

Generative Adversarial Style Transfer Networks for Face Aging

Sveinn Palsson
D-ITET, ETH Zurich
spalsson@ethz.ch

Eirikur Agustsson
D-ITET, ETH Zurich
aeirikur@ethz.ch

Radu Timofte
D-ITET, ETH Zurich
Merantix

Luc Van Gool
D-ITET, ETH Zurich
ESAT, KU Leuven

Abstract

How somebody looked like when younger? What could a person look like when 10 years older? In this paper we look at the problem of face aging, which relates to processing an image of a face to change its apparent age. This task involves synthesizing images and modeling the aging process, which both are problems that have recently enjoyed much research interest in the field of face and gesture recognition. We propose to look at the problem from the perspective of image style transfer, where we consider the age of the person as the underlying style of the image.

We show that for large age differences, convincing face aging can be achieved by formulating the problem with a pairwise training of Cycle-consistent Generative Adversarial Networks (CycleGAN) over age groups. Furthermore, we propose a variant of CycleGAN which directly incorporates a pre-trained age prediction model, which performs better when the desired age difference is smaller.

The proposed approaches are complementary in strengths and their fusion performs well for any desired level of aging effect. We quantitatively evaluate our proposed method through a user study and show that it outperforms prior state-of-the-art techniques for face aging.

1. Introduction

Face aging refers to processing an image of a face to change its apparent age. Face aging methods have attracted a good amount of research interest in recent years and they have found applications in a number of domains. They are used for surveillance, entertainment [5] and forensic artists have used them to simulate the appearance of wanted or missing persons [8]. This problem has been explored from several different angles and prior work can be roughly split into two main categories, physical model-based [13, 28] and prototype-based [30, 8] methods. More recent methods [2, 32] make use of Generative Adversarial Networks (GANs) [6] which have become very popular for image synthesis and we describe in more detail in section 2.2.

The main contributions of this paper are as follows:

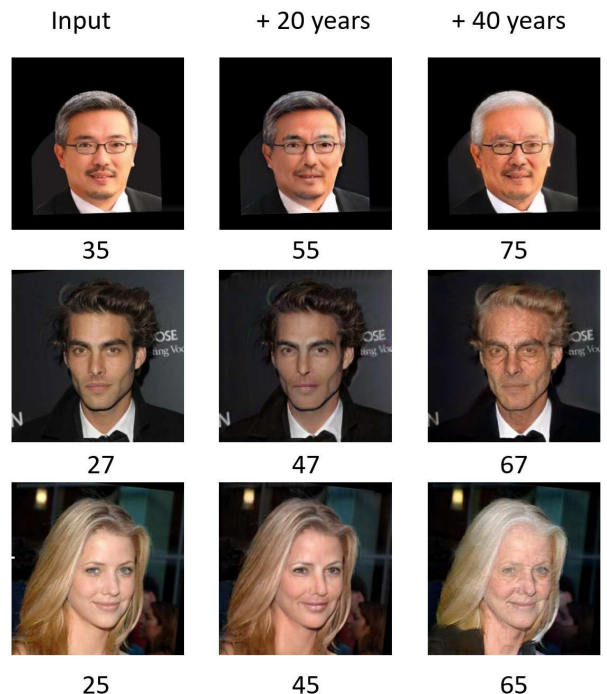


Figure 1. Face aging results of our methods.

1. Viewing the face aging problem as image style transfer, where we consider the age of the person as the underlying style of the image.
2. Group-GAN, formulating the problem with a pairwise training of Cycle-consistent Generative Adversarial Networks (CycleGAN) over age groups.
3. Face Aging GAN (FA-GAN), a variant of CycleGAN which directly incorporates a pre-trained age prediction model, instead of using predefined age groups.
4. F-GAN, a fusion of Group-GAN and FA-GAN, that outperforms prior state-of-the-art face aging methods.

The remainder of the paper is structured as follows. In section 2 we discuss related work. In section 3 we introduce our methods, Group-GAN and FA-GAN. In section 4 we describe the implementation of these methods. In section 5

we provide the experimental setup and the achieved results. We show that the performance of Group-GAN and FAGAN is complementary on the level of desired age change and they can be combined into a fusion model, F-GAN, that has better performance than prior state-of-the-art methods. Section 6 draws the conclusions.

2. Related Work

2.1. Face Aging

The two main categories of face aging approaches are prototype based and model based. Prototype based methods of face aging perform group based learning where within each age group an average face is estimated. They capture the differences in shape and texture between these age groups and then apply them to images for age progression [8]. The main problem with this approach to aging is that personalized features of the face tend to get lost. Preserving identity of the face is a basic requirement for a face aging method in practical applications, *e.g.* in cross-age face identification. Model based approaches such as [29] however aim at creating parametric models for aging various parts of the face. Such models typically suffer from high complexity and computational cost [32]. More recent work in face aging [2, 32] use Generative Adversarial Networks (GANs) with an autoencoder where the input images are first encoded into a latent space and then reconstructed with a different age. This approach to face aging requires learning an age transformation in the latent space and to train a generator that can produce realistic images from any point in the latent space which alone is a difficult task [20]. These approaches are also limited to low resolution images.

It is hard to quantitatively compare face aging systems since a fair comparison depends on application. One way is through subjective tests where human participants vote on the best result [32, 30]. Another is to compare the improved performance of a cross-age face verification system with synthesized images [30].

2.2. Generative Adversarial Networks

Creating a generative model of natural images is a challenging task. Such a model needs to be able to capture the rich distributions from which natural images come from. Generative Adversarial Networks (GANs) [6] have proven to be excellent for this task and can produce images of high visual fidelity [20, 25, 24, 7, 3, 21, 12]. GANs consist of a pair of models, a generative model G and a discriminative model D . D is trained to estimate the probability that a sample x comes from the training data distribution $p_{data}(x)$, while simultaneously G is trained to maximize the probability of D making a mistake. This training procedure is referred to as adversarial training and corresponds to a min-max two-player game between G and D . G is a function

that maps any vector z , sampled from the prior $p(z)$, to the data space. This can be formulated as

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (1)$$

The training procedure is executed in two alternating steps. First D is trained to distinguish between real samples and generated ones and then the generator is trained to fool D with generated samples. GANs have found applications in many computer vision related domains, *e.g.* image super-resolution [14], style transfer [10], colorization [4] and image inpainting [31].

2.3. Image-to-Image Translation

Many tasks in computer vision and graphics can be thought of as translation problems where an input image x is to be translated from domain X to \hat{y} in some other domain Y . Isola *et al.* [10] introduced an image-to-image translation framework that uses GANs in a conditional setting where the generator transforms images conditioned on the input image x . The method requires paired image data from the two domains and the discriminator is trained to estimate the probability that a pair of images (x, y) is a real or fake pair. Introduced by Zhu *et al.* [33], CycleGAN extends this framework to unsupervised image-to-image translation. WESPE [9] and UNIT [16] are other recent unsupervised image-to-image translation frameworks with remarkable results.

CycleGAN [33] consists of two pairs of neural networks, (G, D_X) and (F, D_Y) , where the translators between domains X and Y are $G : X \rightarrow Y$ and $F : Y \rightarrow X$. D_X is trained to discriminate between real images $\{x\}$ and translated images $\{F(x)\}$ while D_Y is trained to discriminate between images $\{y\}$ and $\{G(x)\}$. The system is trained using both an adversarial loss

$$\mathcal{L}_{LSGAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim p_{data}(x)} [D_Y(G(x))^2] \quad (2)$$

and a cycle consistency loss

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1]. \quad (3)$$

The Cycle consistency loss is a way to encourage the mappings G and F to be inverses of each other such that $F(G(x)) \approx x$ and $G(F(y)) \approx y$. The traditional negative log likelihood GAN loss from (1) is however replaced by a least square GAN loss [18] that has been shown to be more stable during training and to produce higher quality results. The full objective of CycleGAN is then expressed as

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{LSGAN}(G, D_Y, X, Y) + \mathcal{L}_{LSGAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \quad (4)$$

Zhu *et al.* [33] produced convincing image translations such as between horses and zebras, between paintings and photographs and between artistic styles. This motivated us to employ CycleGAN for translating images between age groups which we discuss in more detail in section 3.1.

2.4. Face Age Estimation

Relevant to this work is of course the estimation of age from face images. Deep Expectation of Apparent Age (DEX) [22, 23] is a deep convolutional neural network that uses the VGG-16 architecture [27]. The model is trained on a large dataset consisting of over 500,000 images, the IMDB-WIKI dataset, also introduced in [23]. Pre-trained DEX models are publicly available which we borrow for use in our face aging system as described in section 3.2.

3. Proposed methods

3.1. Group-GAN: Group based Face Aging with CycleGAN

As discussed before, CycleGAN [33] has proven to be a useful tool for style transfer with unpaired image data. We propose to use this framework for face aging where the face dataset is split into a few age groups and then one CycleGAN model is trained between each pair of groups. In section 5.2, we discuss our experiments with this method and see that it is effective when the desired age difference is large but performs poorly when it is small. Another issue with this approach is that we need many models and a way to select what model to use. In the next we introduce a method that addresses this issue.

3.2. FA-GAN: Face Aging GAN

One could argue that the ideal face aging model would be one that can take an input image x_0 and a number k and output an image x_k which contains the same face after k years. We would want to be able to perform rejuvenation with the same model so we allow k to be negative as well. Let G be the model that we wish to train for this task. One of the basic requirements for $G(x_0) = \hat{x}_k$ is that it is photo-realistic. Thus, we prefer to have discriminator D that will be trained jointly with G and we express the GAN [6] loss as

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_{x_0 \sim p_{data}(x_0)} [\log D(x_0)] + \mathbb{E}_{x_0 \sim p_{data}(x_0)} [\log(1 - D(G(x_0)))]. \quad (5)$$

Now to force the model to add age to the face we need an appropriate loss function. For this we assume that we have a differentiable age estimator $age(x)$ and define the age loss as the mean absolute error from the target, which is a stan-

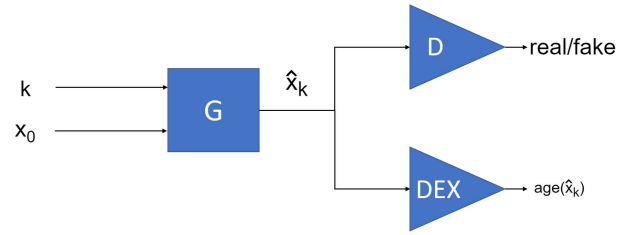


Figure 2. Diagram of proposed FA-GAN model.

dard loss for learning based age estimation techniques:

$$\mathcal{L}_{age}(G) = \mathbb{E}_{k \sim p(k)} \mathbb{E}_{x_0 \sim p_{data}(x_0)} [|age(G(x_0, k)) - age(x_0) - k|] \quad (6)$$

No prior work has directly integrated an age estimator into the loss function for training a face aging model. An essential feature of a face aging system is identity preservation. However, with the combination of these two loss functions there is nothing that forces the system to preserve the identity of the face. This is where CycleGAN [33] comes into play. We can add a cycle consistency loss as in (11) that encourages $G(G(x_0, k), -k) \approx x_0$. Aging a face by k years and then rejuvenating by k years should give us back the original face. We define a cycle loss for our system as

$$\mathcal{L}_{cyc}(G) = \mathbb{E}_{k \sim p(k)} \mathbb{E}_{x_0 \sim p_{data}(x_0)} [\|G(G(x, k), -k) - x_0\|_1]. \quad (7)$$

The full objective function for G and D is:

$$\mathcal{L}(G, D) = \mathcal{L}_{GAN}(G, D) + \lambda_1 \mathcal{L}_{age}(G) + \lambda_2 \mathcal{L}_{cyc}(G), \quad (8)$$

where λ_1 and λ_2 are constants that control the weights of the three objectives. We then solve:

$$G^* = \arg \min_G \max_D \mathcal{L}(G, D). \quad (9)$$

With this formulation of the problem we are requiring G to learn how to take an image of a face and render with an aging effect of k years. But aging a young person by k years is not the same as aging an old person by the same amount, the way people change with age depends on their age. For G to be successful it needs to learn to take the age of the input image into account. To help with this issue, we feed G the estimated age of the input image. The diagram of this method is shown in Fig. 2. We name it Face Aging GAN or, shortly, FA-GAN.

4. Implementation

The Group-GAN is implemented by training a CycleGAN [33] model between each pair of age groups. We use the pytorch [1] implementation of CycleGAN that is available on Github. The remainder of this section describes the implementation of FA-GAN. For FA-GAN, we use the

same implementation for our generator G and discriminator D functions. The generator network architecture comes from Johnson *et al.* [11] which was applied to neural style transfer and super-resolution with impressive results. The architecture has two strided convolution layers, followed by 9 residual blocks, followed by two fractionally strided convolution layers. The discriminator is a 70×70 PatchGAN [15, 10] that does not look at the image as a whole but rather on many 70×70 patches of the image to classify them as real or fake. For the age estimator we use DEX from Rothe *et al.* as a pre-trained model [23]. DEX uses the VGG-16 architecture with an output layer of 101 neurons followed by a softmax expectation. In our preliminary experiments we observed that separating the generator G , described in the previous section, into G_+ and G_- where $G_+(x_0, k) = \hat{x}_k$ does the age progression and $G_-(x_0, k) = \hat{x}_{-k}$ does age regression, performs better than having one generator for both. We also add a loss term for cycle consistency in ageing, *i.e.* $DEX(G_-(\hat{x}_k, k)) \approx DEX(x_0)$ and $DEX(G_+(\hat{x}_{-k}, k)) \approx DEX(x_0)$, which we express as

$$\begin{aligned} \mathcal{L}_{ac}(G_+, G_-) = & \mathbb{E}_{k \sim p(k)} \mathbb{E}_{x_0 \sim p_{data}(x_0)} \left[\left| DEX(G_-(\hat{x}_k, k)) - DEX(x_0) \right| \right. \\ & \left. + \left| DEX(G_+(\hat{x}_{-k}, k)) - DEX(x_0) \right| \right]. \end{aligned}$$

As mentioned in section 2.3, Zhu *et al.* [33] suggest using a LSGAN [18] (2) instead of the traditional loss (5) so we replace \mathcal{L}_{GAN} with

$$\begin{aligned} \mathcal{L}_{LSGAN}(G, D) = & \mathbb{E}_{x_0 \sim p_{data}(x_0)} [(D(x_0) - 1)^2] \\ & + \mathbb{E}_{x_0 \sim p_{data}(x_0)} [D(G(x_0))^2] \end{aligned}$$

For completeness we redefine the age loss and cycle loss as:

$$\begin{aligned} \mathcal{L}_{age}(G_+, G_-) = & \mathbb{E}_{k \sim p(k)} \mathbb{E}_{x_0 \sim p_{data}(x_0)} \left[\left| DEX(G_+(x_0, k)) - DEX(x_0) - k \right| \right. \\ & \left. + \left| DEX(G_-(x_0, k)) - DEX(x_0) + k \right| \right] \end{aligned}$$

$$\begin{aligned} \mathcal{L}_{cyc}(G_+, G_-) = & \mathbb{E}_{k \sim p(k)} \mathbb{E}_{x_0 \sim p_{data}(x_0)} \left[\left| \|G_+(\hat{x}_{-k}) - x_0\|_1 \right| \right. \\ & \left. + \left| \|G_-(\hat{x}_k, k) - x_0\|_1 \right| \right]. \end{aligned}$$

The full objective now becomes

$$\begin{aligned} \mathcal{L}(G_+, G_-, D) = & \mathcal{L}_{LSGAN}(G_+, G_-, D) + \lambda_1 \mathcal{L}_{age}(G_+, G_-) \\ & + \lambda_2 \mathcal{L}_{cyc}(G_+, G_-) + \lambda_3 \mathcal{L}_{ac}(G_+, G_-). \end{aligned}$$

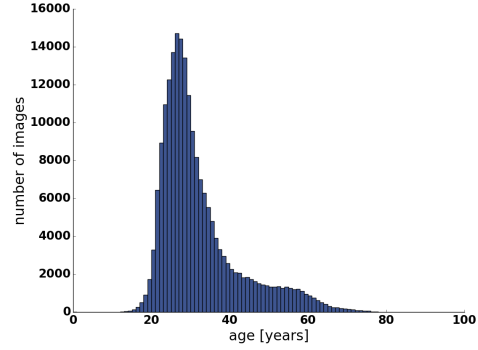


Figure 3. CelebA age distribution as estimated by DEX.

We discuss how we choose values for weight factors $\lambda_1, \lambda_2, \lambda_3$ in section 5.3. We now aim to solve the optimization problem:

$$G_+^*, G_-^* = \arg \min_{G_+, G_-} \min_D \mathcal{L}(G_+, G_-, D) \quad (10)$$

5. Experiments

In this section, we describe the experimental setup and then discuss the experimental results of the proposed methods: Group-GAN and FA-GAN. We evaluate the performance of these two methods on different levels of desired age change and show that the Group-GAN can achieve better performance for large age changes while the FA-GAN performs better for smaller changes. We then show how the combination of these two methods into a fusion model, F-GAN, performs robustly well for any desired age change and significantly better than a couple of existing methods. For this a user study was conducted.

5.1. Dataset

For our experiments we employ CelebA dataset [17] which consists of 202,599 face images of 10,177 different people. Of primary concern, of course, is the distribution of the data over age groups. As shown in figure 3 CelebA best covers the interval of 20-40 year olds but there is little data for persons under 20 and over 60 years. It should be noted that these labels are not the true ages but are obtained by running DEX (trained on IMDB-Wiki [23]) on the images. We use a version of this dataset that has been cropped using an off-the-shelf face detector by Mathias *et al.* [19]. In all our experiments, we discarded images with lower resolution than 256×256 and reserved 10% of the data for testing which leaves us with 129,852 training images.

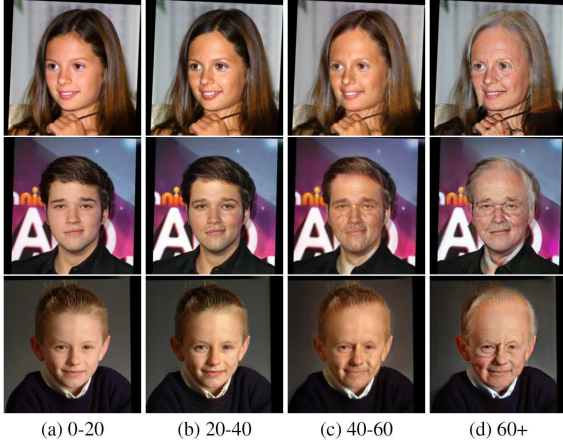


Figure 4. Group-GAN face aging results. The real input images are shown in column (a). The other columns show the output images after translating from group (a) to other age groups.

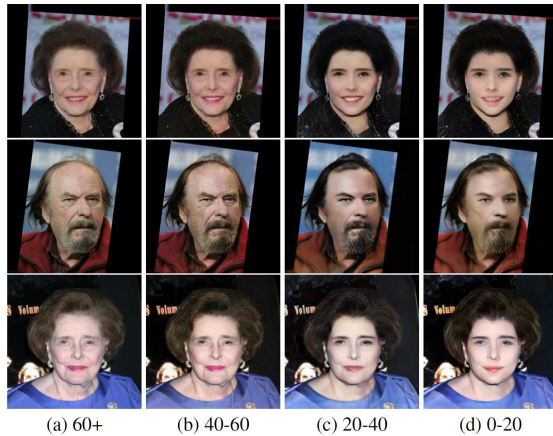


Figure 5. Group-GAN rejuvenation results. The real input images are shown in column (a). The other columns show the output images after translating from group 60+ to other age groups.

5.2. Group based CycleGAN method (Group-GAN)

We employ the CycleGAN framework [33], briefly reviewed in section 2.3, to translate images between age groups. We start by defining the following four age groups and partition the dataset accordingly:

- A) ages 00-20 C) ages 40-60
- B) ages 20-40 D) ages 60+

For each pair of groups we train a CycleGAN model to translate between these groups. The objective function is described in eq. 4 with $\lambda = 10$ as suggested in [33]. Each training process was run for about 24 hours on a GPU. The process is remarkably stable and only a few times we had to restart training due to failure such as mode collapse. When translating between well separated groups, such as A and

D, we observed that the system tends to invert background colors. To prevent this from happening we added a weak L1 loss between the input and output of each generator by adding to the objective

$$0.5(\mathbb{E}_{x \sim p_{data}(x)}[||G(x) - x||_1] + \mathbb{E}_{y \sim p_{data}(y)}[||F(y) - y||_1]) \quad (11)$$

The proposed Group-GAN achieves impressive aging results as shown in figure 4 for age progression and in figure 5 for rejuvenation.

However, the small differences between images in the first two columns in figures 4 and 5 raises the question whether this method is really effective for small age changes. To demonstrate that this is in fact a limitation of the Group-GAN we design a new experiment where we attempt to induce smaller aging effects, *e.g.* translating between age groups of 30-40 and 40-50 year olds. We now define the following, smaller age groups:

- A) ages 30-40 C) ages 50-60
- B) ages 40-50 D) ages 60-70

As before, we train translators between each age group. We also explore the possibility of only training translators between adjacent age groups and then to cascade them to produce smoother aging. From the results, displayed in figure 6, we do not observe any aging between groups A and B. Even between groups A and C the changes are minor but as expected we see radical changes when translating between A and D. In figure 6 we also compare the effect of directly translating between groups and cascading the generators, *i.e.* instead of going from A to C we first translate from A to B and the output is then translated from B to C. Comparing columns (d) and (f), we see that the aging effect is much weaker when we do the aging in smaller steps. This is not surprising since during the training process, the system tries to learn the differences between the two groups of images and the differences are much harder to detect if the age gap is small.

5.3. Face Aging GAN (FA-GAN)

The FA-GAN model we described in section 3.2 takes a number k as input and given an input image x_0 , the system $G(x_0, k)$ should produce a prediction of what that face will look like after k years, \hat{x}_k . This model addresses the issue with Group-GAN that it requires groups and many models. The main idea is to separate the discriminator into two parts, one that classifies the image as real or fake and one that determines the age. The loss function is (8) and the weight factors are experimentally chosen to be $\lambda_1 = 1/k$, $\lambda_2 = 10$ and $\lambda_3 = 1/k$. We implement the inputs to the model by adding two new color channels to each input image. One channel has k in each pixel and the other has $DEX(x_0)$

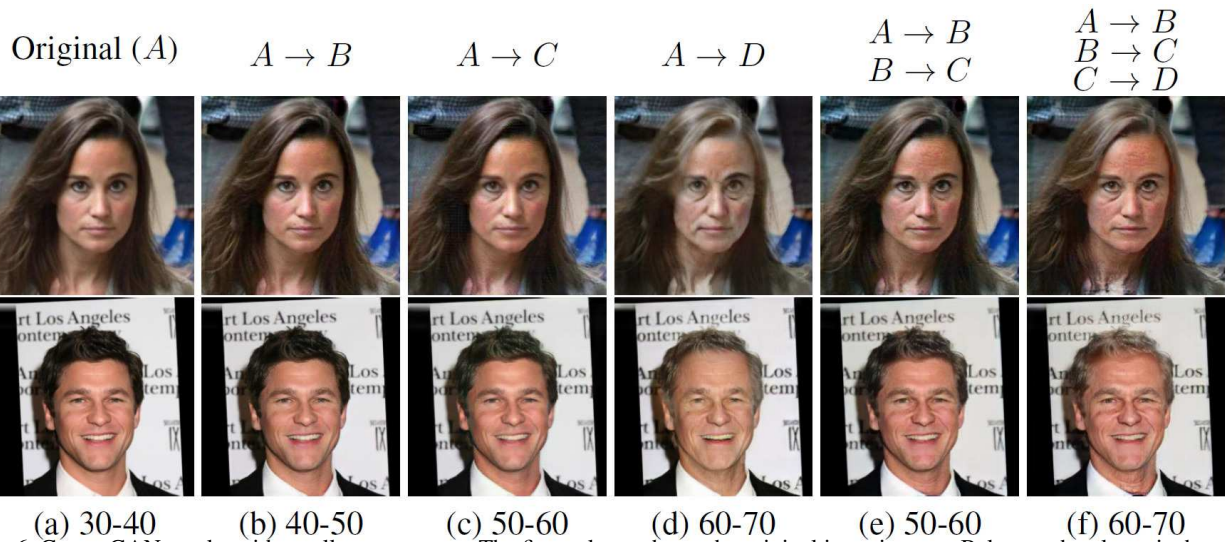


Figure 6. Group-GAN results with smaller age groups. The first column shows the original input images. Below each column is the target age interval and above each column are the transformations that have been applied to the original image. Age groups are: A: 30-40, B: 40-50, C: 50-60, D: 60-70. For example, we can compare the effect of translating an image from group A to group D directly (column d) and translating the image first from A to B, then from B to C and then from C to D (column f).

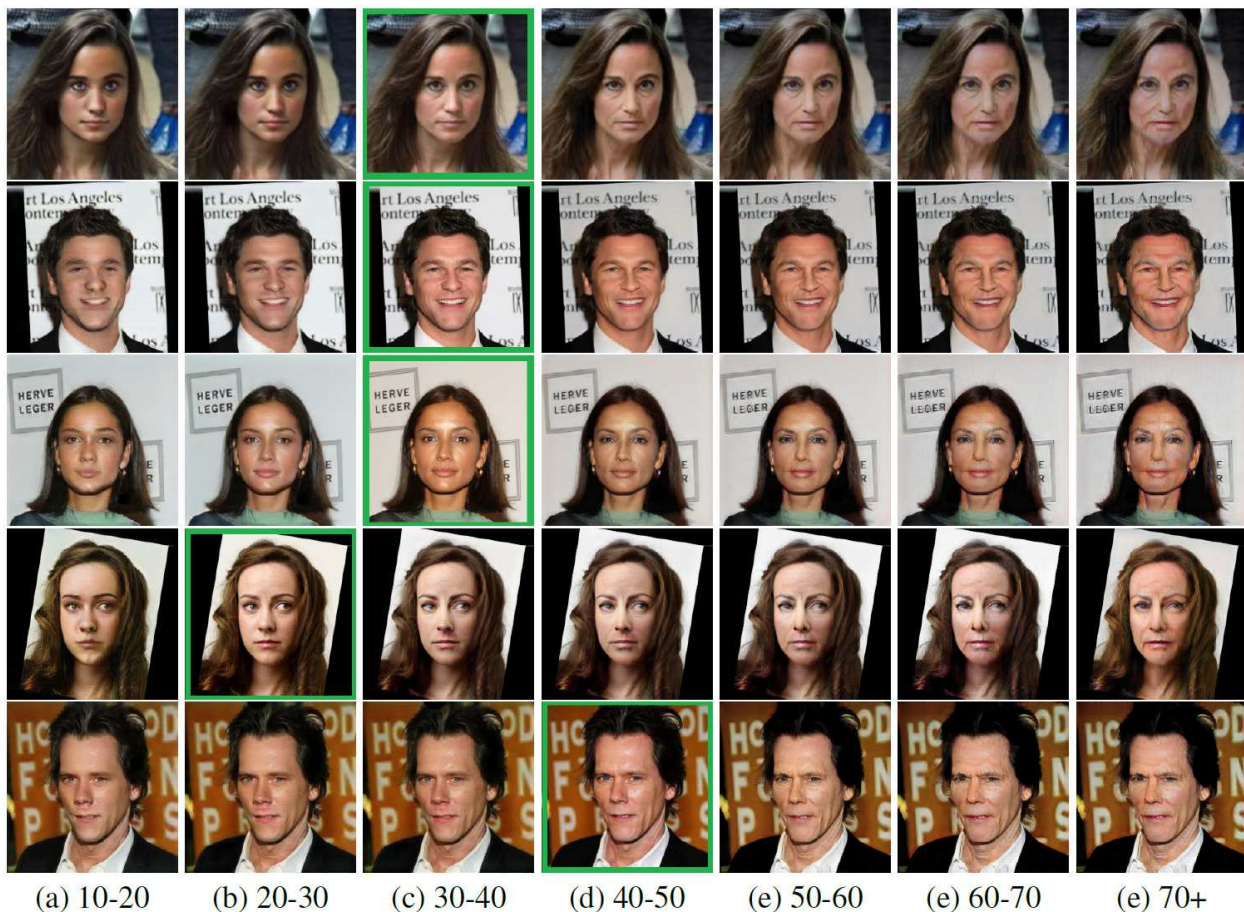


Figure 7. FA-GAN rejuvenation and aging results. The original images are shown in a green frame. To the left of the original image are rejuvenated images and to the right are aged images. In comparison with results in figure 6 we see that the FA-GAN method is mainly focused on the face while the Group-GAN method changes the hair color as well.

Question 5 of 10 or fewer:

On the left is an image of a person at age of 30. On the right is the same image with an age filter applied. Rate this age filter. The target age is 60+



Figure 8. Sample of survey question used for result evaluation.

in each pixel. We chose this implementation because of its simplicity, although it may be computationally inefficient. To increase the training stability we: (i) update the generators twice while we update the discriminator once, and (ii) sample k from $\{0, 10, 20, 30, 40\}$.

In figure 7, we show the results of FA-GAN on a few face images, both for age progression and regression. We see that for small desired age changes, we do get an aging effect but on the other hand, we observe poor quality for large effects. Note that in this setup, the generators get two types of information on how to produce good images: From D they learn how to be photorealistic and from DEX they learn to have the right characteristics to make DEX classify them with the right age. This means that any aging effects we now introduce to the images depend only on how DEX recognizes age. This could be the main limitation of this approach to aging. Both DEX and FA-GAN are trained using datasets (IMDB-Wiki and CelebA) with low coverage of children and old people, which is very likely limiting the quality of the outcomes for large desired age changes.

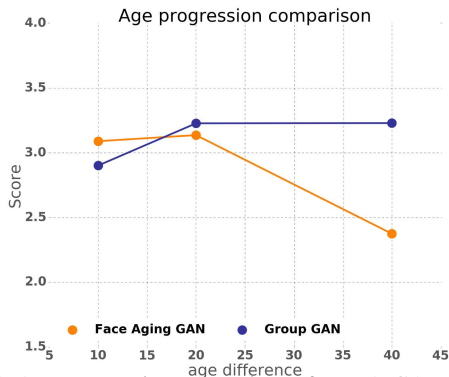


Figure 9. Age progression comparison of our FA-GAN and Group-GAN methods through survey mean opinion scores.

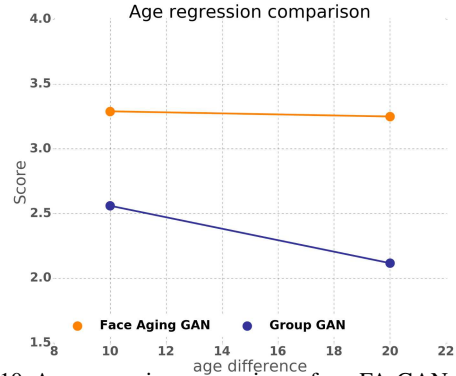


Figure 10. Age regression comparison of our FA-GAN and Group-GAN methods through survey mean opinion scores.

5.4. Quantitative results

To quantitatively compare the proposed Group-GAN and FA-GAN methods, we design a survey where participants are asked to rate image pairs. The image pairs consist of an original image and a corresponding age-progressed or age-regressed image. Participants are given the age of the original image and the target age that the method tries to achieve and gives the pair a score. The score is an integer from 1 to 5 (1 to 5 stars) where a score of 1 represents a very bad result and 5 a very good result. All images used in the survey are sampled at random from a test set that was not used for training. For FA-GAN we used 25 image pairs and for Group-GAN we used 60 image pairs. We also included 30 image pairs from prior work that we discuss in the next section. From the samples, we exclude people of ages 0-20 because our dataset does not have sufficient number of samples in that age interval. About 280 people rated 10 image pairs each, in total 2800 responses. A sample from the survey is shown in figure 8. The average score for different levels of aging for both these methods are shown in figures 9 for age progression and 10 for age regression. As expected, the results show that FA-GAN has better performance for small age changes and that Group-GAN performs well for larger age changes. We thus suggest a fusion of the methods, F-GAN, where for age changes of less than 20, we select FA-GAN, while for larger changes we select the Group-GAN.

5.5. Comparison with prior work

We now compare our methods, Group-GAN and FA-GAN, with a couple of recent state-of-the-art works. First in figure 11 we see our methods compared against a Coupled Dictionary Learning (CDL) face aging method of Shu *et al.* [26] and Recurrent Face Aging method of Wang *et al.* [30]. The first three columns of the figure are borrowed from [30]. Note that these test images are quite different from the CelebA images we used to train our

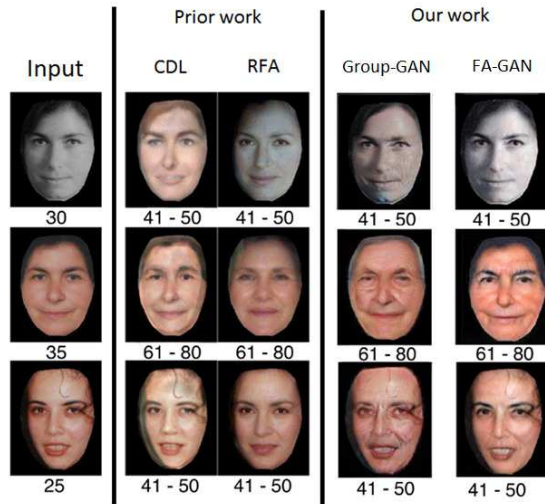


Figure 11. Comparison between prior work cited from [30] and our methods. The numbers show the real age for the inputs and the target age intervals for the outputs. Here we are comparing Coupled Dictionary Learning method [26], Recurrent Face Aging method [30] and our Group-GAN and FA-GAN methods.

models. However, we still observe that both our methods achieve competitive results. We also quantitatively compare with these methods through the survey described in section 5.4. We collected 30 image pairs from three prior works to compare with our F-GAN method. They are the two just mentioned, RFA and CDL, and another GAN-based method, the very recent Conditional Adversarial Autoencoder (CAAE) method of Zhang *et al.* [32]. Note that these works did not release their results so we used the results they picked to present in their papers. Our samples are picked at random and therefore also indicate the robustness of our method. The results are shown in figures 12 and 13 for age-progression and age-regression, respectively. The results show that our F-GAN method has superior performance on a large interval. Only the CAAE method reported results for rejuvenation, which our F-GAN method clearly outperforms. In tables 1 and 2 we show the average scores for each method in age progression and age regression, respectively.

Table 1. Age progression: survey mean opinion scores.

Method	Score \pm SE
F-GAN (ours)	3.17\pm0.016
RFA [30] (CVPR 2016)	2.79 \pm 0.027
CAAE [32] (CVPR 2017)	2.52 \pm 0.037
CDL [26] (ICCV 2015)	2.44 \pm 0.018

Table 2. Age regression: survey mean opinion scores.

Method	Score \pm SE
F-GAN (ours)	3.28\pm0.075
CAAE [32] (CVPR 2017)	3.03 \pm 0.011

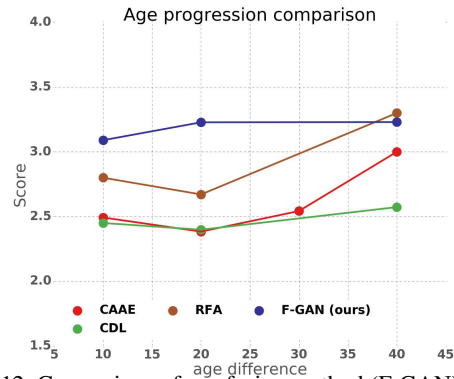


Figure 12. Comparison of our fusion method (F-GAN) with prior work. The figure shows the average scores of image pairs against the level of desired age change. The scores are based on a user study responses from over 280 people, which each rated 10 image pairs.

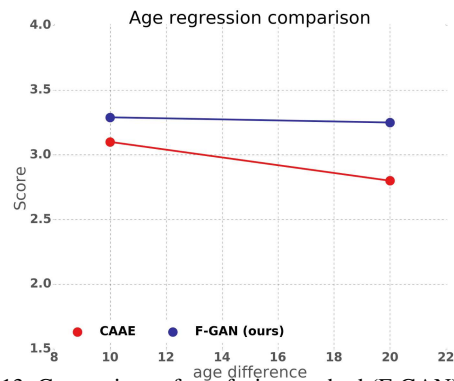


Figure 13. Comparison of our fusion method (F-GAN) with prior work in age regression. The figure shows the average scores of image pairs against the level of desired age change. The scores are based on a user study responses from over 280 people, which each rated 10 image pairs.

6. Conclusion

We have proposed two style transfer approaches to face aging: Group-GAN based on age groups and Face Aging GAN (FA-GAN) that does not require predefined age groups. We explored the performance of these methods both qualitatively and quantitatively through a survey and compared them with prior work. The key findings are that these methods perform differently based on the desired level of aging and they can be combined into a robust fusion model, F-GAN, that achieves better performance on average than prior state-of-the-art. Future work will involve using a dataset with a better coverage of all ages, in particular children which are not significantly present in CelebA.

Acknowledgements

This work was supported by ETH Zurich General Fund (OK) and by NVIDIA through a GPU hardware grant.

References

- [1] pytorch. <https://github.com/pytorch>. 3
- [2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face Aging With Conditional Generative Adversarial Networks. *ArXiv e-prints*, Feb. 2017. 1, 2
- [3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017. 2
- [4] Y. Cao, Z. Zhou, W. Zhang, and Y. Yu. Unsupervised Diverse Colorization via Generative Adversarial Networks. *ArXiv e-prints*, Feb. 2017. 2
- [5] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, Nov 2010. 1
- [6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *ArXiv e-prints*, June 2014. 1, 2, 3
- [7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30*. 2
- [8] S. M. S. I. Kemelmacher-Shlizerman, S. Suwajanakorn. Illumination-aware age progression. In *CVPR*, 2014. 1, 2
- [9] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. WESPE: weakly supervised photo enhancer for digital cameras. *CoRR*, abs/1709.01118, 2017. 2
- [10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *ArXiv e-prints*, Nov. 2016. 2, 4
- [11] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *ArXiv e-prints*, Mar. 2016. 4
- [12] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *CoRR*, abs/1710.10196, 2017. 2
- [13] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, Apr 2002. 1
- [14] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *ArXiv e-prints*, Sept. 2016. 2
- [15] C. Li and M. Wand. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. *ArXiv e-prints*, Apr. 2016. 4
- [16] M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 700–708. Curran Associates, Inc., 2017. 2
- [17] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015. 4
- [18] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley. Least Squares Generative Adversarial Networks. *ArXiv e-prints*, Nov. 2016. 2, 4
- [19] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. *Face Detection without Bells and Whistles*, pages 720–735. Springer International Publishing, Cham, 2014. 4
- [20] A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv e-prints*, Nov. 2015. 2
- [21] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1060–1069, New York, New York, USA, 20–22 Jun 2016. PMLR. 2
- [22] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, December 2015. 3
- [23] R. Rothe, R. Timofte, and L. Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*, July 2016. 3, 4
- [24] A. Sage, E. Agustsson, R. Timofte, and L. Van Gool. Logo synthesis and manipulation with clustered generative adversarial networks. *CoRR*, abs/1712.04407, 2017. 2
- [25] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen. Improved techniques for training gans. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2234–2242. Curran Associates, Inc., 2016. 2
- [26] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan. Personalized age progression with aging dictionary. In *ICCV*, 2015. 7, 8
- [27] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv e-prints*, Sept. 2014. 3
- [28] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):385–401, Mar. 2010. 1
- [29] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):385–401, Mar. 2010. 2
- [30] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2378–2386, June 2016. 1, 2, 7, 8
- [31] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic Image Inpainting with Deep Generative Models. *ArXiv e-prints*, July 2016. 2
- [32] Z. Zhang, Y. Song, and H. Qi. Age Progression/Regression by Conditional Adversarial Autoencoder. *CVPR 2017*, 2017. 1, 2, 8
- [33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *ArXiv e-prints*, Mar. 2017. 2, 3, 4, 5