FACE AGING WITH CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

Grigory Antipov⋆†   *Moez Baccouche*⋆   *Jean-Luc Dugelay*†

⋆ Orange Labs, 4 rue Clos Courtel, 35512 Cesson-Sévigné, France
† Eurecom, 450 route des Chappes, 06410 Biot, France

ABSTRACT

It has been recently shown that Generative Adversarial Networks (GANs) can produce synthetic images of exceptional visual fidelity. In this work, we propose the first GAN-based method for automatic face aging. Contrary to previous works employing GANs for altering of facial attributes, we make a particular emphasize on preserving the original person's identity in the aged version of his/her face. To this end, we introduce a novel approach for "Identity-Preserving" optimization of GAN's latent vectors. The objective evaluation of the resulting aged and rejuvenated face images by the state-of-the-art face recognition and age estimation solutions demonstrate the high potential of the proposed method.

Index Terms—
Face Aging, GAN, Deep Learning, Face Synthesis

1. INTRODUCTION

Face aging, also known as age synthesis [1] and age progression [2], is defined as aesthetically rendering a face image with natural aging and rejuvenating effects on the individual face [1]. It has plenty of applications in various domains including cross-age face recognition [3], finding lost children and entertainment [4]. Traditional face aging approaches can be roughly split between the prototyping ones [5, 6] and the modeling ones [7, 8]. Prototyping approaches estimate average faces within predefined age groups. The differences between these faces constitute the aging patterns which are further used to transform an input face image into the target age group. The prototype approaches are simple and fast, but since they are based on general rules, they totally discard the personalized information which results in unrealistic images. On the contrary, modeling approaches employ parametric models to simulate the aging mechanisms of muscles, skin and skull of a particular individual. However, these approaches often require face aging sequences of the same person with wide range of ages which are very costly to collect.

The presented traditional face aging approaches are limited to modeling of the aging patterns missing the global comprehension of a human face (its personality traits, facial expression, possible facial accessories etc.) However, in many real-life use cases, face aging must be combined with other face alterations, such as adding sunglasses or beard. These non-trivial modifications require global generative models of human faces. Natural image generation has been thoroughly studied for years, but it has not been until 2014 when advances in deep learning has allowed to produce image samples and interpolations of very high visual fidelity. The key model which has made it possible is called Generative Adversarial Network (GAN) [9] and is presented in Subsection 2.1. Contrary to autoencoders (for example, variational autoencoders [10]) which optimize the L_2 reconstruction loss and hence produce blurry images, GANs are explicitly trained to generate the most plausible and realistic images which are hard to distinguish from real data. GANs have already been employed to perform modifications on human faces, such as changing the hair color, adding sunglasses or even binary aging (i.e. simply making face look older or younger without precising particular age categories) [11, 12]. A common problem of previous GAN-based methods of face modification [11, 12] is the fact that the original person's identity is often lost in the modified image. Therefore, in the present study, we focus on *identity-preserving* face aging. In particular, our contributions are as following:

1. We design Age-cGAN (Age Conditional Generative Adversarial Network), the first GAN to generate high quality synthetic images within required age categories.

2. We propose a novel latent vector optimization approach which allows Age-cGAN to reconstruct an input face image preserving the original person's identity.

The rest of the paper is composed of Section 2 where we present our face aging method, Section 3 where we evaluate the proposed identity-preserving face reconstruction and age synthesis approaches, and Section 4 which concludes this study and gives the directions for the future work.

2. PROPOSED METHOD

Our face aging method is based on Age-cGAN, a generative model for synthesis of human faces within required age categories. The design of Age-cGAN is detailed in Subsection 2.1. Once Age-cGAN is trained, the face aging is done in two steps (cf. Figure 1):
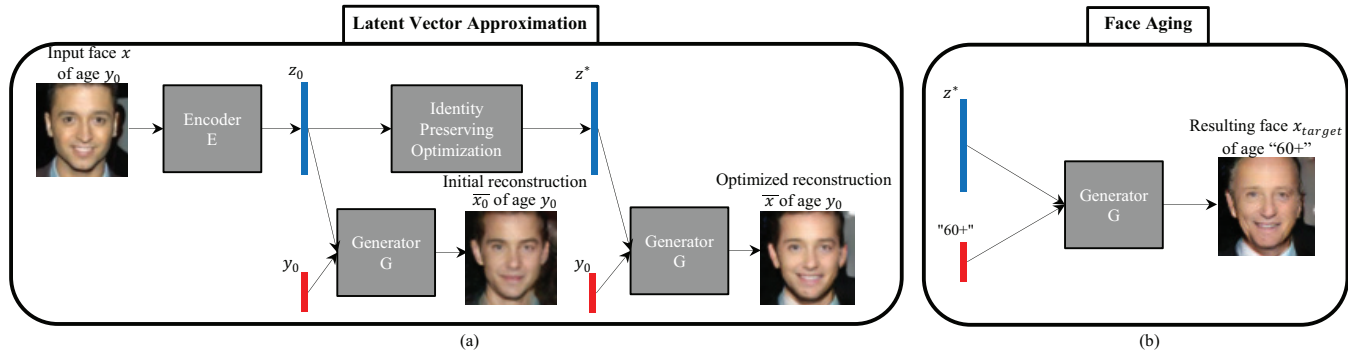
Fig. 1. Our face aging method. (a) approximation of the latent vector to reconstruct the input image; (b) switching the age condition at the input of the generator G to perform face aging.

1. Given an input face image x of age y_0, find an optimal latent vector z^* which allows to generate a reconstructed face $\bar{x} = G(z^*, y_0)$ as close as possible to the initial one (cf. Figure 1-(a)).

2. Given the target age y_{target}, generate the resulting face image $x_{target} = G(z^*, y_{target})$ by simply switching the age at the input of the generator (cf. Figure 1-(b)).

The first step of the presented face aging method (i.e. input face reconstruction) is the key one. Therefore, in Subsection 2.2, we present our approach to approximately reconstruct an input face making a particular emphasize on preserving the original person's identity in the reconstructed image.

2.1. Age Conditional Generative Adversarial Network

Introduced in [9], GAN is a pair of neural networks (G, D): the generator G and the discriminator D. G maps vectors z from the noise space N^z with a known distribution p_z to the image space N^x. The generator's goal is to model the distribution p_{data} of the image space N^x (in this work, p_{data} is the distribution of all possible face images). The discriminator's goal is to distinguish real face images coming from the image distribution p_{data} and synthetic images produced by the generator. Both networks are iteratively optimized against each other in a minimax game (hence the name "adversarial").

Conditional GAN (cGAN) [13, 14] extends the GAN model allowing the generation of images with certain attributes ("conditions"). In practice, conditions $y \in N^y$ can be any information related to the target face image: level of illumination, facial pose or facial attribute. More formally, cGAN training can be expressed as an optimization of the function $v(\theta_G, \theta_D)$, where θ_G and θ_D are parameters of G and D, respectively:

$$
\begin{aligned}
\min_{\theta_G} \max_{\theta_D} v(\theta_G, \theta_D) = {} & \mathbf{E}_{x,y \sim p_{data}} [\log D(x, y)] \\
& + \mathbf{E}_{z \sim p_z(z), \widetilde{y} \sim p_y} [\log (1 - D(G(z, \widetilde{y}), \widetilde{y}))]
\end{aligned}
\tag{1}
$$

The Age-cGAN model proposed in this work uses the same design for the generator G and the discriminator D as in [15]. Following [12], we inject the conditional information at the input of G and at the first convolutional layer of D. Age-cGAN is optimized using the ADAM algorithm [16] during 100 epochs. In order to encode person's age, we have defined six age categories: 0-18, 19-29, 30-39, 40-49, 50-59 and 60+ years old. They have been selected so that the training dataset (cf. Subsection 3.1) contains at least $5,000$ examples in each age category. Thus, the conditions of Age-cGAN are six-dimensional one-hot vectors.

2.2. Approximative Face Reconstruction with Age-cGAN

2.2.1. Initial Latent Vector Approximation

Contrary to autoencoders, cGANs do not have an explicit mechanism for inverse mapping of an input image x with attributes y to a latent vector z which is necessary for image reconstruction: $x = G(z, y)$. As in [12, 17], we circumvent this problem by training an encoder E, a neural network which approximates the inverse mapping.

In order to train E, we generate a synthetic dataset of 100K pairs $(x_i, G(z_i, y_i))$, $i = 1, \ldots, 10^5$, where $z_i \sim N(0, I)$ are random latent vectors, $y_i \sim U$ are random age conditions uniformly distributed between six age categories, $G(z, y)$ is the generator of the priorly trained Age-cGAN, and $x_i = G(z_i, y_i)$ are the synthetic face images. E is trained to minimize the Euclidean distances between estimated latent vectors $E(x_i)$ and the ground truth latent vectors z_i.

Despite GANs are arguably the most powerful generative models today, they cannot *exactly* reproduce the details of all real-life face images with their infinite possibilities of minor facial details, accessories, backgrounds etc. In general, a natural input face image can be rather *approximated* than *exactly reconstructed* with Age-cGAN. Thus, E produce *initial latent approximations* z_0 which are good enough to serve as initializations of our optimization algorithm explained hereafter.

2.2.2. Latent Vector Optimization

Face aging task, which is the ultimate goal of this work, assumes that while the age of a person must be changed, his/her identity should remain intact. In Subsection 3.3, it is shown that though initial latent approximations z_0 produced by E result in visually plausible face reconstructions, the identity of the original person is lost in about 50% of cases (cf. Table 1). Therefore, initial latent approximations z_0 must be improved.

In [17], the similar problem of image reconstruction enhancement is solved by optimizing the latent vector z to minimize the pixelwise Euclidean distance between the ground truth image x and the reconstructed image $\bar{x}$. However, in the context of face reconstruction, the described "Pixelwise" latent vector optimization has two clear downsides: firstly, it increases the blurriness of reconstructions and secondly (and more importantly), it focuses on unnecessary details of input face images which have a strong impact on pixel level, but have nothing to do with a person's identity (like background, sunglasses, hairstyle, moustache etc.)

Therefore, in this paper, we propose a novel "Identity-Preserving" latent vector optimization approach. The key idea is simple: given a face recognition neural network FR able to recognize a person's identity in an input face image x, the difference between the identities in the original and reconstructed images x and $\bar{x}$ can be expressed as the Euclidean distance between the corresponding embeddings $FR(x)$ and $FR(\bar{x})$. Hence, minimizing this distance should improve the identity preservation in the reconstructed image $\bar{x}$:

$$z^*_{IP} = \underset{z}{\arg\min} ||FR(x) - FR(\bar{x})||_{L_2} \qquad (2)$$

In this paper, FR is an internal implementation of the "FaceNet" CNN [18]. The generator $G(z, y)$ and the face recognition network $FR(x)$ are differentiable with respect to their inputs, so the optimization problem 2 can be solved using the L-BFGS-B algorithm [19] with backtracking line search. The L-BFGS-B algorithm is initialized with initial latent approximations z_0. Here and below in this work, we refer to the results of "Pixelwise" and "Identity-Preserving" latent vector optimizations as *optimized latent approximations* and denote them respectively as z^*_{pixel} and z^*_{IP}. In Subsection 3.3, it is shown both subjectively and objectively that z^*_{IP} better preserves a person's identity than z^*_{pixel}.

3. EXPERIMENTS

3.1. Dataset

Age-cGAN has been trained on the *IMDB-Wiki_cleaned* dataset [20] of about 120K images which is a subset of the public *IMDB-Wiki* dataset [21]. More precisely, 110K images have been used for training of Age-cGAN and the remaining 10K have been used for the evaluation of identity-preserving face reconstruction (cf. Subsection 3.3).

3.2. Age-Conditioned Face Generation

Figure 2 illustrates synthetic faces of different ages generated with our Age-cGAN. Each row corresponds to a random latent vector z and six columns correspond to six age conditions y. Age-cGAN perfectly disentangles image information encoded by latent vectors z and by conditions y making them independent. More precisely, we observe that latent vectors z encode person's identity, facial pose, hair style, etc., while y encodes *uniquely* the age.
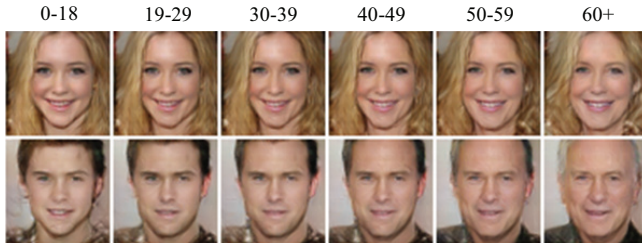

Fig. 2. Examples of synthetic images generated by our Age-cGAN using two random latent vectors z (rows) and conditioned on the respective age categories y (columns).

In order to objectively measure how well Age-cGAN manages to generate faces belonging to precise age categories, we have used the state-of-the-art age estimation CNN described in [20]. We compare the performances of the age estimation CNN on real images from the test part of *IMDB-Wiki_cleaned* and on 10K synthetic images generated by Age-cGAN. Despite the age estimation CNN has never seen synthetic images during the training, the resulting mean age estimation accuracy on synthetic images is just 17% lower than on natural ones. It proves that our model can be used for generation of realistic face images with the required age.

3.3. Identity-Preserving Face Reconstruction and Aging

As explained in Subsection 2.2, we perform face reconstruction (i.e. the first step of our face aging method) in two iterations: firstly, (1) using initial latent approximations obtained from the encoder E and then (2) using optimized latent approximations obtained by either "Pixelwise" or "Identity-Preserving" optimization approaches. Some examples of original test images, their initial and optimized reconstructions are presented in Figure 3 ((a), (b) and (c), respectively).

It can be seen in Figure 3 that the optimized reconstructions are closer to the original images than the initial ones. However, the choice is more complicated when it comes to the comparison of the two latent vector optimization approaches. On the one hand, "Pixelwise" optimization better reflects superficial face details: such as the hair color in the first line and the beard in the last line. On the other hand, the identity traits (like the form of the head in the second line or the form
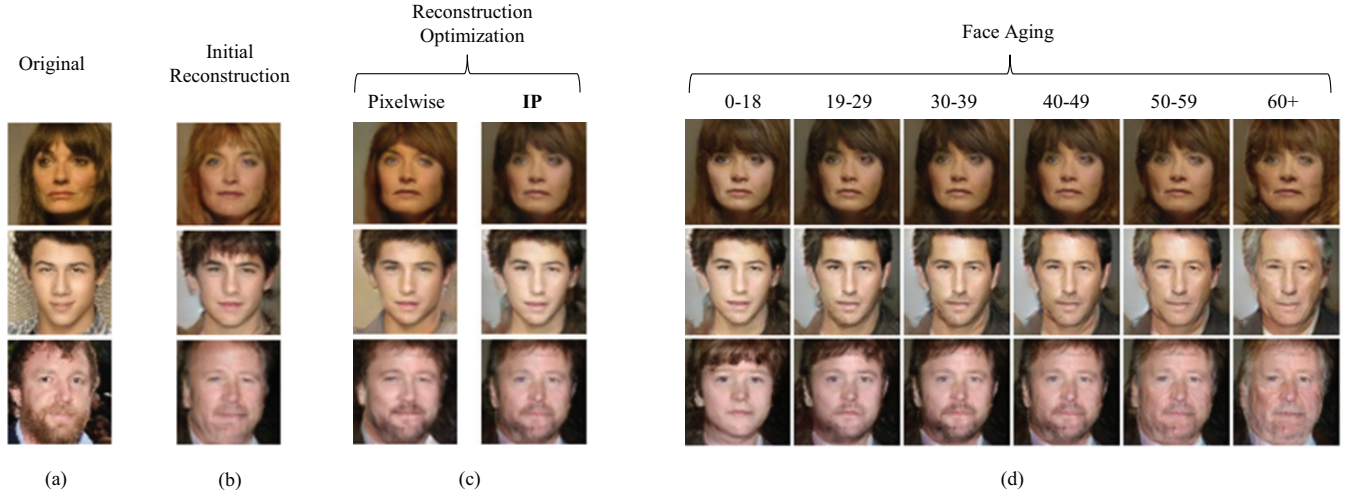
Fig. 3. Examples of face reconstruction and aging. (a) original test images, (b) reconstructed images generated using the initial latent approximations: z_0, (c) reconstructed images generated using the "Pixelwise" and "Identity-Preserving" optimized latent approximations: z^*_{pixel} and z^*_{IP}, and (d) aging of the reconstructed images generated using the identity-preserving z^*_{IP} latent approximations and conditioned on the respective age categories y (one per column).

Reconstruction type	FR score
Initial Reconstruction (z_0)	53.2%
"Pixelwise" Optimization (z^*_{pixel})	59.8%
"Identity-Preserving" Optimization (z^*_{IP})	**82.9%**

Table 1. "OpenFace" Face Recognition (FR) scores on three compared types of face reconstruction.

of the eyes in the first and last lines) are better represented by "Identity-Preserving" optimization.

For the sake of objective comparison of the two approaches for the identity-preserving face reconstruction, we employ "OpenFace" [22] software which is currently one of the best open-source face recognition solutions. Given two face images, "OpenFace" decides whether they belong to the same person or not. Using both approaches, we have reconstructed 10K test images of the *IMDB-Wiki_cleaned* dataset and fed the resulting images alongside corresponding originals to "OpenFace". Table 1 presents the percentages of "OpenFace" positive outputs (i.e. when the software believed that a face image and its reconstruction belong to the same person). The results confirm the visual observations presented above. Initial reconstructions allow "OpenFace" to recognize the original person only in half of test examples. This percentage is slightly increased by "Pixelwise" optimization but the improvement is marginal. On the contrary, "Identity-Preserving" optimization approach preserves the person's identities much better demonstrating by far the best face recognition performance of 82.9%.

Finally, once an identity-preserving face reconstruction

$\bar{x} = G(z^*_{IP}, y_0)$ of the original image x is obtained, we can simply substitute the initial age category y_0 by the target age category y_{target} in order to obtain the output of our face aging method: $x_{target} = G(z^*_{IP}, y_{target})$. Figure 3-(d) illustrates how it works in practice. Our method manages both to realistically age an originally young face into a senior one as in the 2nd line of Figure 3 and vice-versa as in the last line.

4. CONCLUSIONS AND FUTURE WORK

In this work, we have proposed a new effective method for synthetic aging of human faces based on Age Conditional Generative Adversarial Network (Age-cGAN). The method is composed of two steps: (1) input face reconstruction requiring the solution of an optimization problem in order to find optimal latent approximation z^*, (2) and face aging itself performed by a simple change of condition y at the input of the generator. The cornerstone of our method is the novel "Identity-Preserving" latent vector optimization approach allowing to preserve the original person's identity in the reconstruction. This approach is universal meaning that it can be used to preserve identity not only for face aging but also for other face alterations (e.g. adding a beard, sunglasses etc.)

Our face aging method can be used for synthetic augmentation of face datasets and for improving the robustness of face recognition solutions in cross-age scenarios. It is part of our future work. Moreover, we believe that the face reconstruction part of our method can be further improved by combining "Pixelwise" and "Identity-Preserving" approaches into one optimization objective. We are also planning to explore this path in our further studies.

5. REFERENCES

[1] Yun Fu, Guodong Guo, and Thomas S Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[2] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan, "Personalized age progression with aging dictionary," in *Proceedings of International Conference on Computer Vision*, Santiago, Chile, 2015.

[3] Unsang Park, Yiying Tong, and Anil K Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010.

[4] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe, "Recurrent face aging," in *Proceedings of Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016.

[5] Bernard Tiddeman, Michael Burt, and David Perrett, "Prototyping and transforming facial textures for perception research," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 42–50, 2001.

[6] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz, "Illumination-aware age progression," in *Proceedings of Computer Vision and Pattern Recognition*, Columbus, USA, 2014.

[7] 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*, vol. 32, no. 3, pp. 385–401, 2010.

[8] Yusuke Tazoe, Hiroaki Gohara, Akinobu Maejima, and Shigeo Morishima, "Facial aging simulator considering geometry and patch-tiled texture," in *Proceedings of ACM SIGGRAPH*, Los Angeles, USA, 2012.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Proceedings of advances in Neural Information Processing Systems*, Montreal, Canada, 2015.

[10] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," in *Proceedings of International Conference on Learning Representations*, Banff, Canada, 2014.

[11] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther, "Autoencoding beyond pixels using a learned similarity metric," in *Proceedings of International Conference on Machine Learning*, New York, USA, 2016.

[12] Guim Perarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M Álvarez, "Invertible conditional gans for image editing," in *Proceedings of advances in Neural Information Processing Systems Workshops*, Barcelona, Spain, 2016.

[13] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," in *Proceedings of advances in Neural Information Processing Systems*, Montreal, Canada, 2014.

[14] Jon Gauthier, "Conditional generative adversarial nets for convolutional face generation," *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition*, 2014.

[15] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proceedings of International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.

[16] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros, "Generative visual manipulation on the natural image manifold," in *Proceedings of European Conference on Computer Vision*, Amsterdam, Netherlands, 2016.

[18] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of Computer Vision and Pattern Recognition*, Boston, USA, 2015.

[19] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[20] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, Las Vegas, USA, 2016.

[21] Rasmus Rothe, Radu Timofte, and Luc Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of International Conference on Computer Vision Workshops*, Santiago, Chile, 2015.

[22] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," Tech. Rep., CMU-CS-16-118, CMU School of Computer Science, 2016.