Supplementary Material for
Re-Aging GAN: Toward Personalized Face Age Transformation

Farkhod Makhmudkhujaev, Sungeun Hong, and In Kyu Park
Dept. of Information and Communication Engineering, Inha University, Incheon 22212, Korea
{farhodfm, csehong, pik}@inha.ac.kr

In this supplementary material, we provide a detailed information on GAN architectures utilized in our work, evaluation protocols, and more results. We initially describe the architectural construction of our proposed framework. Then, we describe the evaluation protocols, followed by additional results on images of different datasets.

1. Networks Architectures

Identity Encoder (Table 1) The identity encoder starts with a 3×3 convolutional layer that transforms the input image to the feature domain. These features are then further masked out to separate the face region from the background. To obtain the mask of the input image, we use DeepLabV3 [3] pre-trained on CelebAMask-HQ [11] as done in [14]. The masking operation is followed by four downsampling blocks and two intermediate blocks, all of which in the form of residual units [5], to extract identity features f_{id}. We use instance normalization [7, 8] and Leaky-ReLU activation within these blocks. Downsampling is achieved by applying the average pooling operation.

Age Modulator (Table 2) This network adapts identity features f_{id} to the target age y'; thereby producing age-aware features (f_{aw}). To this end, we replace the batch normalization layers in the residual blocks with conditional batch normalization (CBN) layers [13, 12, 18]. Given that the target age is presented as an integer label $y' \in [0, n]$, where n indicates the upper age-bound, we use the embedding layers to obtain its scaling and shifting parameters. To adapt age, applying affine transformations after normalizing each identity feature is sufficient. At the last layer, we use a 4×4 convolutional layer that generates the final age-aware features; we also found that applying a 1×1 convolutional layer followed by adaptive average pooling also applicable to use.

Decoder (Table 3) The decoder contains two intermediate and four upsampling blocks. Similarly, all inherit pre-activation residual units [5]. We upsample the intermediate features by applying nearest-neighbor interpolation. We apply adaptive instance normalization (AdaIN) to all the blocks. Through this normalization technique, we inject age-aware features f_{aw} in order to self-guide the decoder. Similar to existing work [4], we do not set a hyperbolic tangent at the last activation layer. Instead, we force the model to learn the color-range of image by itself. Thus, the last layer is a 1×1 convolutional layer that maps the final features to the image (RGB) domain. Prior to this mapping, we use background features separated in the encoder and add it to the final features to maintain the background information and achieve better visual perception.

Discriminator (Table 4) We use discriminator [4] architecture with multiple linear output branches. At first, a 3×3 convolutional layer is applied to generate a feature representation of the input. Then, six pre-activation residual blocks with Leaky-ReLU activation downsample the feature maps. At the last layer, D fully connected layers are used to predict the validity (*i.e.*, real or fake) of each age-domain. We do not use any normalization in the discriminator.

2. Evaluation Protocol

Identity Preservation As discussed in the paper, we use two metrics, namely, Frechét inception distance (FID) [6] and Kernel inception distance (KID) [1], for estimating the capability of the model on identity preservation. Both metrics are used to measure the discrepancy between two image distributions (*i.e.* real p_R and generated p_G). Note that, in our experiments, we evaluate the model in each age-group

Layer	Resample	Norm	Output Shape
Image x	-	-	$256 \times 256 \times 3$
Conv 3×3	-	-	$256 \times 256 \times 64$
Res. Block	AvgPool	IN	$128 \times 128 \times 128$
Res. Block	AvgPool	IN	$64 \times 64 \times 256$
Res. Block	AvgPool	IN	$32 \times 32 \times 512$
Res. Block	AvgPool	IN	$16 \times 16 \times 512$
Res. Block	-	IN	$16 \times 16 \times 512$
Res. Block	-	IN	$16 \times 16 \times 512$

Table 1. Identity encoder architecture

Layer	Resample	Norm	Output Shape
Identity f_{id}	-	-	$16 \times 16 \times 512$
Target age y'	-	-	$y' \in [0, n]$
Res. Block	AvgPool	CBN	$8 \times 8 \times 512$
Res. Block	AvgPool	CBN	$4 \times 4 \times 512$
Conv 4×4	-	-	$1 \times 1 \times 512$

Table 2. Age modulator

Layer	Resample	Norm	Output Shape
Identity f_{id}	-	-	$16 \times 16 \times 512$
Age-aware f_{aw}	-	-	$1 \times 1 \times 512$
Res. Block	-	AdaIN	$16 \times 16 \times 512$
Res. Block	-	AdaIN	$16 \times 16 \times 512$
Res. Block	Upsample	AdaIN	$32 \times 32 \times 512$
Res. Block	Upsample	AdaIN	$64 \times 64 \times 256$
Res. Block	Upsample	AdaIN	$128 \times 128 \times 128$
Res. Block	Upsample	AdaIN	$256 \times 256 \times 64$
Conv 1×1	-	-	$256 \times 256 \times 3$

Table 3. Decoder architecture

Layer	Resample	Norm	Output size
Image x	-	-	$256 \times 256 \times 3$
Conv 3×3	-	-	$256 \times 256 \times 64$
Res. Block	AvgPool		$128 \times 128 \times 128$
Res. Block	AvgPool	-	$64 \times 64 \times 256$
Res. Block	AvgPool	-	$32 \times 32 \times 512$
Res. Block	AvgPool	-	$16 \times 16 \times 512$
Res. Block	AvgPool	-	$8 \times 8 \times 512$
Res. Block	AvgPool	-	$4 \times 4 \times 512$
LReLU	-	-	$4 \times 4 \times 512$
Conv 4×4	-	-	$1 \times 1 \times 512$
LReLU	-	-	$1 \times 1 \times 512$
Reshape	-	-	512
Linear$*D$	-	-	$1 * D$

Table 4. Discriminator architecture

separately which is similar to calculating intra-FID scores.

FID [6] relies on a pretrained Inception-V3 [17] model that transforms each image of given distributions to the vector space. Afterward, FID measures similarities between two vector space by $\text{FID}(p_R, p_G) = \|\mu_R - \mu_G\|_2^2 + \text{Tr}\left(\Sigma_R + \Sigma_G - 2\left(\Sigma_R \Sigma_G\right)^{\frac{1}{2}}\right)$, where μ and Σ are the empirical mean and covariance, respectively.

Similar to the FID, KID [1] is also operated on the feature-space of p_R and p_G extracted using the Inception model. However, KID computes squared maximum mean discrepancy (MMD) between features by means of polynomial kernel function $k(x, y) = \left(\frac{1}{d}x^T y + 1\right)^3$, where d is feature dimension. It has been argued [1] that KID is an unbiased metric than FID. Nevertheless, as FID became a standard metric on GANs, we report scores for both metrics in our paper.

FFHQ - Original: https://github.com/NVlabs/ffhq-dataset - Labeled: https://github.com/royorel/FFHQ-Aging-Dataset	
CelebA-HQ - From [8]: https://github.com/tkarras/progressive_growing_of_gans - with Mask: https://github.com/switchablenorms/CelebAMask-HQ	
StyleGANv2 [10] - Official model: https://github.com/NVlabs/stylegan2 - Simple model: https://github.com/lucidrains/stylegan2-pytorch - Ready to use images: https://thispersondoesnotexist.com/	
CACD http://bcsiriuschen.github.io/CARC/	

Table 5. Links to the datasets.

LATS [14] - https://github.com/royorel/Lifespan_Age_Transformation_Synthesis	
HRFAE [20] - https://github.com/InterDigitalInc/HRFAE	
IPCGAN [19] - https://github.com/dawei6875797/Face-Aging-with-Identity-Preserved-Conditional-Generative-Adversarial-Networks	

Table 6. Links to the implementations.

Age recognition We use age recognition accuracy (%) to quantitatively measure the correctness of age transformation. We consider test images of age-group 20-29 to be the source images for transforming their ages into 0-2, 3-6, 7-9, 10-14, 30-39, 40-49, and 50+ groups. We chose this particular age-group as an anchor to assess the age transformation in aging and rejuvenating tasks. Similar to FID calculation, we perform recognition for each age-group separately. We use VGG16 [16] trained on age dataset [15] to predict the age of generated images. We report the accuracy based on the ratio of the number of samples recognized correctly to the total number of samples.

3. Datasets and Implementations

In our experiment, we use images of FFHQ [9], CelebA-HQ [8], CACD [2] datasets, as well as synthesized images by StyleGANv2 [10]. In Table 5, we provide links to these image sets.

In our performance comparisons, we mainly consider two recent works: LATS [14] and HRFAE [20]. In addition, we compare our results against IPCGAN [19]. Table 6 provides links to the implementations of these methods.

4. Additional Results

While performing an experiment on CelebA-HQ dataset, we found that the dataset contains old-looking images.

Figure 1. Performance of our model on old-looking images of CelebA-HQ. Left and right are young and old-aged versions of the input at the middle, respectively. Note that images are compressed.

Hence, we additionally demonstrate the performance of our model on such images in Figure 1 with a few exemplar age transformations. As shown, the generated images have a few color differences between the input (*e.g.*, eyes and chicks). We consider such occurrence as the effect of model generalization on unseen data where model is attempting to introduce new information according to the given age. Nevertheless, the overall results exhibit the generalizability of our model on old-looking images.

We also provide additional age transformation results on FFHQ [9] and CelebA-HQ [8] datasets, as well as on synthesized images by StyleGANv2 [10]. Figures 2, 3 and 4 demonstrate the performance of our method in aging and rejuvenating tasks, thereby exhibiting continuous age transformation. Note that the results for the images of CelebA-HQ and StyleGANv2 are generated by our model trained on the FFHQ dataset.

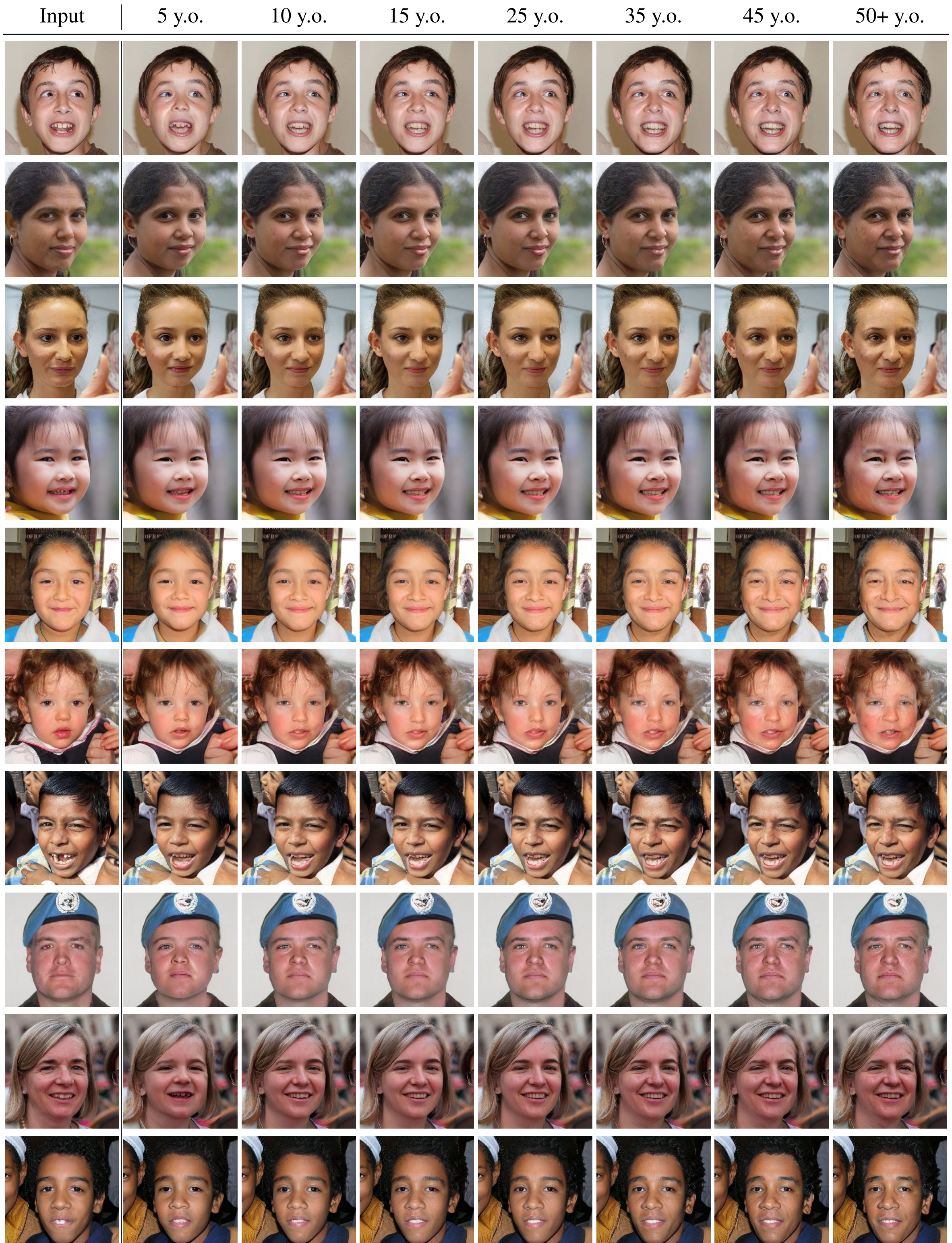

Figure 2. Performance of our model on the FFHQ dataset. The first column is the input, whereas the others are our results (y.o. denotes years old). The input images are from the test set, and the model has never seen them. Note that images are compressed.

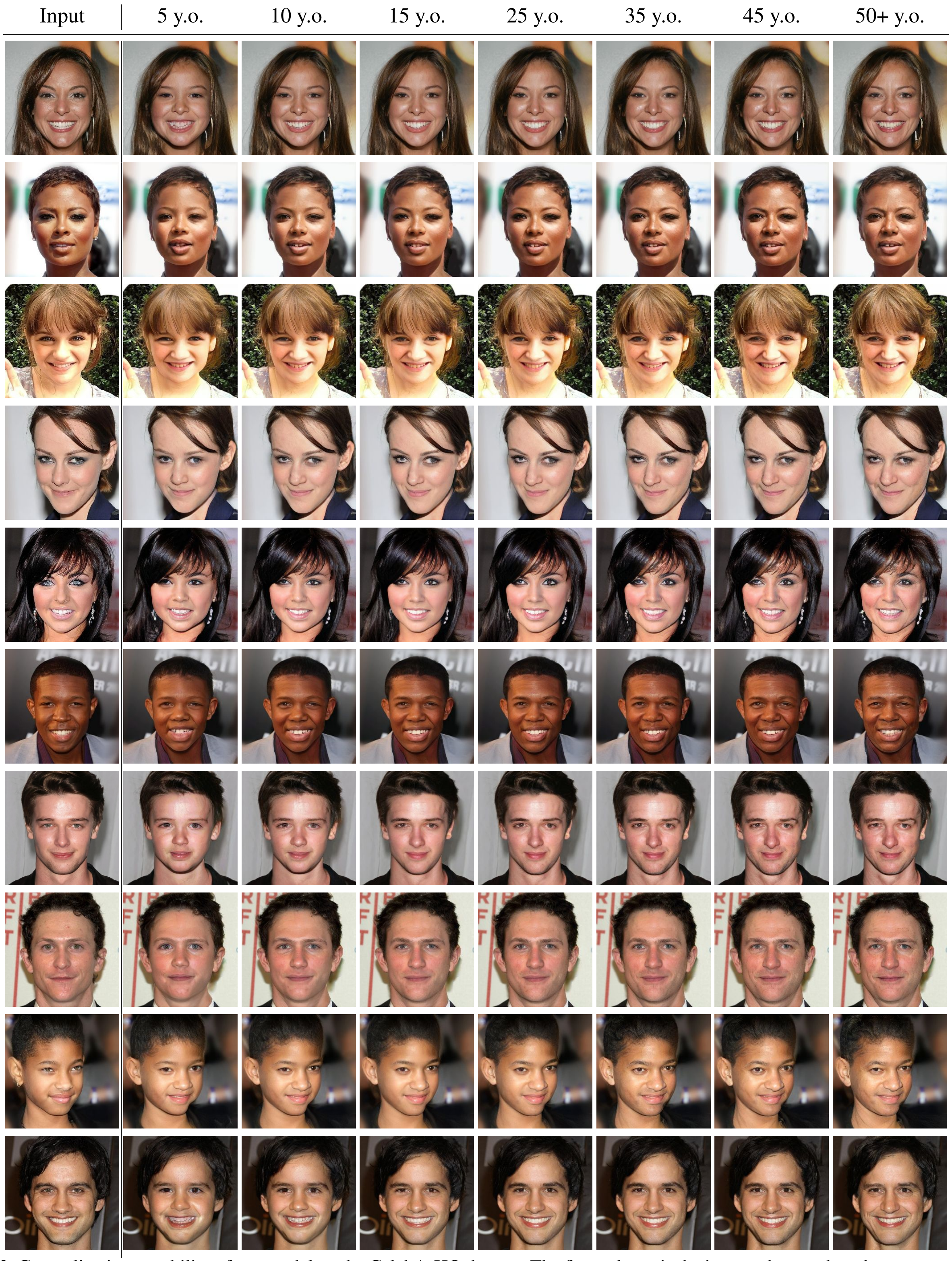

Figure 3. Generalization capability of our model on the CelebA-HQ dataset. The first column is the input, whereas the others are our results (y.o. denotes years old). Note that images are compressed.

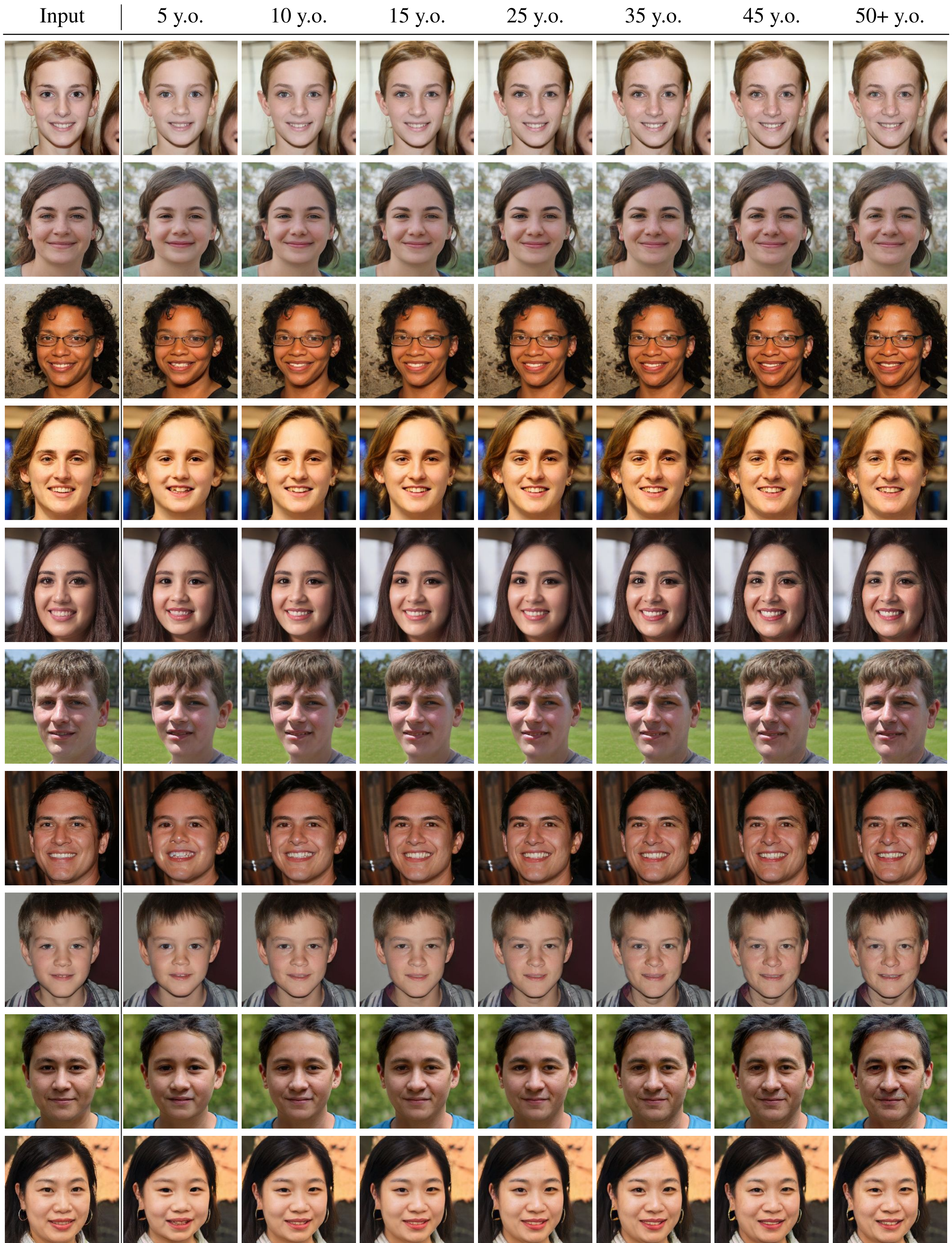

Figure 4. Generalization capability of our model on synthesized images of StyleGANv2. The first column is the input, whereas the others are our results (y.o. denotes years old). Note that images are compressed.

References

[1] Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018. 1, 2

[2] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proc. of the European Conference on Computer Vision*, pages 768–783. Springer, 2014. 2

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017. 1

[4] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020. 1

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Proc. of the European Conference on Computer Vision*, pages 630–645, 2016. 1

[6] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017. 1, 2

[7] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pages 1501–1510, 2017. 1

[8] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. 1, 2, 3

[9] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 2, 3

[10] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020. 2, 3

[11] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. MaskGAN: Towards diverse and interactive facial image manipulation. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5549–5558, 2020. 1

[12] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. 1

[13] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In *International Conference on Learning Representations*, 2018. 1

[14] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *Proc. of the European Conference on Computer Vision*, 2020. 1, 2

[15] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2):144–157, 2018. 2

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015. 2

[17] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016. 2

[18] Manjunath Kudlur Vincent Dumoulin, Jonathon Shlens. A learned representation for artistic style. In *International Conference on Learning Representations*, 2017. 1

[19] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018. 2

[20] Xu Yao, Gilles Puy, Alasdair Newson, Yann Gousseau, and Pierre Hellier. High resolution face age editing. In *2020 25th International Conference on Pattern Recognition*, pages 8624–8631, 2021. 2