FACE AGING AS IMAGE-TO-IMAGE TRANSLATION USING SHARED-LATENT SPACE GENERATIVE ADVERSARIAL NETWORKS

Evangelia Pantraki and Constantine Kotropoulos

Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki 54124, GREECE
epantrak@csd.auth.gr, costas@aiia.csd.auth.gr

ABSTRACT

Here, a novel approach is proposed to generate age progression (i.e., future looks) and regression (i.e., previous looks) of persons based on their face images. The proposed method addresses face aging as an unsupervised image-to-image translation problem where the goal is to translate a face image belonging to an age class to an image of a different age class. To address this problem, we resort to adversarial training and extend the UNsupervised Image-to-image Translation (UNIT) framework to multi-domain image-to-image translation, since several age classes are considered. Due to the shared-latent space constraint of UNIT, the faces belonging to each age class/domain are forced to be mapped to a shared-latent representation. Low-level features are used to perform the transitions between the domains and to generate age progressed/regressed images. In addition, the most personal and abstract features of faces are preserved. The proposed Aging-UNIT framework is compared to state-of-the-art techniques and the ground truth. Promising results are demonstrated, which are attributed to the ability of the proposed method to capture the subtle aging transitions.

Index Terms— face aging, adversarial training, latent space, image-to-image-translation

1. INTRODUCTION

Aging is a long-term process that gradually affects human face. The effects of aging on human face include changes in both hard and soft facial tissues, such as the skeletal structure, skin, facial musculature, and lines [1]. Aging patterns that are common among individuals are the alteration of skin texture and the formation of lines and wrinkles around eyes and/or mouth. Although face aging effects are distinct across long periods of time, the short-term effects of aging are inconspicuous and therefore, difficult to model. Nevertheless, knowing these aging patterns is necessary in order to predict how a person's face might look in the future (face progression) or how a person's face might have looked when he/she was younger (face regression or rejuvenation). Predicting future and previous face looks is extremely important for age-invariant face recognition. Moreover, it can significantly assist the search for missing or wanted persons. In order to achieve realistic results, face age progression and regression need to fulfil the crucial constraint of preserving the personalized features of the face (i.e., personality). The great challenge is to estimate face aging effects, while maintaining the subject-specific characteristics of each face.

Face aging has attracted great research interest over the years. Face progression approaches can be divided in two main categories, model-based and prototype-based. Model-based approaches model the biological patterns of human aging, using parametric or non-parametric learning. In [2], the face aging process is modeled with a hierarchical graph whose nodes correspond to facial parts that are crucial for age perception, such as eyes, mouth, and nose. In [3], a face aging simulation model is presented. According to this model, the face is represented by three layers, namely global, local, and texture layer, which are fused to simulate the facial aging process.

Prototype-based approaches divide training data to different age classes and investigate how the faces images differ among age classes. These differences actually represent the aging patterns, which can be subsequently used to make transitions among classes and generate age progressed or rejuvenated face images. In [4], the aging patterns across different age groups form a dictionary for the aging process. In [5], a deep learning-based approach for face aging is presented that is based on a Recurrent Neural Network (RNN). The age progressed face images are generated by referring to the memory of the previous faces, making the age transitions across different ages groups smoother.

The main limitation of the model-based approaches is that they require images of the same person over a long age span in order to effectively model the aging procedure. On the other hand, the prototype-based approaches still need paired samples of face images in order to capture the transitions across age classes.

The advent of Generative Adversarial Networks (GANs) [6] has introduced an adversarial framework for training generative models and had an astonishing effect on the quality of generated images. In [7], a Conditional Adversarial AutoEncoder (CAAE) is proposed for age progression and regression based on a face manifold. Traversing on the manifold generates transitions across age classes, while personality is preserved by mapping the faces to a latent space using convolutional encoders before projecting to the face manifold. In [8], age conditional GANs are also utilized for age progression, along with a latent vector optimization approach that aims to reconstruct the input face and preserve the person's identity. The approach presented in [9] utilizes a pyramid of GANs for face aging that incorporates face verification and age estimation techniques. Other GAN-based approaches for face aging are presented in [10, 11, 12, 13], where effort has been devoted in the challenging task of identity preservation.

In this paper, we investigate the age progression/regression problem from the perspective of generative modeling. Here, the face al-

This research has been financially supported by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) (Scholarship Code: 81).

terations related to aging are addressed as an image-to-image translation problem, i.e., as a problem of mapping an image belonging to one domain to a corresponding image in another domain. Since each age class is regarded as a single domain, the mappings across domains constitute the transitions across age classes. In order to learn the mappings/transitions across domains, we utilize the UNsupervised Image-to-image Translation (UNIT) network proposed in [14] and extend it to multiple domains. The proposed Aging-UNIT framework is presented in detail in Section 2. In Section 3, the experimental evaluation of the proposed framework is discussed. Finally, Section 4 concludes the paper and recommends future work.

2. PROPOSED FRAMEWORK

Here, we present the pipeline of the proposed Aging-UNIT framework for generating realistic face images belonging to different age classes. The proposed Aging-UNIT emerged from the novel idea of considering face aging as an image-to-image translation problem. To implement image-to-image translations, we adopt the UNIT framework presented in [14]. The UNIT framework [14], as well as its predecessor, the Coupled GAN [15] extend GANs to joint image distribution learning tasks.

When performing image-to-image translation, the goal is to learn a joint distribution of images in different domains by using images from the marginal distributions in individual domains. Since there is an infinite set of joint distributions that can produce the given marginal distributions, finding the joint distribution is not feasible unless a key assumption is made. To this end, the UNIT framework makes the shared-latent space assumption, according to which, a pair of corresponding images in different domains can be mapped to the same latent representation in a shared-latent space. Here, we propose the Aging-UNIT, which extends the UNIT framework to multiple domains and applies it to the face age progression/regression problem.

Motivation: The UNIT framework employs GANs [6, 15] and Variational Autoencoders (VAEs) [16, 17] and is based on two vital assumptions: the shared-latent space assumption and the cycle-consistency assumption. The key motivation of UNIT [14] is to exploit the hierarchical way that deep neural networks learn feature representations. By enforcing a weight sharing constraint on the layers that bear the most high-level semantic information, the most abstract and personalized features are encoded/decoded in the same way for each domain. Subsequently, in order to learn transitions among domains, this shared representation across domains is mapped to images in individual domains in an attempt to fool the domain discriminators.

Sub-networks: According to our approach, each age class corresponds to a single domain. The Aging-UNIT framework consists of three sub-networks for each domain: an encoder $\mathbf{E}$, a generator $\mathbf{G}$ and an adversarial discriminator $\mathbf{D}$. If N age classes are considered, the framework consists of N encoders $\mathbf{E}_n$, N generators $\mathbf{G}_n$, and N adversarial discriminators $\mathbf{D}_n$, $n = 1, 2, \dots, N$. The encoders $\mathbf{E}_n$ and generators $\mathbf{G}_n$ are represented by Convolutional Neural Networks (CNNs). The framework learns the bidirectional translations among all N domains simultaneously.

VAE: For each domain $\mathbf{X}_n$, the encoder-generator pair $\{\mathbf{E}_n, \mathbf{G}_n\}$ constitutes a VAE for the $\mathbf{X}_n$ domain, termed VAE_n. The VAE_n firstly maps the input image $x_n \in \mathbf{X}_n$ to a latent representation z via the encoder $\mathbf{E}_n$. Subsequently, the generator $\mathbf{G}_n$ of VAE_n decodes a random-perturbed version of the latent code z provided by the encoder $\mathbf{E}_n$, in order to reconstruct the input image.

Shared-latent space assumption: The shared-latent space assumption is necessary in order to estimate the joint distribution of samples drawn from different domains in an unsupervised way. This assumption is vital to the Aging-UNIT framework, since images across age classes should share the same high-level representations in order to maintain personalized features. This assumption is implemented by a weight sharing constraint applied to both encoders $\mathbf{E}_n$ and generators $\mathbf{G}_n$.

Since each domain encoder $\mathbf{E}_n$ is implemented by CNNs, the network learns more complex and abstract features as the number of layers increases. If the weights of the last few layers are shared among $\mathbf{E}_n$, $n = 1, 2, \dots, N$, then each domain encoder learns to encode domain images to the same high-level semantic information. On the other hand, the generators $\mathbf{G}_n$, $n = 1, 2, \dots, N$ do the inverse work of decoding the latent representations back to images. Therefore, since the latent representations capture the most abstract high-level information for each image, it is the weights of their first layers that should be tied. Due to adversarial training, the encoders $\mathbf{E}_n$ learn to encode tuples of corresponding images $x_n, n = 1, 2, \dots, N$ to a common latent representation z, while generators $\mathbf{G}_n$ learn to decode the shared-latent representation z to tuples of corresponding images.

In order to make image-to-image translations, the layers that decode low-level details map the shared-latent representation z to images in individual domains. For example, an image $x_1 \in \mathbf{X}_1$ that is encoded to $z_1 \sim q_1(z_1|x_1)$ can be translated to domain $\mathbf{X}_2$ through applying $\widetilde{x}_1^{1\to 2} = \mathbf{G}_2(z_1 \sim q_1(z_1|x_1))$. Correspondingly, the image x_1 can be reconstructed by $\widetilde{x}_1^{1\to 1} = \mathbf{G}_1(z_1 \sim q_1(z_1|x_1))$. If N domains are considered, the transitions among all pairs of individual domains $\mathbf{X}_n$ are implemented accordingly.

Cycle-consistency constraint: According to the Aging-UNIT framework, each image x_k that belongs to domain $\mathbf{X}_k$ is translated to all other domains $\mathbf{X}_l, l = 1, \dots, N, l \neq k$. The cycle-consistency constraint hypothesizes the existence of a cycle-consistency mapping so that each of these translated images, if mapped back to domain $\mathbf{X}_k$, can reliably reconstruct input image x_k. In other words, the initial images $x_k \in \mathbf{X}_k$ are translated to domains $\mathbf{X}_l, l = 1, \dots, N,\ l \neq k$ and yield the translated images $\widetilde{x}_k^{k\to l}$. The translated images $\widetilde{x}_k^{k\to l}$ are translated back to domain $\mathbf{X}_k$ and yield images $\widetilde{x}_k^{k\to l\to k}$. The twice translated images $\widetilde{x}_k^{k\to l\to k}$ that were generated by the translation cycle between domains $\mathbf{X}_k \to \mathbf{X}_{l\neq k} \to \mathbf{X}_k$ should reconstruct initial images $x_k \in \mathbf{X}_k$. With this constraint imposed on the UNIT framework, the ill-posed unsupervised image-to-image translation problem is further regularized.

GANs: A GAN [6] consists of two models that are trained simultaneously: a generative model $\mathbf{G}$ and a discriminative model $\mathbf{D}$. The generative model $\mathbf{G}$ tries to learn the distribution of the training image samples and synthesize images resembling the original ones. The objective of the discriminative model $\mathbf{D}$ is to distinguish the original images from the generated samples. $\mathbf{G}$ and $\mathbf{D}$ compete with each other, using a min-max game

$$\min_G \max_D \mathbb{E}_{x\sim p_x(x)}[\log(\mathbf{D}(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-\mathbf{D}(\mathbf{G}(z))] \quad (1)$$

where z denotes a vector randomly sampled from certain distribution $p_z(z)$ (e.g., Gaussian or uniform), and the data distribution is $p_x(x)$, i.e., the training data $x \sim p_x$. The two models $\mathbf{G}$ and $\mathbf{D}$ are trained alternatively. Goodfellow et al. [6] showed that if enough training iterations are completed and enough capacity is given to $\mathbf{G}$ and $\mathbf{D}$, the distribution of the generative model $p_G(z)$ converges to the real data distribution $p_x(x)$. In other words, from a random vector $\mathbf{z}$, the generative network $\mathbf{G}$ can synthesize an image $\widetilde{x} = \mathbf{G}(z)$ that resembles one that is drawn from the true distribution, $p_x(x)$.

In the Aging-UNIT framework, each domain has each own generator $\mathbf{G}_n$ and discriminator $\mathbf{D}_n, n = 1, \dots, N$. In each GAN_n, generator $\mathbf{G}_n$ tries to fool discriminator $\mathbf{D}_n$ by generating images that reliably resemble the original images of age class $\mathbf{X}_n$, while discriminator $\mathbf{D}_n$ tries to understand which images actually belong to age class $\mathbf{X}_n$.

Objective function: The adversarial training of Eq. (1) can be regarded as a two player zero-sum game, where the first player consists of the team of encoders and generators and the second player comprises the team of adversarial discriminators. In addition to defeating the second player, the first player has to minimize the VAE loss and the cycle-consistency loss. The objective function aims at jointly solving the learning problems of VAE and GAN subject to the cycle-consistency constraint for each domain. The objective function for domain $\mathbf{X}_k$ is given in Eq. (2) for translations from domain $\mathbf{X}_k$ to all other domains $\mathbf{X}_l,\ l = 1, \dots, N,\ l \neq k$:

$$\begin{aligned} \min_{\mathbf{E}_k, \mathbf{E}_l, \mathbf{G}_k, \mathbf{G}_l} \max_{\mathbf{D}_k, \mathbf{D}_l} \Big\{ &\mathcal{L}_{VAE_k}(\mathbf{E}_k, \mathbf{G}_k) + \mathcal{L}_{GAN_k}(\mathbf{E}_l, \mathbf{G}_k, \mathbf{D}_k) \\ &+ \mathcal{L}_{CC_k}(\mathbf{E}_k, \mathbf{G}_k, \mathbf{E}_l, \mathbf{G}_l), \\ &k \in 1, \dots, N,\ l = 1, \dots, N,\ l \neq k \Big\} \end{aligned} \tag{2}$$

$\mathcal{L}_{VAE_k}$ in Eq. (2) corresponds to the objective function of VAE_k training. VAE_k training for domain $\mathbf{X}_k$ aims at minimizing the loss function

$$\begin{aligned} \mathcal{L}_{VAE_k}(\mathbf{E}_k, \mathbf{G}_k) &= \lambda_0\ KL\ (q_k(z_k|x_k)\ ||\ p_p(z)) \\ &- \lambda_1\ \mathbb{E}_{z_k \sim q_k(z_k|x_k)} \big[\log p_{G_k}(\widetilde{x}_k|z_k)\big] \\ &= \lambda_0\ KL\ (q_k(z_k|x_k)\ ||\ p_p(z)) \\ &+ \lambda_1\ \|x_k - \widetilde{x}_k^{k \to k}\|_{\ell_1}, \quad k \in 1, \dots, N \end{aligned} \tag{3}$$

The Kullback-Leibler (KL) divergence penalizes any deviation of the distribution q_k of the latent code from the prior zero mean Gaussian distribution $p_p(z) = \mathcal{N}(z|0, I)$. Distributions p_{G_n} for $n = 1, \dots, N$ are modeled using Laplacian distributions. Hence, minimizing the negative log-likelihood term is equivalent to minimizing the absolute distance between image x_k and the reconstructed image $\widetilde{x}_k^{k \to k}$. Hyper-parameters λ_0 and λ_1 control the weights of each term of the $\mathcal{L}_{VAE_k}$ objective function.

$\mathcal{L}_{GAN_k}$ in Eq. (2) penalizes the image translation stream from domains $\mathbf{X}_l,\ l = 1, \dots, N,\ l \neq k$ to domain $\mathbf{X}_k$. The objective function of GAN training is conditional to age class and ensures that the images translated to an age class $\mathbf{X}_k$ resemble the images that truly belong to that age class. The conditional objective function of GAN training for domain $\mathbf{X}_k$ is given by

$$\begin{aligned} \mathcal{L}_{GAN_k}(\mathbf{E}_l, \mathbf{G}_k, \mathbf{D}_k) &= \lambda_2\ \mathbb{E}_{x_k \sim p_{x_k}} \big[\log \mathbf{D}_k(x_k)\big] \\ &+ \lambda_2\ \mathbb{E}_{z_l \sim q_l(z_l|x_l)} \big[\log\big(1 - \mathbf{D}_k(\mathbf{G}_k(z_l))\big)\big], \\ &k \in 1, \dots, N,\ l = 1, \dots, N,\ l \neq k \end{aligned} \tag{4}$$

Hyper-parameter λ_2 controls the impact of generative loss for domain $\mathbf{X}_k$. In our experiments, the same value of hyper-parameter λ_2 was set for all domains $\mathbf{X}_n, n = 1, \dots, N$.

$\mathcal{L}_{CC_k}$ in Eq. (2) penalizes the cycle-consistency loss for domain $\mathbf{X}_k$. For the translation cycle $\mathbf{X}_k \to \mathbf{X}_l \to \mathbf{X}_k, l = 1, \dots, N,\ l \neq k$, the objective function for the cycle-consistency constraint is given in Eq. (5). The KL divergence terms penalize any deviation of the distribution q_k of the latent codes of the translation stream $\mathbf{X}_k \to \mathbf{X}_l$ and the distribution q_l of the translation stream $\mathbf{X}_l \to \mathbf{X}_k$ from the prior distribution $p_p(z)$. The negative log-likelihood objective term forces the twice translated image $\widetilde{x}_k^{k \to l \to k}$ to resemble the input image x_k. Hyper-parameters λ_3 and λ_4 control the weights of the two objective terms of the cycle-consistency constraint.

$$\begin{aligned} \mathcal{L}_{CC_k}(\mathbf{E}_k, \mathbf{G}_k, \mathbf{E}_l, \mathbf{G}_l) &= \lambda_3\ KL\ (q_k(z_k|x_k)\ ||\ p_p(z)) \\ &+ \lambda_3\ KL\ (q_l(z_l|\widetilde{x}_k^{k \to l})\ ||\ p_p(z)) \\ &- \lambda_4\ \mathbb{E}_{z_l \sim q_l(z_l|\widetilde{x}_k^{k \to l})} \big[\log p_{G_k}(x_k|z_l)\big], \\ &k \in 1, \dots, N,\ l = 1, \dots, N,\ l \neq k \end{aligned} \tag{5}$$

Benefits: The benefits of the proposed Aging-UNIT framework for age progression can be summarized in four major aspects. First, the proposed novel approach reduces the burden of investigations for the exact aging patterns across age classes. By performing age progression and regression via image-to-image translation, the Aging-UNIT is capable of learning the effects of aging on human face in an unsupervised manner. Secondly, personality is preserved across age transitions due to the shared-latent space representation. The shared-latent space is enforced by imposing a weight sharing constraint in both the generative and the discriminative model and no further regularization has to be imposed to the GAN objective function to preserve personality. Thirdly, the proposed framework does not require paired samples in order to learn aging transitions. It can learn a joint distribution of multi-domain images without existence of corresponding images in different domains in the training set. Only a set of images drawn separately from the marginal distributions of the individual domains is required. Fourthly, the Aging-UNIT framework learns transitions in both age directions and accomplishes realistic results in both age progression and rejuvenation simultaneously.

3. EXPERIMENTS

3.1. Datasets

In order to train our framework, we first create our training set by collecting a subset of images from the CACD [18] and the UTKFace [7] datasets. Similar to [10], we define 7 age classes: 0-10, 11-18, 19-29, 30-39, 40-49, 50-59, and 60+ years old. The oldest person belonging to the last age class is 80 years old. Effort has been devoted to create a balanced training dataset across age classes with respect to gender. In total, the training dataset consists of 21,267 face images. In order to evaluate the proposed Aging-UNIT framework, the FGNET dataset [19] is used for testing. The FGNET dataset consists of 1002 images of 82 subjects aging from 0 to 69.

3.2. Experimental evaluation

To evaluate whether the proposed Aging-UNIT framework generates photo-realistic face aging results, we test our method on the FGNET dataset, similar to [7]. The effectiveness of the proposed method in capturing the effects of aging on human faces is demonstrated in Figure 1, where seven input face images from the FGNET dataset and the generated faces for the seven age classes are presented. The red boxes indicate the generated images that belong to the ground truth age class of each input face. It is clear that as age progresses, the skin texture alternates, while subtle changes to cheeks, eyes, and mouth simulate face aging. Both progression and rejuvenation yield realistic results, while the aging effects preserve personalized features. It should be noted that although no gender information is included, the Aging-UNIT framework succeeds to capture abstract face aging effects appropriate to the gender of the depicted person.

Fig. 1: Age progression and regression results admitted by the Aging-UNIT framework for images of the FGNET dataset. The first column depicts input faces, and the rest columns depict the admitted results from both age progression and regression. The red boxes indicate the generated images that belong to the ground truth age class of each input image.

Comparison to ground truth: In Figure 2, we compare the face images generated by the proposed Aging-UNIT framework to the ground truth images of the FGNET dataset. More specifically, we present generated samples of FGNET images translated to different age classes and compare them to the ground truth images of the persons at that specific age interval. The proposed framework demonstrates appealing and realistic results for both age progression and rejuvenation tasks.

Comparison to prior works: The performance of the proposed face aging framework in producing realistic faces at different age intervals is also compared to prior works. Our method is compared to [5, 7, 10, 12] for age progression. In order to compare our results to the face aging images generated by the aforementioned approaches, we use the same input images and perform age progression. The comparative demonstration is presented in Figure 3. The proposed Aging-UNIT demonstrates smooth age transitions and admits more realistic results when compared to generated images admitted by other methods, e.g., for input image in the first row of Figure 3 the image generated by the Aging-UNIT framework is more realistic when compared to the generated image by method [12]. Aging-UNIT succeeds remarkably in maintaining personality, since the unique characteristics of each face that make the person recognizable remain unaltered by aging effects, especially when compared to the generated images of the method in [5] (fourth column of Figure 3).

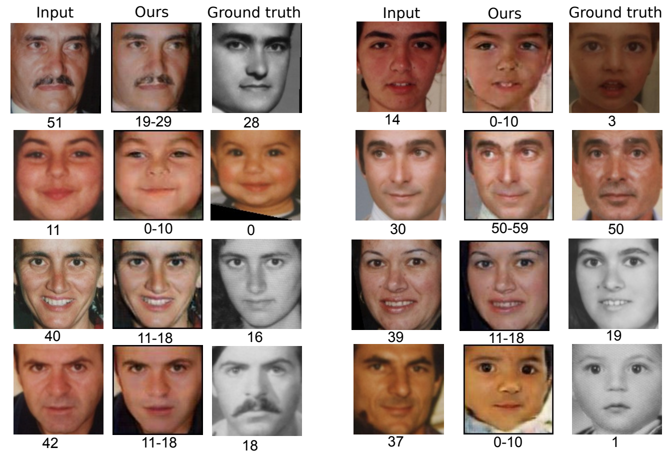

Fig. 2: Comparison to the ground truth images of the FGNET dataset.

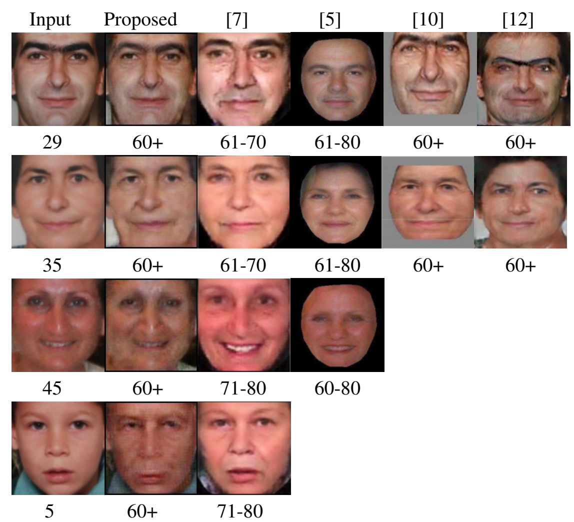

Fig. 3: Comparison of age transitions admitted by the Aging-UNIT framework and prior works evaluated on images of the FGNET dataset.

4. CONCLUSION AND FUTURE WORK

In this paper, face aging is addressed as an unsupervised image-to-image translation problem. To our knowledge, this is the first time that aging is considered as a problem of translation between images. Our goal is to achieve age progression or regression while translating an image to another that represents an older/younger person. The most important condition when translating among images is to maintain personality and preserve personalized features of face. To achieve this, we adopt the UNIT image-to-image translation framework [14] that is based on a shared-latent space constraint among translation domains. This constraint forces the images to share encoded representations and via adversarial training leads to the generation of corresponding images in multiple domains. Experimental results demonstrate the effectiveness of the proposed Aging-UNIT framework to translate face images to different age classes while producing realistic results and preserving personality. Future work will focus on further regularizing the generative training in order to reduce blurriness and improve the quality of the generated images. Moreover, we aim to investigate how translations between distant age classes could be facilitated, since they perform the most drastic aging effects.

5. REFERENCES

[1] M. S. Zimbler, M. S. Kokoska, and J. R. Thomas, "Anatomy and pathophysiology of facial aging," *Facial plastic surgery clinics of North America*, vol. 9, no. 2, pp. 179–187, May 2001.

[2] J. Suo, S. C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 385–401, March 2010.

[3] Y. Li and Y. Li, "Face aging effect simulation model based on multilayer representation and shearlet transform," *Journal of Electronic Imaging*, vol. 26, no. 5, 2017.

[4] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized age progression with aging dictionary," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 3970–3978.

[5] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, 2016, pp. 2378–2386.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[7] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. of Computer Vision and Pattern Recognition Conf.* IEEE, 2017, vol. 2, pp. 2672–2680.

[8] G. Antipov, M. Baccouche, and J. L. Dugelay, "Face aging with conditional generative adversarial networks," in *arXiv preprint arXiv:1702.01983*, 2017.

[9] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," *arXiv preprint arXiv:1711.10352*, 2017.

[10] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan, "Face aging with contextual generative adversarial nets," in *Proc. 2017 ACM on Multimedia Conf.* ACM, 2017, pp. 82–90.

[11] H. Zhu, Q. Zhou, J. Zhang, and J. Z. Wang, "Facial aging and rejuvenation by conditional multi-adversarial autoencoder with ordinal regression," *arXiv preprint arXiv:1804.02740*, 2018.

[12] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, "Global and local consistent age generative adversarial networks," *arXiv preprint arXiv:1801.08390*, 2018.

[13] S. Zhou, W. Zhao, J. Feng, H. Lai, Y. Pan, J. Yin, and S. Yan, "Personalized and occupational-aware age progression by generative adversarial networks," *arXiv preprint arXiv:1711.09368*, 2017.

[14] M. Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 700–708.

[15] M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 469–477.

[16] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[17] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and variational inference in deep latent gaussian models," in *Proc. Int. Conf. Machine Learning*, 2014, vol. 2.

[18] B. C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proc. European Conf. on Computer Vision (ECCV)*, 2014.

[19] 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*, vol. 24, no. 4, pp. 442–455, 2002.