

Production-Ready Face Re-Aging for Visual Effects - Supplemental

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

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 - Supplemental. *ACM Trans. Graph.* 41, 6, Article 237 (December 2022), 6 pages. <https://doi.org/10.1145/3550454.3555520>

This supplemental document contains more details on our implementation and additional results and comparisons with other methods.

1 NETWORK ARCHITECTURE

FRAN. We setup FRAN to use 3×3 convolutions with LeakyReLU non-linear activations. Following Pandey et al. [2021], we use blur-pooling [Zhang 2019] layers for downsampling and upsampling, which have been shown to accommodate small shifts in wrinkles and lead to sharper outputs with more spatial detail. We show the details of our down and up sampling blocks in Table 1 and Table 2 respectively. We show the details of our generator in Table 3. Note that we concatenate the output of the same level down layer with the output of the previous up layer in the generator (skip links).

Layer	Output
Input	$w \times h \times c$
MaxBlurPool	$w/2 \times h/2 \times c$
Conv 3×3	$w/2 \times h/2 \times 2c$
LeakyReLU	
Conv 3×3	$w/2 \times h/2 \times 2c$
LeakyReLU	
Output	$w/2 \times h/2 \times 2c$

Table 1. Down Layer.

Layer	Output
Input	$w \times h \times c$
BlurUpSample	$2w \times 2h \times c$
Conv 3×3	$2w \times 2h \times c/2$
LeakyReLU	
Conv 3×3	$2w \times 2h \times c/2$
LeakyReLU	
Output	$2w \times 2h \times c/2$

Table 2. Up Layer.

Layer	Output
Image	$512 \times 512 \times 3$
Conv 3×3	$512 \times 512 \times 64$
LeakyReLU	
Conv 3×3	$512 \times 512 \times 64$
LeakyReLU	
DownLayer	$256 \times 256 \times 128$
DownLayer	$128 \times 128 \times 256$
DownLayer	$64 \times 64 \times 512$
DownLayer	$32 \times 32 \times 1024$
UpLayer	$64 \times 64 \times 512$
UpLayer	$128 \times 128 \times 256$
UpLayer	$256 \times 256 \times 128$
UpLayer	$512 \times 512 \times 64$
Conv 1×1	$512 \times 512 \times 3$
Output	$512 \times 512 \times 3$

Table 3. Generator Architecture.

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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM. 0730-0301/2022/12-ART237 \$15.00

<https://doi.org/10.1145/3550454.3555520>

Discriminator. The discriminator is designed as a three-layer convolutional PatchGAN [Isola et al. 2017] classifier with LeakyReLU activations. Each element of the final layer has a receptive field of 70×70 pixels. The discriminator is trained with a binary cross entropy loss on the real/fake score.

2 TRAINING

To train FRAN and the discriminator, we use the dataset described in the main text. More specifically, we sample pairs of source and target 1024×1024 images, representing the same individual at a source and a target age (both age maps are filled in with a spatially-uniform age value during training). We randomly add color jitter and slight rotations to both source and target images and randomly

extract a crop. In our experiments, we found that using crops of 512×512 provides a good tradeoff between batchsize and context. We optimize for the parameters of FRAN and the discriminator using Adam [Kingma and Ba 2015] with a learning rate of 0.0001, $\lambda_{L1} = 1.0$, $\lambda_{LPIPS} = 1.0$, and $\lambda_{Adv} = 0.05$. We use a batch size of 8 and train on one Nvidia RTX 3090 for 2 days.

3 NUKE PLUGIN

To make FRAN really practical, we created a Nuke Plugin which allows artists to control the input parameters while editing a video and get feedback at interactive framerates. Please see our supplemental video for a screen recording of the plugin.

4 QUALITATIVE EVALUATION

In addition to the qualitative evaluation available in the main text, we also compare qualitatively against additional methods.

4.1 RAGAN

In Fig. 1 we compare against RAGAN [Makhmudkhujaev et al. 2021]. While FRAN can successfully re-age images from 18 to 85, we restrict this comparison figure to only the overlapping range between RAGAN and FRAN. This figure also shows that FRAN can successfully extrapolate to ages slightly outside the training age range.

4.2 DLFS and LATS

We show comparisons of FRAN against DLFS [He et al. 2021] and LATS [Or-El et al. 2020] in Fig. 2 and Fig. 3. We randomly sampled identities using StyleGAN2 [Karras et al. 2020] and used a pre-trained age prediction network [Rothe et al. 2018] to automatically compute the input age. While both DLFS and LATS fail to offer precise age control, we pick the closest age class to compare their results with FRAN.

4.3 SAM and HRFAE

Figures 4 and 5 show additional comparison of HRFAE, SAM and FRAN. We randomly sampled identities using StyleGAN2 [Karras et al. 2020] and used a pre-trained age prediction network [Rothe et al. 2018] to automatically compute the input age.

4.4 Additional Results

Figure 6 shows some additional results produced by our method. We randomly sampled identities using StyleGAN2 [Karras et al. 2020] and used a pre-trained age prediction network [Rothe et al. 2018] to automatically compute the input age. We then used FRAN to compute re-aged version of the input image to 7 different target ages.

REFERENCES

- Sen He, Wentong Liao, Michael Ying Yang, Yi-Zhe Song, Bodo Rosenhahn, and Tao Xiang. 2021. Disentangled Lifespan Face Synthesis. In *IEEE ICCV*.
 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, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of StyleGAN. In *IEEE CVPR*. 8110–8119.
 Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

- Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park. 2021. Re-Aging GAN: Toward Personalized Face Age Transformation. In *IEEE ICCV*. 3908–3917.
 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.
 Rohit Pandey, Sergio Orts-Escolano, Chloe LeGendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total Relighting: Learning to Relight Portraits for Background Replacement. 40, 4.
 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.
 Richard Zhang. 2019. Making Convolutional Networks Shift-Invariant Again.

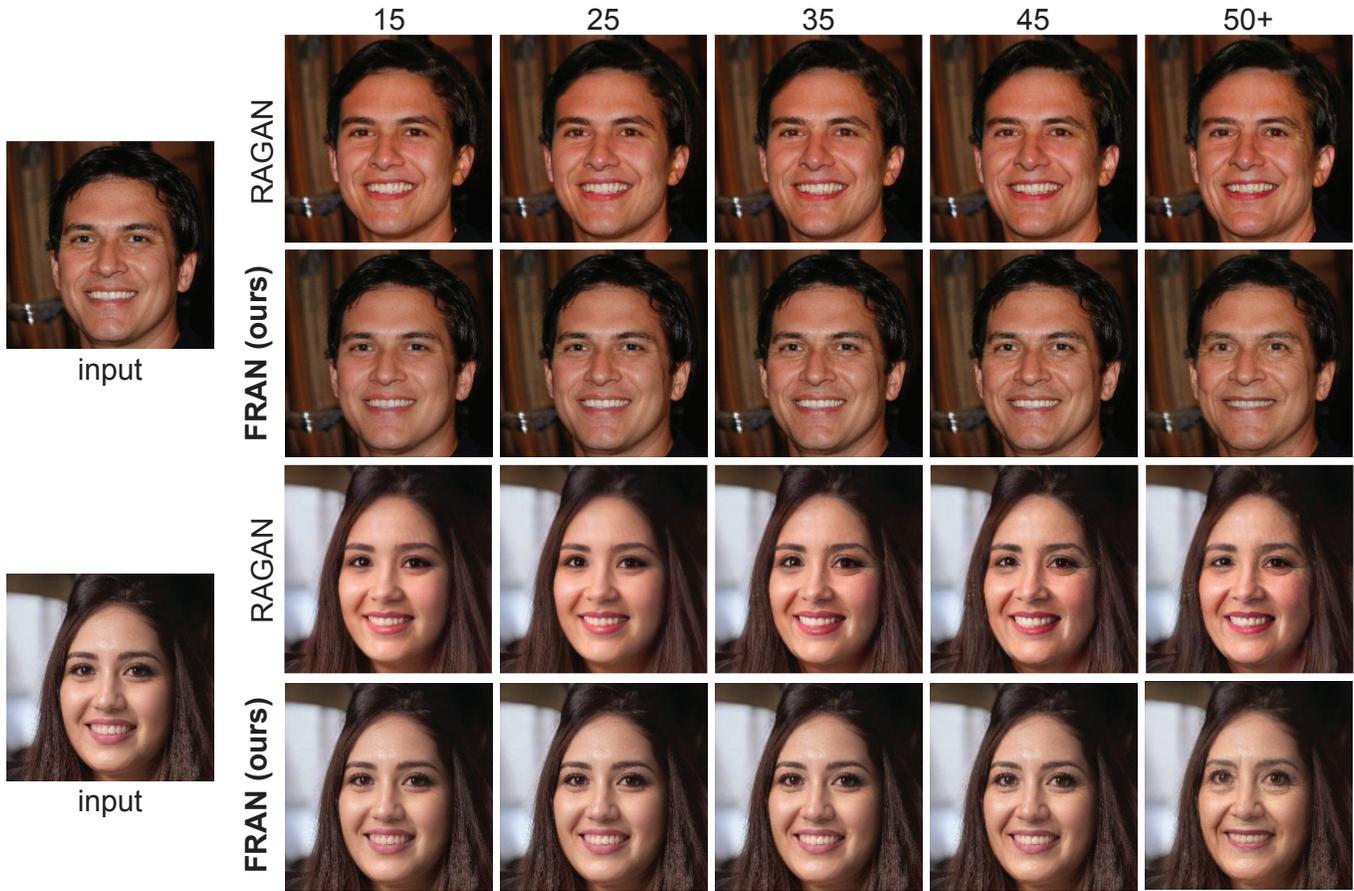


Fig. 1. Comparison between FRAN and RAGAN

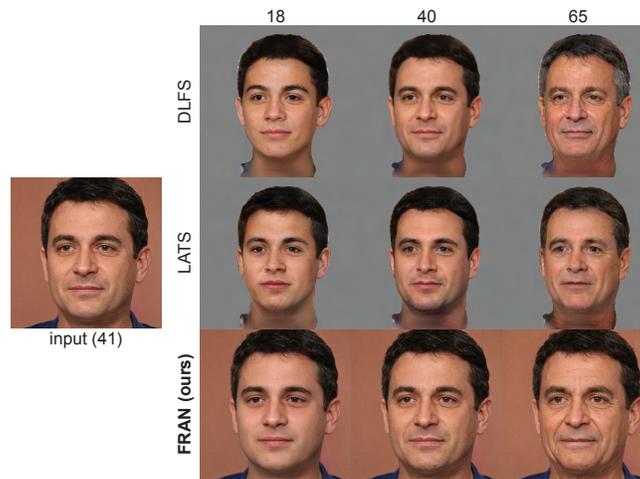


Fig. 2. Comparison between FRAN, LATS and DLFS for one subject.

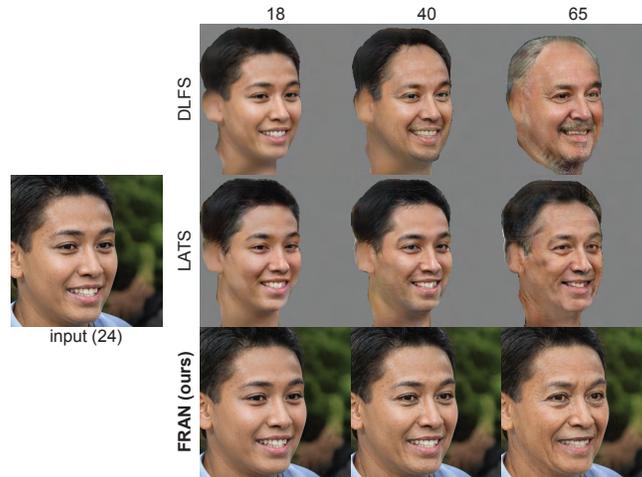


Fig. 3. Comparison between FRAN, LATS and DLFS for another subject.

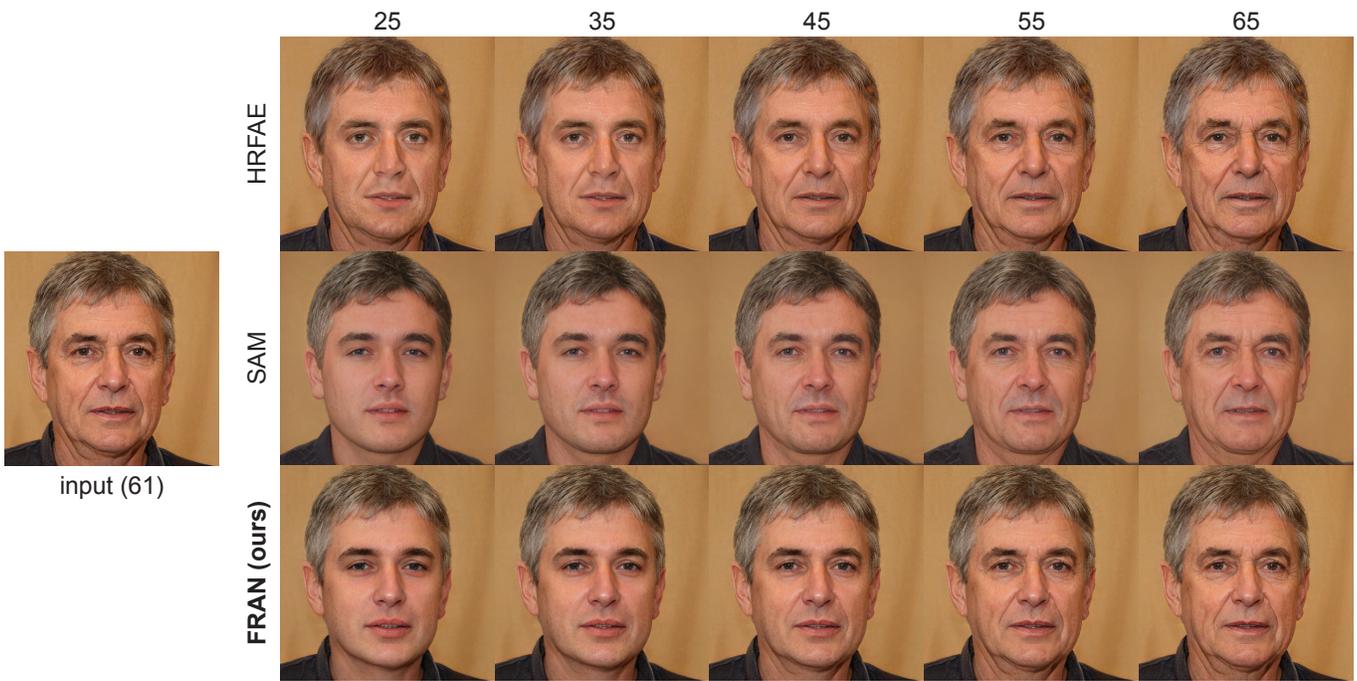


Fig. 4. Additional comparison between HRFAE, SAM and FRAN for one subject.

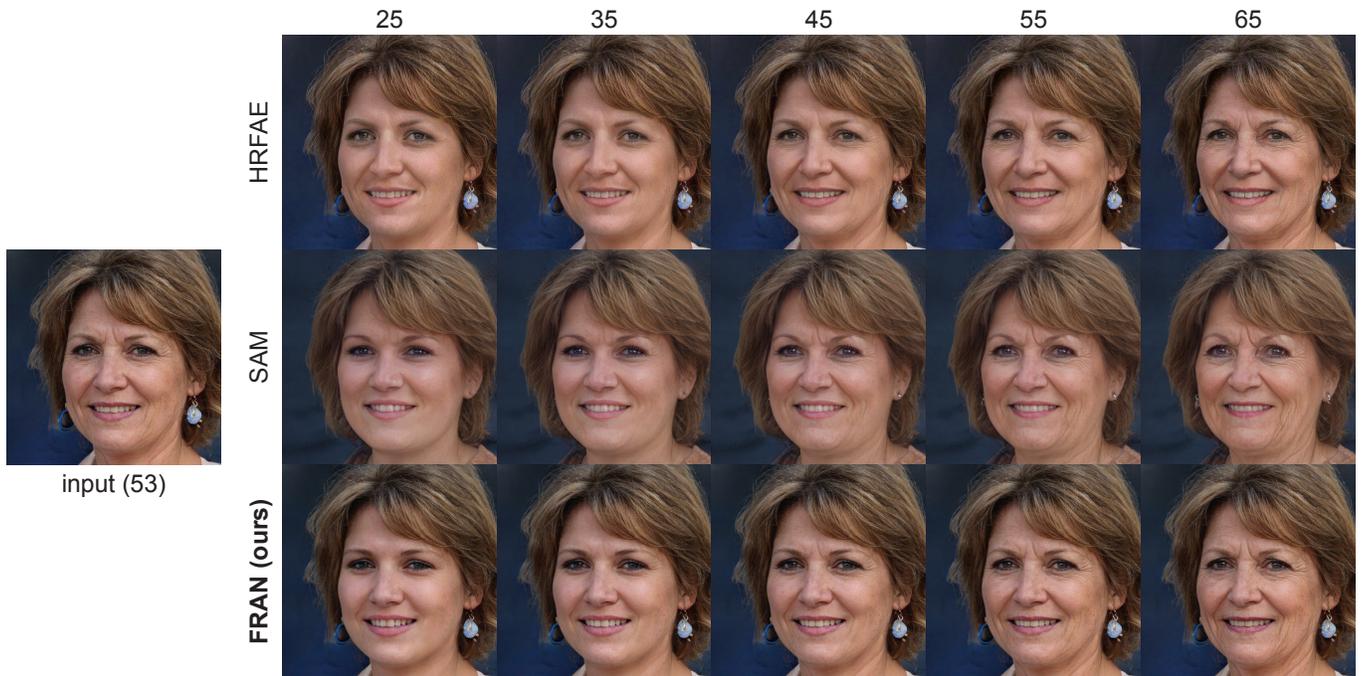


Fig. 5. Additional comparison between HRFAE, SAM and FRAN for another subject.



Fig. 6. Additional results of FRAN on SyleGAN2 generated samples.