Production-Ready Face Re-Aging for Visual Effects

GASPARD ZOSS, DisneyResearch|Studios, Switzerland
PRASHANTH CHANDRAN, DisneyResearch|Studios, Switzerland and ETH Zurich, Switzerland
EFTYCHIOS SIFAKIS, University of Wisconsin-Madison, USA and DisneyResearch|Studios, Switzerland
MARKUS GROSS, DisneyResearch|Studios, Switzerland and ETH Zurich, Switzerland
PAULO GOTARDO, DisneyResearch|Studios, Switzerland
DEREK BRADLEY, DisneyResearch|Studios, Switzerland

Fig. 1. We present FRAN, a production-ready face re-aging network that can automatically re-age face images in video without identity loss and with great artistic control. FRAN provides temporally stable results on videos depicting faces under free viewpoint, depth, motion, and illumination conditions. Here we show several video frames of an individual (age 35), re-aged to 65 (top row) and 18 (bottom row).

Photorealistic digital re-aging of faces in video is becoming increasingly common in entertainment and advertising. But the predominant 2D painting workflow often requires frame-by-frame manual work that can take days to accomplish, even by skilled artists. Although research on facial image re-aging has attempted to automate and solve this problem, current techniques are of little practical use as they typically suffer from facial identity loss, poor resolution, and unstable results across subsequent video frames. In this paper, we present the first practical, fully-automatic and production-ready method for re-aging faces in video images. Our first key insight is in addressing the problem of collecting longitudinal training data for learning to re-age faces over extended periods of time, a task that is nearly impossible to accomplish for a large number of real people. We show how such a longitudinal dataset can be constructed by leveraging the current state-of-the-art in facial re-aging that, although failing on real images, does provide photoreal re-aging results on synthetic faces. Our second key insight is then to leverage such synthetic data and formulate facial re-aging as a practical image-to-image translation task that can be performed by training a well-understood U-Net architecture, without the need for more complex network designs. We demonstrate how the simple U-Net, surprisingly, allows us to advance the state of the art for re-aging real faces on video, with unprecedented temporal stability and preservation of facial identity across variable expressions, viewpoints, and lighting conditions. Finally, our new face re-aging network (FRAN) incorporates simple and intuitive mechanisms that provides artists with localized control and creative freedom to direct and fine-tune the re-aging effect, a feature that is largely important in real production pipelines and often overlooked in related research work.

CCS Concepts: • **Computing methodologies → Image manipulation**.

Additional Key Words and Phrases: Facial re-aging, Image and Video editing

ACM Reference Format:
Gaspard Zoss, Prashanth Chandran, Eftychios Sifakis, Markus Gross, Paulo Gotardo, and Derek Bradley. 2022. Production-Ready Face Re-Aging for Visual Effects. *ACM Trans. Graph.* 41, 6, Article 237 (December 2022), 12 pages. https://doi.org/10.1145/3550454.3555520

Authors' addresses: Gaspard Zoss, DisneyResearch|Studios, Switzerland, gaspard.zoss@disneyresearch.com; Prashanth Chandran, DisneyResearch|Studios, Switzerland and ETH Zurich, Switzerland, prashanth.chandran@disneyresearch.com; Eftychios Sifakis, University of Wisconsin-Madison, USA and DisneyResearch|Studios, Switzerland, sifakis@cs.wisc.edu; Markus Gross, DisneyResearch|Studios, Switzerland and ETH Zurich, Switzerland, gross@disneyresearch.com; Paulo Gotardo, DisneyResearch|Studios, Switzerland, paulo.gotardo@disneyresearch.com; Derek Bradley, DisneyResearch|Studios, Switzerland, derek.bradley@disneyresearch.com.

0730-0301/2022/12-ART237 $15.00
https://doi.org/10.1145/3550454.3555520

1 INTRODUCTION

The use of digitally aged or de-aged human characters in movie productions and advertising has increased dramatically over the last few years. Whether for making Robert De Niro younger in *The Irishman* [2020], or David Beckham older in an ad campaign against malaria [2020], digital re-aging is quickly becoming an essential tool in every visual effects studio. The re-aging process typically starts by filming a stand-in actor, who is often the target actor at the present day age, and then digitally re-aging the performance in post-production. In general, there are two different approaches commonly used for CG digital re-aging. One approach follows the traditional 3D facial modeling pipeline, where a complete 3D re-aged face rig is modeled, animated and rendered to replace the original likeness in the scene. Normally, this rendered rig is driven by the captured performance of the stand-in actor, thus often requiring a second complete 3D face rig of the present-day actor (before re-aging) to facilitate performance retargeting. Due to the overwhelming complexity and time required for this approach, it is often only considered for hero characters that occupy a lot of screen time, with many close-up shots. The second approach to digital re-aging follows a purely 2D "painting" workflow. Given the filmed performance of the present-day actor, the goal is to consistently edit each video frame to synthetically change the actor's age. While this method offers less overall control than the full 3D approach (*e.g.*, large changes to viewpoint or scene illumination are typically out of the question), there is a large benefit in the simplicity of the method over the 3D approach, as there is no requirement to scan the actor ahead of time and build a face rig, fit the face rig to new performances, model a second re-aged face rig, transfer the performance to the re-aged version and realistically render the result back into the scene. Therefore, in-camera 2D digital re-aging has grown in popularity and has been used on a number of blockbuster film productions such as re-aging Michael Douglas in *Ant-Man* and Samuel L. Jackson in *Captain Marvel* [2019].

While the 2D workflow for digital re-aging is the more straightforward one, it still comes at the cost of manually editing a performance video, often frame by frame. For instance, when making an actor older, each frame must integrate the expected growth of ears and nose, the loss of muscle tone and the sagging of facial skin, the addition of dynamic wrinkles, and even changes in skin pigmentation and blood flow. This process requires skilled artists to achieve high-quality photorealistic and temporally coherent results, and it can take several days to re-age even a single shot. In the following, we present a new automatic and controllable method for high-resolution facial re-aging that is directly applicable to video images and can be used to re-age complete shots, taking under just five seconds per re-aged video frame.

In this work, we aim for a practical, production-ready solution to digital face re-aging that complements traditional re-aging techniques that already work well in film production. For example, when shooting an actor who is meant to be playing a different age, production teams use dedicated costumes and hairstyles to depict the new age, with the goal of practical re-aging of every part of the character except the face, which is to be re-aged digitally in post-production. A concrete example is shown in Fig. 2, where a present-day actor at age 46 is transformed to a character at age 20, through practical hair and costume changes followed by digital face re-aging. This highlights the goal of our method, which is to digitally re-age skin regions in a way that blends naturally with the rest of the image. Furthermore, in production settings, extreme re-aging to very young ages (*e.g.*, children) already has a well established, plausible solution that is to cast stand-in actors of the target age, which also properly addresses the large changes in size and appearance of the whole body. Therefore, we formulate our digital re-aging solution without such extreme changes, focusing only on adult ages. Within this scope, we still face an extremely challenging problem that has no single "ground truth" solution. Nevertheless, we present a method for a tractable solution with high quality and degree of automation.

Fig. 2. Our method targets the production use-case of practically re-aging the costume and makeup (center) for a present day actor (left) and then digitally re-aging only the face (right).

The main challenge researchers face when modeling the facial aging process is in obtaining high-quality, longitudinal data of the same individuals, from a variety of ethnicities, over a long period of time spanning several years or decades; collecting such data for a large number of real people is nearly impossible. For this reason, the current state of the art in facial re-aging often leverages powerful neural face models that are pre-trained on thousands of images of different faces without any age pairing [Karras et al. 2020b]. The parameter space of the model is then traversed along highly elaborate "semantic dimensions" to provide realistic edits such as re-aging [Abdal et al. 2021; Alaluf et al. 2021; Härkönen et al. 2020; Shen et al. 2020]. However, these re-aging approaches only perform optimally on synthetic images with static, nearly frontal faces. When re-aging real faces in video, we inevitably see a loss in the identity features, since the neural model cannot always faithfully represent the characteristic facial features that are unique to the person at hand. Even with little or no re-aging at all, these models generate inconsistent identities across subsequent video frames with different expressions, viewpoints, and lighting (see supplemental video).

Here, we leverage these recent advances in neural face modeling while taking a different approach for re-aging. Our first key insight is to show that, although current state-of-the-art re-aging approaches often fail on real face images, one can still leverage these methods to generate rich longitudinal datasets comprised of fully synthetic faces with photorealistic and consistent aging effects. In such a dataset, although the identity is synthetic it remains consistent across ages. This solves the first main challenge that is to obtain paired training images that depict the same person and background, under the same viewpoint and lighting, and with the same facial expression, differing only in age (see suplemental material).

As our second key contribution, we leverage this synthetic, paired dataset and formulate facial re-aging as a practical image-to-image translation task, for which a natural solution is to train a well-understood U-Net architecture, without the need for more complex network designs. We demonstrate how such simple U-Net, surprisingly, helps us improve upon the state of the art for re-aging real faces on video, providing unprecedented temporal stability and preservation of facial identity across variable expressions, viewpoints, and lighting conditions. Our new face re-aging network, *FRAN*, allows for photorealistic and continuous face re-aging within 18 and 85 years of age, both forward and backwards in time. FRAN also incorporates simple and intuitive mechanisms to provide artists with localized control and creative freedom to direct and fine-tune the re-aging effect, a feature that is largely important in real production pipelines and often overlooked in related research work.

We believe FRAN has the potential to drastically reduce time and costs involved in digital re-aging for the entertainment and advertising industries. We demonstrate our method by automatically re-aging both still images and videos of actor performance, in comparison to the current state-of-the-art in re-aging.

2 RELATED WORK

The problem of modeling aging effects and age estimation from face images has been studied for several years and for different application domains, ranging from entertainment to advertisement, medical, cosmetics, and forensics, to name a few. Here, we focus on the more recent, closely related work on digital face re-aging, and refer the reader to the excellent in-depth surveys on facial age estimation [Angulu et al. 2018] and modeling [Georgopoulos et al. 2018; Ramanathan et al. 2009].

As mentioned, collecting paired longitudinal data of many real people, over years or decades, is a nearly impossible task. Often, the few available samples have to be partitioned into a small number of discrete age groups [Kemelmacher-Shlizerman et al. 2014], which prevents the modeling of the continuous re-aging process. This issue is further exacerbated by the fact that the most promising deep learning tools for building data-driven models are also data-hungry, with a large number of tunable parameters. To help alleviate this problem, considerable progress has been made in recent years by deriving new strategies for training on unpaired data, using self-supervision via cycle consistency [Zhu et al. 2017] or using adversarial training. Generative Adversarial Networks (GANs) [Goodfellow et al. 2014] can be trained in an unsupervised way over a very large corpus of nearly uncurated face images, to learn highly expressive neural face models [Gal et al. 2021; Karras et al. 2020a, 2019, 2020b]. These models can synthesize an unlimited number of photo-realistic human portraits with male and female faces of different identities, ethnicities, ages, expressions, viewpoint, lighting, hair styles, accessories (glasses, ear rings) and backgrounds. Today, end-consumer software such as *Adobe Photoshop* [Adobe 2022] and even smartphone applications such as *FaceApp* [2022] can plausibly re-age a facial portrait in seconds. However, control over this process is still very limited or even non-existent; often the user is simply left with a few options for a handful of target ages. Furthermore, it can be extremely challenging to re-age video sequences consistently, including facial performance dialog and head pose changes, without temporal jitter and artifacts.

For low resolution re-aging, we also refer the reader to to the survey by Shu et al. [2016]. In this context, Hsu et al. [2021] proposed an adversarial approach where a discriminator guides the re-aging network to mimic aging effects on the face, while two other networks ensure that facial identity and other attributes are retained during the age transformation. Fang et al. [2020] propose a triple translation approach to learn common aging modalities across multiple identities in an adversarial fashion. Liu et al. [2021] use facial attributes that are common across identities to guide an attribute-aware, attention-based generator to re-age the input; their network is trained in an adversarial fashion alongside a wavelet-based discriminator that determines if the output of the generator corresponds to that of an individual with similar facial attributes. A multi-task learning approach is taken by Huang et al. [2021] to learn a unified embedding of age progression and facial identity; they show how learning such a unified embedding of age and identity can not only help in synthesizing re-aged images, but also in face recognition. Duong et al. [2019] demonstrate aging results on short videos using a deep reinforcement learning method. Despite the great progress achieved in these works, the quality and resolution of the re-aged images are still quite limited, with the output showing artifacts and instability across video frames.

Among the techniques that handle high resolution images, Li et al. [2021] approach re-aging by embedding an age estimator into a generative network and by training both the age estimator and the generator simultaneously. Their method supports high resolution and continuous re-aging that also preserves the input identity. Similarly, Yao et al. [2021] propose a network for continuous high-resolution re-aging that is conditioned by the target age using an age modulation layer. This network is trained in an adversarial fashion with a cycle consistency loss and an age classifier. Makhmudkhujaev et al. [2021] also use an age modulation layer coupled with an encoder-decoder architecture. He et al. [2021] explicitly model the variation in shape, texture and identity in their lifespan face synthesis method. Unfortunately, the applicability of these methods to video sequences is not demonstrated. Like us, Despois et al. [2020] also explore convolutional image-to-image translation to provide localized control on re-aging: the input image is presented with an associated map of aging scores that modulate image decoding via SPADE blocks [Park et al. 2019]. Instead of longitudinal data, their dataset was captured in studio and comprises 6000 frontal, neutral faces of different ages, genders and ethnicities, with manually labeled aging scores. Training is done using a cycle-consistency loss on re-aging, and an adversarial loss for photo-realism. This method performs well in preserving subject identity, but results are only shown for static, nearly frontal faces, in studio conditions. In comparison, our method provides similar level of control while also generalizing well across different facial expressions, viewpoints and lighting conditions, with temporal consistency across video frames.

Recent re-aging research has also turned to exploring the latent space of powerful pre-trained face GANs. Leveraging the semantics learned by the neural model, this body of research work seeks to re-age a face, represented as a particular latent point, either by interpolating an age code explicitly or by traversing the latent space along a linear or non-linear path (a "semantic dimension") as steered by a pre-trained age classifier [Abdal et al. 2021; Alaluf et al. 2021; Antipov et al. 2017; Härkönen et al. 2020; Or-El et al. 2020; Shen et al. 2020; Yang et al. 2021]. This strategy has provided realistic re-aging *for synthetic faces that are perfectly represented in the latent space.* Typically, for real face images, a loss in identity is observed when the real image is encoded into the GAN's latent space and then "filtered" through the neural generator [Tov et al. 2021]. That is, although these models can generate an unlimited number of realistic faces that do not exist in reality, they cannot accurately represent all the particular skin detail and identity features that are unique to a particular real individual not seen during training. Even when the model is optimized further to overfit the new face [Tzaban et al. 2022], re-aging performance may be lost. Another drawback of latent-space traversal methods is that control over re-aging is very limited, lacking any form of localized edits on the different areas of the re-aged face. Furthermore, temporal continuity when re-aging video sequences is hard to achieve. Closer in spirit to our approach, Viazovetsky et al. [2020] investigate speeding up such semantic edits in the latent space by training a feed-forward "student" network for similar tasks. They demonstrate results on different operations but do not properly evaluate re-aging in particular, nor consider identity preservation, consistency across video frames, and level of control. In the following, we demonstrate how our longitudinal data sampling and U-Net design allow us to derive a production-ready approach that advances the state of the art for re-aging real faces on video, with unprecedented performance and artistic control.

3 METHOD

This section presents our identity-preserving, controllable re-aging approach for video images depicting faces in arbitrary expressions, viewpoints, and lighting. We formulate the problem as image-to-image translation with a fully convolutional neural network architecture with skip links. Our network is trained in a supervised fashion on a large number of face image pairs showing the same synthetic and photorealistic person, labeled with the corresponding source and target ages. The first key component of our solution is thus to derive an effective strategy for bypassing the seemingly impossible task of acquiring annotated, longitudinal image datasets depicting a variety of identities, ages, and ethnicities in different viewpoints (Section 3.1). A second key factor is in designing an appropriate parameter space for our solutions, to allow for identity preservation over changing expressions and viewpoints, with good consistency across video frames, as detailed in Section 3.2.

3.1 Synthesizing high-quality, longitudinal aging data

To train our re-aging network in a simple and fully supervised fashion, our goal here is to generate a large number of input-output image pairs, where the images in each pair depict the same (arbitrary) identity, with the same facial expression, pose, lighting, and background, but at two different and known ages. Clearly, this task is impossible to accomplish if the dataset must contain real people. We thus seek to achieve this goal using photorealisc synthetic faces, taking inspiration in recent work that leverages semantic manipulations within the latent space of powerful neural face models pre-trained on thousands of real faces [Abdal et al. 2021; Alaluf et al. 2021; Härkönen et al. 2020; Shen et al. 2020]. As we demonstrate in Section 4, manipulations within these models often perform poorly on real faces from video since real-world faces have to be projected into the latent space, an operation that is only approximate, and leads to identity drift; even when the model is further optimized (fine tuned) to overfit newly given images [Tzaban et al. 2022], temporal artifacts are still clearly noticeable, especially across viewpoints, even with little to no re-aging.

Our first key insight is that these methods above are nevertheless powerful re-aging solutions *for synthetic faces* that are already perfectly represented within the model's latent space. And the re-aged synthetic faces do seem to capture the semantics of the aging process nearly as convincingly as real training images would. We highlight that this property makes such latent space traversal methods suitable for synthesizing a high-quality, longitudinal aging dataset on which we can train a simpler network for re-aging real faces. Given this insight, and using any of the traversal approaches above, we can consider an arbitrary point in latent space, representing a particular identity and age, and then begin traversing the latent space along a path that is steered by a pre-trained age regressor, under the combined influence of identity consistency constraints. Following this traversal both forwards and backwards in time generates a continuous age progression for the particular identity at hand, leading to a large number of image pairs for training. This process can be repeated for a virtually unlimited number of identities, which can also be sampled under different viewpoints, facial expressions, lighting conditions, and backgrounds (see supplemental video).

Among the many methods recently proposed to carry out such guided latent space traversal, here we chose to sample our training dataset using the recent method for Style-based Age Manipulation (SAM) by Alaluf *et al.* [2021]. The main reason for this choice is the method's superior ability to follow a non-linear path in latent space that alters age exclusively, with little side-effect on the other facial attributes, thus matching our goal of maximizing the quality of digital re-aging on facial skin areas. Following the strategy above, we have created a training dataset for face re-aging comprising 2000 identities, each with 14 different ages in the range of 18 to 85 years, thus providing a total of 196 training pairs per sampled identity (including same age pairs). This solves the first main challenge, which is to acquire high-quality training data that would be impossible to capture from real people. Since data is extremely important in deep learning, this achievement goes a long way into solving our re-aging problem, by enabling a simple solution as described next.

3.2 High-quality Face Re-Aging Network (FRAN)

We now turn back to our main goal of identity-preserving, high-resolution face re-aging. This section presents our fully convolutional, controllable image-to-image translation solution that is supervised by the synthetic dataset described in Section 3.1. We refer to our network as the face re-aging network, FRAN.

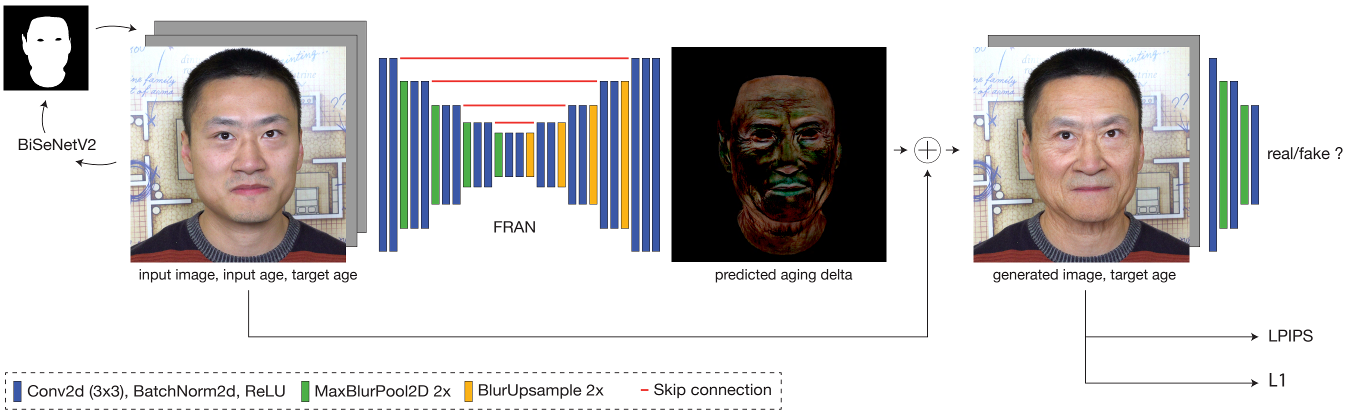

Fig. 3. The U-Net architecture of the proposed Face Re-Aging Network (FRAN) takes as input a 5-channel tensor with the RGB image to be re-aged and two additional channels indicating the current and target age of each pixel. Optionally, a pre-trained face segmentation network (BiSeNetV2 [Yu et al. 2021]) can be used to limit re-aging to skin areas and to set localized input and output age values. We use *blur-pooling* [Zhang 2019] in both FRAN and the discriminator to accommodate small shifts in the positions of wrinkles and other high frequency details.

Given the image-to-image translation nature of our task, we adopt the familiar and proven U-Net architecture design for FRAN and make small adjustments for translation quality and re-aging control. FRAN takes as input a 5-channel tensor comprising the RGB image to be re-aged and two single-channel age maps that specify the input and output age for each image pixel (*e.g.*, two uniform channels with two different ages). The U-Net then predicts per-pixel RGB deltas (offsets) that are added on top of the input image to create the final re-aged result (Fig. 3). FRAN is trained using paired, synthetic data with L1, perceptual, and adversarial losses as detailed below.

The two spatial age maps that are provided as input to FRAN are single channel images at the same spatial resolution as the input RGB image. The pixel values in these age maps are normalized between 0 and 1, representing a continuous age interval (years/100). By providing not only the target age, but also the input age, we allow FRAN to focus on the re-aging task (predicting aging deltas), rather than spend its capacity to try and estimate the current age of the input, which is already done well by existing, pre-trained age regressors. Furthermore, these input age channels do not necessarily need to be spatially constant and, thus, can be filled with non-homogeneous values to control different amounts of re-aging on different areas of the face to provide spatially-varying control on re-aging (*e.g.*, see Fig. 14). Even the input age map can be edited non-homogeneously to create different re-aging effects (by altering the subjectively perceived input age), thus providing more creative freedom to artists who wish to direct and fine tune the re-aging result (see supplemental material). Using these age maps to control FRAN also makes it easy to integrate pre-trained face segmentation networks (*e.g.* BiSeNetV2 [Yu et al. 2021]) that can automate control over the produced effect, limiting it to specific areas of the face.

Our formulation of re-aging as an image-to-image translation task has several benefits. The first of these is identity preservation which is a result of a U-Net architecture that is well known for preserving the spatial layout of the input; this is due to the U-Net's skip links providing the output layers with direct access to input image features at high-resolution. Additionally, the network does not need to learn how to fully generate faces of different identities, under different expressions, viewpoints, and lighting; it only needs to predict re-aging output as RGB offsets on top of the input image, which also prevents substantial loss on the input identity. And the temporal smoothness over the input video frames naturally contributes for the good temporal consistency in FRAN's output. Combined, these factors make FRAN an excellent, production-ready solution for re-aging real faces on video.

Discriminator. The discriminator (top-right of Fig. 3) provides additional adversarial supervision to our re-aging network. We use a modified implementation of the PatchGAN discriminator [Isola et al. 2017], which takes as input the re-aged RGB image and the input target age map. The task of the discriminator is to judge whether or not the generated re-aged appearance looks consistent with the target age, given our training dataset. The discriminator is trained alongside the main re-aging U-Net; samples from our synthetic dataset with the correct age label are provided as 'real' examples and those generated by our re-aging network are provided as 'fake' examples. We also provide real images with incorrect age maps as additional 'fake' examples.

Losses. Let I and O denote an input-output image pair in our training dataset, with known age labels a_i and a_o, and let $\tilde{O}$ denote the re-aging output of our network. We train our network using a combination of L1, perceptual, and adversarial losses,

$$\mathcal{L} = \lambda_{L1}\mathcal{L}_{L1}(\tilde{O},O)+\lambda_{P}\mathcal{L}_{LPIPS}(\tilde{O},O)+\lambda_{adv}\mathcal{L}_{adv}(\tilde{O},a_o),$$

using the VGG [Simonyan and Zisserman 2015] variant of the popular LPIPS perceptual loss [Zhang et al. 2018]. The effect of each individual loss is analyzed in the supplemental material, along with more detailed information about our network architecture and training procedure.

4 RESULTS AND EVALUATION

We now present results of our face re-aging network (Section 4.1), with qualitative and quantitative comparisons to previous approaches

Fig. 4. FRAN consistently re-ages input photos realistically, while maintaining the target identity. The age progression is temporally continuous and smooth.

(Section 4.2 and Section 4.3), including a user study evaluation (Section 4.4). Finally, we demonstrate interesting artistic age manipulations enabled by the control mechanisms in FRAN (Section 4.5).

4.1 Face Re-Aging

We start by re-aging several still photos in Fig. 4. For each input photo with the estimated original age (left), we ask the network to individually output images at varying ages from 20 to 80 years old, in 10 year steps. Not only are our results smooth and continuous in age, but they also very closely match the identity of the input image. Furthermore, our method works well in both *increasing* and *decreasing* the age realistically.

Next, we demonstrate the temporal stability of aging video sequences in Fig. 5, which shows the aging of two people in two different videos (3 frames per video are shown). Another example is shown in Fig. 1, with several more in the supplemental video. Our method can robustly handle varying head pose and extreme light conditions, and produces temporally consistent re-aging results. We further evaluate re-aging consistency in Fig. 6 by applying our method to controlled variations in expression, head pose, and illumination. In all cases, the face is successfully re-aged with consistency.

We wish to point out that our method works well even in the presence of typical motion effects such as blur, as shown in Fig. 7. This benefit allows us to readily apply our network on video frames without the need for pre- or post-processing steps, such as pre-sharpening or post-applying blur filters to match the original footage.

Finally, we compare our re-aging results to real world imagery in Fig. 8. It goes without saying that it is not possible to obtain pixel-aligned real-world images of the same person taken several years apart, which could be used as a reference result in the absence

Fig. 5. When applied on video frames, FRAN produces consistent re-aging results, and can seamlessly deal with varying depth and position of the head in the frame, plus different head poses and changes in lighting conditions. Please refer to complete result in the supplemental video.

Fig. 6. Results of FRAN under extreme conditions, with controlled variations in expression (row 1), head pose (row 2) and illumination (row 3). The first column shows one original image used as input.

Fig. 7. When re-aging video frames, FRAN provides realistic results even in the presence of typical video effects such as motion blur.

of real ground truth. However, reference photography of a person across different ages with similar expression, head pose, and lighting allows us to qualitatively evaluate our network's performance. Fig. 8 shows such a reference, of a subject taken in 2007 and again in 2022. Our FRAN result, de-aging the 2022 face to the time period of 2007 shows that our method achieves very plausible re-aged faces.

4.2 Qualitative comparison to Previous Work

We now compare the results of FRAN against those by three other recent methods representing the state of the art in facial image re-aging: LATS [Or-El et al. 2020], HRFAE [Yao et al. 2021], and SAM [Alaluf et al. 2021]. Figure 9 shows the results of all these methods side by side, for two different subjects illustrating outdoors and studio-like conditions, respectively. The column corresponding to age 35 clearly shows that LATS and SAM introduce a large identity loss even with little to no re-aging (even the background is not

Present Day (Age 46) FRAN Re-Aged (Age 31) Reference 2007 (Age 31)

Fig. 8. Starting with a present-day image (left), we compare FRAN de-aged results (middle) to a real-world reference image (right) from 2007.

preserved in the result). In contrast, our new method better preserves the input identity and the characteristic skin detail of the particular subject, followed by HRFAE whose results show some attenuation of skin detail. The figure also demonstrates that our method and SAM do not model the characteristic graying of the hair, which is slightly apparent in the results of HRFAE and LATS. However, note that hairstyle is not a focus of our assumed scenario for digital re-aging (see discussion in Section 5). Overall, our method compares favorably, with the additional advantage of demonstrated stable results when re-aging video images. We provide a more in depth comparison with previous work and a quantitative evaluation of our method in the supplemental material.

4.3 Quantitative Evaluation

An important contribution of this work is in preserving facial identity while re-aging, as already demonstrated qualitatively by the results in Fig. 9 and the additional results in the supplemental material. To provide a more objective evaluation of facial identity preservation, Fig. 10 presents a quantitative comparison of FRAN against the state-of-the-art re-aging method SAM [Alaluf et al. 2021] that was used to synthesize our training dataset and HRFAE [Yao et al. 2021]. In this evaluation, we used a pre-trained face recognition network [Schroff et al. 2015] to compute an identity loss (distance) between the original image and the re-aged output of each method. This was done for 100 different test subjects across a number of age groups (14 for FRAN and SAM and 9 for HRFAE). We picked randomly generated StyleGAN2 subjects as input images for all the methods (examples are shown in the supplemental material). As shown in Fig. 10 (left), the results of FRAN consistently remain closer to the identity of the original image for all age groups. This result quantitatively demonstrates that our method is able to preserve the identity of the original image better than state-of-the-art re-aging methods. We additionally ran the same experiment with LATS [Or-El et al. 2020] and DLFS [He et al. 2021], which only allow selecting a target age class instead of a continuous target age. For these two methods, we compute the metric by masking out the constant gray background from the output re-aged image. We show the mean identity loss across all subjects and ages for all methods in Table 1. As discussed in Section 3.2, our choice of architecture and training strategy helps preserve the input identity better than the other methods. Larger identity distances between the input and

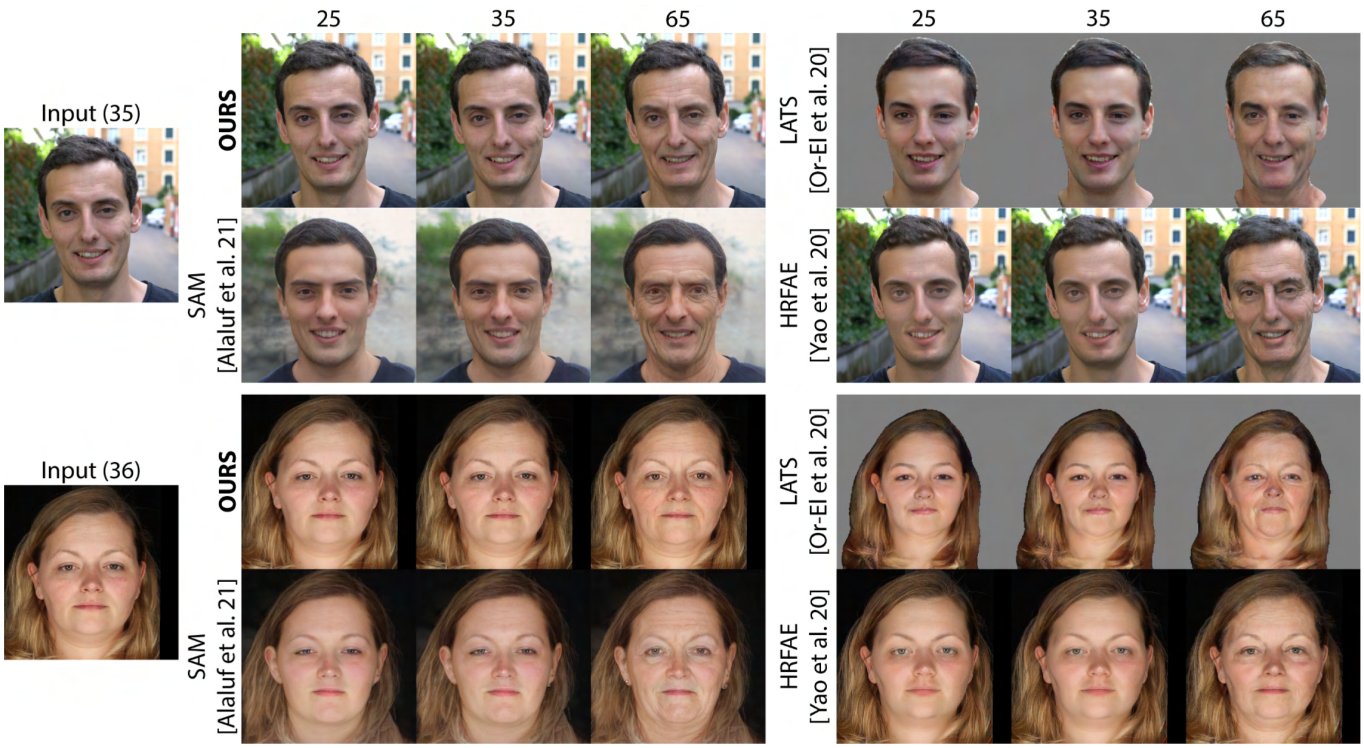

Fig. 9. Side-by-side comparison between the results of FRAN (OURS) and those by other recent methods, representing the state of the art in facial image re-aging, on two different subjects under outdoors and studio-like conditions, respectively. The column corresponding to age 35 is closest to the actual input ages and demonstrates that our method better preserves the input identity and the characteristic skin detail of the particular subject, followed by HRFAE whose results show some attenuation of skin detail. SAM and LATS largely alter the original facial identity (and even the background). Please refer to the supplemental material for additional comparisons.

Table 1. Average identity loss (distance) between original and re-aged images, as given by a pre-trained face recognition network (lower is better).

Method	Mean Identity Score ↓
LATS [Or-El et al. 2020]	0.856
DLFS [He et al. 2021]	0.703
HRFAE [Yao et al. 2021]	0.646
SAM [Alaluf et al. 2021]	0.839
FRAN (ours)	**0.616**

the result re-aged by FRAN are observed for very old target ages, Fig. 10 (left). For the sake of illustration, the result with the largest identity distance in this test is shown in Fig. 10 (right).

Note that a perfect score at maintaining the identity could be easily achieved by a method that consistently produced no re-aging at all. Thus, we must also quantitatively evaluate how well each method is achieving the desired output age. To this end, we computed an estimate of the facial age by feeding the output of each method though a pre-trained age prediction network [Rothe et al. 2018]. The results of this evaluation are plotted in Fig. 11, as an average (percentage) error in achieving the desired target age. While for older ages (50+) FRAN is under-aging the results slightly (in comparison to SAM), FRAN also consistently reaches the target age better for younger ages. For all methods, we do expect re-aging performance to drop at some very old age. For FRAN, in particular, we aim at balancing several performance metrics; the small degree of under-aging is a very small trade-off for the better performance in terms of identity preservation and stability on video, which make all the difference in achieving a practical, production-ready solution for visual effects workflows. Finally, in an artist-driven production environment, another essential factor is the level of control and freedom to achieve the envisioned *look* for the face, rather than a simple absolute age value (see artistic intensified re-aging in Section 4.5).

As FRAN is trained exclusively using synthetic, photorealistic data (Section 3.1), it is also important to evaluate the performance of FRAN on real data. We thus computed the same identity and age metrics as above for (*i*) a test set of 50 real photographs randomly extracted from FFHQ and (*ii*) a set of 50 random synthetic images generated using StyleGAN2. We used the pre-trained age prediction network [Rothe et al. 2018] to make sure all images contained subjects exclusively in the range of 18 to 85 years old. We then ran FRAN on all images and computed both the mean identity distance and average age error. The results are reported in Table 2, and show that FRAN operates nearly as well on real images as synthetic ones.

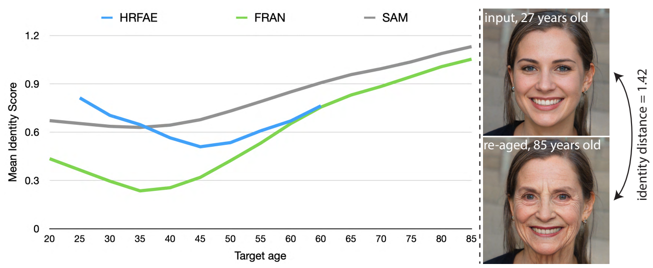

Fig. 10. Identity-loss between the original image and the re-aged outputs produced by our method (FRAN), SAM, and HRFAE, using latent features of a pre-trained face recognition network [Schroff et al. 2015]: (left) identity loss averaged over 100 subjects at 14 age targets, lower is better; (right) FRAN result presenting the largest identity distance. On average, FRAN presents lower identity loss than the other methods at all 14 target ages.

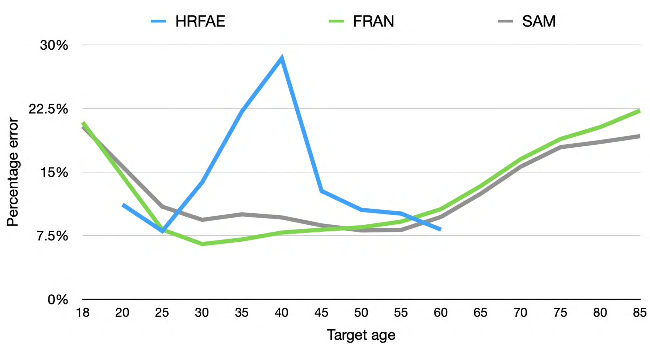

Fig. 11. We used a pre-trained age prediction network [Rothe et al. 2018] to compute an estimate of the age of the output of each methods.

Table 2. We compute the average identity distance between the original and re-aged images, as given by a pre-trained face recognition network [Schroff et al. 2015] and the average age error, as given by a pre-trained age prediction network [Rothe et al. 2018] for both a test dataset of synthetically generated faces (StyleGAN2) and a test dataset of real faces (FFHQ).

Dataset	Mean Identity Score ↓	Age Percentage Error ↓
Synthetic	0.64	13%
Real	0.40	21%

4.4 User Study

As an additional comparison, we ran a user study with 32 anonymous participants. The study included a total of 12 input images: 9 images from FFHQ, one studio-lit portrait, and 2 "in-the-wild" images extracted from a video. We performed the FFHQ alignment on all 12 images, picked a random target age (between 18 and 65 years of age), before running our method (FRAN), HRFAE [Yao et al. 2021], DLFS [He et al. 2021], LATS [Or-El et al. 2020], and SAM [Alaluf et al. 2021] on all 12 images. We chose a common upper limit of 65 years because only FRAN and SAM can re-age beyond that. Also, since LATS and DLFS cannot do continuous re-aging, we selected their output age group closest to the desired target age.

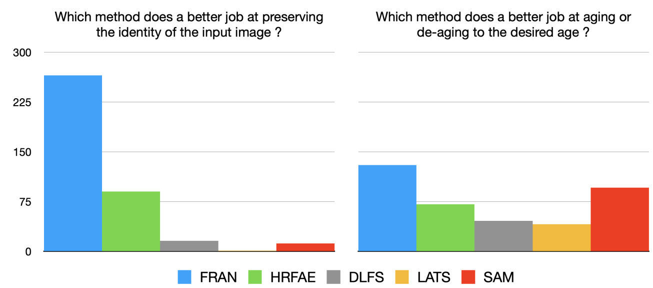

Fig. 12. User study comparison of our method (FRAN) against HRFAE, DLFS, LATS, and SAM, with 32 participants and a total of 384 answers per question.

The corresponding outputs of the methods were grouped in random order and each anonymous participant was asked: (1) *Which method does a better job at preserving the identity of the input image?* and (2) *Which method does a better job at aging or de-aging to the desired age?* Participants were instructed to ignore the output image background and only take into consideration the re-aging effect on the face. The result comprises a total of 384 answers for each question, as shown in Fig. 12. The clear majority of the participants found that our method FRAN does a better job both at preserving the input facial indentity and at reaching the target facial age.

4.5 Artistic Control

Artistic control is often neglected in research work, but is a fundamental feature in real production pipelines (which often include super-human characters and fantastic events). FRAN provides intuitive and localized artistic control over the re-aging result. Since our primary mode of control is the desired output age, and we process video frames independently, this allows us to artistically vary the output age over time in a video sequence. For example, in Fig. 13 we continuously age an actor while they perform on camera, starting at a young age and ending at an old age (please refer to the full result in the supplemental video). Note how the aging is continuous, smooth, and stable over time.

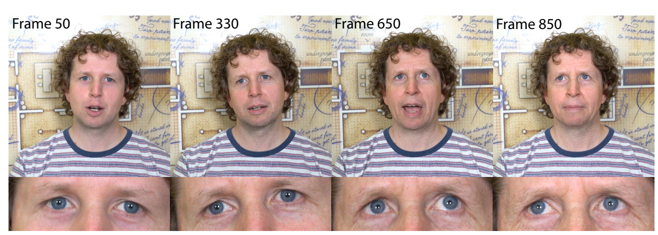

Fig. 13. Example of applying FRAN to artistically and progressively increase the age over time in a video, during a performance (please refer to full performance in the supplementary video).

We can further decompose the effect of aging beyond the temporal domain and into the spatial domain, leveraging the target age map that is input to our network. As hinted at earlier, this age map does not need to contain a homogeneous value but can be filled with different ages to control different amounts of re-aging on different areas of the face. An example is shown in Fig. 14, where a painted

Fig. 14. Painting a target age map allows for re-aging different parts of the face with different age targets. Here we use a mask for the target age map and re-age that image area to ages 18 and 85, while keeping the other pixels fixed to the input age.

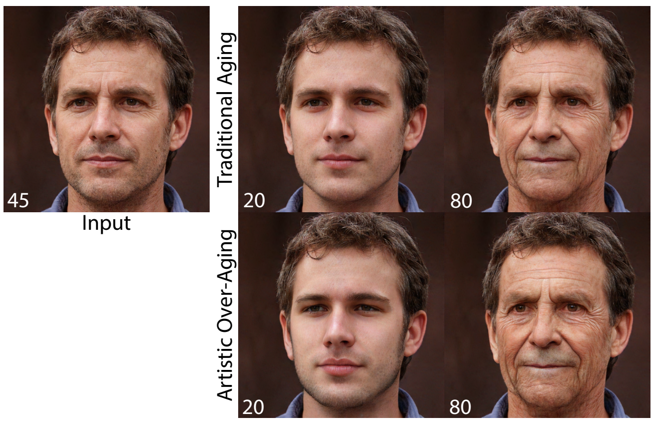

Fig. 15. Artistically intensified re-aging by modifying the *input age* so that it does not correspond to the actual apparent age in the input image.

mask is used to re-age only parts of the input image around the eyes. We show re-aging the eye region to be either younger or older.

Finally, FRAN also takes a 2D age map indicating the *input age*. While we train the network using valid age maps that correspond to the detected age of the input image, at test time there is no restriction that the input age map matches the input image, yielding interesting artistic effects when combining old input age maps with young input images or vice versa. We experiment with these artistic effects in Fig. 15. It begins with traditional re-aging, where the input age map was uniformly set to 45 for the 45 year-old man (top row) and we illustrate re-aged images at 20 and 80 years. On the second row, we show the corresponding output given the same input image but with an input age map altered to 80 (second column) and 20 (third column). The effect is *artistically intensified re-aging*, where the new 20 year-old looks younger than in the original young result, and the new 80 year-old looks even older than in the original result.

4.6 Extent of Re-aging

As discussed in Section 1, FRAN focuses only on adults and does not aim to introduce changes in scalp hair or large changes on the shape of the head, such as when re-aging to very young ages (children). While textural changes (*e.g.*, wrinkles) are easier to notice in FRAN's results, FRAN does alter other facial features and introduces geometric changes that are also important for realistic re-aging. Some of these changes are indicated in Fig. 16 by per-pixel differences and

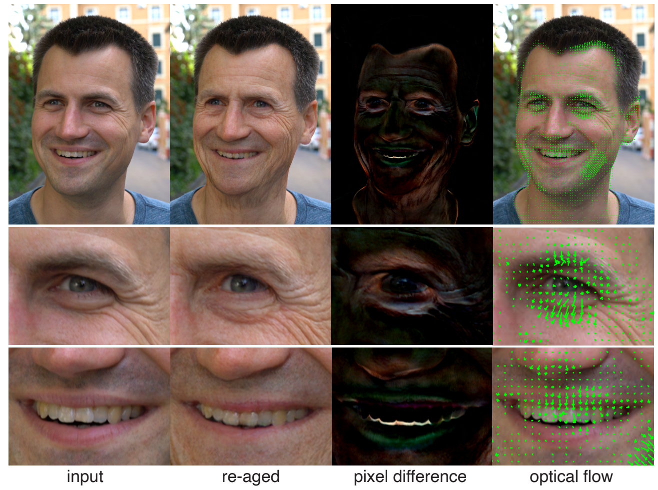

Fig. 16. We demonstrate how FRAN changes the input image by visualizing the per-pixel difference between the input and output (predicted re-aging deltas) and by visualizing the optical flow between the input and output. We highlight the eye and mouth areas (best viewed by zooming in).

an optical flow visualization, both computed between the input and the re-aged result. We highlight the changes introduced in the eye region (such as wrinkling, sagging of the eyebrow and eye bag), in the mouth region (such as lips becoming thinner), and in the altered size of cartilage tissue (on ear lobes and nose).

4.7 Segmentation at Inference

We also evaluate the effect of using a segmentation mask to guide the aging only on the skin pixels of the face, with results shown in Fig. 17. Here, we first set the input age to 35 and target age to 65 uniformly for all pixels of the input image (i.e. the result "without" the mask) and show the difference between the input and output, aged, images. Because the target age is set to be different than the input age also for the background pixels, the background contains some undesired changes. We repeat the experiment setting the target age to 65 only on the skin pixels defined by the segmentation mask automatically computed using [Yu et al. 2021] and set the target age to 35 (same as input age) for all pixels outside this mask (the result labeled "with" the mask). The corresponding output image does not have the unwanted changes in the background, as show by the difference image.

4.8 Losses Ablation

As an ablation study, we trained multiple variants of FRAN, each time enabling only a subset of the losses. More specifically, we used a fixed budget of 20 epochs and trained FRAN with 100 subjects and all 14 age classes, maintaining all the hyperparameters fixed except for the losses. We demonstrate the effect of our loss terms both individually, and by also incrementally adding them in the order of their computational complexity and show the result of this ablation in Fig. 18. With the L1 loss alone, the results lack sharpness due to the loss being per-pixel only. The adversarial loss alone is not sufficient for FRAN to produce aging effects with a fixed budget of

Fig. 17. We show the effect of applying a segmentation mask [Yu et al. 2021] vs. not applying it when aging an input image from 25 years old to 75 years old. Without using the mask (center), the target age is set to be different than the input age also for the background pixels, and thus the background contains some undesired changes. When using the mask (right), the target age in the background region is set to 25 (same as input age), which prevents changes in the background, as seen in the difference images.

20 epochs. Training only with LPIPS gives sharper results but adding L1 gives a stronger aging effect. Finally, training with L1, LPIPS and the adversarial loss increases the aging effect and sharpness.

5 CONCLUSION

This paper presents our practical, production-ready face re-aging network, FRAN. Inspired by recent advances in facial re-aging within the latent space of face GANs, we propose FRAN as a controllable, temporally-stable, identity-preserving alternative to latent space traversal methods. FRAN reformulates re-aging as a simple image-to-image translation task that is naturally and effectively solved using the familiar U-Net architecture. As with most deep learning approaches, a large amount of high-quality training data is key in enabling FRAN. Here, we overcome the impossible task of collecting longitudinal aging dataset from real people by leveraging the current state of the art in digital face re-aging. Our key insight is that, while these methods fail on real faces, they provide highly realistic solutions on *fully synthetic faces* that are already represented perfectly within the latent space of powerful GANs.

FRAN is trained in a simple supervised fashion by leveraging a large longitudinal dataset of photo-realistically re-aged, synthetic face pairs. As a result, FRAN provides realistic and continuous re-aging within a range of 18 and 85 years of age. As far as we know, FRAN is the first method to provide high-resolution, temporally stable re-aging results on videos showing faces in different expressions and under free viewpoint, depth, motion, and illumination conditions. Finally, its design provides artists with intuitive, localized control for directing and fine tuning the resulting re-aging effect. FRAN is a valuable, production-ready tool in different application domains such as entertainment and advertising. Given that artistic intervention is always desirable, no result is ever perfect on the first try. Thus, FRAN has the potential to improve existing re-aging workflows, reducing the time it takes to re-age complete shots from a matter of days to just a few hours or even minutes, facilitating the creation of high-quality visual effects at scale.

Ethical Impact. FRAN can synthetically modify images and videos of human faces to make them look younger or older. It is designed with entertainment applications in mind, but it is important to acknowledge potential for misuse. While techniques for detecting altered images do exist, it is also important to broadly educate people about the dangers of image and video modification algorithms. By publishing the details of our algorithm we hope to support the efforts around automatic deep fakes detection.

Limitations. FRAN is, of course, not without its limitations. As with typical U-Net architectures, large image changes are more difficult to generate, making it challenging to re-age to and from very young ages, which is not included in our application scenario. For such cases (*e.g.*, important applications involving missing children, in which stability on video is not a focus), other methods such as LATS [Or-El et al. 2020] and DLFS [He et al. 2021] may be preferable. Another limitation is the fact that the graying of scalp hair is not captured in our current training data, and is therefore not reflected in the outputs of FRAN. This is an additional aging effect that could be included into our training dataset; we leave it for future work as current production pipelines already have a suitable alternative for controlling the re-aged hairstyle in a traditional way. Similarly, re-aging can also introduce variation in body mass index (BMI) whose effects on the face we currently cannot control. In addition, FRAN does not yet provide artists with a means for specifying what structures are actually added to or removed from the face during re-aging (*e.g.*, moles, specific wrinkling patterns, or other particular skin signs). Finally, FRAN currently only improves upon the predominant 2D re-aging workflow, whereas in particular cases a 3D re-aging solution may be preferable for more elaborate levels of control including, for instance, relighting and other physically based manipulations. We believe that these limitations also represent exciting opportunities for additional improvements in future work.

ACKNOWLEDGMENTS

We thank all of our test subjects who let us re-age their faces, as well as Naruniec *et al.* [2020] for sharing additional test videos.

REFERENCES

Rameen Abdal, Peihao Zhu, Niloy J. Mitra, and Peter Wonka. 2021. StyleFlow: Attribute-Conditioned Exploration of StyleGAN-Generated Images Using Conditional Continuous Normalizing Flows. *ACM TOG* 40, 3, Article 21 (2021).

Adobe. 2022. Photoshop. https://www.adobe.com/products/photoshop.html. [Online; accessed 20-May-2022].

Yuval Alaluf, Or Patashnik, and Daniel Cohen-Or. 2021. Only a Matter of Style: Age Transformation Using a Style-Based Regression Model. *ACM TOG* 40, 4, Article 45 (2021).

Raphael Angulu, Jules R. Tapamo, and Aderemi O. Adewumi. 2018. Age estimation via face images: a survey. *EURASIP J. Image and Video Proc.* 1, 42 (2018).

Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. 2017. Face aging with conditional generative adversarial networks. In *ICIP*. 2089–2093.

Netflix Film Club. 2020. How The Irishman's Groundbreaking VFX Took Anti-Aging To the Next Level | Netflix. Retrieved May 13, 2022 from https://www.youtube.com/watch?v=OF-lElIlZM0

Julien Despois, Frederic Flament, and Matthieu Perrot. 2020. AgingMapGAN (AMGAN): High-Resolution Controllable Face Aging with Spatially-Aware Conditional GANs. In *ECCV*.

Fig. 18. We age an input image (18 years old) from the test set using multiple variations of FRAN trained with a subset of the losses, trying to hit the target image (85 years old). We show the effect of L1, Adversarial and LPIPS losses alone and their combination.

Malaria Must Die. 2020. A World Without Malaria. Retrieved May 13, 2022 from www.youtube.com/watch?v=0l4eTfpIsKw

C N Duong, K Luu, K G Quach, N Nguyen, E Patterson, T D Bui, and N Le. 2019. Automatic Face Aging in Videos via Deep Reinforcement Learning. In *IEEE CVPR*. 10005–10014.

H Fang, W Deng, Y Zhong, and J Hu. 2020. Triple-GAN: Progressive Face Aging with Triple Translation Loss. In *CVPR Workshops*. 3500–3509.

Joe Fordham. 2019. The Power Within Her. *Cinefex* 164 (2019), 37–64.

Rinon Gal, Dana Cohen, Amit Bermano, and Daniel Cohen-Or. 2021. SWAGAN: A Style-based Wavelet-driven Generative Model. *ACM TOG* 40, 4, Article 134 (2021).

Markos Georgopoulos, Yannis Panagakis, and Maja Pantic. 2018. Modeling of facial aging and kinship: A survey. *Image and Vis. Comp.* 80 (Dec. 2018), 58–79.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *NeurIPS* 27 (2014), 2672–2680.

Sen He, Wentong Liao, Michael Ying Yang, Yi-Zhe Song, Bodo Rosenhahn, and Tao Xiang. 2021. Disentangled Lifespan Face Synthesis. In *IEEE ICCV*.

Gee-Sern Hsu, Rui-Cang Xie, and Zhi-Ting Chen. 2021. Wasserstein Divergence GAN With Cross-Age Identity Expert and Attribute Retainer for Facial Age Transformation. *IEEE Access* 9 (2021), 39695–39706.

Zhizhong Huang, Junping Zhang, and Hongming Shan. 2021. When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework. In *IEEE CVPR*.

Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. 2020. GANSpace: Discovering Interpretable GAN Controls. In *NeurIPS*.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *IEEE CVPR*.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020a. Training Generative Adversarial Networks with Limited Data. In *NeurIPS*.

Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *IEEE CVPR*. 4401–4410.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020b. Analyzing and improving the image quality of StyleGAN. In *IEEE CVPR*. 8110–8119.

Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M. Seitz. 2014. Illumination-Aware Age Progression. In *IEEE CVPR*. 3334–3341.

Zeqi Li, Ruowei Jiang, and Parham Aarabi. 2021. Continuous Face Aging via Self-estimated Residual Age Embedding. In *IEEE CVPR*.

FaceApp Technology Limited. 2022. FaceApp. https://www.faceapp.com. [Online; accessed 20-May-2022].

Yunfan Liu, Qi Li, Zhenan Sun, and Tieniu Tan. 2021. A3GAN: An Attribute-Aware Attentive Generative Adversarial Network for Face Aging. *IEEE TIFS* 16 (2021), 2776–2790.

Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park. 2021. Re-Aging GAN: Toward Personalized Face Age Transformation. In *IEEE ICCV*. 3908–3917.

Jacek Naruniec, Leonhard Helminger, Christopher Schroers, and Romann Weber. 2020. High-Resolution Neural Face Swapping for Visual Effects. *Comput. Graph. Forum* 39, 4 (2020), 173 – 184.

Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. 2020. Lifespan Age Transformation Synthesis. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). 739–755.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *IEEE CVPR*.

Narayanan Ramanathan, Rama Chellappa, and Soma Biswas. 2009. Computational Methods for Modeling Facial Aging: A Survey. *J. Visual Lang. and Comp.* 20, 3 (2009), 131–144.

Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *IJCV* 126, 2-4 (2018), 144–157.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *IEEE CVPR*.

Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. 2020. Interpreting the Latent Space of GANs for Semantic Face Editing. In *IEEE CVPR*.

Xiangbo Shu, Guo-Sen Xie, Zechao Li, and Jinhui Tang. 2016. Age progression: Current technologies and applications. *Neurocomputing* 208 (2016), 249–261.

Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*.

Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. 2021. Designing an Encoder for StyleGAN Image Manipulation. 40, 4 (2021).

Rotem Tzaban, Ron Mokady, Rinon Gal, Amit H. Bermano, and Daniel Cohen-Or. 2022. Stitch it in Time: GAN-Based Facial Editing of Real Videos. arXiv:2201.08361 [cs.CV]

Yuri Viazovetskyi, Vladimir Ivashkin, and Evgeny Kashin. 2020. StyleGAN2 Distillation for Feed-Forward Image Manipulation. In *ECCV*. 170–186.

Hongyu Yang, Di Huang, Yunhong Wang, and Anil K. Jain. 2021. Learning Continuous Face Age Progression: A Pyramid of GANs. *IEEE TPAMI* 43, 2 (2021), 499–515.

Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. 2021. High Resolution Face Age Editing. In *ICPR*.

Changqian Yu, Changxin Gao, Jingbo Wang, Gang Yu, Chunhua Shen, and Nong Sang. 2021. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation. *IJCV* 129 (2021), 3051–3068.

Richard Zhang. 2019. Making Convolutional Networks Shift-Invariant Again.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE CVPR*.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *IEEE ICCV*.