MyTimeMachine: Personalized Facial Age Transformation

Luchao Qi[1]   Jiaye Wu[2]   Bang Gong[1]   Annie N. Wang[1]   David W. Jacobs[2]   Roni Sengupta[1]
[1]University of North Carolina at Chapel Hill   [2]University of Maryland, College Park
{lqi, gongbang, awang13, ronisen}@cs.unc.edu   {jiayewu, djacobs}@cs.umd.edu
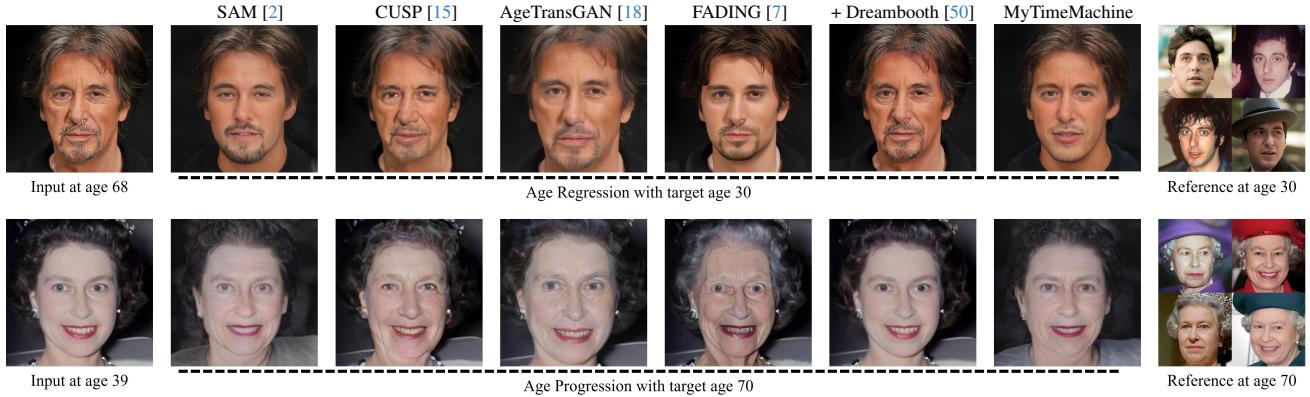
Figure 1. We introduce MyTimeMachine to perform personalized age regression (top) and progression (bottom) by training a person-specific aging model from a few (∼50) personal photos spanning over a 20-40 year range. Our method outperforms existing age transformation techniques to generate re-aged faces that closely resemble the characteristic facial appearance of the user at the target age.

Abstract

Facial aging is a complex process, highly dependent on multiple factors like gender, ethnicity, lifestyle, etc., making it extremely challenging to learn a global aging prior to predict aging for any individual accurately. Existing techniques often produce realistic and plausible aging results, but the re-aged images often do not resemble the person's appearance at the target age and thus need personalization. In many practical applications of virtual aging, e.g. VFX in movies and TV shows, access to a personal photo collection of the user depicting aging in a small time interval (20∼40 years) is often available. However, naive attempts to personalize global aging techniques on personal photo collections often fail. Thus, we propose MyTimeMachine (MyTM), which combines a global aging prior with a personal photo collection (using as few as 50 images) to learn a personalized age transformation. We introduce a novel Adapter Network that combines personalized aging features with global aging features and generates a re-aged image with StyleGAN2. We also introduce three loss functions to personalize the Adapter Network with personalized aging loss, extrapolation regularization, and adaptive w-norm regularization. Our approach can also be extended to videos, achieving high-quality, identity-preserving, and temporally consistent aging effects that resemble actual appearances at target ages, demonstrating its superiority over state-of-the-art approaches.

1. Introduction

What makes face aging so challenging? Virtual age transformation algorithms aim to digitally simulate the physical aging process of an individual's face. The goal of these methods [16, 18, 29, 33, 40, 63, 69] is to modify the shape and texture of the face to create the desired re-aging effect, while preserving the individual's unique identity, along with the pose, lighting, and style of the input image. However, facial aging is often highly dependent on several factors, such as ethnicity, gender, genetics, lifestyle, and health conditions [34, 53], which makes it challenging to model.

Existing age transformation algorithms [2, 15, 18] learn a generative global prior, modeling how an average face ages, using datasets like FFHQ [21]. While these methods have advanced in generating visually pleasing re-aging effects, they often fail to accurately depict how an individual actually ages. For instance, when provided with an image of *Al Pacino* at 68 years old, state-of-the-art techniques [2, 7, 15, 18] can produce a realistic yet inaccurate version of his appearance at 30, as shown in Fig. 1. For many practical applications, e.g. re-aging actors in films, it is crucial to accurately re-age individuals to the target age, as viewers are often familiar with the actor's appearance at various stages of life. However, accurately predicting an individual's re-aged appearance from a single image is highly ill-posed and challenging, since aging is person-specific [11].

1

In this paper, we show that accurate age synthesis can be performed when an algorithm has access to as few as 50 photos of an individual across a 20∼40 year time range. Personal photo collections are often available in many practical applications of virtual aging, and utilizing them can significantly improve the result, see MyTimeMachine in Fig 1. For example, consider de-aging effects often used in movies where a particular actor at 60 years old is shooting a scene where they need to be rendered as 30 years old. We can easily access the past 20∼40 years of photos of the actor to learn an accurate aging model. Similarly, consider an individual interested in simulating how a photograph of their loved one at 40 years would appear at 60 years old or beyond. We can also easily access the past 10-20 years of photo collection of their loved ones to understand the aging process and more accurately simulate their future appearance at 60 years and beyond. We therefore create a personalized aging method that can transform an input image to any target age, both within and beyond the age range represented in the personal photo collection used for training.

Simply personalizing a generic global age transformation algorithm, e.g. FADING [7], with Dreambooth [50] is ineffective. Personal photo collections often cover a limited range of age, pose, lighting, and style variations compared to large-scale facial datasets like FFHQ. Consequently, naive fine-tuning typically results in overfitting, limiting the model's ability to generalize to unseen ages, poses, styles, and lighting conditions, as shown by the extrapolation failure of FADING + Dreambooth in row 4 of Fig. 4. Additionally, FADING is built on diffusion, facing the typical inversion-editability trade-off problem [17, 58]. Specifically, re-aging requires both high fidelity to the input face at similar ages and high editability as the target age diverges. In contrast, such trade-offs have been more well explored in StyleGAN2's well-trained latent space [5, 47, 58, 62]. Therefore, we demonstrate an effective approach to personalized age transformation based on StyleGAN2.

Our proposed personalized age transformation network, MyTimeMachine (MyTM), introduces a novel adapter network that updates global facial aging features with personalized aging characteristics, trained on a personal photo collection using custom loss functions. Built on top of SAM [2], a global age transformation network capable of continuous aging without per-image optimization, MyTM enhances SAM's global age encoder, which projects an input image into StyleGAN2's latent space with a specified target age. We design a personalized adapter network that learns to adjust the global aging features. To train this adapter, we introduce three loss functions: personalized aging loss, extrapolation regularization, and adaptive w-norm regularization. The personalized aging loss ensures that identity-preserving features of the reaged image closely resemble those in a reference image from the personal photo col-

lection at a similar target age. Extrapolation regularization controls aging effects beyond the training age range using global priors, while adaptive w-norm regularization addresses StyleGAN's inversion-editability trade-off, ensuring distinct shape and texture changes due to aging while preserving identity. We then extend MyTM to video re-aging by utilizing face-swapping techniques to generate temporally consistent and identity-preserving results.

We curated a longitudinal aging dataset comprising high-quality images of 12 celebrities, captured under diverse conditions, including varying poses, expressions, and lighting. Inspired by real-world applications of personalized aging, we train our model on this dataset and establish two experimental frameworks to evaluate its performance: one for age regression, where a 70-year-old is rendered younger, and another for age progression, where a 40-year-old is rendered older. Our method outperforms existing global age transformation and naive personalization techniques, delivering high-quality, identity-preserving aging effects in both images and videos that closely resemble each individual's actual appearance at the target age.

In summary, our contributions are as follows: (i) We demonstrate that with access to a few (∼50) personal images spanning a few decades (20∼40 years), we can achieve high-quality, identity-preserving facial age transformations. These transformations accurately reflect the person's appearance at the target age while maintaining the style of the input image. (ii) We introduce several key technical advancements that integrate a global aging prior with a personal photo collection to enable personalized aging. Our approach trains an adapter network to adjust the global aging prior, utilizing three custom loss functions: personalized age loss, extrapolation regularization, and adaptive w-norm regularization. (iii) We show that MyTM can also be extended to perform temporally consistent and identity-preserving reaging in videos, which is important for many VFX applications.

2. Related Work

Traditional age transformation methods fall into two categories: prototype-based [24, 57] and physical model-based approaches [52, 55]. For a detailed overview, we refer readers to the survey by [13]. Recently, generative models have shown impressive results in synthesizing and editing high-resolution face images, inspiring their use in aging tasks.

Global Age Editing. Global age editing refers to age transformation without personal data. Prior works [38, 51] leverage StyleGAN2's well-trained latent space, identifying and traversing a linear age editing direction within it. However, this assumption often fails with larger age changes, especially across a lifespan, and can entangle other attributes (e.g., gender or glasses) [19]. To address this, recent methods [40, 66] introduce nonlinear aging paths in the latent space by training separate age encoders.

In diffusion models, several methods [4, 26, 28] perform age editing through latent manipulation guided by CLIP [43]. However, these methods continue to struggle with attribute entanglement in the latent space. FADING [7] improves disentanglement by projecting the input face into the diffusion model's latent space using NTI [35] and applying age editing through p2p [17].

However, FADING focuses on textural changes to achieve re-aging, neglecting the broader facial shape changes that occur over a person's lifespan [15]. This limitation arises because p2p identifies age-related pixels through attention control, resulting in localized edits rather than facial structural changes [49]. To address this, we build on StyleGAN2, leveraging its well-trained latent space for both fine-grained textural control and structural changes.

Personalization of Generative Models. Personalization involves tuning face models with personal images. PTI [47] fine-tunes a StyleGAN2 generator anchored by an inverted latent code. Other approaches [37, 42, 67] adapt the generator on a small set of personal images (50∼100) to create a personalized prior. In diffusion models, Dreambooth [50] optimizes the weights of the text-to-image network to adapt to a specific subject through a prompt identifier.

In the context of lifespan age transformation, these personalization techniques often overfit to the few training images in a limited range (e.g., ages 50 to 70), making it challenging to extrapolate to ages beyond the training range (e.g., 20 years old). We demonstrate that MyTM produces personalized face aging results within the training age range and generalizes to ages beyond it.

Video Re-aging. A recent video re-aging approach, FRAN [72], applies facial masks to predict age-related changes within masked regions per frame. However, similar to FADING, such method often overlooks structural changes in facial shape that naturally occur over a person's lifetime, such as the widening of a previously narrow face due to bone growth and shifts in facial fat distribution with age [15]. It also suffers with temporal consistency since it is trained on static images. To address this, Muqeet et al. [36] propose a re-aging model trained on synthetic video data, generating re-aged keyframes and interpolating between them to enhance temporal consistency. However, neither of these approaches is open-sourced or supports personalized video re-aging. Recent work [20] seeks to enhance temporal consistency in identity-specific face-swapping by personalizing models for individual users. However, this approach demands around 5,000 images of the person's face captured under various conditions, all at a similar age, limiting its effectiveness for lifespan aging transformations.

To address this, we follow face-swapping techniques [6, 65], using our personalized re-aged face as the source for swapping. This approach eliminates the need for training a dedicated model on a large number of personal images.

3. Method
In Sec. 3.1, we begin by outlining the fundamentals of global aging, SAM [2]. Next, Sec. 3.2 introduces our personalized aging adapter, MyTM, and explains how we integrate personal aging with global aging. Finally, we present the training losses for MyTM in Sec. 3.3.

3.1. Preliminaries
Here we provide a brief overview of SAM [2], a global age transformation network that forms the building block of our proposed personalized network, MyTM. SAM trains an age encoder (E_θ) that maps an input image (x) into the latent space $\mathcal{W}^+$ of StyleGAN, aligning with the desired target age (a_{tgt}). The latent code is then processed through the pre-trained StyleGAN (D_θ) to generate the age-transformed face (y_{tgt}). SAM is trained on the FFHQ dataset [22], where the training procedure involves producing an age-transformed output $y_{\text{tgt}} = D_\theta(E_\theta(x, a_{\text{tgt}}))$ in a forward pass. This process is supervised by the loss $\mathcal{L}_{\text{forward}}$, encouraging the re-aged image to be similar to the input image:

$$\mathcal{L}_{\text{forward}}(y_{\text{tgt}}) = \lambda_{l2}\mathcal{L}_2(y_{\text{tgt}}) + \lambda_{lpips}\mathcal{L}_{\text{LPIPS}}(y_{\text{tgt}}) \\ + \lambda_{id}\mathcal{L}_{\text{ID}}(y_{\text{tgt}}) + \lambda_{age}\mathcal{L}_{\text{age}}(y_{\text{tgt}}) \quad (1)$$

$\mathcal{L}_2$, $\mathcal{L}_{lpips}$ and $\mathcal{L}_{\text{ID}}$ matches age-transformed image (y_{tgt}) to input image (x) in pixel space, LPIPS feature space [68] and ArcFace [9] feature space respectively. $\mathcal{L}_{\text{age}}$ matches the predicted age by a face age detector, $DEX(\cdot)$ [48], with target age a_{tgt}:

$$\mathcal{L}_{\text{age}}(y_{\text{tgt}}) = ||a_{\text{tgt}} - DEX(y_{\text{tgt}})||_2 \quad (2)$$

After the forward pass, SAM encourages the transformed image (y_{tgt}) to be re-transformed back to the input image. This process helps ensure cycle consistency [71] and can be formally described as $y_{\text{cycle}} = D_\theta(E_\theta(y_{\text{tgt}}, a_x))$ with the same loss: $\mathcal{L}_{\text{cycle}}(y_{\text{cycle}}) = \mathcal{L}_{\text{forward}}(y_{\text{cycle}})$. The complete training loss is then given by:

$$\mathcal{L}_{\text{sam}} = \mathcal{L}_{\text{forward}}(y_{\text{tgt}}) + \mathcal{L}_{\text{cycle}}(y_{\text{cycle}}) \quad (3)$$

3.2. MyTM: Designing a Personalized Age Adapter
Training a personalized aging prior from scratch is suboptimal due to the limited availability of personal aging data. To address this, we introduce MyTM, which personalizes a pre-trained age encoder by combining two components: 1) SAM, a pre-trained age encoder that captures a shared global aging prior learned from a diverse set of identities, and 2) Age Adapter, a personalized age adapter network trained exclusively on an individual. Specifically, we assume the individual has a personal photo collection of N RGB images, x_i, each with an associated ground truth age, a_i, represented as $\mathcal{D} = \{(x_i, a_i)\}_{i=1}^N$.

Our key idea is to update the age-transformed latent code $\mathcal{W}_{\text{tgt}}^+$ produced by the global age transformation network, SAM, using a personalized adapter network, $AN(\cdot)$. The adapter takes the latent vector $\mathcal{W}_{\text{tgt}}^+$ predicted by SAM and computes an offset latent vector $\Delta\mathcal{W}_{\text{tgt}}^+ =$
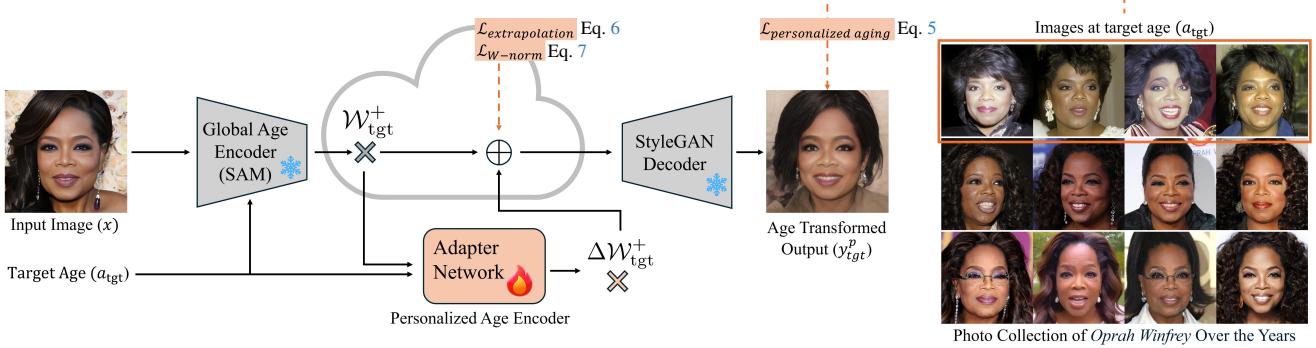
Figure 2. Given an input face of *Oprah Winfrey* at 70 years old, our adapter re-ages her face to resemble her appearance at 30, while preserving the style of the input image. To achieve personalized re-aging, we collect ∼50 images of an individual across different ages and train an adapter network that updates the latent code generated by the global age encoder SAM. Our adapter preserves identity during interpolation when the target age falls within the range of ages seen in the training data, while also extrapolating well to unseen ages.

$AN(E_\theta(x, a_{tgt}), a_{tgt})$. As shown in Fig. 2, we then combine this personalized latent adaptation $\Delta\mathcal{W}^+_{tgt}$ with the global latent code $\mathcal{W}^+_{tgt}$ and pass the result through the pre-trained StyleGAN2 decoder. Formally,

$$y^p_{tgt} = D_\theta(\underbrace{E_\theta(x, a_{tgt})}_{\text{global aging}} + \underbrace{AN(E_\theta(x, a_{tgt}), a_{tgt})}_{\text{personal aging}}) \quad (4)$$

By doing so, MyTM enables the integration of partially observed, personalized aging information of an individual—using only images of that person—into the global aging trajectory. Our adapter is based on an MLP architecture, with detailed implementation of the age adapter network available in the supplementary material.

3.3. MyTM: Loss Funtions

Our adapter is trained on a personal photo collection $\mathcal{D} = \{(x_i, a_i)\}^N_{i=1}$. We introduce three loss functions—personalized aging loss, extrapolation regularization, and adaptive w-norm regularization—to integrate global priors with personal data. Additionally, we also use the loss function $\mathcal{L}_{sam}$ from Sec. 3.1, based on SAM, to mitigate the forgetting of global priors [25, 39].

Personalized Aging Loss. After examining the loss formulation of SAM [2], we notice that the primary source of aging information is the aging loss $\mathcal{L}_{age}$, which relies on a pre-trained age classifier. The problem with this global aging loss is that the age classifier is not robust across different ethnicities, styles, and individual aging patterns [30]. It is often impossible to train a robust aging detector that works well for every individual. Rather than relying on the power of a global age classifier, we propose a personalized aging loss that encourages facial features of the transformed face to be similar to reference images in a similar age range in the training dataset. This encourages the re-aged image to closely resemble how the person looked at that age, ignoring the pose, lighting, and style variations.

We denote the minimum and maximum age of the training dataset $\mathcal{D} = \{(x_i, a_i)\}^N_{i=1}$ as $a_{min} = \min(a_1...a_n)$

and $a_{max} = \max(a_1...a_n)$. During training, we randomly sample a target age a_{tgt} between minimum and maximum age $a_{tgt} \sim \mathcal{U}(a_{min}, a_{max})$. We create a reference set, $\mathcal{D}_{tgt} = \{(x_j, a_j)\}^M_{j=1}$, which contains actual images of the individual near the target age ($a_{tgt} \pm 3$-years). We then employ a facial recognition network, Arcface [10], to extract identity features and compute the similarity between the age-transformed image y_{tgt} and all images in the reference set $\mathcal{D}_{tgt}$, and only consider the maximum similarity. Considering maximum similarity over the reference set ensures that the identity recognition networks are not significantly influenced by stylistic differences between images. Formally, we define the personalized aging loss as:

$$\mathcal{L}_{pers-age} = 1 - max\left\{\left\langle R(y^p_{tgt}), R(x_j)\right\rangle\right\}^M_{j=1} \quad (5)$$

$R(\cdot)$ is a pretrained ArcFace [9] network for facial feature recognition, $\langle\cdot, \cdot\rangle$ computes the cosine similarity between its argument [41], and M is the number of images in a reference set $\mathcal{D}_{tgt}$ with faces near the target age ($a_{tgt} \pm 3$-years).

Extrapolation Regularization. When training the adapter network with personalized age loss, we observe that the network's performance can degrade when a_{tgt} falls outside the training age range $[a_{min}, a_{max}]$. Specifically, this degradation manifests as the generated images continuing to resemble the appearance at the boundaries of the training age range (a_{min}, a_{max}), rather than appropriately aging or de-aging. For instance, as illustrated in row 4 of Fig. 4 (FADING + Dreambooth), when the training set covers faces aged 30 to 70, the model may overfit, generating faces that still resemble a 30-year-old when $a_{tgt} = 10$.

To prevent this extrapolation failure, we enforce the preservation of the pre-trained SAM's output during extrapolation. We apply experience replay [39, 50], which encourages the output of our personalized age encoder (y^p_{tgt}) to be similar to that produced by the pre-trained SAM (y_{tgt}):

$$\mathcal{L}_{reg-extra} = \lambda_{l2}\mathcal{L}_2(y^p_{tgt}, y_{tgt}) + \lambda_{LPIPS}\mathcal{L}_{LPIPS}(y^p_{tgt}, y_{tgt})$$
$$+ \lambda_{ID}\mathcal{L}_{ID}(y^p_{tgt}, y_{tgt}) \quad (6)$$

4

Adaptive w-norm regularization. During personalization, we observed that SAM struggles to capture distinct facial feature changes across ages, as illustrated in row 2 of Fig. 4. We attribute this issue to the inversion-editability trade-off [47, 58]. Specifically, the latent codes predicted by SAM are distant from the training distribution and the center of the latent space, $\overline{\mathcal{W}}$, reducing their editing capacity and making personalization challenging. This trade-off is particularly relevant in facial aging tasks. When the target age is close to the input age, we encourage the latent codes close to SAM's pre-trained output, preserving inversion accuracy while staying distant from the average latent code $\overline{\mathcal{W}}$. As the target age diverges from the input age, greater deformations in head shape and facial texture are needed, requiring latent codes nearer to $\overline{\mathcal{W}}$ to facilitate editing.

To address this, we propose adaptive W-norm regularization inspired by [46], where $\mathcal{L}_{reg} = \lambda_{reg}\|\mathcal{W}_{tgt}^+ - \overline{\mathcal{W}}\|$ is used to constrain the latent codes. We further enhance this by making λ_{reg} a cosine function, $\langle \cdot, \cdot \rangle$, of the difference between input and target age $\Delta_{age} = |a_i - a_{tgt}|$:

$$\mathcal{L}_{reg} = \lambda_{reg}(\Delta_{age})\|\mathcal{W}_{tgt}^+ - \overline{\mathcal{W}}\|$$
$$\lambda_{reg}(\Delta_{age}) = 1 - \langle \pi \cdot \Delta_{age}/100 \rangle \qquad (7)$$

4. Experiments

In Sec. 4.1 we first introduce our experimental setup, including datasets, experimental framework, state-of-the-art aging algorithms, and evaluation metrics. In Sec. 4.2 we present comparisons with state-of-the-art baselines, followed by the application of MyTM for video re-aging in Sec. 4.3. Finally, we present ablation studies in 4.4.

4.1. Experimental Setup

Dataset. We curated a dataset of images featuring 12 celebrities spanning a wide age range, including 7 males and 5 females from diverse ethnic backgrounds such as Caucasian, African American, Hispanic, and Asian. For further details, please refer to Table 3 in the supplementary material. For each celebrity, we train MyTM using 50 images, as discussed in Sec. 4.4. We then evaluate the performance of MyTM using test images of the same celebrity at either 40 or 70 years old, depending on the task outlined later in the experimental framework.

Experimental Framework. We consider the following two real-world scenarios where age transformation techniques are heavily used and demand high quality. **1) Age regression or de-aging** renders images of an individual to go back in time and is heavily used in VFX for movies and TV shows [59]. Motivated by this, we design an experimental setup where we personalize our aging model by training on images from either a 30∼70 or 50∼70 age range and then evaluate de-aging performance on an unseen image at 70 years old to a target age ($a_{tgt} \leq 70$ years old). We sample the target age every 10 years where $a_{tgt} \in \{0, 10, 20, 30, 40, 50, 60, 70\}$. **2) Age progression**

or aging renders images of an individual going forward in time and is used for forensic investigations, missing person searches, or as an emotional support tool to visualize departed loved ones. We design an experimental framework where we personalize our aging model by training on images from 20∼40 years old and evaluate on unseen faces at 40 years old to generate a target age ($a_{tgt} \geq 40$), where $a_{tgt} \in \{40, 50, 60, 70, 80, 90, 100\}$.

State-of-the-art Aging Algorithms. We compare our results with the following state-of-the-art aging methods: (i) SAM [2], which uses a pre-trained StyleGAN2 decoder and trains an age encoder on FFHQ [23]. It treats aging as a continuous process, enabling fine-grained control over transformations. (ii) CUSP [15], which jointly trains both an age encoder and decoder on FFHQ-Aging [40]. While effective for age transformations, it lacks fine-grained control due to its reliance on predefined age group-based transformations, limiting editing capabilities and supporting resolutions only up to 256×256. (iii) AgeTransGAN [18], an encoder-decoder architecture, also limited by age group-based transformations, similar to CUSP. (iv) FADING [7], which inverts images into the latent space of a face diffusion model using NTI [35], then edits them through p2p [17], allowing the aging diffusion model to be personalized to the input image. We also introduce additional personalization baselines: (v) SAM Pers. f.t., naively fine-tune SAM on personal images; (vi) FADING + Dreambooth, personalizes FADING by following the Dreambooth approach [50] with the prompt *"photo of a [a_{tgt}] year old [sks] man/woman"*.

Evaluation Metrics. Following the evaluation protocols in prior aging baselines [2, 7, 15, 18], we evaluate our personalized age transformation results in terms of age accuracy and identity preservation. We use the following metrics to evaluate the re-aged results quantitatively:

• *Age Accuracy (Age_{MAE})*. Following previous works [2, 15], we define age mean absolute error as $Age_{MAE} = |\hat{a_{tgt}} - a_{tgt}|$, where $\hat{a_{tgt}}$ is predicted by FP-Age [30].

• *Identity Preservation (ID_{sim})*. Previous works [56, 72] evaluate identity preservation by comparing the re-aged face to the input face. However, facial recognition systems, such as ArcFace [9], demonstrate a strong dependence on age and thus favor age consistency between the re-aged face and the input face [2], favoring algorithms that perform small changes. We address this problem by creating reference image sets of the individual near the target age which are not used in training, and then calculate the identity similarity to the reference images at the target age, in contrast to using the input image. Formally,

$$ID_{sim}(y_{tgt}) = max\{\langle R(y_{tgt}), R(x_j)\rangle\}_{j=1}^M \qquad (8)$$

where $R(\cdot)$ is a pretrained ArcFace [9] network for facial feature recognition and x_j belongs in reference image set near the target age ($a_{tgt} \pm$ 3-years). We report the average ID_{sim} across all sampled target ages.
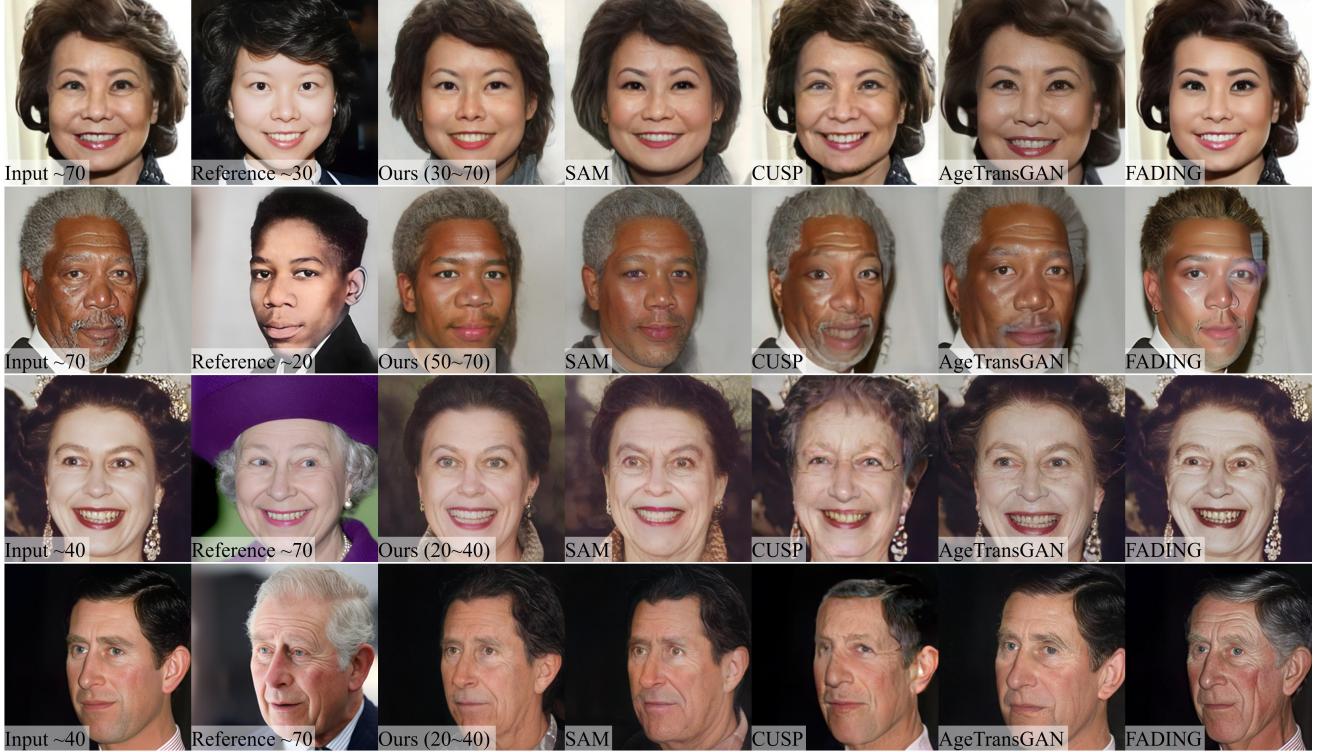
Figure 3. Performance of age transformation techniques for age regression (first two rows) and age progression (last two rows). The first column shows the input image, and the second column provides a reference image of the same person at the target age. MyTM (Ours) is compared against other state-of-the-art methods including SAM [2], CUSP [15], AgeTransGAN [18], and FADING [7].

4.2. Comparison with Age Transformation Methods

Age Regression. We use two age ranges of personal photos—40 years (ages 30∼70) and 20 years (ages 50∼70)—to examine the impact of training age span. Results are presented in Table 1, with visual examples of other pre-trained methods in Fig. 3 and other personalized methods in Fig. 4. For a detailed visual comparison across all ages (0∼100), please refer to Fig. 12 in the supplementary material.

Compared to pre-trained baselines (SAM, CUSP, Age-TransGAN, and FADING), our method achieves superior identity preservation (ID_{sim}), with an 11.7% improvement (0.67 vs. 0.60) in ID_{sim} over the best-performing method, FADING. This improvement is also maintained during interpolation (e.g., when $a_{\text{tgt}} \in 30 \sim 70$), producing a 9.0% increase in ID_{sim} (0.72 vs. 0.66) compared to FADING, even when FADING overfits to the input image via NTI [35], favoring its ID_{sim} score for smaller age gaps.

Compared to other personalized methods, our approach achieves both high age accuracy (Age_{MAE}) and strong identity preservation. SAM + Pers. f.t. shows minimal improvement over SAM alone, underscoring the effectiveness of our proposed loss function in Sec. 3.3. While FADING + Dreambooth [50] (50∼70) records a slight improvement over ours in ID_{sim} (0.78 vs. 0.76), it fails to maintain age accuracy (MAE 25.9 vs. 7.7) and overfits to the training age range, limiting its ability to generalize to unseen ages.

Method	$Age_{MAE}(\downarrow)$	$ID_{sim}(\uparrow)$		
		$a_{\text{tgt}} \leq 70$	$a_{\text{tgt}} \in 50 \sim 70$	$a_{\text{tgt}} \in 30 \sim 70$
SAM [2]	8.1	0.49	0.58	0.53
+ Pers. f.t. (50∼70)	8.2	0.48	0.58	-
+ Pers. f.t. (30∼70)	9.2	0.49	-	0.53
CUSP [15]	11.0	0.39	0.44	0.42
AgeTransGAN [18]	11.1	0.53	0.65	0.58
FADING [7]	8.9	0.60	0.72	0.66
+ Dreambooth [50] (50∼70)	25.9	0.63	**0.78**	-
+ Dreambooth [50] (30∼70)	23.0	0.64	-	<u>0.70</u>
Ours (50∼70)	**7.7**	<u>0.65</u>	<u>0.76</u>	-
Ours (30∼70)	<u>7.8</u>	**0.67**	-	**0.72**

Table 1. Performance of age regression where an input test image at 70 years old is de-aged to a target age $a_{tgt} \leq 70$. We also evaluate MyTM (Ours) using 20-year ($a_{\text{tgt}} \in 50 \sim 70$) and 40-year ($a_{\text{tgt}} \in 30 \sim 70$) age ranges in the training data. **Bold** indicates the best results, while <u>underlined</u> denotes the second-best.

Age progression. We perform age progression with a 20-year (ages $20 \sim 40$) range of personal photos. Age progression specifically evaluates the extrapolation ability of our technique to ages not seen in training. Quantitative results are presented in Table 2, with visual comparisons to other pre-trained methods shown in Fig. 3, and a full visual comparison provided in Fig. 12 in the supplement.

Our model outperforms pre-trained baselines, achieving the highest age accuracy (6.3) and best identity preservation (0.70 for $a_{\text{tgt}} \geq 40$ and 0.78 for $a_{\text{tgt}} \in 40 \sim 60$), due to the benefits of personalization. As shown in Fig. 12, FADING often produces poor results when the target age differs greatly from the input age, due to NTI + p2p editing [49].
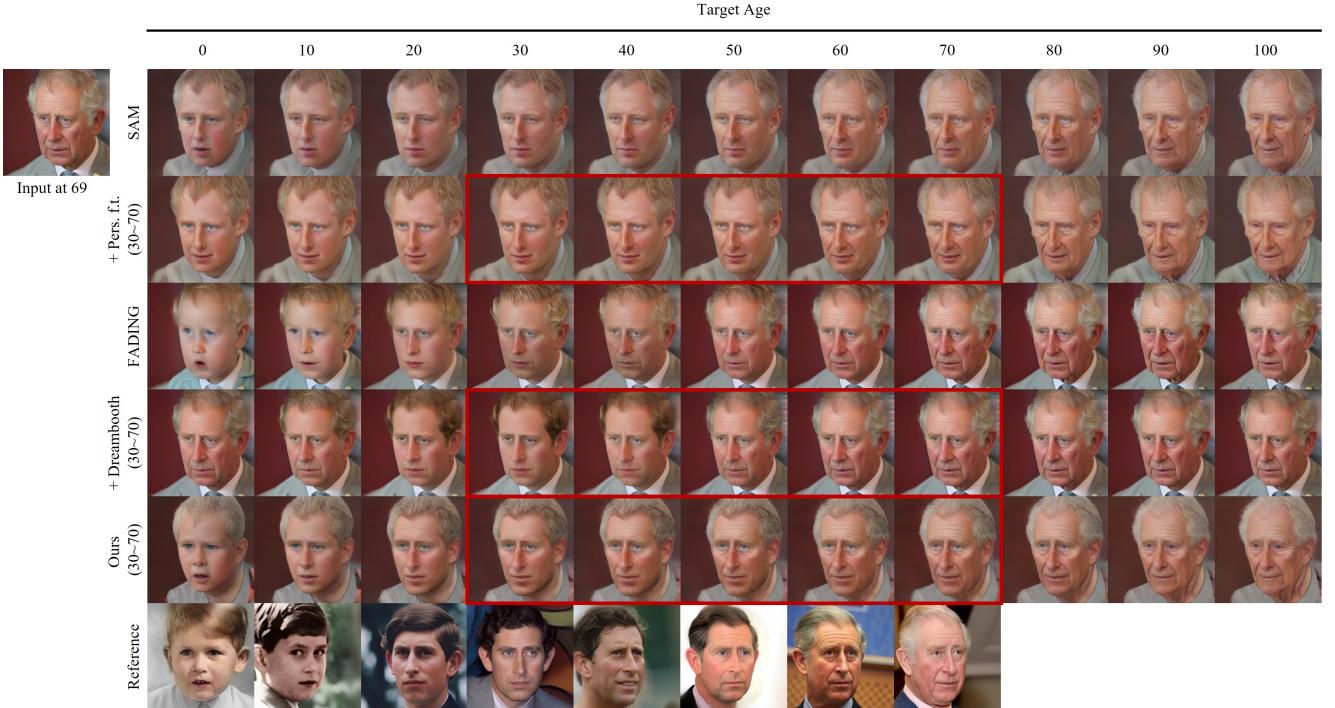
Figure 4. Performance of age transformation techniques for age regression, where an input test image around 70 is transformed to all target ages between 0 and 100. We show MyTM (Ours) trained on 40 years of data (ages 30 ∼ 70), with the age range included in the personal training data highlighted in red. An example image of the same person within 3 years of the target age is provided as a reference at bottom.
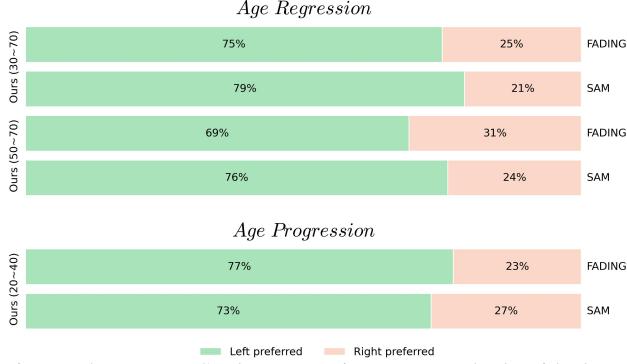

Figure 5. User Study comparing our method with baselines—FADING and SAM—for age regression ($a_{\text{tgt}} \leq 70$) and age progression ($a_{\text{tgt}} \geq 40$). We present the percentage of user preference for our method over the baselines.

Compared to other personalized methods, FADING + Dreambooth achieves slightly better ID_{sim} than our model (0.72 vs. 0.70). However, it struggles to extrapolate to unseen ages, resulting in a high Age_{MAE} of 20.2.

User Studies. We conduct user studies to qualitatively evaluate our method through pairwise human evaluations. In each pair, users see the original input image alongside two re-aged results at the target age a_{tgt}—one generated by an existing method and the other by ours, presented in random order. Users also receive reference images showing the person's face near a_{tgt} and are asked to select the result that best matches the reference images while preserving the style of the original input image.

Method	$Age_{MAE}(\downarrow)$	$ID_{sim}(\uparrow)$	
		$a_{\text{tgt}} \geq 40$	$a_{\text{tgt}} \in 40 \sim 60$
SAM [2]	6.9	0.54	0.58
+ Pers. f.t. (20∼40)	10.3	0.56	0.59
CUSP [15]	7.3	0.44	0.48
AgeTransGAN [18]	8.5	0.61	0.65
FADING [7]	7.6	0.62	0.71
+ Dreambooth [50] (20∼40)	20.2	**0.72**	0.77
Ours (20∼40)	**6.3**	0.70	**0.78**

Table 2. Performance of age progression where an input test image at 40 years old is aged to a target age $a_{\text{tgt}} \geq 40$. We also evaluate MyTM (Ours) using 20-year ($a_{\text{tgt}} \in 40 \sim 60$) and $a_{\text{tgt}} \geq 40$ age ranges in the training data. **Bold** indicates the best results, while underlined denotes the second-best. Note that FADING + Dreambooth has the lowest aging accuracy, as measured by Age_{MAE}.

We evaluate our method across two age regression tasks (30∼70 and 50∼70) and one age progression task (20∼40), totaling three tasks. We then sample one input and re-aged image pair per celebrity, resulting in 10 pairs for each age regression task and 8 pairs for the age progression task. For each pair, we collect 24 user responses for comparisons with FADING and 29 user responses for comparisons with SAM. As shown in Fig. 5, our method is significantly preferred over the baselines across all re-aging tasks.

4.3. MyTM for Video Re-aging

Having established a personalized aging prior, we extend our focus to video re-aging. Instead of training a separate re-aging model specifically for video re-aging, we build
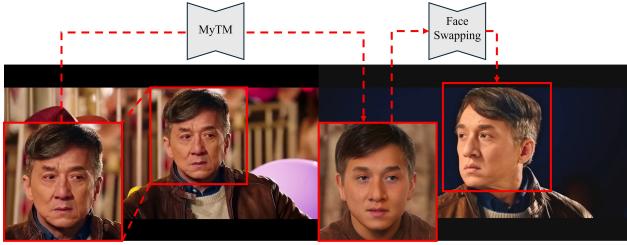
Figure 6. We apply video re-aging on a video of *Jackie Chan* from the movie *Bleeding Steel*. Left: The keyframe from the source video that we re-age with MyTM. Right: The re-aged face is mapped onto other frames of the source video via face-swapping.

upon face-swapping techniques by utilizing Inswapper[1], a widely adopted black-box model for face swapping [60].

Given a source video, we manually select a keyframe in a near-frontal pose with minimal occlusion and motion blur as the basis for re-aging. MyTM is then applied to this keyframe to transform the face image to the desired target age, generating a personalized re-aged face. Next, for each video frame, the face in the current frame and the re-aged face are input into the swapping model to generate the final re-aged result. This re-aged face is then pasted back onto the current frame using landmark-based warping. This framework requires only a single re-aged face for swapping, ensuring strong temporal consistency while preserving personalized facial identity. Our video re-aging pipeline is illustrated in Fig. 6. For further details on temporal consistency, please refer to the supplement.

4.4. Ablation Study

Effect of Dataset Size. We investigate the impact of training dataset size on MyTM by sampling subsets of images for each celebrity, with sizes of 10, 50, and 100. We then assess MyTM's performance on the age regression task (ages 30∼70), which demands the largest training age range. We report the average ID_{sim} in Fig. 7. Results indicate a significant performance improvement from 10 to 50 images, with minimal gains from 50 to 100 images. Consequently, we use 50 images for personalization, unless otherwise noted.
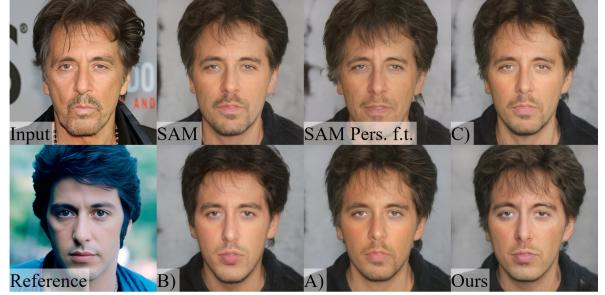
Effect of Proposed Loss Functions and Architecture. We analyze the effectiveness of our proposed network architecture and loss functions by conducting an ablation study in Fig. 8. We begin with SAM and progressively introduce each proposed component, including custom loss terms and the adapter network. For the age regression task, we train MyTM on ages spanning 30 to 70, testing with target ages $a_{tgt} \leq 70$. Our proposed Personalized Aging Loss yields the most improvement in ID_{sim}.

5. Conclusion

We present MyTimeMachine, a personalized facial age transformation technique that combines an individual's longitudinal photo collection (as few as 50 images) with global

Dataset $\mathcal{D}$ Size Ablations	Metric	SAM	$\mathcal{N}=10$	$\mathcal{N}=50$	$\mathcal{N}=100$
	Age$_{MAE}(\downarrow)$	8.1	8.5	7.8	8.0
	ID$_{sim}(\uparrow)$	0.49	0.58	0.67	0.67

Figure 7. Effect of training dataset size $\mathcal{D}$ on personalization. MyTM is trained on ages 30∼70 and tested for $a_{tgt} \leq 70$. Visual examples of *Robert De Niro* are shown at the top, with quantitative results displayed below.

Model	Adapter Network	Extrapolation Reg	Personalized Aging Loss	Adaptive W-norm Reg	ID$_{sim}(\uparrow)$
Ours	✓	✓	✓	✓	0.67
A)	-	✓	✓	✓	0.65
B)	-	-	✓	✓	0.65
C)	-	-	-	✓	0.55
SAM Pers. f.t.	-	-	-	-	0.46
SAM	-	-	-	-	0.45

Figure 8. Contributions of our proposed loss functions and the adapter network for the age regression task, trained on ages 30∼70 and tested for $a_{tgt} \leq 70$ on *Al Pacino*.

aging priors. Our main technical contributions, which enable the learning of a personalized aging prior, include a novel Adapter Network architecture, personalized aging Loss, extrapolation regularization, and adaptive w-norm regularization. Extensive quantitative and qualitative evaluations demonstrate that our method outperforms existing age transformation and naive personalization approaches.

Limitations. While our model effectively performs age transformations, as shown in Fig. 13 in the supplement, it can struggle with accessories (e.g. glasses) due to the limitations of the e4e encoder [58]. Furthermore, the pre-trained SAM model has difficulty in 1) modifying hair color to or from white—a common challenge in other aging works [29, 54]. 2) producing red-eye artifacts when generating older faces. Although our proposed w-norm regularization mitigates these issues, they are not fully resolved.

Ethical Considerations. Facial aging is a complex and inherently challenging problem, and even with personalization, our model may lack robustness across all underrepresented populations. Our approach also has the potential to produce manipulated images of real individuals, which poses a significant societal risk. This issue is common across generative models, underscoring the need for future research in detecting synthetic image composites.

References

[1] Mehran Aghabozorgi, Shichong Peng, and Ke Li. Adaptive IMLE for Few-shot Pretraining-free Generative Modelling. 2023. 2

[2] Yuval Alaluf, Or Patashnik, and Daniel Cohen-Or. Only a matter of style: age transformation using a style-based regression model. *ACM Transactions on Graphics*, 40(4):1–12, 2021. TLDR: An image-to-image translation method that learns to directly encode real facial images into the latent space of a pre-trained unconditional GAN subject to a given aging shift, which learns a more disentangled, non-linear path. 1, 2, 3, 4, 5, 6, 7

[3] David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun-Yan Zhu, and Antonio Torralba. Semantic Photo Manipulation with a Generative Image Prior. *ACM Transactions on Graphics*, 38(4):1–11, 2019. arXiv:2005.07727 [cs] TLDR: This paper adapts the image prior learned by GANs to image statistics of an individual image and can accurately reconstruct the input image and synthesize new content, consistent with the appearance of theinput image. 1

[4] Stefan Andreas Baumann, Felix Krause, Michael Neumayr, Nick Stracke, Vincent Tao Hu, and Björn Ommer. Continuous, Subject-Specific Attribute Control in T2I Models by Identifying Semantic Directions, 2024. arXiv:2403.17064 [cs] TLDR: It is shown that there exist directions in the commonly used token-level CLIP text embeddings that enable fine-grained subject-specific control of high-level attributes in text-to-image models and it is demonstrated that these directions can be used to augment the prompt text input with fine-grained control over attributes of specific subjects in a compositional manner. 3

[5] Anand Bhattad, Viraj Shah, Derek Hoiem, and D. A. Forsyth. Make It So: Steering StyleGAN for Any Image Inversion and Editing, 2023. arXiv:2304.14403 [cs]. 2

[6] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. SimSwap: An Efficient Framework For High Fidelity Face Swapping, 2021. arXiv:2106.06340. 3

[7] Xiangyi Chen and Stéphane Lathuilière. Face Aging via Diffusion-based Editing, 2023. arXiv:2309.11321 [cs]. 1, 2, 3, 5, 6, 7

[8] Jun Myeong Choi, Max Christman, and Roni Sengupta. Personalized Video Relighting With an At-Home Light Stage, 2023. arXiv:2311.08843 [cs]. 3

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019. 3, 4, 5

[10] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, 2022. arXiv:1801.07698 [cs]. 4

[11] Julien Despois, Frederic Flament, and Matthieu Perrot. AgingMapGAN (AMGAN): High-Resolution Controllable Face Aging with Spatially-Aware Conditional GANs, 2020. arXiv:2008.10960. 1

[12] Amil Dravid, Yossi Gandelsman, Kuan-Chieh Wang, Rameen Abdal, Gordon Wetzstein, Alexei A. Efros, and Kfir Aberman. Interpreting the Weight Space of Customized Diffusion Models, 2024. arXiv:2406.09413 [cs] TLDR: It is shown that inverting a single image into this space reconstructs a realistic identity, even if the input image is out of distribution (e.g., a painting). 2

[13] 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. 2

[14] Gege Gao, Huaibo Huang, Chaoyou Fu, Zhaoyang Li, and Ran He. Information Bottleneck Disentanglement for Identity Swapping. pages 3404–3413, 2021. 3

[15] Guillermo Gomez-Trenado, Stéphane Lathuilière, Pablo Mesejo, and Óscar Cordón. Custom Structure Preservation in Face Aging, 2022. arXiv:2207.11025 [cs]. 1, 3, 5, 6, 7

[16] Zhenliang He, Meina Kan, Shiguang Shan, and Xilin Chen. S2GAN: Share Aging Factors Across Ages and Share Aging Trends Among Individuals. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9439–9448, Seoul, Korea (South), 2019. IEEE. TLDR: The proposed method can achieve continuous face aging with favorable aging accuracy, identity preservation, and fidelity, and befitted from the effective design, a unique model is capable of all ages and the prediction time is significantly saved. 1

[17] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-Prompt Image Editing with Cross Attention Control, 2022. arXiv:2208.01626 TLDR: This paper analyzes a text-conditioned model in depth and observes that the cross-attention layers are the key to controlling the relation between the spatial layout of the image to each word in the prompt, and presents several applications which monitor the image synthesis by editing the textual prompt only. 2, 3, 5

[18] Gee-Sern Hsu, Rui-Cang Xie, Zhi-Ting Chen, and Yu-Hong Lin. AgeTransGAN for Facial Age Transformation with Rectified Performance Metrics. In *Computer Vision – ECCV 2022*, pages 580–595. Springer Nature Switzerland, Cham, 2022. Series Title: Lecture Notes in Computer Science. 1, 5, 6, 7

[19] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: Discovering Interpretable GAN Controls, 2020. arXiv:2004.02546 [cs]. 2

[20] iperov. iperov/DeepFaceLive, 2024. original-date: 2020-12-15T12:19:22Z. 3

[21] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks, 2019. arXiv:1812.04948 [cs, stat]. 1

[22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN, 2020. arXiv:1912.04958 [cs, eess, stat]. 3

[23] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*, 2020. 5, 1

[24] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M. Seitz. Illumination-Aware Age Progression. pages 3334–3341, 2014. 2

[25] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. arXiv:1612.00796 [cs, stat]. 4

[26] Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. Diffusion Models already have a Semantic Latent Space, 2023. arXiv:2210.10960 [cs] TLDR: Asyrp is proposed which discovers the semantic latent space in frozen pretrained diffusion models, named h-space, which has nice properties for accommodating semantic image manipulation: homogeneity, linearity, robustness, and consistency across timesteps. 3

[27] Jason Lee, Kyunghyun Cho, and Douwe Kiela. Countering Language Drift via Visual Grounding, 2019. arXiv:1909.04499 [cs]. 2

[28] Peipei Li, Rui Wang, Huaibo Huang, Ran He, and Zhaofeng He. Pluralistic Aging Diffusion Autoencoder, 2023. arXiv:2303.11086 [cs]. 3

[29] Zeqi Li, Ruowei Jiang, and Parham Aarabi. Continuous Face Aging via Self-estimated Residual Age Embedding. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15003–15012, Nashville, TN, USA, 2021. IEEE. 1, 8

[30] Yiming Lin, Jie Shen, Yujiang Wang, and Maja Pantic. FP-Age: Leveraging Face Parsing Attention for Facial Age Estimation in the Wild, 2022. arXiv:2106.11145 [cs]. 4, 5, 1

[31] Mingcong Liu, Qiang Li, Zekui Qin, Guoxin Zhang, Pengfei Wan, and Wen Zheng. BlendGAN: Implicitly GAN Blending for Arbitrary Stylized Face Generation, 2021. arXiv:2110.11728 [cs]. 1

[32] Yuchen Lu, Soumye Singhal, Florian Strub, Aaron Courville, and Olivier Pietquin. Countering Language Drift with Seeded Iterated Learning. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6437–6447. PMLR, 2020. ISSN: 2640-3498. 2

[33] Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park. Re-Aging GAN: Toward Personalized Face Age Transformation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3888–3897, Montreal, QC, Canada, 2021. IEEE. TLDR: This work proposes Re-Aging GAN (RAGAN), a novel single framework considering all the critical factors in age transformation that achieves state-of-the-art personalized face age transformation by compelling the input identity to perform the self-guidance of the generation process. 1

[34] Bryan Mendelson and Chin-Ho Wong. Changes in the Facial Skeleton With Aging: Implications and Clinical Applications in Facial Rejuvenation. *Aesthetic Plastic Surgery*, 36 (4):753–760, 2012. 1

[35] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text Inversion for Editing Real Images using Guided Diffusion Models, 2022. arXiv:2211.09794 [cs]. 3, 5, 6, 2

[36] Abdul Muqeet, Kyuchul Lee, Bumsoo Kim, Yohan Hong, Hyungrae Lee, Woonggon Kim, and KwangHee Lee. Video Face Re-Aging: Toward Temporally Consistent Face Re-Aging, 2023. arXiv:2311.11642 [cs] TLDR: A novel synthetic video dataset that features subjects across a diverse range of age groups is proposed that outperforms existing approaches in age transformation accuracy and temporal consistency and the development of novel metrics tailored explicitly for evaluating the temporal consistency of video re-aging techniques is developed. 3

[37] Yotam Nitzan, Kfir Aberman, Qiurui He, Orly Liba, Michal Yarom, Yossi Gandelsman, Inbar Mosseri, Yael Pritch, and Daniel Cohen-Or. MyStyle: A Personalized Generative Prior. *ACM Transactions on Graphics*, 41(6):206:1–206:10, 2022. TLDR: This work introduces MyStyle, a personalized deep generative prior trained with a few shots of an individual that allows to reconstruct, enhance and edit images of a specific person, such that the output is faithful to the person's key facial characteristics. 3, 2

[38] Yotam Nitzan, Rinon Gal, Ofir Brenner, and Daniel Cohen-Or. LARGE: Latent-Based Regression through GAN Semantics. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19217–19227, New Orleans, LA, USA, 2022. IEEE. TLDR: This work proposes a novel method for solving regression tasks using few-shot or weak supervision for GANs, and makes the simple observation that distances traversed along such directions are good features for downstream tasks - reliably gauging the magnitude of a property in an image. 2

[39] Yotam Nitzan, Michaël Gharbi, Richard Zhang, Taesung Park, Jun-Yan Zhu, Daniel Cohen-Or, and Eli Shechtman. Domain Expansion of Image Generators, 2023. arXiv:2301.05225 [cs]. 4

[40] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan Age Transformation Synthesis, 2020. arXiv:2003.09764 [cs]. 1, 2, 5

[41] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery, 2021. arXiv:2103.17249 [cs]. 4, 1

[42] Luchao Qi, Jiaye Wu, Annie N. Wang, Shengze Wang, and Roni Sengupta. My3DGen: A Scalable Personalized 3D Generative Model, 2023. arXiv:2307.05468 [cs]. 3, 2

[43] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision, 2021. arXiv:2103.00020 [cs]. 3

[44] Xiaohang Ren, Xingyu Chen, Pengfei Yao, Heung-Yeung Shum, and Baoyuan Wang. Reinforced Disentanglement for Face Swapping without Skip Connection, 2023. arXiv:2307.07928 [cs]. 3

[45] K. Ricanek and T. Tesafaye. MORPH: a longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345, 2006. 1

[46] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2287–2296, Nashville, TN, USA, 2021. IEEE. TLDR: This work presents a generic image-to-image translation framework, pixel2style2pixel (pSp), based on a novel encoder network that directly generates a series of style vectors which are fed into a pretrained StyleGAN generator, forming the extended latent space. 5

[47] Daniel Roich, Ron Mokady, Amit H. Bermano, and Daniel Cohen-Or. Pivotal Tuning for Latent-based Editing of Real Images, 2021. arXiv:2106.05744 [cs]. 2, 3, 5

[48] Rasmus Rothe, Radu Timofte, and Luc Van Gool. DEX: Deep EXpectation of Apparent Age from a Single Image. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 252–257, Santiago, Chile, 2015. IEEE. TLDR: The proposed method, Deep EXpectation (DEX) of apparent age, first detects the face in the test image and then extracts the CNN predictions from an ensemble of 20 networks on the cropped face, significantly outperforming the human reference. 3

[49] Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Semantic Image Inversion and Editing using Rectified Stochastic Differential Equations, 2024. arXiv:2410.10792. 3, 6, 2

[50] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation, 2023. arXiv:2208.12242 [cs]. 1, 2, 3, 4, 5, 6, 7

[51] Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs, 2020. arXiv:2005.09635 [cs, eess]. 2

[52] 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, 2010. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2

[53] Arthur Swift, Steven Liew, Susan Weinkle, Julie K Garcia, and Michael B Silberberg. The Facial Aging Process From the "Inside Out". *Aesthetic Surgery Journal*, 41(10):1107–1119, 2020. 1

[54] Xu Tang, Zongwei Wang, Weixin Luo, and Shenghua Gao. Face Aging with Identity-Preserved Conditional Generative Adversarial Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, Salt Lake City, UT, USA, 2018. IEEE. 8

[55] Yusuke Tazoe, Hiroaki Gohara, Akinobu Maejima, and Shigeo Morishima. Facial aging simulator considering geometry and patch-tiled texture. In *ACM SIGGRAPH 2012 Posters*, page 1, New York, NY, USA, 2012. Association for Computing Machinery. 2

[56] Qianrui Teng, Rui Wang, Xing Cui, Peipei Li, and Zhaofeng He. Exploring 3D-aware Lifespan Face Aging via Disentangled Shape-Texture Representations, 2023. arXiv:2312.16881 [cs]. 5

[57] B. Tiddeman, M. Burt, and D. Perrett. Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, 21(5):42–50, 2001. Conference Name: IEEE Computer Graphics and Applications. 2

[58] Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. Designing an encoder for StyleGAN image manipulation. *ACM Transactions on Graphics*, 40(4): 133:1–133:14, 2021. 2, 5, 8

[59] Ben Travis. The Irishman: Robert De Niro Recreated Iconic GoodFellas Scene To Test De-Ageing Tech – Exclusive, 2019. 5

[60] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, and Anthony Chen. InstantID: Zero-shot Identity-Preserving Generation in Seconds, 2024. arXiv:2401.07519 [cs]. 8

[61] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards Real-World Blind Face Restoration with Generative Facial Prior, 2021. arXiv:2101.04061 [cs]. 1

[62] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. GAN Inversion: A Survey, 2022. arXiv:2101.05278 [cs]. 2

[63] Jiu-Cheng Xie, Jun Yang, Wenqing Wang, Feng Xu, and Hao Gao. Diverse and Lifespan Facial Age Transformation Synthesis with Identity Variation Rationality Metric, 2024. arXiv:2401.14036 [cs]. 1

[64] Chao Xu, Jiangning Zhang, Yue Han, Guanzhong Tian, Xianfang Zeng, Ying Tai, Yabiao Wang, Chengjie Wang, and Yong Liu. Designing One Unified Framework for High-Fidelity Face Reenactment and Swapping. In *Computer Vision – ECCV 2022*, pages 54–71, Cham, 2022. Springer Nature Switzerland. TLDR: This paper proposes an effective end-to-end unified framework to achieve both face reenactment and swapping, and sufficiently transfers identity and attribute based on learned disentangled representations to generate high-fidelity faces. 3

[65] Zhiliang Xu, Zhibin Hong, Changxing Ding, Zhen Zhu, Junyu Han, Jingtuo Liu, and Errui Ding. MobileFaceSwap: A Lightweight Framework for Video Face Swapping, 2022. arXiv:2201.03808. 3

[66] Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. High Resolution Face Age Editing, 2020. arXiv:2005.04410 [cs]. 2

[67] Libing Zeng, Lele Chen, Yi Xu, and Nima Kalantari. MyStyle++: A Controllable Personalized Generative Prior, 2023. arXiv:2306.04865 [cs]. 3

[68] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 3

[69] Zhifei Zhang, Yang Song, and Hairong Qi. Age Progression/Regression by Conditional Adversarial Autoencoder, 2017. arXiv:1702.08423 [cs]. 1

[70] Yan Zheng and Lemeng Wu. InverseMeetInsert: Robust Real Image Editing via Geometric Accumulation Inversion

in Guided Diffusion Models, 2024. arXiv:2409.11734 version: 1 TLDR: GEO is introduced, an exceptionally versatile image editing technique designed to cater to customized user requirements at both local and global scales and is driven by a novel geometric accumulation loss that enhances DDIM inversion to faithfully preserve pixel space geometry and layout. 2

[71] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017. 3

[72] Gaspard Zoss, Prashanth Chandran, Eftychios Sifakis, Markus Gross, Paulo Gotardo, and Derek Bradley. Production-Ready Face Re-Aging for Visual Effects. *ACM Transactions on Graphics*, 41(6):1–12, 2022. TLDR: This paper presents the first practical, fully-automatic and production-ready method for re-aging faces in video images, and demonstrates how the simple U-Net allows for unprecedented temporal stability and preservation of facial identity across variable expressions, viewpoints, and lighting conditions. 3, 5

MyTimeMachine: Personalized Facial Age Transformation

Supplementary Material

Along with this supplemental PDF, we provide additional visual materials (e.g., images and videos) in an HTML file, accessible via <mark>*index.html*</mark>. We highly recommend readers to refer to the accompanying videos for a comprehensive examination of the visual outcomes.

A. Overview of Appendices

Our appendices contain the following additional details:

- Sec. B provides an overview of our data preprocessing pipeline, the curated dataset (summarized in Table 3), and additional details about the celebrities used in experiments Sec. 4.2.
- Sec. C provides implementation details of our personalized adapter network, including hyperparameters and training configurations.
- Sec. D presents benchmarking results against other pretrained aging methods, with qualitative results shown in Fig. 12.
- Sec. E includes benchmarking results against alternative naive personalization techniques, with both quantitative and qualitative results displayed in Fig. 9 and Fig. 10.
- Sec. F explains our choice of using StyleGAN2's aging encoder for personalization over encoder-decoder GAN models or diffusion models.
- Sec. G discusses the design rationale behind our video re-aging pipeline.

B. Dataset Curation

Existing in-the-wild aging datasets [30, 45, 69] lack longitudinal data for individual subjects, as they do not offer multiple high-quality images of the same person over several decades. To address this limitation, we collected a new dataset as summarized in Table 3. To better illustrate the age distributions, we count the number of images within the age ranges 20–40, 40–60, and 60–80, and report these numbers in the table. These ranges differ from the training ranges of 20–40, 50–70, and 30–70. For each celebrity, we first gathered facial images, then enhanced older images to improve visual quality, compensating for the limitations of earlier camera technology and image processing methods. Following [61], we restored grayscale or low-quality images to ensure a more consistent and enhanced visual representation over time. Faces were then cropped and aligned according to the FFHQ [23] standard. While downloading publicly available images, we extracted metadata, such as the time of capture, to calculate each subject's age.

For re-aging tasks in Sec. 4.2, the available age distribution of the collected celebrities varies; for instance, some celebrities have fewer than 50 images in the 20 to 40 age range. Therefore, we conduct age regression tasks for the

Celebrity	Age range	20~40	40~60	60~80
Al Pacino	21~84	89	56	198
Charles III	01~76	219	409	530
Elizabeth II	03~96	65	116	539
Robert De Niro	27~81	121	340	286
Jennifer Aniston	02~55	375	322	-
Oprah Winfrey	24~70	163	529	315
Morgan Freeman	20~87	4	136	290
Jackie Chan	21~70	31	444	201
Chow Yun-fat	20~68	91	109	60
Elaine Chao	16~71	14	117	123
Diego Maradona	17~60	165	301	-
Margaret Thatcher	20~87	70	270	268

Table 3. A longitudinal facial aging dataset featuring images of 12 celebrities. The number of images for each celebrity is reported across different age ranges. **Unless stated otherwise, 50 images are selected for training.**

following 10 celebrities: *Al Pacino, Charles III, Elizabeth II, Robert De Niro, Oprah Winfrey, Morgan Freeman, Jackie Chan, Chow Yun-fat, Elaine Chao, and Margaret Thatcher*, and age progression tasks for the following 8 celebrities: *Al Pacino, Charles III, Elizabeth II, Jennifer Aniston, Oprah Winfrey, Chow Yun-fat, Diego Maradona, and Margaret Thatcher*. These results also correspond to the number of pairs used in user studies, as discussed in Sec. 4.2. For dataset size ablation studies in Sec. 4.4, we use the same celebrities selected in the regression task.

C. Implementation Details

Personalized Age Adapter Network. Inspired by [3, 31, 41], our adapter network is built on a multi-layer perceptron (MLP) architecture that takes as input the latent vector $\mathcal{W}_{\text{tgt}}^+$ and the target age (a_{tgt}), and outputs the offset vector $\Delta\mathcal{W}_{\text{tgt}}^+$. Specifically, the 18×512 dimensional latent code $\mathcal{W}_{\text{tgt}}^+$ is first processed through a Global MLP, which produces a down-sampled global representation $\mathcal{W}_{\text{global}}$ of dimension 18×32, flattened to 1×512. Next, we design an Aging MLP that takes the scalar target age as input and generates a 1×16 dimensional age feature, $a_{\text{tgt-feat}}$. We then train 18 independent Style MLPs, each operating on one of the $k \in [1, 18]$ styles in the W+ space, to produce an offset vector for each style, $\Delta\mathcal{W}_{\text{tgt}}^+(k)$. Each Style MLP receives the 1×512 dimensional age-transformed latent code from SAM, $\mathcal{W}_{\text{tgt}}^+(k)$, the 1×512 dimensional global representation $\mathcal{W}_{\text{global}}$, and the target aging feature $a_{\text{tgt-feat}}$, and then outputs the per-style offset code $\Delta\mathcal{W}_{\text{tgt}}^+(k)$. Both the Global, Aging, and 18 Style MLPs are designed as 2-layer neural networks with ReLU activation. This architecture enables the network to subtly and effectively adjust

the latent representation, preserving the individual's identity while incorporating personalized aging characteristics.

For each celebrity, we train our adapter network on a GPU A6000 for 10,000 iterations, which takes approximately 4 hours. We inherit SAM's hyperparameters, including its original loss weights. Additionally, we set $\lambda_{\text{pers-age}} = 1$ for Eq. 5, $\lambda_{\text{reg-extra}} = 1$ for Eq. 6, and $\lambda_{\text{reg}} = 1$ for Eq. 7.

D. Comparison with SOTA Methods without Personalization

As discussed in Sec. 4.2, we benchmark our approach against all other pre-trained baselines, including SAM, CUSP, AgeTransGAN, and FADING, as shown in Fig. 12. For baseline methods like CUSP and AgeTransGAN, which utilize pre-defined age groups based on FFHQ-Aging [40], we interpolate between these age groups to demonstrate continuous aging, following the approach used by SAM [2].

E. Comparison with Naive Personalization Techniques.

We perform additional ablation studies using alternative personalization approaches on data for *Al Pacino* aged $30 \sim 70$ years, with results shown in Fig. 9. SAM Pers. f.t. behaves similarly to the pre-trained SAM, as the latent codes are far from the latent center, limiting its editing capabilities. This aligns with the inversion-editability trade-off discussed in Sec. 3.3. SAM Pers. ft. + MyStyle [37] first personalizes the SAM encoder, then tunes the decoder following the PTI pipeline [47]. However, this introduces significant artifacts due to changes in the latent distribution, which diverges from the pre-trained StyleGAN2 distribution. In SAM, global aging knowledge is learned with a fixed StyleGAN2 decoder, and modifying decoder weights distorts the latent space distribution, compromising the aging knowledge and introducing decoding artifacts. FADING + Dreambooth [50] overfits the aging results to the input image, especially when the target age lies outside the training age range. Additionally, this approach neglects age-related facial shape transformations, such as a toddler's rounder face or proportional changes in facial features over time, which are caused by NTI + p2p as discussed in Sec. 2. Additionally, as discussed in Sec. 4.2, we show visual comparisons against other naive personalization methods for age progression in Fig. 10.

F. Why Personalizing the Encoder SAM?

Finetuning encoder-decoder GAN with limited personal data often leads to overfitting, mode collapse [1], and data drift [27, 32], preventing the model from generalizing to unseen test images of an individual [42]. Therefore, encoder-decoder GAN structures, like AgeTransGAN, necessitate a substantial amount of paired data to achieve effective personalization in aging transformations. For instance, personalizing the appearance of a celebrity such as *Al Pacino*

Experiment	SAM	SAM Pers. ft.	SAM Pers. ft. + MyStyle	FADING + Dreambooth	Ours (30~70)
$\text{ID}_{sim}(\uparrow)$	0.45	0.49	0.60	0.64	0.66
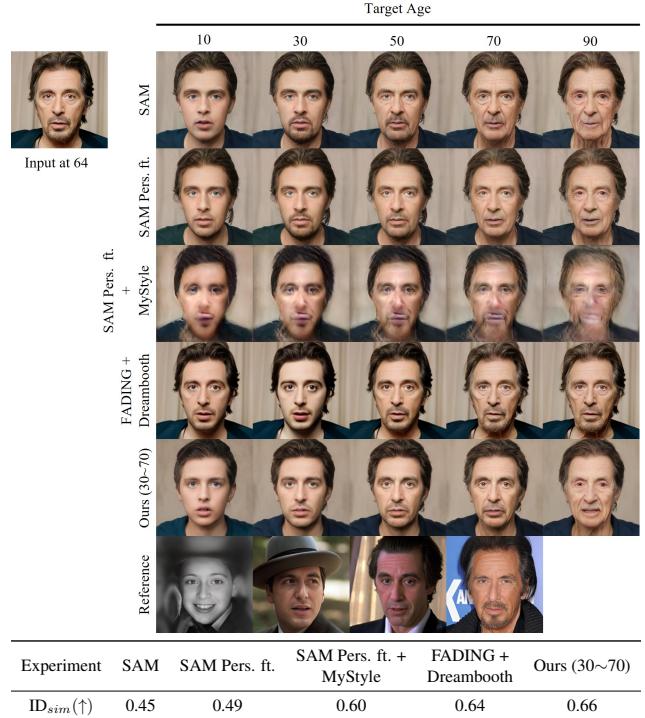
Figure 9. We compare MyTM (Ours) with naive personalization techniques: SAM Pers. ft., SAM Pers. ft. + MyStyle, and FADING + Dreambooth, trained on ages 30~70 and tested within the same age range for *Al Pacino*. While SAM Pers. ft. + MyStyle achieves a high ID_{sim} score, it suffers from poor visual quality, resulting in adversarial examples for ArcFace.

would necessitate images of him at both ages 20 and 70, with consistent pose, lighting, and expression. However, acquiring such data in real-world conditions is extremely challenging, as it demands rare and specific longitudinal images that capture individuals across a wide age span under controlled settings. This limitation makes encoder-decoder GANs less practical for applications where personalized aging transformations are desired.

For diffusion models, there are several limitations in re-aging tasks: (1) They lack the W latent space, which enables fine-grained continuous aging control and editing [12]. (2) Models like FADING, which use NTI + p2p for age editing, often struggle with the trade-off between inversion accuracy and editability [49]. Additionally, FADING frequently produces unstable results, as shown in Fig. 11, which we attribute to the unstable NTI optimization [49, 70]. New stable optimization-free methods could be explored for diffusion models in the future. (3) VQ auto-encoders, commonly used in diffusion to encode images, can introduce artifacts, particularly in the human face domain [35]. These issues highlight the need for an alternative approach, such as utilizing StyleGAN2's well-trained latent space and optimization-free e4e encoder [58], to achieve high-quality, artifact-free re-aging transformations.

Figure 10. Performance of personalized age transformation techniques for age progression, where an input test image is transformed to all target ages between 0 and 100. MyTM (Ours) is trained on 20 years of data (ages 20∼40). The age range included in the personal training data is highlighted in red. We also provide an example image of the same person within 3 years of the target age as a reference.
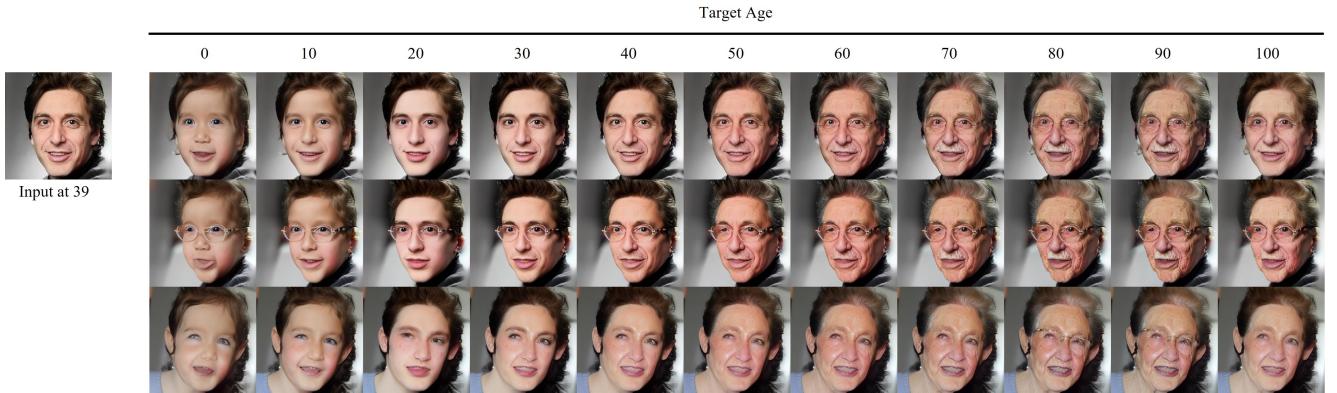

Figure 11. Visual results of FADING using identical input and inference code. The instability in age transformation arises from the optimization of NTI [35], leading to inconsistencies.

G. Why Not Use a Reference Image for Face-Swapping?

Firstly, obtaining images of a person at any arbitrary age is often challenging, particularly high-quality images comparable to our synthesized faces at 1024x1024 resolution. Even if reference images at the target age are available, face-swapping techniques [6, 14, 64] generally yields optimal results when the source and target faces share similar styles, such as pose, expression, and lighting. Signifi-

cant style differences between the source and target faces can cause artifacts like flickering, particularly in real-world video scenarios [8, 44].
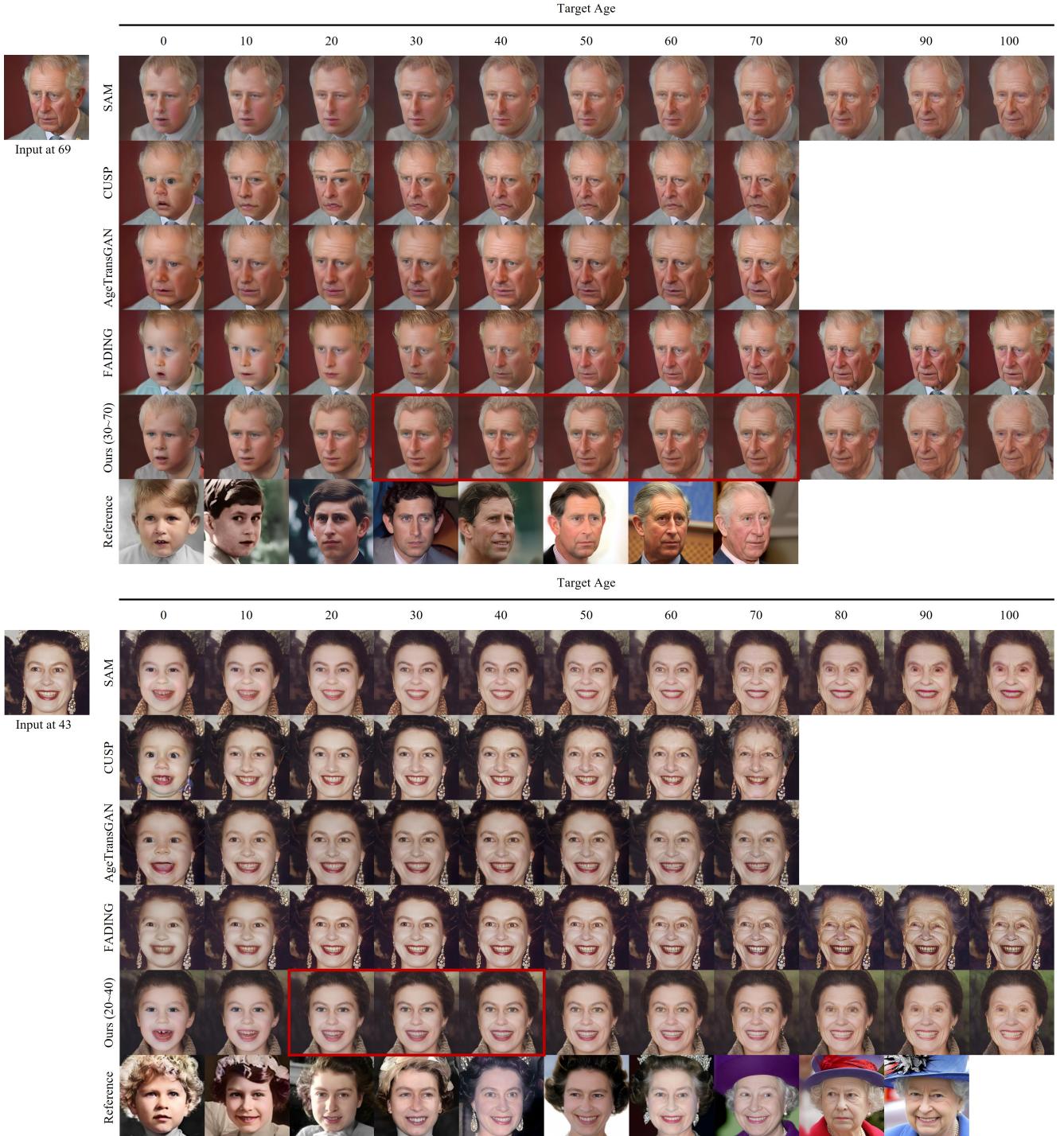
Figure 12. Performance of age transformation techniques for age regression (top) and age progression (bottom). The input test images match those in Fig. 4 and Fig. 10 for consistency. For age regression, MyTM (Ours) is trained across a 40-year range (ages 30 to 70), while for age progression, it is trained over a 20-year range (ages 20 to 40). Personalized training data age ranges are marked in red. A reference image of the same person, taken within three years of the target age, is included for comparison.
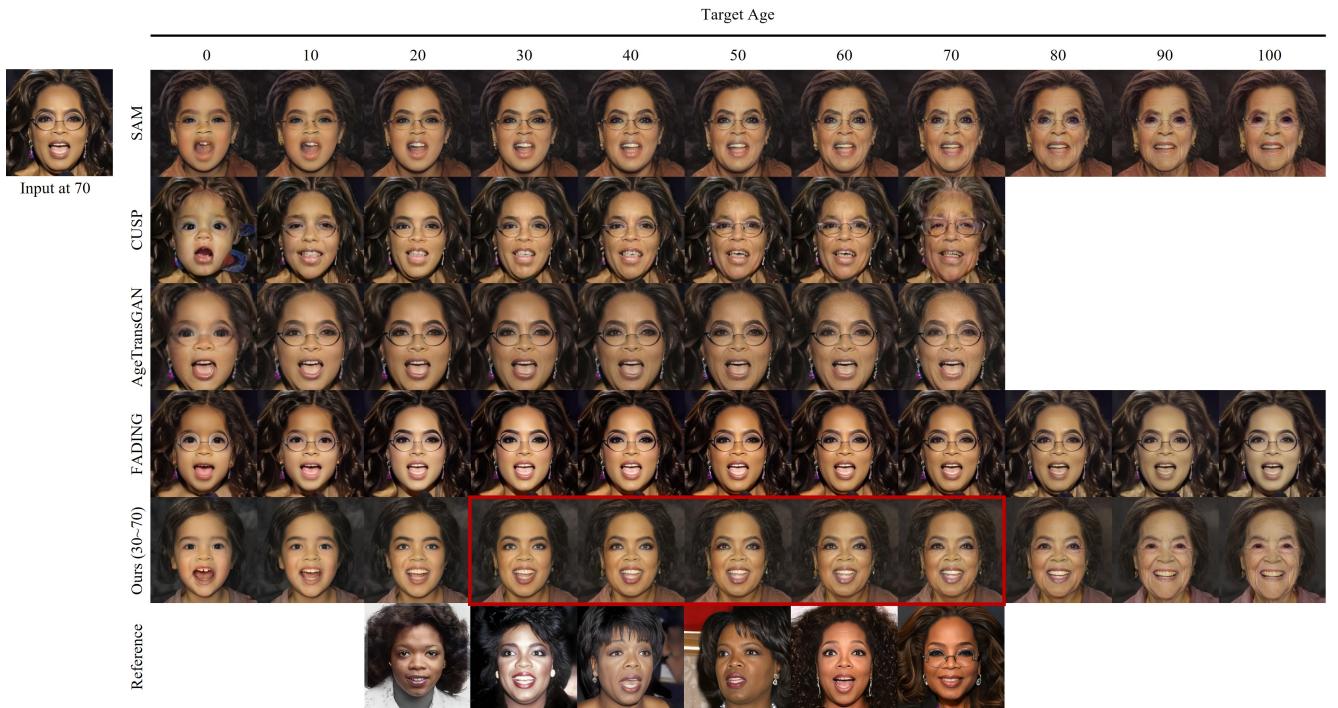
Figure 13. Limitations of MyTM. Our method may struggle with accessories (e.g., glasses), as these elements are not consistently handled by the e4e encoder [58].