and possible ages, respectively. Given a face image  $x \in \mathcal{X}$  and target age randomly drawn from  $y' \in \mathcal{Y}$ , our goal is to train a single generator  $G$  such that it can generate a face image  $x'$  of a particular age  $y'$  corresponding to the identity in  $x$ . In addition, we introduce an age modulator within  $G$  to reshape identity features by considering the target age and utilize it as self-guiding information.

In comparison to prior works, our main objective is to robustly transform face age as well as maximally retain age-unrelated information in  $x'$ , such as background, hairstyle, expression, etc. This means that our framework should preserve the age-unrelated information contained in the input image during the age transformation process. Therefore, we consider a simple strategy on encoder and decoder networks to share some of the valuable information.

#### 3.2. Proposed Framework

Our GAN-based framework is divided into the generator consisting of an encoder, a modulator, and a decoder, and the discriminator. Since our discriminator part follows the existing approaches [7, 29, 26] and is not the main focus of this study, only the generator part is described in the overall flow of Figure 2. Our generator makes use of the encoder-decoder architectural concept for image generation and is made of an identity encoder  $Enc$ , an age modulator  $AM$ , and a decoder  $Dec$ . In a superficial view, our encoder and decoder networks perform the same procedure as existing works, yet having a few modifications (see details below).

One of the differences between ours and other works is the integration of an additional sub-network at the bottleneck of the generator. By this network, we can obtain features providing information on how a particular person should look like at the age under the consideration. Given that such age-aware features are learned based on a given input image, then it can be used as self-guidance information in the further generation process.

#### 3.3. Identity Encoder

Given an image  $x$  for age transformation, our identity encoder  $Enc$  extracts identity-related features  $f_{id}$  of the image, where  $f_{id} = Enc(x)$ . Particularly, the encoder provides such features that supply facial structures at the local level and general information on face shape. These features are necessary to generate the same-looking face and thus have high importance, as discussed earlier. In turn, this importance intuitively leads us to focus on the face region only. Therefore, we propose to perform a masking operation after transferring  $x$  into the feature domain. To this, we utilize a network [4] trained on sophisticated mask-based dataset [21]. We deliberately perform masking at this stage such that we can obtain face as well as background-related features to operate on the face region only while maintaining background information simultaneously. At the architecture level, the identity encoder is designed to have an image-to-feature level convolutional layer followed by downsampling residual blocks [11].

#### 3.4. Age Modulator

Age modulator  $AM$  is constructed in the form of CNN, which is widely used in learning a low-dimensional vec-

tor of an input. It takes identity features  $f_{id}$  from the encoder and, by considering given age information  $y'$ , outputs its reshaped version  $f_{aw} = AM(f_{id}, y')$ , where  $f_{aw}$  is an element age-aware vector. To embed target age into  $AM$ , we add conditional batch normalization (CBN) layers used as a way to incorporate label information into the network [41, 8, 32, 31]. By doing so,  $AM$  learns optimal age-aware features for input identity, which enables satisfying both identity and age properties. Given that the network is integrated into  $G$ , it can be trained alongside the generator in an end-to-end manner. We implement  $AM$  as a set of downsampling layers with CBN technique producing a compact feature-vector used to modulate decoder layers.

### 3.5. Decoder

The decoder network  $Dec$  takes an identity feature alongside the age-aware features and produces age-transformed face image by  $x' = Dec(f_{id}, f_{aw})$ . Identity preservation should go along with a change in age characteristics. Thus, to satisfy both of these properties, we make use of age-aware features to self-guide the decoding process through modulation operations on unshaped identity features. The identity features  $f_{id}$  are modulated by adaptive instance normalization (AdaIN) layers [14, 16]. As discussed above, performing masking operation in  $Enc$  supplies background-related features through ad hoc connection. That is, we can use it to generate images with better visual perception prior to map learned feature representation into the image domain. However, instead of directly adding this information to face-related features, we perform a masking operation on learned face representations again to remove minor variations that occurred in the background due to the normalization and activation layers. Subsequently, we combine face- and background-related features and map them back to the image domain using the last convolutional layer. We construct this network by mirroring the encoder architecture with the replacement of downsampling to upsampling layers.

### 3.6. Discriminator

Our discriminator  $D$  follows works presented in [7, 29, 26] and performs a multi-task classification. Hence, the last fully connected layer has a number of output branches to classify multiple age classes. By performing binary classification, each of the branches learns to determine the validity of the image being real  $x$  or fake  $x'$  of its age domain.

## 4. Optimization

Our framework is designed to produce images where the identity of the input is preserved, whereas a target age is accurately represented. To ensure such ideal transform, three types of losses are used in our framework: adversarial, reconstruction, and cycle-consistency. The framework oper-

ates on three input information, an input image  $x$  and its corresponding age label  $y$ , and randomly sampled target age  $y'$  into which input should be transformed. Subsequently, given this information,  $G$  will produce age-transformed  $x'$ , reconstructed  $x_{rec}$ , and cycle-consistency images  $x_{cycle}$  as

$$x' = G(x, y'), \quad x_{rec} = G(x, y), \quad x_{cycle} = G(x', y) \quad (1)$$

**Adversarial Loss** As mentioned above, the output of the discriminator corresponds to the particular age domain. Hence, we can think that adversarial loss is conditioned on the age class. We use an adversarial loss formulated as:

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{x,y} [\log D_y(x)] + \mathbb{E}_{x,y'} [\log (1 - D_{y'}(x'))], \quad (2)$$

where  $D_y(\cdot)$  is the only output of  $D$  corresponding to the age  $y$ .

**Reconstruction Loss** While training  $G$ , we have to consider the case when the age  $y$  of the input image  $x$  and the target age  $y'$  belong to the same age-group ( $y = y'$ ). In this case, the age-transformed image  $x'$  should be as close as possible to the input  $x$ . Therefore, we add reconstruction loss which forces  $G$  to pay attention to such cases by

$$\mathcal{L}_{rec}(G) = \|x - x_{rec}\|_1 \quad (3)$$

**Cycle-Consistency Loss** We can train  $G$  to generate images that are realistic and accurate in terms of target age by minimizing the adversarial and reconstruction losses. However, such minimization (*i.e.*, (2) and (3)) cannot help to force  $G$  to learn maintaining the identity all over the age transformation process while changing the age-characteristics of input  $x$ . To address it, we include cycle-consistency loss [18, 49, 6] in the form of given below:

$$\mathcal{L}_{cyc}(G) = \|x - x_{cycle}\|_1 \quad (4)$$

**Full Objective** The overall objective function considered to optimize  $G$  and  $D$  is as follows:

$$\min_G \max_D \lambda_{adv} \mathcal{L}_{adv}(G, D) + \lambda_{rec} \mathcal{L}_{rec}(G) + \lambda_{cyc} \mathcal{L}_{cyc}(G) \quad (5)$$

where  $\lambda_{adv}$ ,  $\lambda_{rec}$ , and  $\lambda_{cyc}$  are the weights for each loss to balance their influences on training.

## 5. Experimental Setup

We train our model with a batch size of 8 for 30 epochs on a single NVIDIA Titan RTX GPU. As an optimizer, we use Adam [19] with momentum parameter settings  $\beta_1 = 0.0$  and  $\beta_2 = 0.99$  and a learning rate of  $10^{-4}$ . We also add R1 regularization [29] in our training stage. In addition, we apply the learning rate scheduler for both the generator

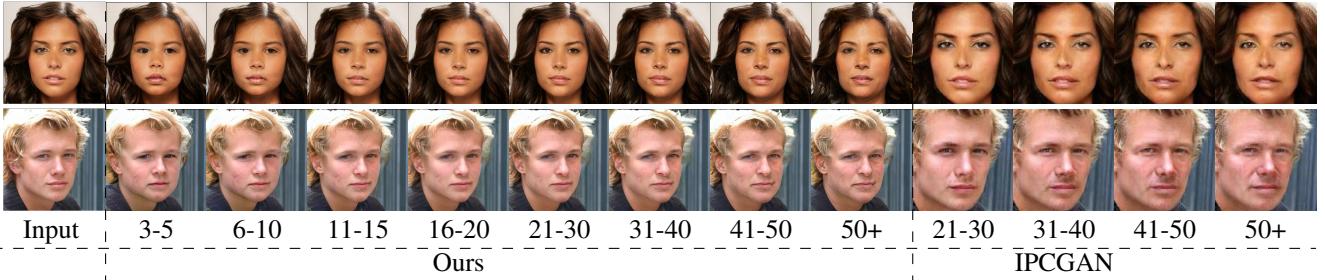


Figure 3: Qualitative comparison with IPCGAN [43].

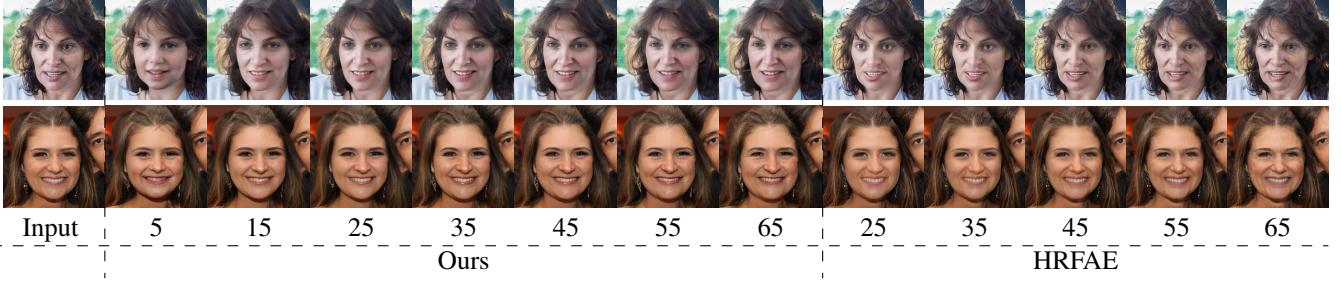


Figure 4: Qualitative comparison with HRFAE [47].

and the discriminator. In the beginning of 10 epochs, we train our model with  $\lambda_{rec} = 10$ ,  $\lambda_{cyc} = 1$ , and  $\lambda_{adv} = 1$  for reconstruction, cycle-consistency and adversarial losses, respectively. Thereafter, we reduce  $\lambda_{rec}$  to 1 because such training leads to better results.

**Datasets** For training our framework, we use FFHQ [16] dataset labelled for age transformation task by [33]. This dataset provides images of 70,000 people in 10 age groups. We follow the strategy of [33] to form training and test splits, as well as pruning the dataset images that have a low label-confidence score. In addition, this dataset provides face semantic maps. We use this semantic information to mask out images and so separate face-region and background information for reliable age transformation. To evaluate the generalization capability of our model to unseen images, we use CelebA-HQ [15] and CACD [3] datasets in our qualitative evaluations. Both of these datasets provide face images of celebrities with diversity in age, pose, illumination and expression. All images used in our training and evaluations are at the resolution of  $256 \times 256$  which is a commonly used resolution in existing works.

**Evaluation metrics** In quantitative evaluations, we consider evaluating the identity preservation and age modification of generated images. For identity preservation, we introduce to utilize Frechét inception distance (FID) [13] and Kernel inception distance (KID) [2] metric which evaluates the discrepancy between distributions. Additionally, to assess how accurately age is introduced into the input image (*i.e.*, the correctness of age transformation), we follow existing works and use age recognition accuracy [43, 42].

## 6. Experimental Results

### 6.1. Comparison with Previous Methods

We compare our framework to the recent state-of-the-art methods in face age transformation, namely, HRFAE [47], LATS [33]. In addition, we consider IPCGAN [43] as it was designed to preserve identity information in facial aging.

We start the qualitative evaluation by comparing our results on the CACD dataset with a comparison to IPCGAN. For this purpose, we use IPCGAN model which considers age transformation in these age-groups: 11-20, 21-30, 31-40, 41-50, and 50+. Similarly, along with images of these age categories, we also show our results in 3-5, 6-10 younger classes to demonstrate the potential of our method in performing full age transformation of CACD images as shown in Figure 3. It is noteworthy to mention that we use our model trained on the FFHQ dataset for this qualitative comparison. From Figure 3, we can see our method can generate more realistic transformations whereas IPCGAN has some degradation on the image quality and artifacts especially in the eye-regions of older images.

We compare our method and recently introduced HRFAE on the FFHQ dataset. In comparison to ours, HRFAE performs age transformation on the age range of 20-77 years old. Following their protocol, we perform age editing in given target ages for both; but, we also include our results in younger ages to provide a full sense of continuous transformation. In Figure 4, the results of HRFAE are more in texture level which means only facial appearance is modified. In comparison, we can also deform the head

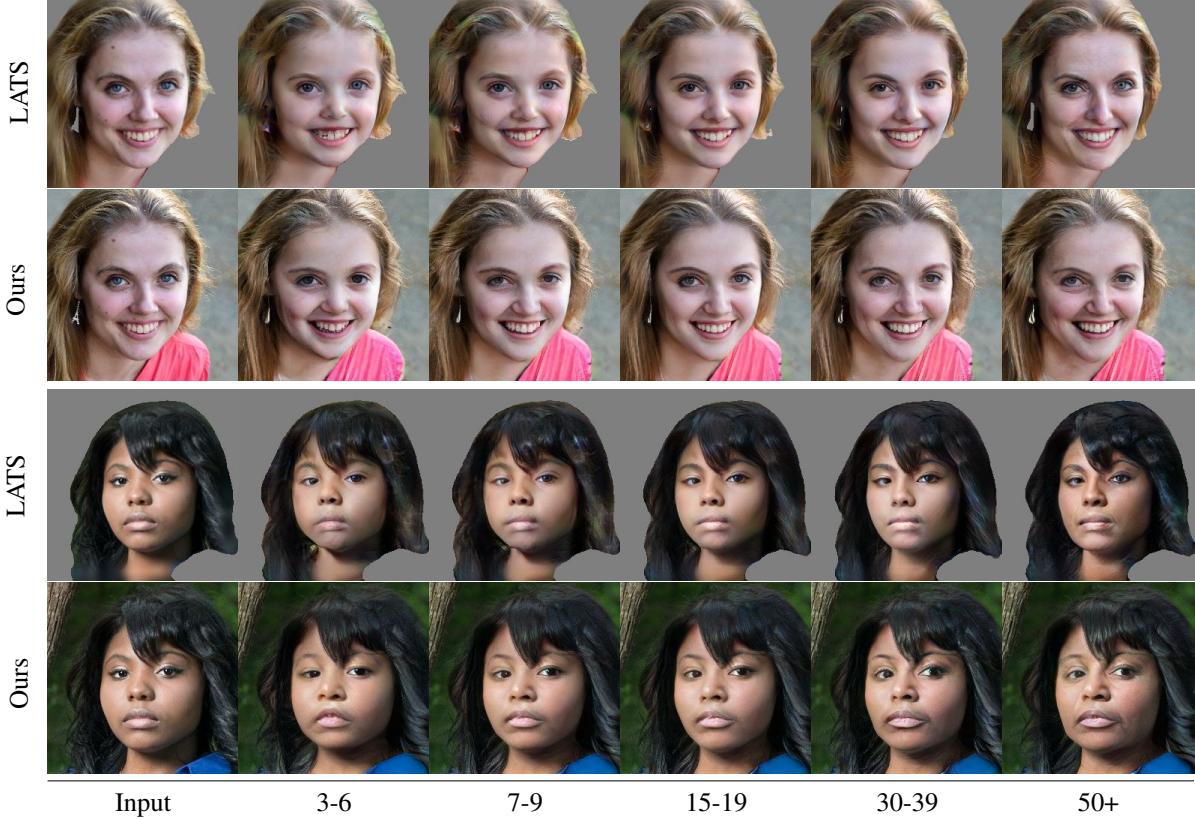


Figure 5: Qualitative comparison with LATS [33].

shape, such that faces align with the human growth process.

We also qualitatively evaluate the performance of our method against LATS, the state-of-the-art transform model as shown in Figure 5. One of the clearest differences between these two methods is the background. As LATS considers only the face region itself, it loses such necessary information that we consider valuable for visual perception. By the rule, losing one helps LATS to stress age transformation on the faces (especially, at the very young ages), but at the additional cost of introducing artifacts on the face borders. From this, we can infer that keeping background not only facilitates better perception but helps to avoid such visible artifacts. Besides, there are visible artifacts on mouth regions of LATS images, whereas our model smoothly generates mouth and teeth regions according to the target age. These results demonstrate the efficiency of self-guidance from  $AM$ , which provides such reliable features conveying both identity and target age information.

## 6.2. Generalization Ability

One of the important aspects of all deep learning-based methods is to have the ability to generalize to unseen data. To this, we perform the experiment on the images of CelebA-HQ [15]. We should stress that our model has

not seen images of this dataset in the training stage, and all the results were generated solely using a model trained on the FFHQ. We demonstrate age transformation results of CelebA-HQ images in Figure 6 (two upper rows). By looking at the generated images, we can see the age transformation flow of a person in a smooth way. It is notable that while preserving the identity and changing age characteristics, the background information is kept that can be seen in the banner behind of person at the top row. We observe the cases where it is difficult to notice age transformations due to the make-up, and as it is one of the characteristics of the distribution, the model also learns it accordingly. Nevertheless, this issue is common in age-related tasks, including conventional age estimation/classification.

Also, we test how the proposed framework could generalize to the fake images. To this end, we use synthesized images<sup>1</sup> by StyleGAN2 [17]. Similarly, the model has never seen such fake images in its training. The performance of our age transformation framework is provided in Figure 6 (two bottom rows). From the figure, we can confirm our model performs well in synthesized images as in CelebA-HQ. Overall results demonstrate the generalization capability of our model to the unseen data.

<sup>1</sup>[www.thispersondoesnotexist.com](http://www.thispersondoesnotexist.com)

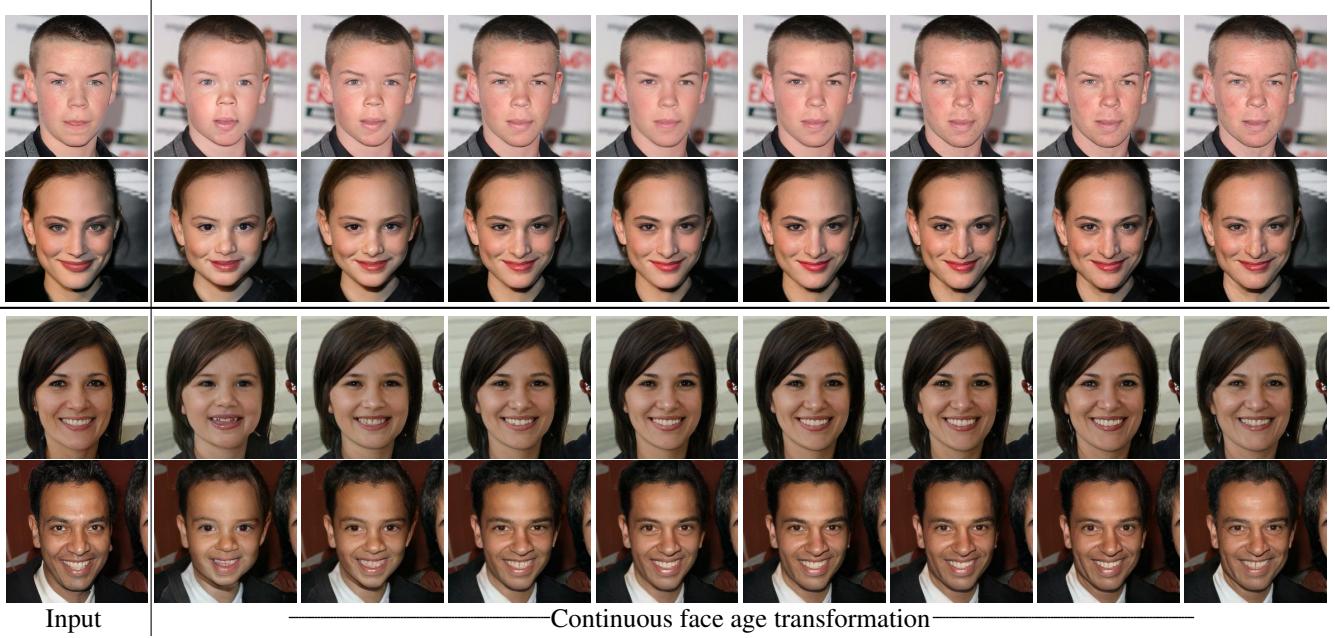


Figure 6: Generalization capability of our model on images of CelebA-HQ [15] dataset (two upper rows) and on synthesized images of StyleGAN2 [17] (two bottom rows). The first column is the input whereas the others are our results.

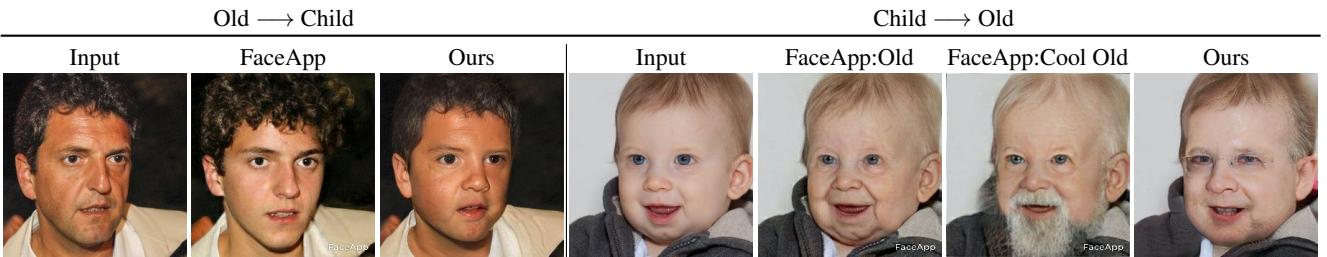


Figure 7: Qualitative comparison with *FaceApp* application in face age regression and progression.

### 6.3. Comparison with *FaceApp*

Face manipulation tools on mobile have become useful for editing facial attributes. One of the representative application that allows age editing is *FaceApp*<sup>2</sup>. This app provides several filters for making a person look younger or older. Here, we compare our results with *FaceApp* in rejuvenating and aging. For the rejuvenating process, we use *FaceApp* filters named Child; whereas for the person aging process, we set the filters of Old and Cool Old. Figure 7 demonstrates side-by-side comparison of our and *FaceApp* results. We can clearly see that the *FaceApp* images have different skin tones compared with the input, the results are brighter. In age rejuvenation, it is noticeable that our results have a better childish view compared to *FaceApp*. In the case of the aging process, except for a few minor details like hair color and adding beard, we do not find any distinct difference between Old and Cool Old filters from

*FaceApp*. Although the *FaceApp* can add aging wrinkles, it conveys less visual perception for such transformation. Interestingly, our transformation introduces a bristle and attempts to align glasses.

### 6.4. Identity Preservation

We take a different strategy on evaluating identity preservation. Specifically, we use FID [13] and KID [2] metrics for quantitative comparison. As mentioned above, the model needs to learn the case where the input and target ages belong to the same group and produce an output close to the input. If the model cannot regenerate the given input, then we can consider that identity information is affected by transformation, and the identity is thus deteriorated or even lost. Based on this assumption, we are interested in estimating the identity preservation/deviation of the model. We can estimate such factors through FID and KID as these metrics assess the discrepancy between two distributions. Also, we can get insight on the quality of generated images.

<sup>2</sup>[www.faceapp.com](http://www.faceapp.com)

Method	Age-groups									
	0-2	3-6	7-9	10-14	15-19	20-29	30-39	40-49	50+	
	FID ↓									
IPCGAN	-	-	-	-	-	45.35	59.12	58.87	52.65	
HRFAE	-	-	-	-	-	41.67	54.90	56.17	48.65	
LATS	75.87	77.14	79.91	60.38	59.43	43.28	57.41	55.76	47.12	
Ours	<b>71.61</b>	<b>74.64</b>	<b>75.87</b>	<b>54.47</b>	<b>57.15</b>	<b>38.27</b>	<b>51.93</b>	<b>50.47</b>	<b>45.63</b>	
	KID ↓									
IPCGAN	-	-	-	-	-	0.010	0.011	0.009	0.008	
HRFAE	-	-	-	-	-	0.006	0.009	0.005	0.006	
LATS	0.016	0.013	0.008	0.007	0.007	0.005	0.005	0.005	0.003	
Ours	<b>0.011</b>	<b>0.010</b>	<b>0.006</b>	<b>0.006</b>	<b>0.005</b>	<b>0.002</b>	<b>0.003</b>	<b>0.002</b>	<b>0.002</b>	

Table 1: Identity preservation scores estimated by FID and KID metrics. Lower scores are better.

To this end, we perform a re-generation process using images of the test set in each age domain and compare them with their real counterparts. Specifically, we evaluate the model in each age-group separately. For each test image from a source age-group, we transform back it into its own age-group. We then extract features from Inception-V3 [40] trained on ImageNet for real  $x$  and the generated images  $x'$ . Afterward, we calculate FID and KID between real and generated distributions for their respective age-group. As IPCGAN and HRFAE operate at higher ages, we provide the results accordingly. Lower FID and KID scores indicate that input images are better regenerated, and thus the model has the capability of preserving the identity.

We present the results of our comparison in Table 1. We found that our method performs well in all age groups by showing lower discrepancy compared to LATS. It is notable that such results also depict the better quality of generated images. In both metrics, our method achieved better results compared to others and thus showing its identity preservation capacity.

## 6.5. Age Recognition

This task is specifically designed to evaluate the reflectance level of the target age on the generated image. By performing such an assessment, we can verify how accurately the transformed image depicts its new-age domain. For this purpose, we use VGG16 [38] network trained on the age dataset [37]. We consider test images of 20-29 age-group from the FFHQ dataset to be source images for transforming their ages into 0-2, 3-6, 7-9, 10-14, 30-39, 40-49, and 50+ groups. We choose this particular age-group to assess the age transformation in both aging and rejuvenating tasks. Similar to the previous calculation, we perform recognition for each age-group separately. We report the accuracy based on the ratio of the number of samples recognized correctly to the total number of samples. Similar to the tests on identity preservation, we provide results of IPCGAN and HRFAE for higher ages. The higher recognition accuracy indicates the better introduction of target age

Method	0-2	3-6	7-9	10-14	15-19	30-39	40-49	50+
LATS	<b>57.98</b>	<b>55.75</b>	53.14	56.49	54.21	58.89	62.27	64.66
Ours	55.62	53.46	<b>56.78</b>	<b>59.11</b>	<b>57.04</b>	<b>59.95</b>	<b>63.16</b>	<b>65.47</b>

Table 2: Quantitative comparison between LATS and our method under wide-range age-group classification. The accuracy (%) is presented.

Method	15-19	30-39	40-49	50+
IPCGAN	52.01	55.41	53.86	55.64
HRFAE	55.37	58.10	57.36	59.47
Ours	<b>57.04</b>	<b>59.95</b>	<b>63.16</b>	<b>65.47</b>

Table 3: Age-group classification accuracy on the generated images of higher age-groups by IPCGAN, HRFAE, and ours. The accuracy (%) is presented.

characteristics in generated images.

Table 2 and Table 3 report our comparison for age-group recognition. As can be seen, our method performs well in all age-groups. LATS and our method attempt to introduce more aging-related characteristics like beard, glasses, grayish hair color, etc. We think this is an obvious factor leading to higher accuracy.

## 7. Conclusion

In this work, we proposed a novel age transformation framework called RAGAN that allowed preserving the identity of a person while changing its age characteristics across various age groups. Our framework leveraged the identity information to learn personalized age features that could self-guide itself for an efficient and reliable face age transformation. Specifically, we adapted the personalized identity features to the target age characteristics for this purpose. Our framework could keep the background of the input for better visual perception of the generated image in comparison to the state-of-the-art method. Overall, our simple yet effective single framework was able to take into account both age characteristics of the facial appearance and identity details in an end-to-end manner. The qualitative and quantitative results supported the superiority of the proposed method compared to the existing approaches.

## Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1A2C1006706 / No. NRF-2021R1F1A1054569). This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-01389, Artificial Intelligence Convergence Research Center (Inha University)).

## 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*, pages 2089–2093, 2017. 1, 2
- [2] Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018. 5, 7
- [3] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proc. of the European Conference on Computer Vision*, pages 768–783. Springer, 2014. 5
- [4] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017. 3
- [5] Ying-Cong Chen, Xiaohui Shen, Zhe Lin, Xin Lu, I Pao, Jiaya Jia, et al. Semantic component decomposition for face attribute manipulation. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9859–9867, 2019. 2
- [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 *Proc. of the IEEE/CVF 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 *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020. 2, 3, 4
- [8] Harm De Vries, Florian Strub, Jérémie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C Courville. Modulating early visual processing by language. In *Advances in Neural Information Processing Systems*, pages 6594–6604, 2017. 4
- [9] Julien Despois, Frédéric Flament, and Matthieu Perrot. AgingMapGAN (AMGAN): High-resolution controllable face aging with spatially-aware conditional gans. In *Proc. of the European Conference on Computer Vision*, pages 613–628. Springer, 2020. 2
- [10] 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*, pages 2672–2680, 2014. 1
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 3
- [12] Zhenliang He, Meina Kan, Shiguang Shan, and Xilin Chen. S<sup>2</sup>GAN: Share aging factors across ages and share aging trends among individuals. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 9440–9449, 2019. 1, 2
- [13] 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. 5, 7
- [14] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 1501–1510, 2017. 4
- [15] 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. 5, 6, 7
- [16] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 4, 5
- [17] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020. 2, 6, 7
- [18] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proc. of the International Conference on Machine Learning*, 2017. 4
- [19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. 4
- [20] Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, and Marc’Aurelio Ranzato. Fader networks: Manipulating images by sliding attributes. In *Advances in Neural Information Processing Systems*, pages 5967–5976, 2017. 2
- [21] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. MaskGAN: Towards diverse and interactive facial image manipulation. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5549–5558, 2020. 3
- [22] Peipei Li, Yibo Hu, Ran He, and Zhenan Sun. Global and local consistent wavelet-domain age synthesis. *IEEE Trans. on Information Forensics and Security*, 14(11):2943–2957, 2019. 2
- [23] 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*, pages 1073–1078, 2018. 2
- [24] Qi Li, Yunfan Liu, and Zhenan Sun. Age progression and regression with spatial attention modules. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 11378–11385, 2020. 1, 2
- [25] 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 *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3673–3682, 2019. 2
- [26] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. In *Proc. of the IEEE/CVF*

- International Conference on Computer Vision*, pages 10551–10560, 2019. 2, 3, 4
- [27] Yunfan Liu, Qi Li, and Zhenan Sun. Attribute-aware face aging with wavelet-based generative adversarial networks. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11877–11886, 2019. 2
- [28] Yunfan Liu, Qi Li, Zhenan Sun, and Tieniu Tan. A<sup>3</sup>GAN: An attribute-aware attentive generative adversarial network for face aging. *IEEE Trans. on Information Forensics and Security*, 16:2776–2790, 2021. 2
- [29] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. Which training methods for GANs do actually converge? In *Proc. of the International Conference on Machine Learning*, 2018. 2, 3, 4
- [30] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv:1411.1784*, 2014. 1, 2
- [31] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. 4
- [32] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In *International Conference on Learning Representations*, 2018. 4
- [33] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *Proc. of the European Conference on Computer Vision*, 2020. 1, 2, 5, 6
- [34] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019. 2
- [35] Albert Pumarola, Antonio Agudo, Aleix M Martinez, Alberto Sanfeliu, and Francesc Moreno-Noguer. Ganimation: Anatomically-aware facial animation from a single image. In *Proc. of the European Conference on Computer Vision*, pages 818–833, 2018. 2
- [36] Shengju Qian, Kwan-Yee Lin, Wayne Wu, Yangxiaokang Liu, Quan Wang, Fumin Shen, Chen Qian, and Ran He. Make a face: Towards arbitrary high fidelity face manipulation. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 10033–10042, 2019. 2
- [37] 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*, 126(2):144–157, 2018. 8
- [38] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015. 8
- [39] Jingkuan Song, Jingqiu Zhang, Lianli Gao, Xianglong Liu, and Heng Tao Shen. Dual conditional GANs for face aging and rejuvenation. In *Proc. of the International Joint Conference on Artificial Intelligence*, pages 899–905, 2018. 1, 2
- [40] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016. 8
- [41] Manjunath Kudlur, Vincent Dumoulin, Jonathon Shlens. A learned representation for artistic style. In *International Conference on Learning Representations*, 2017. 4
- [42] Haoyi Wang, Victor Sanchez, and Chang-Tsun Li. Age-oriented face synthesis with conditional discriminator pool and adversarial triplet loss. *IEEE Trans. on Image Processing*, 2021. 1, 5
- [43] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018. 1, 2, 5
- [44] Taihong Xiao, Jiapeng Hong, and Jinwen Ma. Elegant: Exchanging latent encodings with GAN for transferring multiple face attributes. In *Proc. of the European Conference on Computer Vision*, pages 168–184, 2018. 2
- [45] Daksha Yadav, Naman Kohli, Mayank Vatsa, Richa Singh, and Afzel Noore. Age gap reducer-GAN for recognizing age-separated faces. In *2020 25th International Conference on Pattern Recognition*, pages 10090–10097, 2021. 1
- [46] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. Learning face age progression: A pyramid architecture of GANs. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 31–39, 2018. 1, 2
- [47] Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. High resolution face age editing. In *2020 25th International Conference on Pattern Recognition*, pages 8624–8631, 2021. 1, 2, 5
- [48] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5810–5818, 2017. 1, 2
- [49] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 2223–2232, 2017. 4