CGI2023 manuscript No.
(will be inserted by the editor)

Age-Dependent Face Diversification via Latent Space Analysis

Taishi Ito · **Yuki Endo** · **Yoshihiro Kanamori**

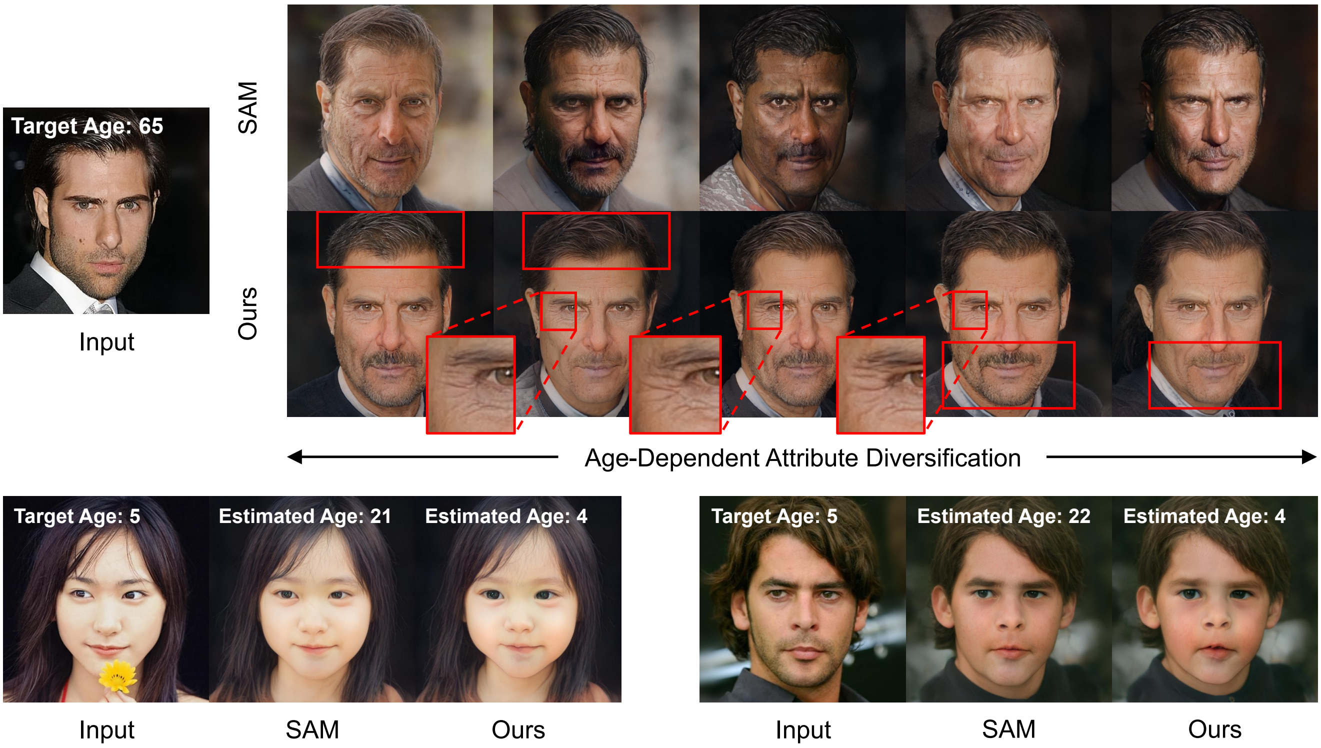

Fig. 1: In the results of age transformation, our method can diversify age-dependent attributes such as hairline, wrinkles, and face shape while preserving the target age and the person's identity compared with SAM [3] (top row). The red rectangles show the differences in age-dependent attributes. We can also more accurately perform age transformation to childhood (bottom row). We estimated age for each output image using Face++ [9].

Abstract Facial age transformation methods can change facial appearance according to the target age. However, most existing methods do not consider that people get older with different attribute changes (e.g., wrinkles, hair volume, and face shape) depending on their circumstances and environment. Diversifying such *age-dependent attributes* while preserving a person's identity is crucial to broaden the applications of age transformation. In addition, the accuracy of age transformation to childhood is limited due to dataset bias. To solve these problems, we propose an age transformation method based on latent space analysis of StyleGAN. Our method obtains diverse age-transformed images by randomly manipulating age-dependent attributes in a latent space. To do so, we analyze the latent space and perturb channels affecting age-dependent attributes. We then optimize the perturbed latent code to refine the age and identity of the output image. We also present an unsupervised approach for improving age transformation to childhood. Our approach is based on the assumption that existing methods cannot sufficiently move a latent code toward a desired direction.

T. Ito
E-mail: itohlee14@gmail.com
University of Tsukuba

Y. Endo
E-mail: endo@cs.tsukuba.ac.jp
University of Tsukuba

Y. Kanamori
E-mail: kanamori@cs.tsukuba.ac.jp
University of Tsukuba

We extrapolate an estimated latent path and iteratively update the latent code along the extrapolated path until the output image reaches the target age. Quantitative and qualitative comparisons with existing methods show that our method improves output diversity and preserves the target age and identity. We also show that our method can more accurately perform age transformation to childhood.

Keywords Age Transformation · Image Editing · Deep learning · GAN · Multimodal

1 Introduction

Age transformation is an image editing technique of changing the appearance of human faces to a given target age while preserving their identity. This technique can be useful for various fields, such as film production and forensics. Previous studies [22,24,18,17] perform age transformation via manipulation along binary age attribute directions (i.e., old/young) in a latent space of pre-trained generative adversarial networks (GANs). Another study [16] allows the user to specify a target age group by considering age transformation as an image-to-image translation task. Recently, Alaluf et al. [3] introduced a pre-trained age classifier [21] into training a latent code mapper for age manipulation in StyleGAN [11,12], and this framework enables specifying arbitrary target ages.

However, these existing methods still have limitations in diversity and accuracy. For diversity, age transformation should have multiple results for a single input because facial appearance varying due to aging is not uniquely determined. For example, depending on circumstances and environment that people experience, they get older with various changes in, e.g., wrinkles, hair, and hair color. Several methods such as style mixing [11,3] can diversify the age transformation results, but they struggle to preserve a target age and a person's identity (see the top row in Figure 1). In addition, in the recent methods [3,7] that use pre-trained age classifiers to specify arbitrary target ages, the transformation accuracy to childhood age is limited due to the bias in the training data for the age classifiers (see the bottom row in Figure 1).

In this paper, we propose a multimodal age transformation method based on latent space analysis of StyleGAN. To diversify the age transformation results, our method identifies and perturbs *age-dependent attributes*, such as wrinkles, hair, hair color, and facial shape. To this end, we first obtain an age-transformed latent code from an input face image and a target age using the pre-trained image-to-latent code mapper for style-based age manipulation (SAM) [3]. To perturb this latent code for age-dependent attributes only, we leverage StyleSpace ($\mathcal{S}$ space) [24], which is one of StyleGAN's latent spaces and well disentangled, that is, each channel controls an independent and fine-grained attribute. We present an approach based on correlation analysis in $\mathcal{S}$ space to identify channels that more strongly affect the age-dependent attributes and less affect the identity of the input person. Because perturbing the identified channels may cause some deviation from the target age and identity, we further refine the latent code via optimization.

In addition, we propose an unsupervised approach for improving the accuracy of age transformation to childhood. We suppose that the reason for low accuracy for younger age is because the existing methods cannot sufficiently move a latent code toward a desired direction. On the basis of this intuition, we extrapolate a latent path connecting estimated latent codes around a target age. We then move the latent code toward the extrapolated direction until the estimated age of the output image matches the target age. We tried various extrapolation methods and found that a simple linear extrapolation was the most effective.

In summary, our main contributions are (i) the multimodal age transformation method based on correlation analysis in $\mathcal{S}$ space and (ii) the unsupervised approach for improving the accuracy of age transformation to childhood via latent path extrapolation. As shown in Figure 1 and quantitative and qualitative evaluations, our method can generate more diversified outputs preserving target age and identity compared with existing methods. We also demonstrate that our method can perform more accurate age transformation for younger ages.

2 Related Work

2.1 Image-to-image translation

Image-to-image translation is the task of transforming images between domains. This task has attracted much attention from the advent of pix2pix [10] using GANs and has been applied to various problems, including age transformation. Several studies [5,6,14,23,16] performed age transformation between pre-defined age groups. For example, Or-El et al.[16] approximate continuous appearance changes due to aging as a multi-domain image-to-image transformation problem. This method trains translation between images in several pre-defined age clusters (e.g., 3-6 and 15-19). However, these age group-based approaches struggle to accurately capture appearance changes between each age group, and the user cannot arbitrarily specify the target age.

Alaluf et al. [3] presented style-based age manipulation (SAM), which handles age transformation as a regression problem, enabling the user to specify the target age directly. They trained the image-to-latent code mapper in an unsupervised manner using pre-trained StyleGAN and the age classification network. This method can also diversify age-transformed images via style mixing. Gomez et al. [7] proposed an age transformation method called CUSP containing two types of encoders that extract style or content information from an input image. Applying a variable mask to the skip connections between the content encoder and the generator allows the user to control the degree of facial structure preservation. However, these methods cannot obtain sufficiently diversified results with various changes of age-dependent attributes while preserving the target age and identity. For more detail, we show and discuss comparisons with our method in Section 4.

2.2 Latent space exploration

Many studies have performed age transformation by analyzing the latent space of GANs and finding latent paths that control specific attributes. Shen et al. [22] computed a hyperplane corresponding to the separation boundary for a binary attribute (e.g., gender) from sampled latent codes. They used the normal vector of the hyperplane as a latent direction for editing the attribute. However, this method often changes attributes unrelated to the specified one. Wu et al. [24] found that $\mathcal{S}$ space is better disentangled than other latent spaces such as $\mathcal{W}$ and $\mathcal{W}+$ spaces. They also proposed methods for detecting locally-active and attribute-specific style channels in $\mathcal{S}$ space. Inspired by their work, we leverage $\mathcal{S}$ space to diversify age-dependent attributes only while preserving the other attributes. Patashnik et al. [18] performed text-based StyleGAN image manipulation. They used CLIP [19] features as input to train a mapping network between latent codes. They also proposed a method for identifying and manipulating channels associated with a text in $\mathcal{S}$ space by comparing images and texts before and after editing in the CLIP feature space. These methods can perform well-disentangled edits but only obtain a single output for a single input and do not account for multimodality. Parihar et al. [17] used a small number of samples manually created by simple image composition to find latent directions for editing binary attributes. This method can diversify the output by manipulating latent codes on a style manifold obtained from positive samples of several different styles for the target attribute. However, because this method can handle binary attributes only, it is unsuitable for age transformation taking arbitrary ages as input.

2.3 Style-based age manipulation (SAM)

This section briefly describes the overview of SAM [3] because our method builds upon it. Given a face image and a desired target age, SAM outputs a face image with the appearance of the target age. SAM first embeds the input image into the StyleGAN [12] latent space using a pre-trained pixel2Style2pixel network [20]. The SAM encoder estimates the residual between latent codes before and after age transformation. We can obtain an output image by adding the residual to the embedded latent code and feeding the combined code to the pre-trained StyleGAN generator. The SAM encoder is trained by minimizing a loss function between the target age and the age estimated from the output image using the pre-trained age classifier [21].

SAM generally assumes only a single result for the target age. Although style mixing provided by StyleGAN can diversify the output, it cannot preserve the target age and identity. In addition, because the editing accuracy depends on the performance of the age classifier trained with imbalanced data, editing accuracy to childhood is limited.

3 Method

We describe our multimodal age transformation and unsupervised latent space exploration in Sections 3.1 and 3.2, respectively.

3.1 Multimodal age transformation

The goal of multimodal age transformation is to diversify age-dependent attributes while maintaining the target age and identity. Figure 2 shows the overall architecture of our method. Given a face image x and a desired target age α_{target} as input, our method aims to obtain a diversified output image y' corresponding to a target age α_{target}. To this end, we perform additional operations on an age-transformed latent code $w \in \mathcal{W}+$ obtained using SAM [3]. We add a random offset to the latent code s in $\mathcal{S}$ space [24] computed from w in $\mathcal{W}+$ space [2]. To manipulate only age-dependent attributes, we preliminarily analyze $\mathcal{S}$ space and identify channels that affect age-dependent attributes and do not affect facial identity (Section 3.1.1). Furthermore, we optimize the diversified latent code to preserve a target age and identity (Section 3.1.2).

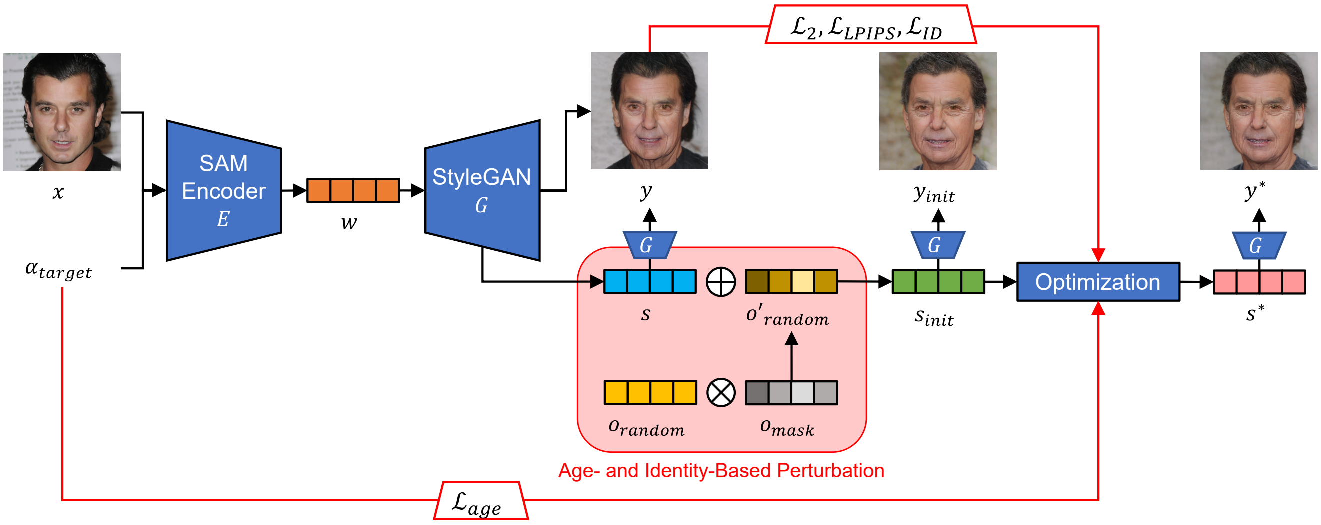

Fig. 2: Network architecture of our multimodal age transformation method. Given a person image x and a target age α_{target} as input, we obtain an age-transformed latent code $w \in \mathcal{W}+$ using the pre-trained SAM encoder E. We then convert w into $s \in \mathcal{S}$ through the affine layers of StyleGAN generator G. To diversify age-dependent attributes, we perform age- and identity-based perturbation for s and obtain s_{init}. Finally, optimizing s_{init} using the four loss functions yields the refined latent code s^*, which is fed into G to produce the final image y^*.

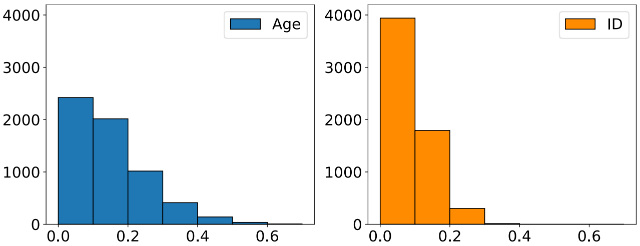

Fig. 3: Histogram of the absolute correlation coefficients between the variations of $\mathcal{S}$ space channels and the variations of the age or identity. The horizontal and vertical axes indicate the absolute correlation coefficient and frequency, respectively.

3.1.1 Correlation analysis and perturbation in $\mathcal{S}$ space

Our method aims to diversify only age-dependent attributes by adding perturbations to an age-transformed latent code w. However, the $\mathcal{W}+$ space latent code w obtained using SAM [3] are highly entangled and difficult to use for perturbing specific attributes. Therefore, we convert the $\mathcal{W}+$ space latent code into the $\mathcal{S}$ space [24] one. $\mathcal{S}$ space is more disentangled than $\mathcal{W}+$ space and has the advantage of controlling an individual and fine-grained attribute via single channel manipulation.

In $\mathcal{S}$ space, we need to identify channels that manipulate age-dependent attributes. We also identify channels that less affect a person's identity to preserve the identity during random manipulation. To do so, we analyze the correlation between each channel in $\mathcal{S}$ space and age or identity of the output image. Specifically, we feed a randomly sampled $\{z_n | z_n \in \mathcal{Z}\}_{n=1}^{100}$, where $\mathcal{Z}$ is a latent space following a normal distribution, to the mapping network and generator of the pre-trained StyleGAN and convert them into $\{s_n | s_n \in \mathcal{S}\}_{n=1}^{100}$. Inspired by StyleCLIP [18], we then compute a vector ${s_n}' = s_n + \alpha_n \Delta s_c$ by perturbing s_n with Δs_c. The vector Δs_c contains the standard deviation at a channel c and zero at the other channels. We preliminarily compute the standard deviation of each channel from 1,000 randomly sampled latent code in $\mathcal{S}$ space. α_n is a random value from a uniform distribution $u(-5, 5)$. For the obtained latent codes s_n and ${s_n}'$ before and after the perturbation, the variations of age and identity are obtained as follows:

$$\Delta_{age,n} = |A(G({s_n}')) - A(G(s_n))|, \tag{1}$$

$$\Delta_{ID,n} = 1 - \langle R(G(s_n)), R(G({s_n}')) \rangle. \tag{2}$$

Here, $G(\cdot)$ denotes the pre-trained StyleGAN [12] generator, $A(\cdot)$ denotes the pre-trained age classifier [21], $R(\cdot)$ denotes the pre-trained ArcFace network [4], and $\langle \cdot, \cdot \rangle$ denotes the inner product. We finally compute correlation coefficients between $\alpha \Delta s_c$ and Δ_{age} or Δ_{ID}.

Figure 3 shows the histograms of absolute correlation coefficients for 6,048 channels except the tRGB channels in $\mathcal{S}$ space. The histograms show that a few hundred channels correlate with age or identity. We can also observe that channels correlated with identity are fewer than those correlated with age. This is likely because identity is a complex metric controlled by multiple attributes. In Figure 4, we further analyze how specific correlated channels affect the output appearance. The scatter plots show that the 435th channel on the 9th layer (9_435) and the 155th channel on the 8th layer (8_155) correlate with age and identity, respectively. In the edited images, we can confirm that the 9_435 channel controls the amount and depth of wrinkles around the eyes and mouth. The 8_155 channel controls the height of the eyebrows that affect identity.

Driven by the above analysis, we use the obtained correlation coefficients as weights for a random offset o'_{random}. We assign larger weights to the channels correlated with age and smaller weights to those correlated with identity. Specifically, using the correlation coefficients σ_c^{age} and σ_c^{ID} for age and identity in a channel c, we compute a soft mask o_{mask} as follows:

$$o_{mask,c} = \frac{\sigma_c - \min_k(\sigma_k)}{\max_k(\sigma_k) - \min_k(\sigma_k)}, \tag{3}$$

$$\sigma_c = |\sigma_c^{age}| + (1 - |\sigma_c^{ID}|), \tag{4}$$

where $o_{mask,c}$ is the value of o_{mask} for a channel c and normalized to $[0, 1]$. Next, we perturb the latent code s using the weighted offset o'_{random} as follows:

$$s_{init} = s + o'_{random}, \tag{5}$$

$$o'_{random} = o_{random} \odot o_{mask}, \tag{6}$$

where $\odot$ denotes the element-wise product, and $o_{random} \in \mathcal{S}$ is a latent code converted from randomly sampled $z \in \mathcal{Z}$ using the pre-trained mapping network and StyleGAN generator.

3.1.2 Latent code refinement via optimization

Although we can diversify age-dependent attributes by latent code perturbation described in Section 3.1.1, this procedure may cause some deviation from the target age α_{target} and the identity of the input image x. To address this issue, we further optimize the diversified latent code. Let s^* be the diversified latent code to be optimized, initialized with s_{init}. Note that we exclude the tRGB channels from the optimized parameters because they control the overall hue and are not related to age-dependent attributes.

For loss functions used for optimization, we first employ the L2 loss and LPIPS [25] loss between an original age-transformed image y before perturbation and the diversified image $G(s^*)$:

$$\mathcal{L}_2(s^*) = \|y - G(s^*)\|_2, \tag{7}$$

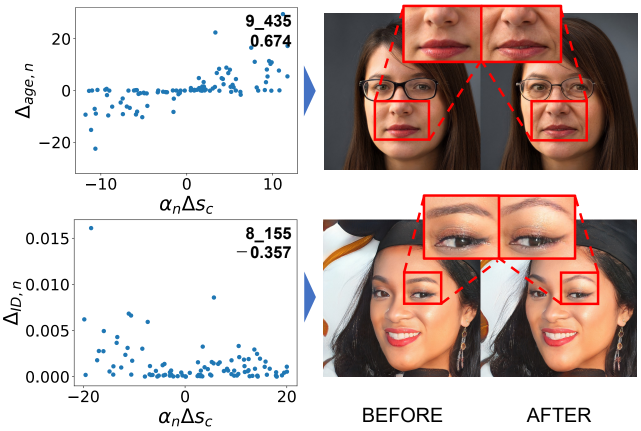

Fig. 4: Scatter plots for the variations of the specific $\mathcal{S}$ space channels correlated with the variations of age or identity (left). The upper right numbers in the scatter plots show the manipulated channel indices and correlation coefficients from top to bottom. As shown in the right images, manipulating these channels affects age- and identity- dependent attributes (e.g., depth of wrinkles and height of eybrows).

$$\mathcal{L}_{LPIPS}(s^*) = \|F(y) - F(G(s^*))\|_2, \tag{8}$$

where $F(\cdot)$ denotes the perceptual feature extractor. These loss functions have a role to rectify identity drifts caused by latent code perturbation. Note that, as discussed in Section 3.1.1, identity is a complex metric controlled by multiple attributes, and our correlation analysis may not perfectly identify identity-related channels. Next, we use the identity loss $\mathcal{L}_{ID}$ between the diversified and the original output images y for identity preservation:

$$\mathcal{L}_{ID}(s^*) = 1 - \langle R(y), R(G(s^*)) \rangle. \tag{9}$$

Note that there might be an alternative loss between the diversified image $G(s^*)$ and the input image x instead of y. However, this loss function did not work because it made the optimization unstable due to the large difference of their appearance. Finally, we introduce the aging loss for target age preservation:

$$\mathcal{L}_{age}(s^*) = |\alpha_{target} - A(G(s^*))|, \tag{10}$$

where $A(\cdot)$ denotes the pre-trained age classifier [21]. The final objective function is defined as follows:

$$\begin{aligned} \mathcal{L}(s^*) = \lambda_{l2}\, \mathcal{L}_2(s^*) + \lambda_{lpips}\, \mathcal{L}_{LPIPS}(s^*) \\ + \lambda_{id}\, \mathcal{L}_{ID}(s^*) + \lambda_{age}\, \mathcal{L}_{age}(s^*), \end{aligned} \tag{11}$$

where λ_{l2}, λ_{lpips}, λ_{id}, and λ_{age} are the weights for each loss function. We empirically determined these weights as $\lambda_{l2} = 0.01$, $\lambda_{lpips} = 0.01$, $\lambda_{id} = 0.1$, and $\lambda_{age} = 5$.

3.2 Latent path extrapolation via guided optimization

As explained in Section 1, the recent age transformation methods [3,7] can use an arbitrary target age as input but struggle to accurately estimate age transformation to childhood due to the dataset bias. A naïve remedy would be to optimize latent codes to reach the target age, but we found it also struggles bad local minima (Section 4.3). To solve this problem, we present an unsupervised approach for latent path extrapolation via guided optimization.

To extrapolate toward the younger direction, we first define a latent path by sampling multiple latent codes. Specifically, we sample four latent codes $\{s_k\}_{k=0}^{3} \subset \mathcal{S}$ by feeding corresponding age targets $\{\alpha_k\}_{k=0}^{3}$ to SAM encoder E. We select $\{\alpha_k\}$ at three-year intervals from a target age α_{target} $(= \alpha_0)$ backward toward the older direction, which is more reliable than the younger direction. For example, for $\alpha_{target} = 5$, we sample $\{\alpha_k\} = \{5, 8, 11, 14\}$. The latent path is then defined by the four samples $\{s_k\}$ and an interpolation function f. If function f is a linear interpolator, i.e.,

$$f(\alpha, k) = s_k + \frac{\alpha - \alpha_k}{\alpha_{k+1} - \alpha_k}(s_{k+1} - s_k), \tag{12}$$

the latent path is then a polyline (the blue polyline in Figure 5). We extrapolate the latent path for younger ages $\alpha < \alpha_{target}$ by $f(\alpha, 0)$ (the orange line with "Ours_5" in Figure 5). For function f, we compared several candidates, such as quadratic and cubic Hermite interpolation, and found that the simple linear interpolator works best among our candidates. Section 4.3 explains the details.

Next, we manipulate the latent code along the extrapolated latent path so that the output image reaches the target age α_{target} via an iterative search. We refer to this technique as *guided optimization.* Let $\hat{\alpha}_t$ be the input age value at iteration t and $\hat{\alpha}_0 = \alpha_{target}$. The iteration terminates when the following condition is met:

$$|A_t - \alpha_{target}| < \beta, \tag{13}$$

where A_t denotes $A(G(f(\hat{\alpha}_t, 0)))$ for simplicity, and $\beta = 0.5$ represents a tolerance parameter. We update $\hat{\alpha}_t$ in each iteration as follows:

$$\hat{\alpha}_{t+1} \leftarrow \hat{\alpha}_t + \Delta m_t, \tag{14}$$

where Δm_t is the step size of α_t defined as:

$$\Delta m_t = \begin{cases} -1 & \text{if } t = 0 \text{ and } A_t \geq \alpha_{target}, \\ 1 & \text{if } t = 0 \text{ and } A_t < \alpha_{target}, \\ -\frac{\Delta m_{t-1}}{2} & \text{if } (A_{t-1} - \alpha_{target})(A_t - \alpha_{target}) < 0, \\ \Delta m_{t-1} & \text{otherwise.} \end{cases}$$

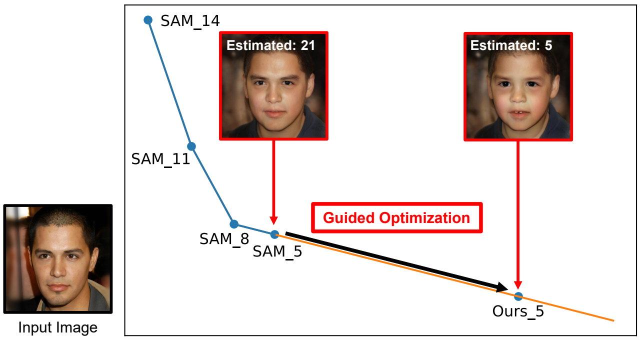

Fig. 5: Overview of our latent path extrapolation via guided optimization. The latent codes sampled using SAM (labeled as SAM_k, where k denotes the target age) define a latent path (blue polyline). SAM's output face for the age of five (i.e., SAM_5) looks older than five, so we optimize the latent code Ours_5 along the extrapolated path (orange line) until the estimated age of the generated face reaches the target age. We employ PCA projection of latent codes for 2D visualization.

(15)

This equation means moving the latent code on the extrapolated path toward the target age direction. If the estimated age A_t passes through the target age α_{target}, we move the latent code the half distance at the previous step in the opposite direction. To deal with the case where this iterative update does not converge, we set the maximum iteration number $T = 100$.

4 Experiments

4.1 Implementation details

We implemented our method using PyTorch and ran our program on NVIDIA RTX A4000. For latent code optimization, we used Adam [13] with a learning rate of 0.01 and set the parameters β, ϵ, and the weight decay as (0.9, 0.999), 1e-8, and 0, respectively. We stopped the optimization when the change in the loss was less than 0.0001 in 7 consecutive steps. The computation time was about 37 seconds for multimodal age transformation and 3 seconds for latent path extrapolation.

4.2 Evaluation of multimodal age transformation

Comparisons with existing methods. In this experiment, we created 10 diverse output images from each of 100 CelebA-HQ [15] test images and random target ages. We evaluated the result images using three metrics: diversity, identity preservation, and target age

Table 1: Quantitative comparison of diversification with SAM [3], CUSP [7], and our method. Boldface indicates the best score for each metric.

Method	LPIPS ↑	ID ↓	AGE_SD ↓
SAM	0.0495	0.355	4.20
CUSP	0.0220	0.165	3.09
Ours	**0.0503**	**0.0164**	**2.84**

preservation. For the diversity score, we take pairs of 10 diversified images generated from each input image and compute the average of LPIPS [25] values for all pairs. The identity preservation score (ID) is the average of ArcFace [4] values for all pairs. The age preservation score (AGE_SD) is the average of the standard deviations of ages estimated from sets of diversified images. For fair evaluation of estimated ages, we did not use the age classifier [21] (used for optimization) but used Face++ [9], which is a Web API service for face recognition. We masked background regions using a semantic face parser [1] to exclude them from evaluation. Table 1 shows the quantitative comparison. To obtain diverse results for SAM [3], we performed style mixing on 8th and 9th layers as described in the paper. In the case of CUSP [7], the variances σ_m and σ_g of Gaussian kernels were sampled randomly. The results show that our method outperforms the other methods in all metrics.

Figure 6 shows the qualitative comparison. The results of SAM using style mixing show changes in eye color, skin color, lighting, and other attributes, which are unrelated to aging, and some results deviate from the target age. The results of CUSP show poor identity preservation due to changes in nose and eye shape, and some outputs appear unnatural. In contrast, our method diversifies mainly age-dependent attributes. For example, as shown in the red rectangles, we can see that the number of wrinkles differs around the eyes and mouths, hair color fades differently, and facial shapes vary around the chins. Hair volume variations (i.e., receding hairlines of the man and hair length changes of the woman) are also possible changes with age variations.

Ablation study. We evaluated the effectiveness of our optimization process for refining diversified latent codes. Table 2 shows the quantitative comparison between our multimodal age transformation methods with and without optimization. The results demonstrate that the optimization somewhat decreases the LPIPS score for diversity but significantly improves the ID and AGE_SD scores, which are essential to preserving the identity and target age.

Table 2: Quantitative comparison between our methods with and without optimization after diversification.

Settings	LPIPS ↑	ID ↓	AGE_SD ↓
w/o optimization	**0.0706**	0.217	5.49
w/ optimization	0.0503	**0.0164**	**2.84**

Table 3: Quantitative comparison with our diversification methods using $\mathcal{W}+$ space or $\mathcal{S}$ space.

Method	LPIPS ↑	ID ↓	AGE_SD ↓
Ours ($\mathcal{W}+$ space)	0.0351	0.253	2.89
Ours ($\mathcal{S}$ space)	**0.0503**	**0.0164**	**2.84**

Comparisons between $\mathcal{S}$ and $\mathcal{W}+$ spaces. We evaluated the effectiveness of using $\mathcal{S}$ space in our method. For comparison, we attempted multimodal age transformation in $\mathcal{W}+$ space [2]. The overall flow using $\mathcal{W}+$ space is similar to our method using $\mathcal{S}$ space. However, instead of weighted offsets, we used randomly sampled $\mathcal{W}+$ latent codes for perturbation because $\mathcal{W}+$ space is not disentangled for each channel, and thus our correlation analysis cannot be applied. Let $w^*, w_{init} \in \mathcal{W}+$ be the diviersified latent code and its initial value in $\mathcal{W}+$ space, similar to s^*, s_{init} defined in Section 3.1.1. For latent code optimization in $\mathcal{W}+$ space, we introduce a regularization loss $\mathcal{L}_{reg}$ in addition to our losses.

$$\begin{aligned}\mathcal{L}(w^*) = \lambda_{l2}\,\mathcal{L}_2(w^*) + \lambda_{lpips}\,\mathcal{L}_{LPIPS}(w^*) \\ + \lambda_{id}\,\mathcal{L}_{ID}(w^*) + \lambda_{age}\,\mathcal{L}_{age}(w^*) \\ + \lambda_{reg}\,\mathcal{L}_{reg}(w^*).\end{aligned} \tag{16}$$

The regularization loss $\mathcal{L}_{reg}$ encourages a latent code to approach the average latent code and improves image quality by removing undesirable artifacts. Note that we did not use $\mathcal{L}_{reg}$ in $\mathcal{S}$ space because we found that this loss tends to produce artifacts. For the weight of each loss, we empirically set $\lambda_{l2} = 1, \lambda_{lpips} = 0.1, \lambda_{id} = 0.1, \lambda_{age} = 5$, and $\lambda_{reg} = 0.005$. We stopped optimization if the loss function value becomes less than 0.5.

Table 3 shows the quantitative results. Compared with SAM using style mixing in Table 1, Ours ($\mathcal{W}+$ space) improves the ID and AGE_SD scores but decreases the LPIPS score for diversity. In contrast, Ours ($\mathcal{S}$ space) shows the best LPIPS and ID scores while obtaining the AGE_SD score comparable with Ours ($\mathcal{W}+$ space).

Figure 7 shows the qualitative results. In the results of $\mathcal{W}+$, the image quality is relatively low. This is probably because latent codes are perturbed with equal

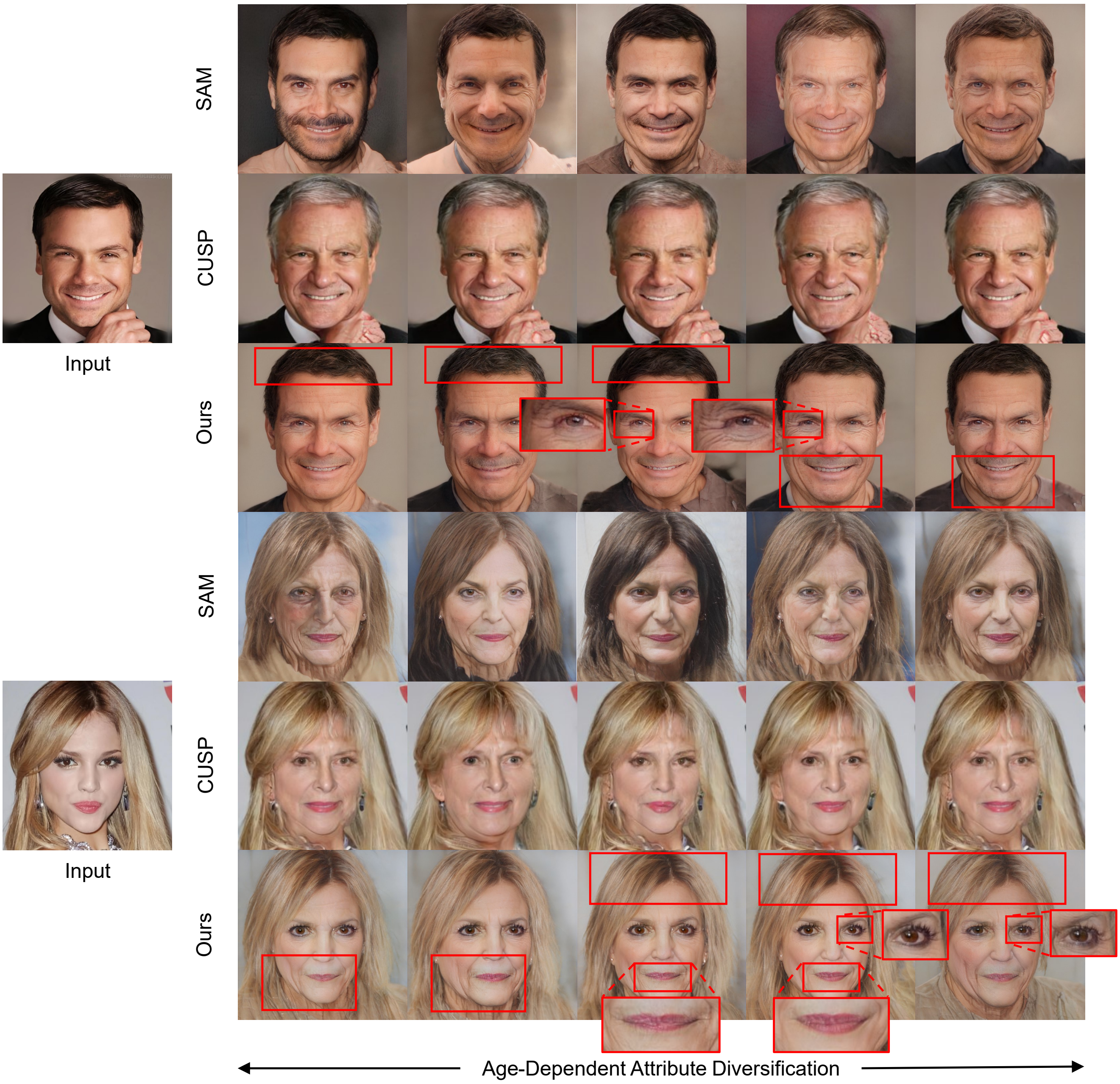

Fig. 6: Qualitative comparison of diversifying age-dependent attributes with the existing methods and our method. We created five different outputs for each person specifying the target age of 65.

weights for all channels and deviate from the real distribution in $\mathcal{W}+$ space. In addition, manipulation in $\mathcal{W}+$ space changes not only age-dependent attributes but also identity. In contrast, our method using $\mathcal{S}+$ space diversifies the results while preserving age and identity thanks to our correlation analysis approach.

4.3 Evaluation of latent path extrapolation

Comparisons with existing methods. To evaluate the effectiveness of our latent path extrapolation for young age, we conducted a quantitative evaluation for the target ages of 3-19 using 2,000 test images from CelebA-HQ [15] as input. As evaluation metrics, we used AGE_MAE for age transformation accuracy and FID [8] for image quality. AGE_MAE is the average absolute difference between the estimated age and the target age

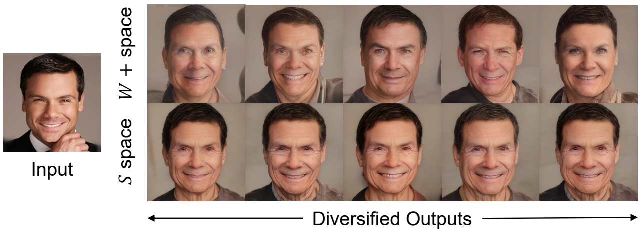

Fig. 7: Qualitative comparison of diversification using $\mathcal{W}+$ space and $\mathcal{S}$ space.

Table 4: Quantitative comparison of age transformation to childhood with SAM and our method.

Target Age	Method	FID ↓	AGE_MAE ↓
5	SAM	149.1	17.7
	Ours	**109.3**	**7.75**
8	SAM	119.0	15.5
	Ours	**113.0**	**13.9**
12	SAM	110.4	12.6
	Ours	**109.3**	**11.6**
17	SAM	**100.7**	10.8
	Ours	101.3	**7.09**

for each output image. We used Face++ [9] for age estimation. As a ground-truth dataset for computing FID, we used FFHQ-Aging [16] images belonging to the 3-6, 7-9, 10-14, and 15-19 age groups. Using SAM and our method, we performed age transformation by specifying the target ages of 5, 8, 12, and 17, which are the median values of age groups. We then computed the FID scores between generated images and ground-truth images corresponding to specified age groups. Note that we did not use CUSP [7] for comparison because the public model (trained with FFHQ-RR) for specifying the arbitrary target age cannot handle ages less than 20.

As shown in Table 4, our method provides better FID and AGE_MAE at the target ages of 5, 8, and 12. For the target age of 17, the FID score of our method is comparable to SAM, but our method is significantly better than SAM in the AGE_MAE score. These results indicate that our method outperforms SAM when the target age is younger.

Figure 8 shows the qualitative comparison of age transformation to the target age of 5. We can confirm that the outputs of our method are closer to the target age compared with SAM. In particular, our method can make the eyes larger and the mouth smaller and change the facial contours. Other qualitative results are shown in Appendix A.

Table 5: Quantitative comparison of age transformation to childhood via our guided optimization for latent path extraporation or a naïve optimization.

Target Age	Method	FID ↓	AGE_MAE ↓
5	naïve opt.	124.9	14.7
	guided opt.	**109.3**	**7.75**
8	naïve opt.	**112.8**	14.5
	guided opt.	113.0	**13.9**
12	naïve opt.	**107.4**	**11.4**
	guided opt.	109.3	11.6
17	naïve opt.	**98.25**	7.20
	guided opt.	101.3	**7.09**

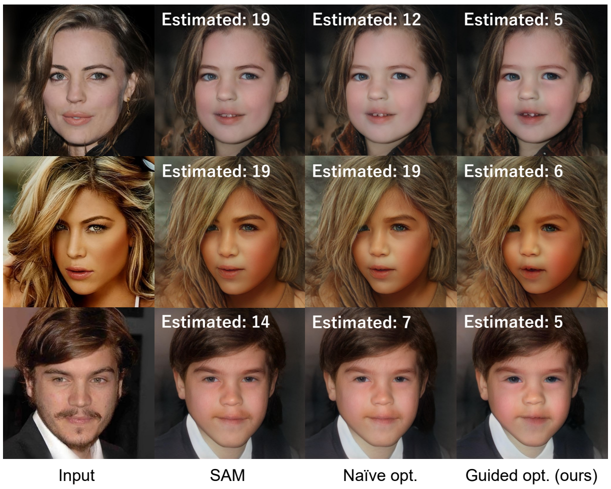

Fig. 8: Qualitative comparison of age transformation to childhood (target age of 5).

Ablation study. We also compared our method with a naïve optimization that does not use guidance by latent path extrapolation. Specifically, we naïvely minimize the age preservation loss in Equation (10) without restricting the moving direction of optimized codes. In Table 5, this optimization method (*naïve opt.*) is quantitatively comparable with our method (*guided opt.*) in the target ages of 8, 12, and 17, whereas it shows significantly lower performance in the target age of 5.

In Figure 8, the naïve optimization only yields qualitatively subtle changes from SAM because of falling into undesirable local minima. In contrast, our guided optimization appropriately represents the characteristics of five-year-olds (e.g., pupil proportion, cheek redness, and dentition) by estimating desired latent directions.

Comparisons with other extrapolation methods. We evaluated the validity of using linear extrapolation for latent path extrapolation. Specifically, we performed age transformation to the target age of 5 for 2,000 test images in CelebA-HQ. As an evaluation metric, we used

Table 6: Quantitative comparison with various extrapolation methods.

Method	Success Rate ↑	FID ↓
linear (Ours)	**1.00**	**109.3**
barycentric	0.487	175.8
Krogh	0.487	175.8
quadratic	0.495	188.7
cubic	0.487	175.8
Akima	0.683	138.0
PCHIP	0.145	244.6

success rates besides FID. We define success rate as the ratio of images that reach the target age during iterative update. In addition to linear extrapolation, we tried barycentric, Krogh, quadratic spline, cubic spline, Akima, and piecewise cubic Hermite (PCHIP) extrapolation methods. Table 6 shows the quantitative comparison with various extrapolation methods. We can confirm that linear extrapolation shows the best success rate and FID score.

Figure 9 shows qualitative results. The results in the first row are relatively good for all extrapolation methods except PCHIP. For the other results, however, all methods except linear extrapolation are unstable, and some images collapse. Linear extrapolation is the most stable and produces the highest quality images. This is probably because the latent path is partially linear in the lower age intervals. Further analysis in the latent space is left for future research.

5 Conclusion

In this paper, we proposed a multimodal facial age transformation method to diversify age-transformed facial images while preserving the target age and identity. Through analysis of $\mathcal{S}$ space [24], we compute a correlation between each $\mathcal{S}$ space channel and age or identity. We diversify age-dependent attributes (e.g., wrinkles and hair) by latent code perturbation using the correlation coefficients as weights of random offsets. Our method refines the perturbed latent codes to improve the age and identity of the output images. In addition, we proposed an unsupervised latent path extrapolation method to improve the accuracy of age transformation to childhood. We extrapolate a latent path obtained from latent codes sampled around a target age. We then obtain new latent codes for the target age via iterative search along the extrapolated path. Evaluation experiments demonstrate that our method quantitatively and qualitatively outperforms the existing methods in diversity and accuracy.

Limitations and future work. For output diversification, our method relies on randomness. In some cases, the user may not obtain desired results that are sufficiently diversified. In addition, as shown in the left of Figure 10, our latent code perturbation based on correlation analysis is not perfect and sometimes changes gender and face orientation besides age-dependent attributes. In the future, we would like to mitigate this problem by incorporating additional losses for preserving such attributes into latent code optimization. Furthermore, our latent path extrapolation may produce collapsed images when the target age is extremely young (see the right in Figure 10). Our method may fail when we move the latent code too far away from its initial position. Future work is to develop more effective latent space exploration methods.

Data Availability Statement

Our source code is available at `https://github.com/shiiiijp/ADFD`.

References

1. zllrunning/face-parsing.PyTorch: Using modified BiSeNet for face parsing in PyTorch. `https://github.com/zllrunning/face-parsing.PyTorch`. (Accessed on 12/23/2022)
2. Abdal, R., Qin, Y., Wonka, P.: Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space? In: Proc. of ICCV, pp. 4432–4441 (2019)
3. Alaluf, Y., Patashnik, O., Cohen-Or, D.: Only a Matter of Style: Age Transformation Using a Style-Based Regression Model. ACM Transactions on Graphics **40**(4), 45:1–45:12 (2021)
4. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proc. of CVPR, pp. 4690–4699 (2019)
5. Fang, H., Deng, W., Zhong, Y., Hu, J.: Triple-gan: Progressive face aging with triple translation loss. In: Proc. of CVPR Workshops, pp. 804–805 (2020)
6. Georgopoulos, M., Oldfield, J., Nicolaou, M.A., Panagakis, Y., Pantic, M.: Enhancing facial data diversity with style-based face aging. In: Proc. of CVPR Workshops, pp. 14–15 (2020)
7. Gomez-Trenado, G., Lathuilière, S., Mesejo, P., Cordón, Ó.: Custom structure preservation in face aging. In: Proc. of ECCV, pp. 565–580 (2022)
8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In: Proc. of NeurIPS, pp. 6626–6637 (2017)
9. Inc., M.: Face++ research toolkit. `http://www.faceplusplus.com`
10. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. In: Proc. of CVPR, pp. 5967–5976 (2017)
11. Karras, T., Laine, S., Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks. In: Proc. of CVPR, pp. 4401–4410 (2019)

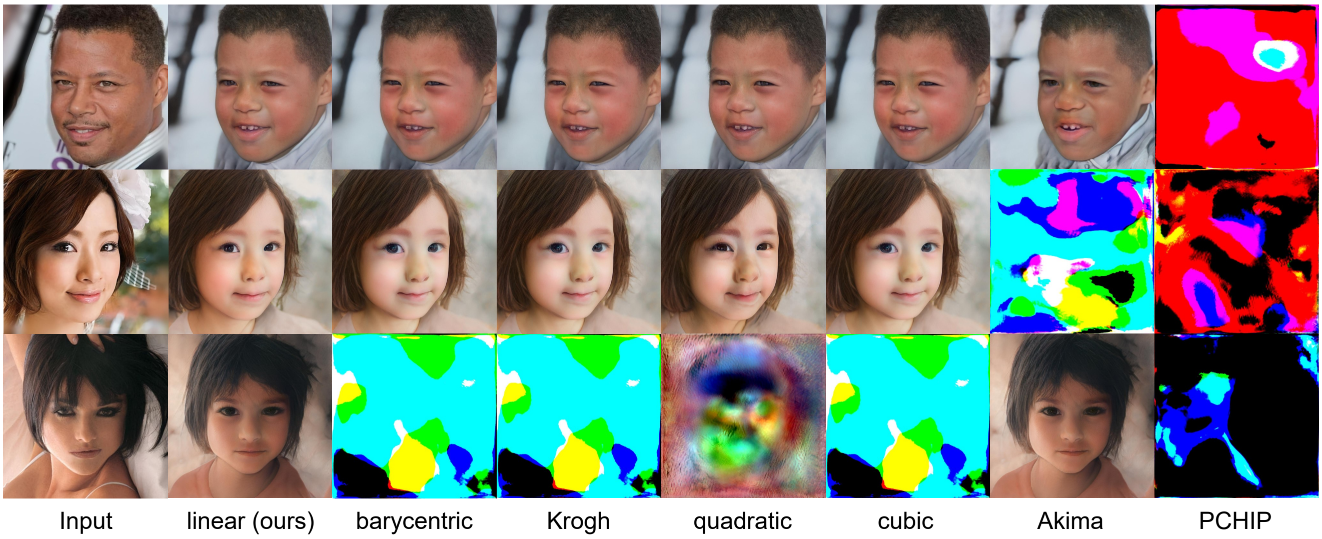

Fig. 9: Qualitative comparison of our method using different extrapolation methods. We set the target age as 5.

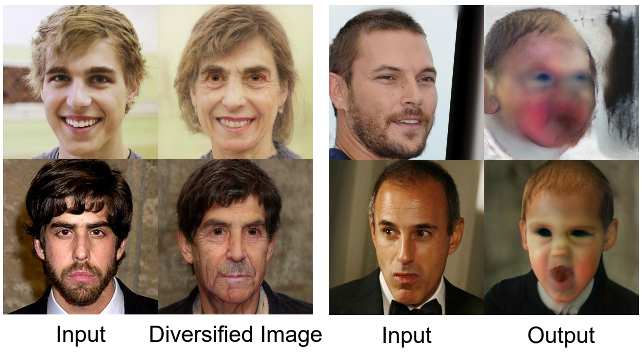

Fig. 10: Failure cases of our method. Our latent code perturbation sometimes changes gender and face orientation besides age-dependent attributes (left). Our latent path extrapolation may also produce artifacts when we specify an extremely young age, e.g., 1 (right).

12. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and Improving the Image Quality of StyleGAN. In: Proc. of CVPR, pp. 8107–8116 (2020)
13. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Li, P., Huang, H., Hu, Y., Wu, X., He, R., Sun, Z.: Hierarchical face aging through disentangled latent characteristics. In: Proc. of ECCV, pp. 86–101 (2020)
15. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep Learning Face Attributes in the Wild. In: Proc. of ICCV, pp. 3730–3738 (2015)
16. Or-El, R., Sengupta, S., Fried, O., Shechtman, E., Kemelmacher-Shlizerman, I.: Lifespan Age Transformation Synthesis. In: Proc. of ECCV, pp. 739–755 (2020)
17. Parihar, R., Dhiman, A., Karmali, T.: Everything is There in Latent Space: Attribute Editing and Attribute Style Manipulation by StyleGAN Latent Space Exploration. In: Proc. of ACMMM, pp. 1828–1836 (2022)
18. Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., Lischinski, D.: StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery. In: Proc. of ICCV, pp. 2065–2074 (2021)
19. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning Transferable Visual Models From Natural Language Supervision. In: Proc. of ICML, pp. 8748–8763 (2021)
20. Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., Cohen-Or, D.: Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In: Proc. of CVPR, pp. 2287–2296 (2021)
21. Rothe, R., Timofte, R., Van Gool, L.: DEX: Deep EXpectation of Apparent Age From a Single Image. In: Proc. of ICCV Workshop, pp. 252–257 (2015)
22. Shen, Y., Gu, J., Tang, X., Zhou, B.: Interpreting the Latent Space of GANs for Semantic Face Editing. In: Proc. of CVPR, pp. 9240–9249 (2020)
23. Viazovetskyi, Y., Ivashkin, V., Kashin, E.: Stylegan2 distillation for feed-forward image manipulation. In: Proc. of ECCV, pp. 170–186 (2020)
24. Wu, Z., Lischinski, D., Shechtman, E.: StyleSpace Analysis: Disentangled Controls for StyleGAN Image Generation. In: Proc. of CVPR, pp. 12,863–12,872 (2021)
25. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In: Proc. of CVPR, pp. 586–595 (2018)

A Additional Qualitative Comparisons

Figure 11 shows the additional qualitative comparisons with SAM [3], CUSP [7], and our method. For diversification, the results of the existing methods often show changes in not only identity but also lighting and background, which are unrelated to age, while our method diversifies age-dependent attributes. Our method also performs age transformation to childhood better than SAM.

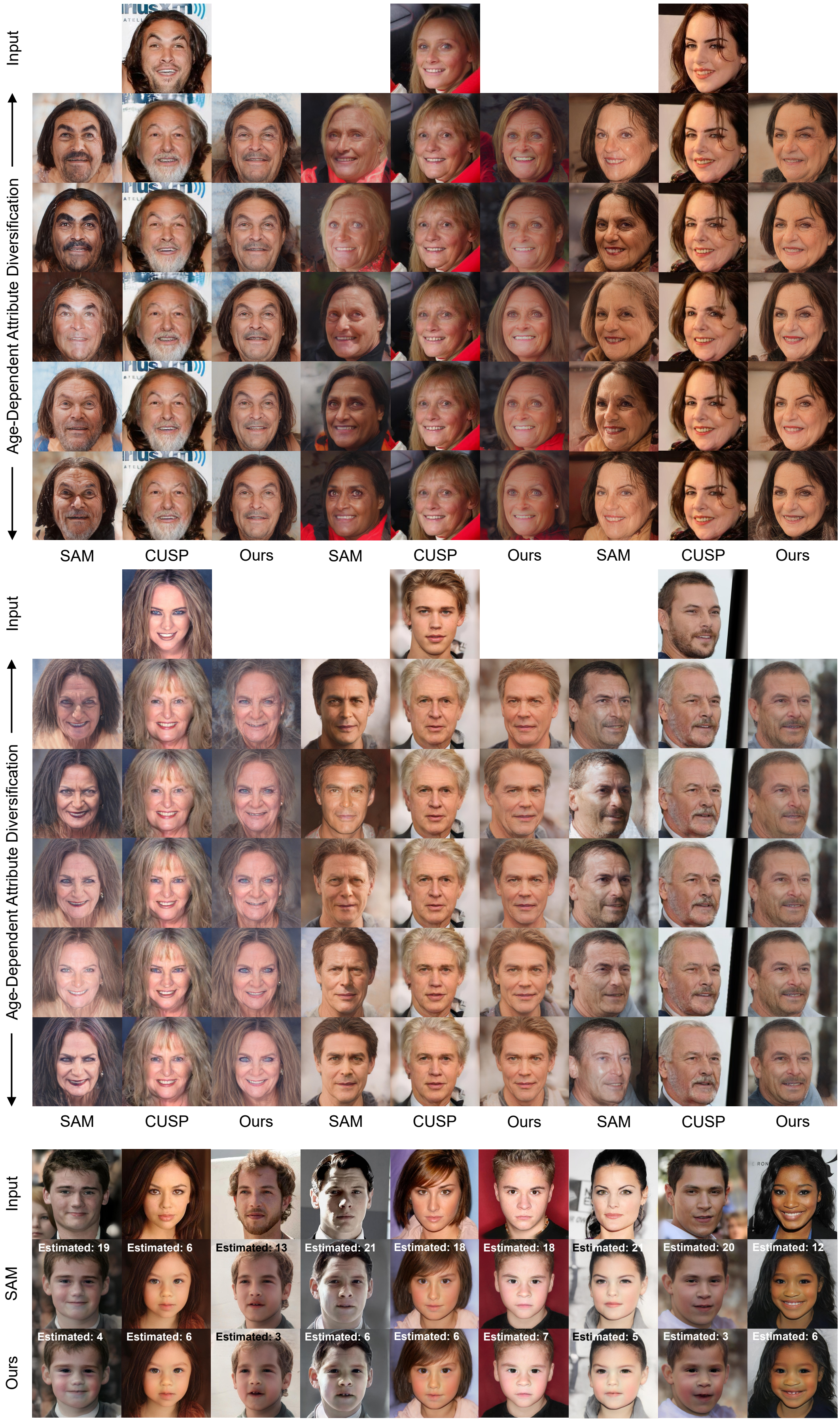

Fig. 11: Additional qualitative comparisons with SAM [3], CUSP [7], and our method for diversification (top and middle) and accuracy (bottom). The target ages are 65 for diversification and 5 for accuracy, respectively.