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

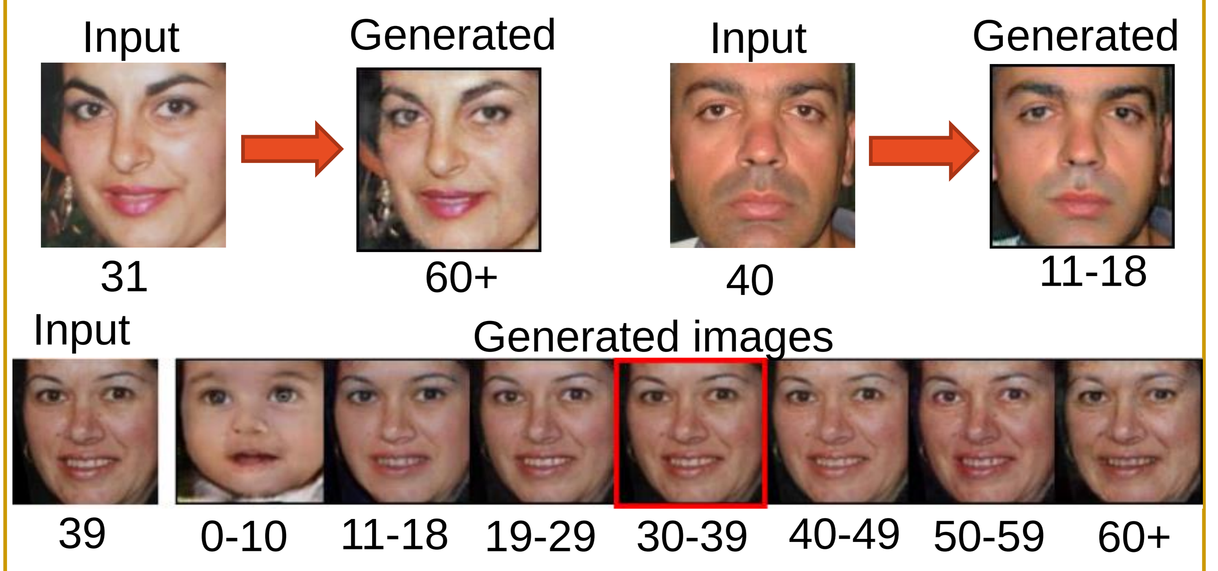
Evangelia Pantraki and Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki, Greece

epantrak@csd.auth.gr, costas@aiia.csd.auth.gr

ABSTRACT

Face aging as an unsupervised **image-to-image translation problem** between face images of different age classes.

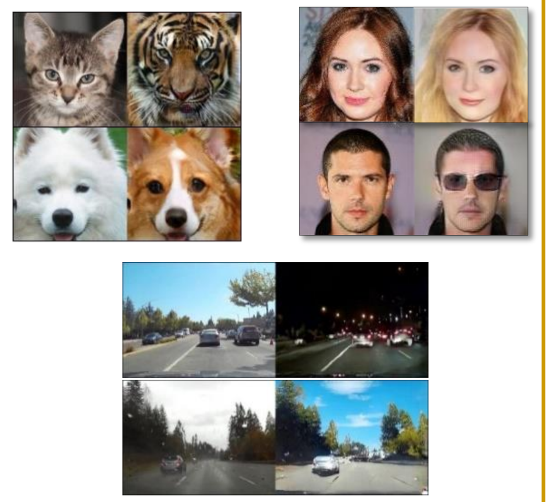


PROPOSED METHOD

Goal: 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**) while preserving **personality**.

Method: age progression/regression formulated as the generative task of **unsupervised translation** from face images of a specific age class to different age classes.

- Exploit the dynamics of **Generative Adversarial Networks** (GANs).
- Apply the UNsupervised Image-to-Image Translation (**UNIT**) network [1], where couples of GANs (**coupled GANs**) are trained in image transformation between two domains.
- Learn the bidirectional pairwise transformations/ **translations between age classes**.



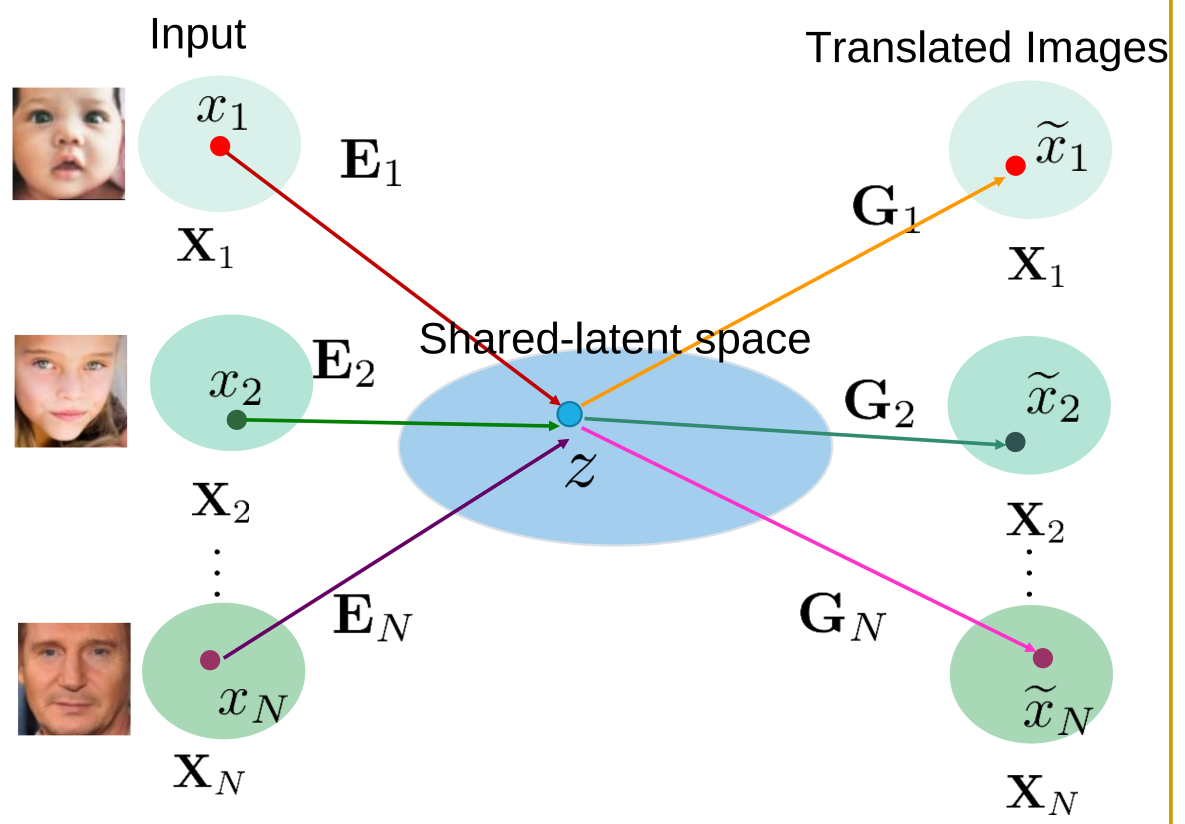
PROPOSED METHOD (CONT.)

Network: The proposed **Aging-UNIT framework** employs GANs and Variational Autoencoders (VAEs). Each age class $\mathbf{X}_n, n = 1, 2, \ldots, N$ is represented by:

1. an encoder $\mathbf{E}_n$
2. a generator/decoder $\mathbf{G}_n$
3. an adversarial discriminator $\mathbf{D}_n$

$$\left.\begin{array}{l}\left.\begin{array}{l}\text{1. an encoder } \mathbf{E}_n\\ \text{2. a generator/decoder } \mathbf{G}_n\end{array}\right\} VAE_n\\ \text{3. an adversarial discriminator } \mathbf{D}_n\end{array}\right\} GAN_n$$

Schematic representation of the proposed framework



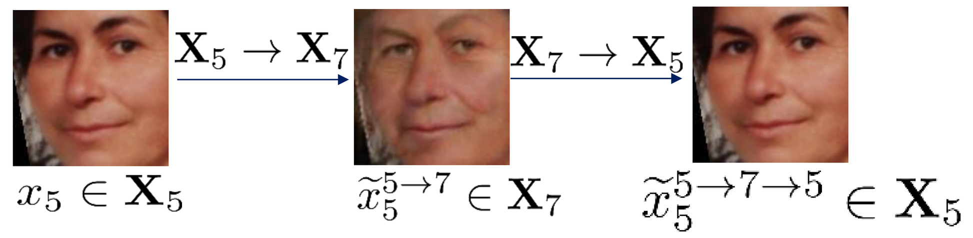
Two generative streams:

1. Translation stream: $$\tilde{x}_k^{k\to l} = \mathbf{G}_l(z_k \sim q_k(z_k|x_k)),\quad k \in 1, \ldots, N,\ l = 1, \ldots, N,\ l \neq k$$

2. Reconstruction stream: $$\tilde{x}_n^{n\to n} = \mathbf{G}_n(z_n \sim q_n(z_n|x_n)),\quad n = 1, \ldots, N$$

Two assumptions:

1. **Shared-latent space assumption** ➡ implemented by a **weight sharing** constraint applied to the **high-level layers** of both encoders and generators across age classes
2. **Cycle-consistency assumption**



PROPOSED METHOD (CONT.)

Adversarial training between two teams:

1. **Discriminators**
2. **Encoders and generators**. To defeat the discriminators and to minimize the encoding loss and the cycle-consistency loss.

Objective function: jointly solve the learning problems of VAE and GAN subject to the cycle-consistency constraint.

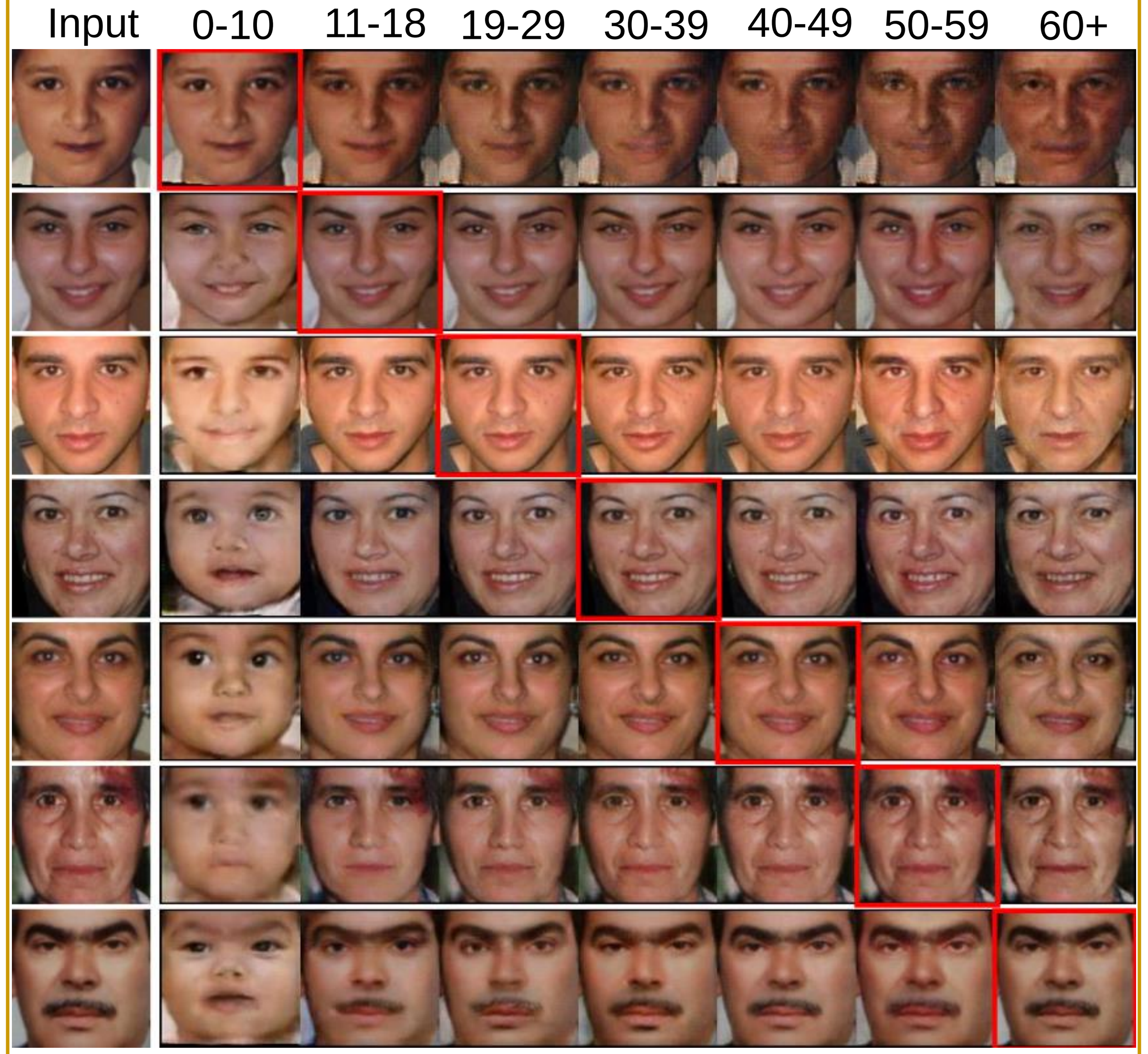
- For translations from age class $\mathbf{X}_k$ to all other age classes $\mathbf{X}_l,\ l = 1, \ldots, N,\ l \neq k$:

$$\min_{\mathbf{E}_k,\mathbf{E}_l,\mathbf{G}_k,\mathbf{G}_l} \max_{\mathbf{D}_k,\mathbf{D}_l} \Big\{ \text{Ł}_{VAE_k}(\mathbf{E}_k, \mathbf{G}_k) + \text{Ł}_{GAN_k}(\mathbf{E}_l, \mathbf{G}_k, \mathbf{D}_k) + \text{Ł}_{CC_k}(\mathbf{E}_k, \mathbf{G}_k, \mathbf{E}_l, \mathbf{G}_l),\quad k \in 1, \ldots, N,\ l = 1, \ldots, N,\ l \neq k \Big\}$$

EXPERIMENTAL EVALUATION

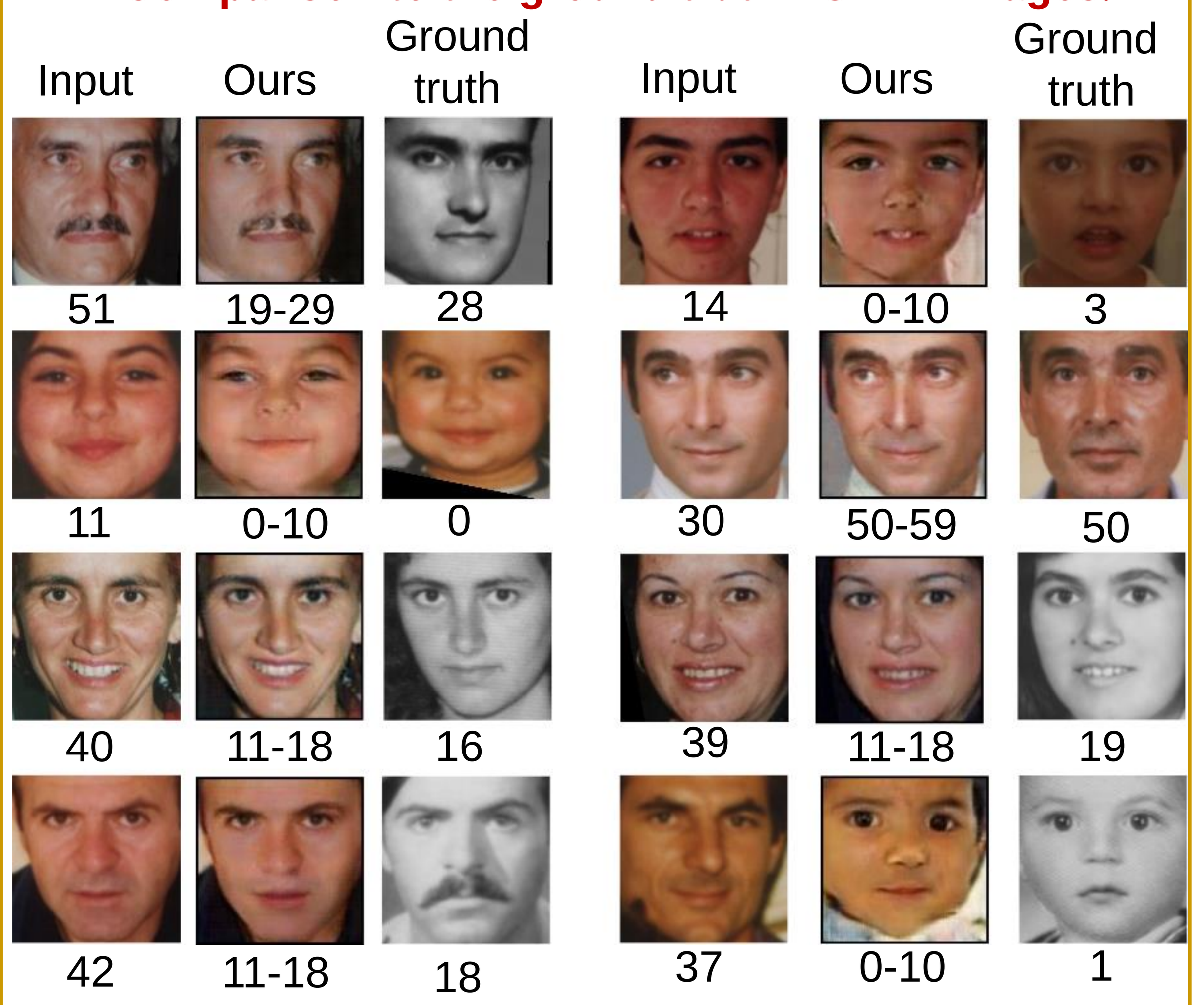
Dataset: a subset of 21,267 images from the Cross-Age Celebrity Dataset [2] and the UTKFace [3] dataset

Age progression and regression results admitted by the Aging-UNIT framework on the FGNET [4] dataset.

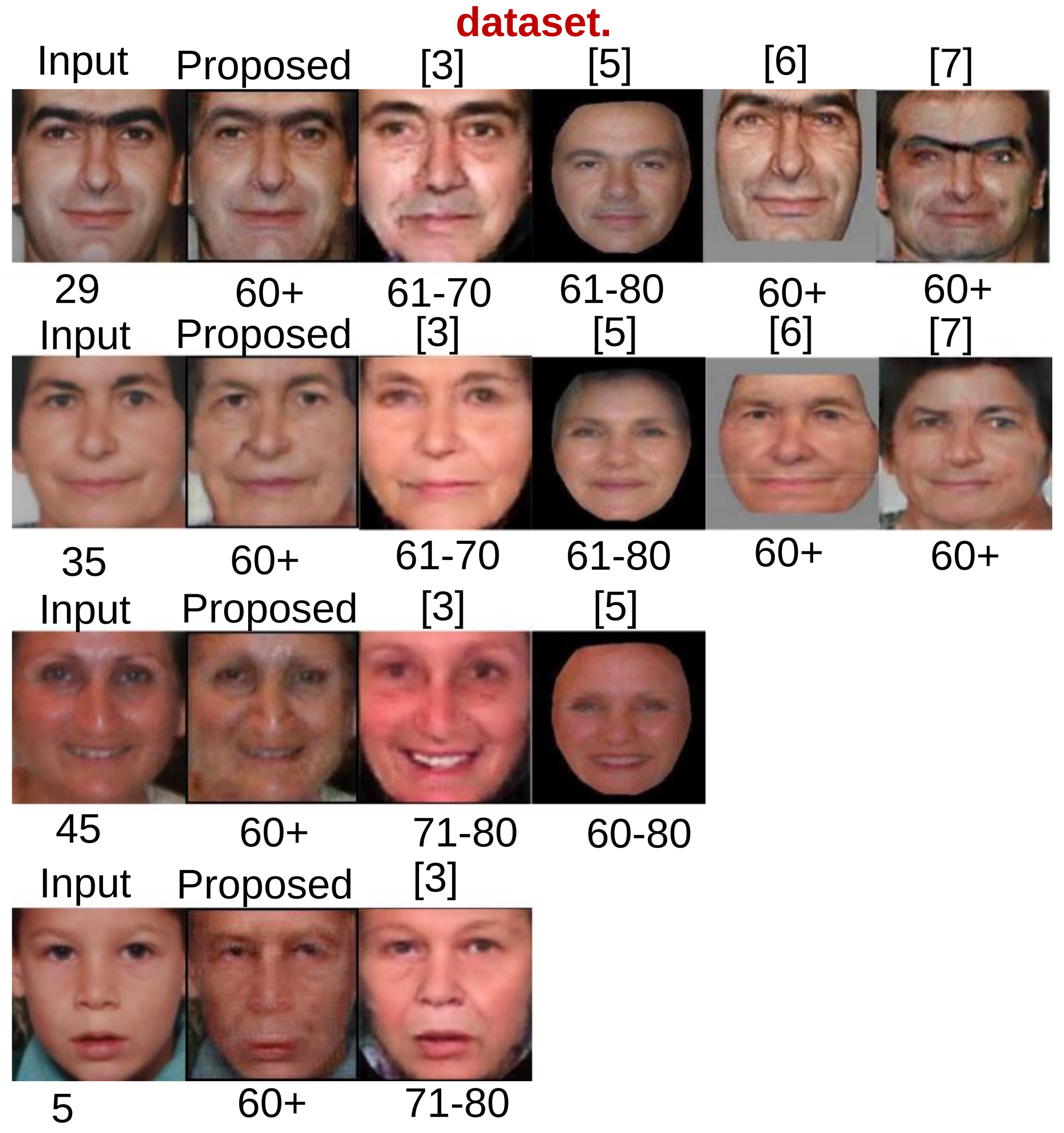


EXPERIMENTAL EVALUATION (CONT.)

Comparison to the ground truth FGNET images.



Comparison to prior works evaluated on the FGNET dataset.



CONCLUSIONS

- **Face aging** is treated as a problem of **translation** between images for first time.
- Realistic results for face age progression and regression simultaneously, while **preserving personalized facial features**.
- Captures abstract **face aging effects appropriate to the gender** of the depicted person, although no gender information is included.
- **Eliminates the need for paired samples** at different ages fro training.

Future work:

1. Improve the quality of the generated images by further regularization.
2. Facilitate translations between distant age classes, causing the most drastic aging effects.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[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] 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.

[7] 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.

ACKNOWLEDGMENTS

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).