

# LEVERAGING IMAGE-TO-IMAGE TRANSLATION GENERATIVE ADVERSARIAL NETWORKS FOR FACE AGING

*Evangelia Pantraki\**, *Constantine Kotropoulos*

Department of Informatics  
Aristotle University of Thessaloniki  
Thessaloniki 54124, Greece  
epantrak@csd.auth.gr, costas@aiaa.csd.auth.gr

*Andreas Lanitis*

Department of Multimedia and Graphic Arts  
Cyprus University of Technology  
Limassol 3040, Cyprus  
andreas.lanitis@cut.ac.cy

## ABSTRACT

Here, face images of a specific age class are translated to images of different age classes in an unsupervised manner that enables training on independent sets of images for each age class. In order to learn pairwise translations between age classes, we adopt the UNsupervised Image-to-image Translation framework that employs Variational AutoEncoders and Generative Adversarial Networks. By mapping face images of different age classes to shared latent representations, the most personalized and abstract facial characteristics are preserved. To effectively diffuse age class information, a pyramid of local, neighbour, and global encoders is employed so that the latent representations progressively cover an increased age range. The proposed framework is applied to the FGNET aging database and compared to state-of-the-art techniques and the ground truth. Appealing experimental results demonstrate the ability of the proposed method to efficiently capture both intense and subtle aging effects.

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

## 1. INTRODUCTION

Predicting the effects of aging on human face is challenging due to the progressive, accumulative, and multi-factorial nature of the aging process. In order to reliably predict face age progression (i.e., future looks) and regression (i.e., previous looks), it is vital to maintain personality, i.e., the unique characteristics of each face that make the person recognizable. Face age progression and regression is very useful for cross-age face recognition, can significantly assist the search for missing or wanted persons, or be exploited for entertainment related applications.

Great research effort has been devoted to face aging with impressive results. The face aging framework presented in [1] comprises of Recurrent Neural Networks (RNNs), which are trained on the aging transformations between adjacent age classes. Since RNNs memorize previous states, the aged faces are gradually generated and face identity is preserved. The limitation of this approach is that it requires face images of the same person for adjacent age classes in order to train the RNNs. In [2], a model-based approach for face aging is proposed that represents a face using three layers, namely global, local, and texture layer. Face aging information from individual layers is fused in order to effectively simulate aging.

\*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).

Age progression/regression can be perceived as the generative task of creating face images that belong to different age ranges. Exploiting the tremendous capabilities of Generative Adversarial Networks (GANs) [3] is nowadays the state-of-the-art in face aging. In [4], a Conditional Adversarial AutoEncoder (CAAE) is employed for age progression/regression based on a face manifold. In order to preserve personality, CAAE firstly maps face images to latent representations, which are subsequently projected to the face manifold conditional on age. Traversing on the manifold generates transitions across age classes and produces age progressed or rejuvenated face images. In [5], age conditional GANs are utilized for age progression along with a latent vector optimization constraint. In order to preserve personality, a face recognition neural network is incorporated into the training process. The approach in [6] employs GANs for face aging and incorporates face verification and age estimation techniques into the objective function in order to enhance aging effects and preserve personality. Moreover, a pyramid of facial feature representations estimated at multiple scales is fed to the discriminator in order to better capture subtle aging transformations. Other GAN-based approaches for face aging are presented in [7, 8, 9, 10], where effort has been devoted to identity preservation.

In this paper, we investigate the age progression/regression problem from the perspective of generative modelling. The main contributions are as follows:

1. A novel framework is built by addressing face aging as an unsupervised image-to-image translation problem, therefore eliminating the need to learn exact aging patterns across age classes as well as prerequisite paired face images of the same person at different age classes.
2. Personality is preserved across age transitions by forcing face images belonging to different age classes to be mapped to shared latent representations.
3. To facilitate transitions between age classes, the proposed Pyramid Aging-GAN framework employs a pyramid shared latent space for both encoding and decoding, the benefit of which is demonstrated by experimental results.
4. Pyramid Aging-GAN accomplishes realistic results for both age progression and rejuvenation, simultaneously.

The proposed framework is described in detail in Section 2. In Section 3, the experimental evaluation of the Pyramid Aging-GAN is conducted. Finally, Section 4 concludes the paper and recommends future work.

## 2. PROPOSED FRAMEWORK

Here, we analyze the pipeline of the proposed Pyramid Aging-GAN framework. To perform image-to-image translation, we adopt the

powerful UNsupervised Image-to-image Translation (UNIT) framework [11] and learn pairwise translations between age classes. In order to infer the joint distribution of images in different age classes based on their marginal distributions, two vital assumptions are made: the shared latent space assumption and the cycle consistency assumption. According to the first assumption, a tuple of images  $x_n$  corresponding to different age classes  $\mathbf{X}_n, n = 1, \dots, N$  (i.e., the input to the proposed framework) is mapped to a shared latent representation  $z$  in a shared latent space. This is essential in order to capture correspondences between age classes. According to the cycle consistency assumption, a cycle consistency mapping should exist so that each image that originally belongs to a specific age class  $\mathbf{X}_n$  can be reconstructed after being translated from age class  $\mathbf{X}_n$  to a different age class  $\mathbf{X}_k$  and then, translated back to age class  $\mathbf{X}_n$ , i.e., after the translation cycle  $\mathbf{X}_n \rightarrow \mathbf{X}_{k \neq n} \rightarrow \mathbf{X}_n$ .

The proposed framework consists of Variational AutoEncoders (VAEs) [12, 13] and GANs [3]. For each age class, a VAE and a GAN are jointly trained in a VAE-GAN [14]. In VAE-GANs, the decoder is adversarially trained against the GAN discriminator. Each age class is represented by a VAE-GAN, which consists of three sub-networks: an encoding network  $\mathbf{E}$ , a generative network  $\mathbf{G}$ , and an adversarial discriminator  $\mathbf{D}$ .  $VAE_n$  for age class  $\mathbf{X}_n$  consists of the encoder-generator pair  $\{\mathbf{E}_n, \mathbf{G}_n\}$ , while  $GAN_n$  consists of the generator-discriminator pair  $\{\mathbf{G}_n, \mathbf{D}_n\}$ . All sub-networks  $\mathbf{E}$ ,  $\mathbf{G}$ , and  $\mathbf{D}$  are implemented by Convolutional Neural Networks (CNNs), while  $\mathbf{E}$  and  $\mathbf{G}$  also employ residual blocks [15].

The shared latent space constraint is implemented by enforcing a weight sharing scheme to encoders  $\mathbf{E}$  and generators  $\mathbf{G}$ . In order to preserve personality, weight sharing is enforced to the layers that bear the most high-level semantic information. Due to the hierarchical way the deep neural networks learn feature representations, these high-level layers are the last few layers of encoders and the first few layers of decoders.

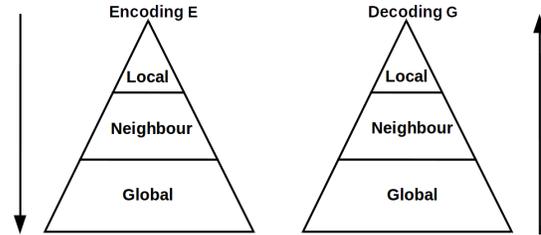
Here, extending [16], a weight sharing scheme is employed, which follows a pyramid structure. Each encoder  $\mathbf{E}$  is regarded as the ensemble of three sub-encoders:  $\mathbf{E} = \mathbf{E}_G \circ \mathbf{E}_A \circ \mathbf{E}_L$ . The first layers of each encoder  $\mathbf{E}$ , denoted by  $\mathbf{E}_L$ , capture local, low-level information related to the specific age class and do not share any weights. The intermediate layers of encoder  $\mathbf{E}$ , denoted by  $\mathbf{E}_A$ , have tied weights with the corresponding layers of the encoder that belongs to an adjacent age class. Therefore, they fuse information from the neighbourhood of each age class. Shared intermediate encoder layers  $\mathbf{E}_A$  map pairs of images  $x_n$  and  $x_{n+1}$  that belong to adjacent age classes to an intermediate neighbour shared latent representation  $z_{n:n+1}$ . Subsequently, all intermediate neighbour shared latent representations are mapped to a global shared latent representation  $z_{1:N}$  by encoder layers  $\mathbf{E}_G$ , which have tied weights across all age classes. This way, the obtained shared latent representation  $z_{1:N}$  progressively fuses information from all age classes.

For a given age class  $\mathbf{X}_k$ , the intermediate layers  $\mathbf{E}_{A,k}$  will share weights with the corresponding layers of the precedent age class  $\mathbf{X}_{k-1}$  or the following age class  $\mathbf{X}_{k+1}$  if age progression or regression is performed. More specifically, for translation to age class  $\mathbf{X}_l$ , age class  $\mathbf{X}_k$  is coupled with its precedent class if  $l < k$  and with its following class if  $l > k$ . When  $l = k$  and reconstruction is performed, age class  $\mathbf{X}_k$  is coupled with its precedent class if  $k \leq N/2$ , otherwise it is coupled with its following age class.

Subsequently, the final shared latent code  $z_{1:N}$  is decoded by generators  $\mathbf{G}$ , which also follow a corresponding pyramid weight sharing structure. The generator  $\mathbf{G}$  for each age class  $\mathbf{X}_n$  is regarded as the ensemble of three sub-generators:  $\mathbf{G} = \mathbf{G}_L \circ \mathbf{G}_A \circ \mathbf{G}_G$ . The first layers of  $\mathbf{G}$ , namely  $\mathbf{G}_G$ , are tied across all age classes, the

intermediate layers  $\mathbf{G}_A$  are shared between adjacent age classes, and the last layers  $\mathbf{G}_L$  are specific to each age class.  $\mathbf{G}_G$  can be regarded as a common high-level generation function that maps global shared latent representation  $z_{1:N}$  to an intermediate decoded global shared high-level representation  $\tilde{z}_{1:N}$  that holds the most abstract information fused from all age classes.  $\mathbf{G}_A$  maps intermediate decoded global shared representation  $\tilde{z}_{1:N}$  to an intermediate decoded neighbour shared representation  $\tilde{z}_{n:n+1}$ , which holds information for the neighbourhood of the adjacent age classes  $\mathbf{X}_n$  and  $\mathbf{X}_{n+1}$ . For each age class  $\mathbf{X}_n$ ,  $\mathbf{G}_L$  can be regarded as a low-level generation function that further decodes the intermediate decoded neighbour shared representation  $\tilde{z}_{n:n+1}$  to generated images  $\tilde{x}_n \in \mathbf{X}_n$ .

The information flow for the pyramid encoding and decoding scheme of the proposed Pyramid Aging-GAN framework is illustrated graphically in Figure 1. During encoding, information is accumulated to the shared latent representations from local to neighbour to global. On the contrary, during decoding, decoded information grows progressively from global to neighbour to local.



**Fig. 1:** Information flow for the encoding and decoding scheme in the Pyramid Aging-GAN. For encoding, information progresses from local to neighbour to global, while the reverse order applies for decoding.

The generator  $\mathbf{G}_n$  is utilized to either reconstruct input images  $x_n$  in age class  $\mathbf{X}_n$  or to translate input images  $x_k, k \neq n, n = 1, \dots, N$  to age class  $\mathbf{X}_n$ . The reconstruction stream generates the images  $\tilde{x}_n^{n \rightarrow n} \in \mathbf{X}_n$ , while the translation stream generates the images  $\tilde{x}_k^{k \rightarrow n} \in \mathbf{X}_n$ . In order to translate image  $x_n$  to a different age class, e.g., age class  $\mathbf{X}_{n+1}$ , the latent code  $z_{1:N}$  is decoded by generator  $\mathbf{G}_{n+1}$  which translates latent codes to age class  $\mathbf{X}_{n+1}$  and the translated image  $\tilde{x}_n^{n \rightarrow n+1}$  is obtained. By doing so, the framework learns the bidirectional translations among all  $N$  age classes simultaneously. To produce realistic translated images, the generators  $\mathbf{G}_n, n = 1, \dots, N$  are trained against adversarial discriminators. The discriminator  $\mathbf{D}_n$  is fed by both original images  $x_n \in \mathbf{X}_n$  as well as images of different age classes that are translated to age class  $\mathbf{X}_n$  in order to learn to discriminate them. In each  $GAN_n$ ,  $\mathbf{G}_n$  tries to fool  $\mathbf{D}_n$  by generating images that resemble the original images in age class  $\mathbf{X}_n$ , while  $\mathbf{D}_n$  tries to understand which images are original and which are generated. A few layers of discriminators  $\mathbf{D}_n$  also have tied weights across age classes in an attempt to capture similarities between age classes.

## 2.1. Objective function

The adversarial training of Pyramid Aging-GAN can be regarded as a two player min-max game, where the team of encoders and generators is trained against the team of adversarial discriminators. The first team except of defeating the second team has to minimize the VAE loss, the cycle consistency loss, and the total variation loss. The inclusion of total variation in the loss function aims at smoothing the noise due to steep differences between pixel values and has

been proved efficient in removing the ghosting artifacts of generated images for face aging [4, 7, 6].

The objective function for translations from age class  $\mathbf{X}_k$  to all other age classes  $\mathbf{X}_l$ ,  $l = 1, \dots, N$ ,  $l \neq k$  is given by

$$\min_{\mathbf{E}_k, \mathbf{E}_l, \mathbf{G}_k, \mathbf{G}_l} \max_{\mathbf{D}_k, \mathbf{D}_l} \left\{ \mathcal{L}_{VAE_k}(\mathbf{E}_k, \mathbf{G}_k) + \mathcal{L}_{GAN_l}(\mathbf{E}_k, \mathbf{G}_l, \mathbf{D}_l) + \mathcal{L}_{CC_k}(\mathbf{E}_k, \mathbf{G}_k, \mathbf{E}_l, \mathbf{G}_l) + \mathcal{L}_{TV_l}(\mathbf{G}_l) \right\} \quad (1)$$

where  $\mathcal{L}_{VAE_k}$  corresponds to the objective function of  $VAE_k$  training. The final shared latent code  $z_{k,1:N}$  is included in  $VAE_k$  training. The intermediate steps of pyramid encoding for obtaining  $z_{k,1:N}$  are only indirectly penalized.  $VAE_k$  training for age class  $\mathbf{X}_k$  aims at minimizing the loss function:

$$\mathcal{L}_{VAE_k}(\mathbf{E}_k, \mathbf{G}_k) = \lambda_0 KL(q_k(z_{k,1:N}|x_k) || p_p(z)) - \lambda_1 \underbrace{\mathbf{E}_{z_{k,1:N} \sim q_k(z_{k,1:N}|x_k)} [\log p_{G_k}(\tilde{x}_k^{k \rightarrow k} | z_{k,1:N})]}_{\|x_k - \tilde{x}_k^{k \rightarrow k}\|_{\ell_1}}. \quad (2)$$

The Kullback-Leibler (KL) divergence in Eq. (2) 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 Laplacians. Minimizing the negative log-likelihood term is equivalent to minimizing the absolute distance between image  $x_k$  and the reconstructed image  $\tilde{x}_k^{k \rightarrow k}$ . Hyper-parameters  $\lambda_0$  and  $\lambda_1$  control each term of the  $\mathcal{L}_{VAE_k}$  objective function.

In Eq. (1),  $\mathcal{L}_{GAN_l}$  penalizes the image translation stream from age class  $\mathbf{X}_k$  to age classes  $\mathbf{X}_l$ ,  $l = 1, \dots, N$ ,  $l \neq k$ . The objective function of GAN training is given by:

$$\mathcal{L}_{GAN_l}(\mathbf{E}_k, \mathbf{G}_l, \mathbf{D}_l) = \lambda_2 \mathbf{E}_{x_l \sim p_{x_l}} [\log \mathbf{D}_l(x_l)] + \lambda_2 \mathbf{E}_{z_{k,1:N} \sim q_k(z_{k,1:N}|x_k)} [\log (1 - \mathbf{D}_l(\mathbf{G}_l(z_{k,1:N})))]. \quad (3)$$

where  $\lambda_2$  controls the impact of generative loss for age class  $\mathbf{X}_l$ . In the experiments, the same value of hyper-parameter  $\lambda_2$  was set for all age classes  $\mathbf{X}_n$ ,  $n = 1, \dots, N$ .

In Eq. (1),  $\mathcal{L}_{CC_k}$  penalizes the cycle consistency loss for age class  $\mathbf{X}_k$ . For the translation cycle  $\mathbf{X}_k \rightarrow \mathbf{X}_l \rightarrow \mathbf{X}_k$ ,  $l = 1, \dots, N$ ,  $l \neq k$ , the objective function for the cycle consistency constraint is:

$$\mathcal{L}_{CC_k}(\mathbf{E}_k, \mathbf{G}_k, \mathbf{E}_l, \mathbf{G}_l) = \lambda_3 KL(q_k(z_{k,1:N}|x_k) || p_p(z)) + \lambda_3 KL(q_l(z_{l,1:N}|\tilde{x}_k^{k \rightarrow l}) || p_p(z)) - \lambda_4 \mathbf{E}_{z_{l,1:N} \sim q_l(z_{l,1:N}|\tilde{x}_k^{k \rightarrow l})} [\log p_{G_k}(x_k | z_{l,1:N})]. \quad (4)$$

In Eq. (4), the KL divergence terms penalize any deviation of the distribution  $q_k$  of the latent codes of the translation stream  $\mathbf{X}_k \rightarrow \mathbf{X}_l$  and the distribution  $q_l$  of the translation stream  $\mathbf{X}_l \rightarrow \mathbf{X}_k$  from the prior distribution  $p_p(z)$ . The negative log-likelihood term forces the twice translated image  $\tilde{x}_k^{k \rightarrow l \rightarrow k}$  to resemble the input image  $x_k$ . The hyper-parameters  $\lambda_3$  and  $\lambda_4$  control the two terms of the cycle consistency constraint.

### 3. EXPERIMENTS

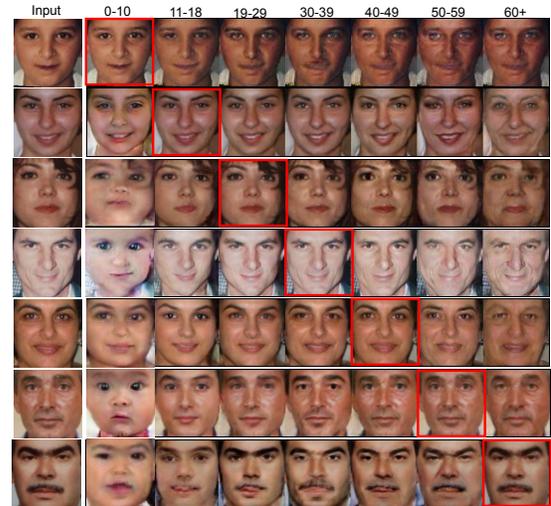
#### 3.1. Dataset

The proposed framework is trained on a dataset that comprises of images collected from the Cross-Age Celebrity Dataset (CACD) [17] and the UTKFace [4] dataset. Similar to [7, 16], 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. Approximately, the same number of face images belongs to each age class. Both genders are equally represented in each age class. In total, 21,267 face images comprise the training set. Similar to [4, 1, 6, 9], the proposed framework is evaluated on the FGNET dataset [18]. The FGNET dataset consists of 1002 images of 82 subjects, aging from 0 to 69.

#### 3.2. Experimental evaluation

Age progression and regression results achieved by the proposed Pyramid Aging-GAN are depicted in Figure 2. 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. Visual results demonstrate that Pyramid Aging-GAN succeeds to generate realistic images illustrative of the progressive effects of aging, while preserving personalized face features. Remarkably, the achieved age progression/regression is compatible to each person's gender, although no gender information is included to the framework. This can be attributed to the ability of the proposed framework to capture the most abstract face aging effects appropriate to both genders. For example, in the last row of Figure 2, the person's moustache is removed in the age class 0-10. The moustache is slightly visible for age class 11-18, while for age class 30-39 a beard appears in the person's face. Beard and moustache are related to person's gender as well as to the depicted age class. All aging effects maintain the personalized characteristics of the input face image. The training of the first, second, and last age class is stopped at an earlier iteration due to earlier collapse compared to the other age classes. The earlier collapse of training for those extreme age classes compared to the intermediate classes may be attributed to the drastic face alterations achieved for translations to these age classes, e.g., from a child to an adult and from a middle-aged person to an elderly person, and vice versa.



**Fig. 2:** Age progression and regression achieved by the Pyramid Aging-GAN framework for sample images of FGNET. The first column depicts input faces, and the rest columns depict results for 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 3, we compare the face images generated by the Pyramid Aging-GAN to the ground truth images of the FGNET dataset. More specifically, generated samples of FGNET images translated to different age classes are presented and compared to the ground truth images of the persons at that specific age interval. The proposed framework achieves realistic results, similar to the ground truth images, for both age progression and rejuvenation tasks.

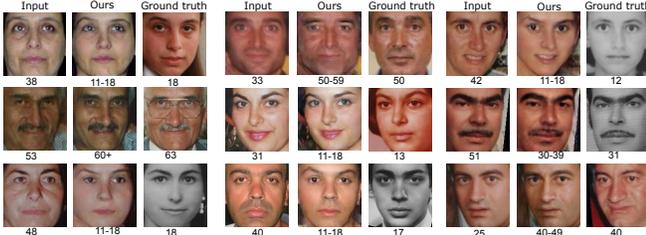


Fig. 3: Comparison to the ground truth images of FGNET.

**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 [16, 4, 1, 7, 9] for age progression and to [16, 4] for age regression. We use the same input images and perform age progression and regression. The comparative demonstration is shown in Figure 4. It can be seen that the face aging effects for age progression are more intense compared to those of the competitive methods. By comparing the results of the proposed method and those in [16], it is attested that the inclusion of the pyramid encoding and decoding scheme enabled the proposed framework to more effectively capture the age-related alterations of face characteristics. For age rejuvenation, the resulted face images are also appealing when compared to the rejuvenation results in [4]. The proposed method succeeds exceptionally in maintaining personality, especially when compared to the generated images of the method in [1].

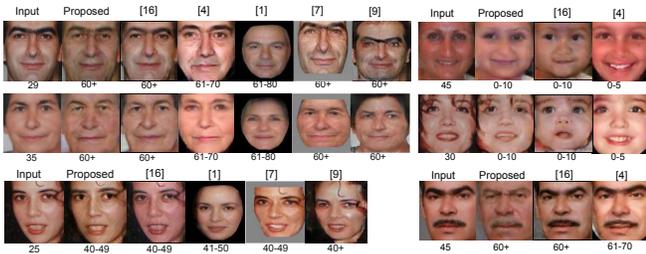


Fig. 4: Comparison of age transitions generated by the Pyramid Aging-GAN and other methods applied to images of FGNET.

**Quantitative comparisons:** To quantitatively compare the Pyramid Aging-GAN with our previous work in [16], we employ the Inception Score (IS) [19] as an evaluation metric. Since our goal is to evaluate the quality of generated face images, IS is computed based on a deep neural network trained on faces, namely the pre-trained VGG Face Model [20]. In Table 1, this metric, namely the VGG Face score, is calculated for the generated images by Pyramid Aging-GAN and compared to those generated by the method in [16], and the ground truth FGNET images. The best values for each row of Table 1 are indicated in bold. The proposed method consistently achieves higher VGG Face scores than [16]. Therefore, it generates

face images of better visual quality and benefits from the incorporation of the pyramid encoding and decoding scheme. Moreover, as depicted in Table 2, the images generated by the proposed framework achieve smaller mean absolute error (in years) for age estimation compared to those generated by the method in [16], using the pre-trained age estimation model in [21]. The best values for each row of Table 2 are indicated in bold.

VGG Face score			
Reconstruction	Proposed	[16]	Ground truth
$X_1 \rightarrow X_1$	<b>14.22 ± 3.10</b>	13.46 ± 2.41	17.90 ± 2.88
$X_2 \rightarrow X_2$	<b>15.52 ± 2.27</b>	14.83 ± 2.11	18.59 ± 2.40
$X_3 \rightarrow X_3$	<b>8.76 ± 1.34</b>	8.08 ± 1.45	12.17 ± 1.53
$X_4 \rightarrow X_4$	6.06 ± 1.17	<b>6.18 ± 0.88</b>	7.11 ± 0.46
$X_5 \rightarrow X_5$	<b>3.78 ± 0.60</b>	3.74 ± 0.70	4.25 ± 0.62
$X_6 \rightarrow X_6$	<b>1.44 ± 0.45</b>	1.42 ± 0.45	1.49 ± 0.49
$X_7 \rightarrow X_7$	<b>1.32 ± 0.45</b>	1.32 ± 0.45	1.33 ± 0.47
$X_k \rightarrow X_k, k \in 1, \dots, 7$	<b>29.16 ± 3.59</b>	26.60 ± 3.59	42.02 ± 4.31
Translation	Proposed	[16]	Ground truth
$X_{k \neq 1} \rightarrow X_1$	<b>9.11 ± 2.36</b>	7.39 ± 0.91	-
$X_{k \neq 2} \rightarrow X_2$	22.78 ± 3.29	<b>25.81 ± 2.52</b>	-
$X_{k \neq 3} \rightarrow X_3$	<b>37.14 ± 4.41</b>	29.24 ± 3.51	-
$X_{k \neq 4} \rightarrow X_4$	<b>34.42 ± 3.12</b>	28.29 ± 2.90	-
$X_{k \neq 5} \rightarrow X_5$	<b>31.07 ± 3.48</b>	23.79 ± 2.62	-
$X_{k \neq 6} \rightarrow X_6$	<b>32.40 ± 2.46</b>	26.96 ± 2.58	-
$X_{k \neq 7} \rightarrow X_7$	<b>31.54 ± 1.49</b>	20.65 ± 1.42	-
$X_k \rightarrow X_l, k \neq l, k, l \in 1, \dots, 7$	<b>61.90 ± 20.30</b>	45.78 ± 15.63	-

Table 1: VGG Face score, i.e., Inception Score [19] evaluated on the pre-trained VGG Face Model [20]. The score admitted by the proposed method is compared to the score admitted by [16] and the score of the ground truth FGNET images.

Age estimation results using [21]		
	Proposed	[16]
Reconstruction	<b>12.281</b>	14.384
Regression	<b>11.833</b>	13.965
Progression	<b>18.427</b>	18.560
All translations	<b>17.161</b>	17.678

Table 2: Mean absolute error in years for age estimation evaluated on the reconstructed, age progressed, and regressed images generated by the proposed method and the method in [16], using the pre-trained age estimation model in [21].

#### 4. CONCLUSION AND FUTURE WORK

A novel approach has been proposed that addresses face age progression and regression as an image-to-image translation problem and learns the pairwise translations between face images belonging to different age classes. To meet this goal, the UNIT framework [11] has been employed and extended to multiple age classes. In order to better capture the effects of aging on face characteristics, we have employed a pyramid of local, neighbour, and global encoders so that the latent representations progressively encapsulate semantic information. Shared latent representations have been fed to GANs, each of which is trained on a single age class and utilizes low-level information to generate images that resemble the original images of the age class. The proposed Pyramid Aging-GAN framework succeeds to capture both subtle and intense face aging effects and perform appealing age progression and regression simultaneously. Future work will focus on facilitating the transitions between distant age classes, since they cause the most drastic aging effects and reducing the blur in the generated images.

## 5. REFERENCES

- [1] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2016, pp. 2378–2386.
- [2] 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, pp. 053011, 2017.
- [3] 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.
- [4] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2017, pp. 4352–4360.
- [5] G. Antipov, M. Baccouche, and J. L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Processing*, 2017, pp. 2089–2093.
- [6] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 2018, pp. 31–39.
- [7] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan, "Face aging with contextual generative adversarial nets," in *Proc. 25th ACM Int. Conf. Multimedia*, 2017, pp. 82–90.
- [8] 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.
- [9] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, "Global and local consistent age generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognition*, August 2018, pp. 1073–1078.
- [10] 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.
- [11] 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.
- [12] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [13] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proc. 31st Int. Conf. Machine Learning*, 2014, vol. 32, pp. 1278–1286.
- [14] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *Proc. 33rd Int. Conf. Machine Learning*, 2016, vol. 48, pp. 1558–1566.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [16] E. Pantraki and C. Kotropoulos, "Face aging as image-to-image translation using shared-latent space generative adversarial networks," in *Proc. 6th IEEE Global Conf. Signal and Information Processing*, 2018, pp. 116–120.
- [17] 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. Computer Vision*. Springer, 2014, pp. 768–783.
- [18] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.
- [19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.
- [20] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. British Machine Vision Conf.*, 2015, pp. 41.1–41.12.
- [21] R. Rothe, R. Timofte, and L. Van Gool, "DEX: Deep expectation of apparent age from a single image," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, December 2015, pp. 10–15.