Supplemental Materials for MeshWGAN: Mesh-to-Mesh Wasserstein GAN with Multi-Task Gradient Penalty for 3D Facial Geometric Age Transformation

Jie Zhang, Kangneng Zhou, Yan Luximon, Tong-Yee Lee, *Senior Member, IEEE*, and Ping Li, *Member, IEEE*

1 Overview

To make our paper self-contained, more information is provided in these supplemental materials, including 3D face dataset establishment (see Section 2), 3D facial texture mapping (see Section 3.1), 3D facial wrinkle prediction (see Section 3.2), detailed user study design (see Section 4), and 3D facial expression processing (see Section 5).

2 Dataset Supplement

To collect the children's head data, a human subjects ethics review (for a period from 1 Jan 2020 to 1 Jan 2023) was approved by The Hong Kong Polytechnic University (Ref.: HSEARS20190222002) and an informed consent to participate in research was obtained from the parents/guardians of each child involved in the experiment. In the informed consent, it was informed that his/her participation in the project was voluntary, the information obtained from this research may be used in future research, published or commercialized; however, his/her privacy will be retained, i.e., his/her personal details will not be revealed. All child subjects are Chinese aged 5-17 years, and their basic information was also recorded, including age, gender, height, and weight.

- *Jie Zhang and Ping Li are with the Department of Computing and the School of Design, The Hong Kong Polytechnic University, Hong Kong. E-mail: peterzhang1130@163.com, p.li@polyu.edu.hk.*
- *Kangneng Zhou is with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China. E-mail: elliszkn@163.com.*
- *Yan Luximon is with the School of Design, The Hong Kong Polytechnic University, Hong Kong, and also with the Laboratory for Artificial Intelligence in Design, Hong Kong. E-mail: yan.luximon@polyu.edu.hk.*
- *Tong-Yee Lee is with the Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan 70101, Taiwan. E-mail: tonylee@ncku.edu.tw.*

Manuscript received 17 Oct. 2022; revised 27 Apr. 2023; accepted 25 May 2023. This work was supported in part by the Research Grants Council of Hong Kong under Grant PolyU 15603419, in part by the National Science and Technology Council under Grant 110-2221-E-006-135-MY3, Taiwan, and in part by The Hong Kong Polytechnic University under Grants P0042740, P0030419, P0043906, and P0044520.
(Corresponding Authors: Yan Luximon and Ping Li.)

(a) (b)

Fig. 1. Examples of head scans. (a) Head scan of one female participant. (b) Head scan of one male participant.

In this experiment, the children's heads were scanned using an Artec Eva 3D commercial scanner (3D accuracy and resolution are up to 0.1mm and 0.2mm, respectively). Following previous approaches [1], [2], to mitigate the surface distortions of hair, each participant was required to wear a tight, custom-designed latex cap during scanning. Two examples of head scans are shown in Fig. 1. To guarantee a consistent head pose during scanning, each participant was also required to sit on a chair to keep still with a neutral expression and open eyes until the scanning was finished.

3 Method Supplement

3.1 Texture Mapping

The 3D facial texture dataset (including children and adults) is limited, which cannot provide enough training dataset for our networks. Fortunately, both 2D face aging and 3D face reconstruction are well researched and understood problems. Hence, we integrated the existing state-of-the-art methods of 2D face aging and 3D face reconstruction to generate and acquire the 3D facial aging texture. SAM

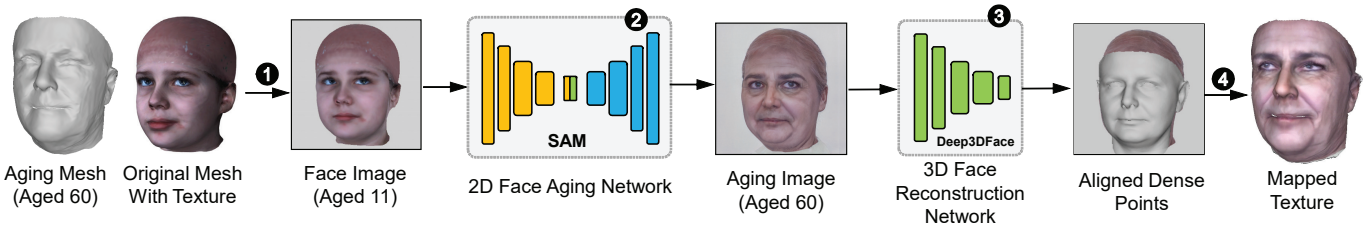

Fig. 2. Pipeline of the automatic 3D facial texture mapping. ❶ 3D facial mesh projection. ❷ 2D aging face generation using SAM [3]. ❸ Dense points-to-pixel correspondence using Deep3DFace [4]. ❹ 3D facial texture retrieval using bilinear interpolation.

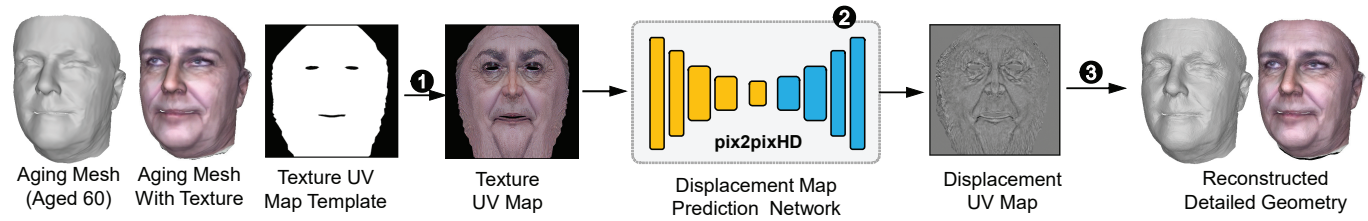

Fig. 3. Pipeline of the automatic 3D facial details generation. ❶ Texture UV map generation. ❷ Displacement UV map prediction via pix2pixHD [5], [6]. ❸ 3D facial details generation.

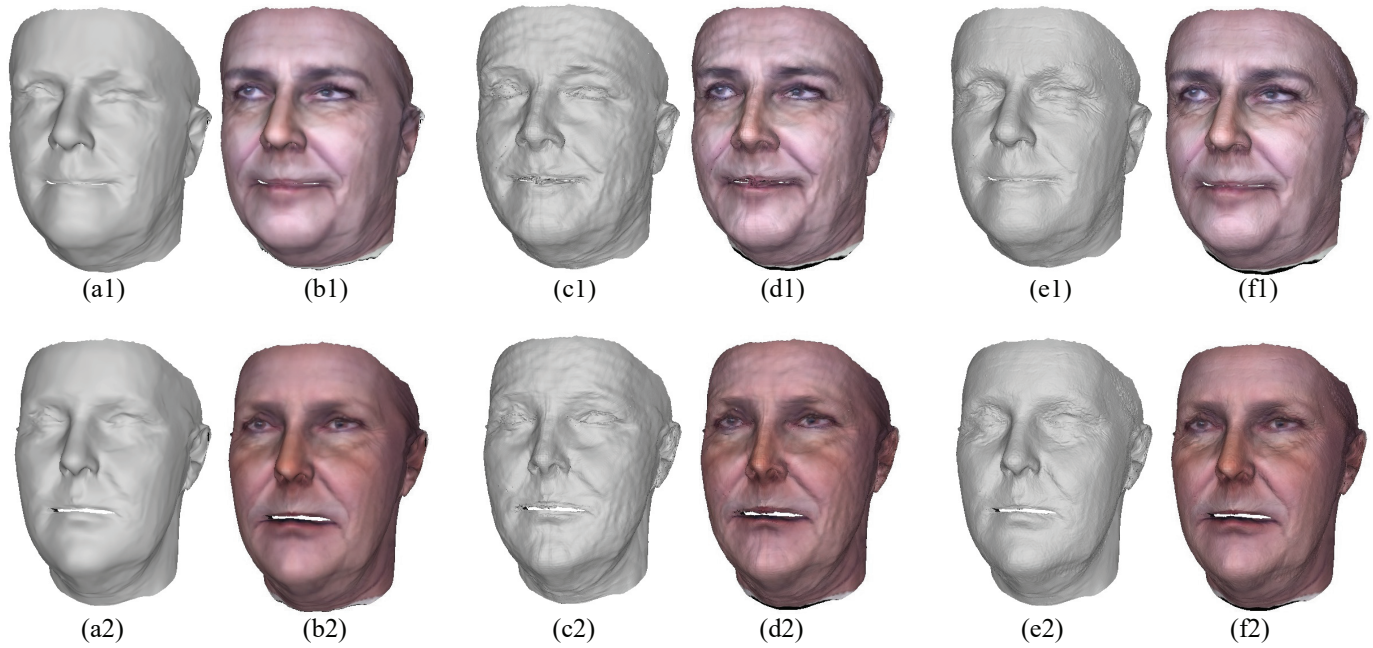

Fig. 4. Qualitative comparison of generated 3D facial meshes (age group: 50-70). (a)/(b) Without geometric details. (c)/(d) With geometric details predicted by using DECA [7]. (e)/(f) With geometric details predicted by using pix2pixHD [5]. Note that it is clear that the synthesized facial detailed geometries from facial texture can produce the facial wrinkles and ensure the mesh-and-texture matches to increase the feature difference of facial meshes between age groups 30-49 and 50-70. Furthermore, the pix2pixHD can predict more realistic wrinkles than DECA.

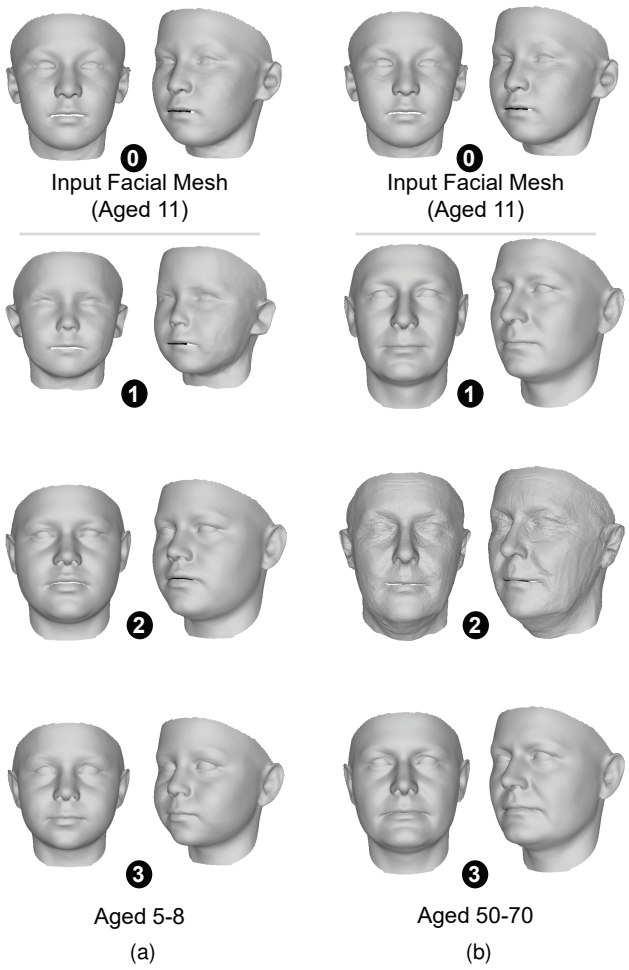

Fig. 5. Examples of identity preservation evaluation for aging facial geometries without textures. (a) Age group: 5-8. (b) Age group: 50-70. For each test, the participant was told that the facial meshes (❶, ❷, and ❸) with two different views at the same age group were generated from three different methods when the facial mesh ❶ was the input, and then asked which facial mesh (❶, ❷, or ❸) mostly resembles the input facial mesh ❶ in a given age (see the bottom texts).

[3] can receive 2D face image and generate the high-resolution (1024×1024 pixels) face image at any desired age. Furthermore, previous survey [8] has demonstrated that Deep3DFace [4] has the highest accuracy of facial shape reconstruction. The Deep3DFace [4] regresses the coefficients α_s of 3D morphable model (3DMM) of facial shapes and parameters $[R|t]$ of camera model with perspective projection using an encoder, and then reconstructs the 3D facial shape $\ddot{S}$ using the model coefficients α_s, as:

$$\ddot{S} = \overline{S} + P_s\alpha_s, \tag{1}$$

where $\overline{S}$ is the average mesh shape, P_s is the facial shape principal basis of 3DMM generated from many globally aligned facial meshes using principal components analysis (PCA). In Deep3DFace [4], the specific 3DMM is Basel Face Model (BFM) [9]. Then, the points-to-pixel correspondence is achieved using the camera model parameters $[R|t]$, as:

$$p = K(R\ddot{S} + t), \tag{2}$$

where $p = (u, v, 1)^T$, (u, v) is the pixel position in image space, R is a rotation matrix $R \in \mathbb{R}^{3\times3}$, t is a translation vector $t \in \mathbb{R}^3$, and K is a matrix of the camera's intrinsic parameters.

Compared with LATS [10] and DLFS [11] (that generate images with size 256×256 pixels), SAM can produce high-resolution face aging images (size 1024×1024 pixels), which can provide more details and higher quality in the retrieved 3D facial textures. Therefore, in our integrated methods as shown in Fig. 2, the pretrained SAM [3] and Deep3DFace [4] are used to generate the 2D aging image and retrieve the 3D facial texture, respectively. The automatic pipeline consists of four main steps: 1. 3D facial mesh projection, 2. 2D aging face generation using SAM [3], 3. dense points-to-pixel correspondence using Deep3DFace [4], and 4. 3D facial texture retrieval using bilinear interpolation.

In our studies, we rendered and projected the (original) 3D head scans into the 2D face images (512×512 pixels) and fed the images into their pretrained models. Since SAM can predict face images at a specific age, we applied SAM to the rendered head image to produce images aged 7, 11, 15, 24, 40, and 60 years, which correspond to the median age of each of our age groups. In the 3D facial texture retrieval, to produce 3D high-quality facial texture, the facial mesh (5,000 vertices and 9,449 triangles) was upsampled to a new high-resolution mesh (847,900 vertices and 1,693,440 triangles) firstly [12]. Such high resolution of 3D facial mesh can also help show the geometric details, especially wrinkles [6].

3.2 Wrinkle Prediction

Because of the scanning accuracy of commercial scanners, most 3D raw scans tend to miss high-frequency facial details. Furthermore, the NICP algorithm adopted for 3D face registration also has limitations in capturing high-frequency details of 3D raw scans due to its regularization and smoothness constraints [13]. Since the elders' faces (e.g., age group 50-70) usually have wrinkles, and our MeshWGAN can only generate the 3D aging facial shape without details, there are easily mesh-and-texture mismatches in the elders' facial meshes. To mitigate these issues, we leverage and predict displacement UV map [6], [7], [14], [15], [16] from 2D face images to express and produce 3D facial geometric details. Fortunately, when receiving a texture UV map, the pix2pixHD [5] was trained as a backbone to predict the high-resolution displacement UV map for describing expression-specific dynamic detailed geometry in previous studies [6], [15]. Thus, to improve the mesh-and-texture matches, we used pix2pixHD [5], [6] to predict the 3D facial details and produce the wrinkles from the facial textures aged 30-70.

The pipeline of automatic 3D facial details generation is shown in Fig. 3. The retrieved facial textures are used to produce texture UV map (size 1024×1024 pixels) based on the predefined points-to-UV coordinates transformation, and then the pretrained pix2pixHD [5], [6] is adopted to predict the displacement UV map d (size: 1024×1024 pixels), finally the 3D facial geometry S_{tra} is updated based on facial normals $\hat{n}$ as:

$$\ddot{S}^k_{tra} = S^k_{tra} + d_{k\rightarrow(i,j)}\hat{n}^k, \tag{3}$$

where i and j is the kth vertex's corresponding UV coordinates, k=1,2,...,N and N is number of vertices. Fig. 4 shows

qualitative comparison of 3D facial meshes (age group: 50-70), including Fig. 4(a)/(b) without geometric details, Fig. 4(c)/(d) with geometric details predicted by using DECA [7], and Fig. 4(e)/(f) with geometric details predicted by using pix2pixHD [5]. It is clear that the synthesized facial detailed geometries from facial texture in Fig. 4(e)/(f) can produce the facial wrinkles, especially on the corner of the eyes and forehead, which can ensure the mesh-and-texture matches and increase the feature difference of facial meshes between age groups 30-49 and 50-70. Moreover, since DECA can also predict the displacement UV map from 2D face image, we also applied it to update our generated facial normals. Compared to DECA, the pix2pixHD can predict more realistic wrinkles, it could be because the resolution (256×256 pixels) of displacement UV map for DECA is much less than that for pix2pixHD (1024×1024 pixels).

4 User Study

In human evaluation, two indexes related to face aging were measured, including identity preservation and age closeness. Thirty respondents aged 18-35 years with experience in 3D graphics/animation design were recruited, and ten facial meshes aged 5-70 years with textures were selected randomly from our testing dataset to produce stimulus. The facial meshes for each of our six age groups were generated using our MeshWGAN, and combined Deep3DFace [4] and LATS [10] / SAM [3]. To evaluate the quality of facial aging geometries without textures, every facial mesh was rendered (with solid blue colors using the same direct lighting environment) into two face images (512×512 pixels) showing front and 30^o-side views, respectively.

To measure identity preservation, one original and three facial meshes were shown as one column on four rows (see Fig. 5). In each test, the participant was told that the facial meshes (❶, ❷, and ❸) with two different views at the same age group were generated from three different methods when the facial mesh ❶ was the input, and then asked which facial mesh (❶, ❷, or ❸) mostly resembles the input facial mesh ❶ in a given age (see the bottom texts); It must be noted that these methods change the apparent age but retain the varying degrees of identity information. In total, there were 60 (10×6) tests. The displaying sequence of all tests was random, and the location sequence of three generated face images was also random, when the original face images were always placed on the first row. The evaluated metric was defined as the percentage of respondents who preferred each method.

To measure age closeness of each method, six facial meshes in different age groups were randomly placed on one column on six rows (see Fig. 6). For each test, the participant was told that the facial meshes with two different views are the same identity at the different age group, and then required to use lines to connect the facial meshes (left) to the corresponding age groups (right). In total, there were 30 (10×3) tests. The displaying sequence of all tests was random. The evaluated metric was defined as the percentage of images correctly assigned by respondents.

To evaluate the quality facial aging geometries with textures, the facial meshes from our MeshWGAN and SAM [3], were rendered (with retrieved textures using the same direct lighting environment) into two color images. In identity preservation measurement, one original and two facial meshes were shown as one column on three rows and the total testes were 60 (10×6). In age closeness measurement, the total testes were 20 (10×2).

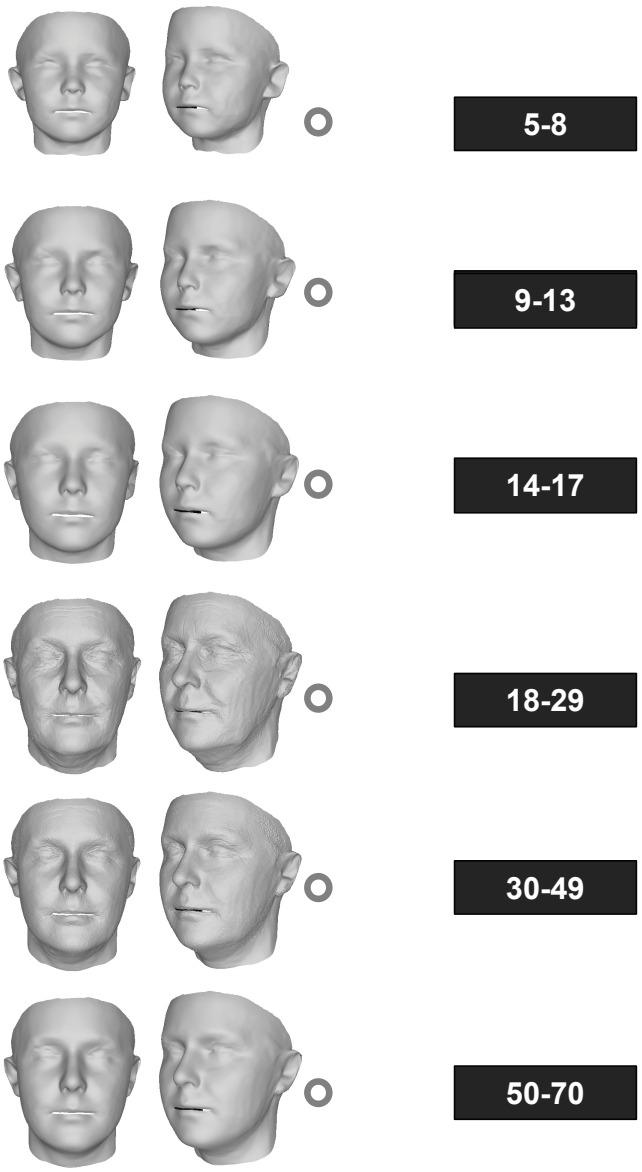

Fig. 6. Examples of age closeness evaluation of facial geometries without textures (using our MeshWGAN). For each test, the participant was told that the facial meshes with two different views are the same identity at the different age group, and then required to use lines to connect the facial meshes (left) to the corresponding age groups (right).

5 Expression Processing

In this study, our main objective is to create 3D aging figures and their expressions are usually neutral. To produce 3D face S_{new} with various expressions, the parameterized 3D facial aging meshes S_{ide} with 3DMMs of facial expressions (e.g., FaceWarehouse [17]) can be used by

$$S_{new} = S_{ide} + \sum_{i=1}^{m} \alpha_{exp,i} P_{exp,i} = S_{ide} + P_{exp}\alpha_{exp}, \quad (4)$$

Fig. 7. Examples of 3D facial aging meshes with different expressions.

where, P_{exp} is the facial expression principal basis, m is the number of the facial expression principal components and α_{exp} is the facial expression representation coefficient. For the input 3D facial meshes S_{new} with various expressions, with the 3DMMs of facial shapes (e.g., 3DCMM [18]), it is only needed to separate the identity S_{ide} and expression S_{exp} shapes by estimating their corresponding model coefficients [19] as:

$$S_{new} = S_{ide} + S_{exp} = \bar{S} + P_{ide}\alpha_{ide} + P_{exp}\alpha_{exp}, \quad (5)$$

where $\bar{S}$ is the average face, P_{ide} is the facial identity principal basis with n principal components, and α_{ide} is the facial identity representation coefficient. Then, the neutral facial shape S_{ide} ($S_{ide} = \bar{S} + P_{ide}\alpha_{ide}$) was input to our generator to generate 3D facial aging meshes S^i_{tra}, and they with the separate expression shape S_{exp} were combined to produce the new faces S^i_{new} with the original expression: $S^i_{new} = S^i_{tra} + S_{exp}$, where i indicates the index of the age group and i=1, 2, ..., 6. Fig. 7 shows some examples of 3D facial aging meshes with different expressions. It can be seen that the facial expressions are consistent in different age groups. This is also the advantage of the parameterized 3D facial meshes that they can be easily manipulated.

REFERENCES

[1] J. Zhang, Y. Luximon, P. Shah, K. Zhou, and P. Li, "Customize my helmet: A novel algorithmic approach based on 3D head prediction," *Computer-Aided Design*, vol. 150, pp. 103 271:1–103 271:10, 2022.

[2] J. Zhang, H. Iftikhar, P. Shah, and Y. Luximon, "Age and sex factors integrated 3D statistical models of adults' heads," *International Journal of Industrial Ergonomics*, vol. 90, pp. 103 321:1–103 321:13, 2022.

[3] Y. Alaluf, O. Patashnik, and D. Cohen-Or, "Only a matter of style: Age transformation using a style-based regression model," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 45:1–45:12, 2021.

[4] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, "Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 285–295.

[5] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.

[6] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao, "FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 598–607.

[7] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, "Learning an animatable detailed 3D face model from in-the-wild images," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 88:1–88:13, 2021.

[8] Z. Chai, H. Zhang, J. Ren, D. Kang, Z. Xu, X. Zhe, C. Yuan, and L. Bao, "REALY: Rethinking the evaluation of 3D face reconstruction," in *Proceedings of the Europeon Conference on Computer Vision*, 2022, pp. 74–92.

[9] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the ACM SIGGRAPH*, 1999, pp. 187–194.

[10] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 739–755.

[11] S. He, W. Liao, M. Y. Yang, Y.-Z. Song, B. Rosenhahn, and T. Xiang, "Disentangled lifespan face synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 3877–3886.

[12] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, "Generating 3D faces using convolutional mesh autoencoders," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 704–720.

[13] J. Zhang, Y. Luximon, P. Shah, and P. Li, "3D statistical head modeling for face/head-related product design: A state-of-the-art review," *Computer-Aided Design*, vol. 159, pp. 103 483:1–103 483:24, 2023.

[14] Q. Deng, L. Ma, A. Jin, H. Bi, B. H. Le, and Z. Deng, "Plausible 3D face wrinkle generation using variational autoencoders," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 9, pp. 3113–3125, 2021.

[15] J. Ling, Z. Wang, M. Lu, Q. Wang, C. Qian, and F. Xu, "Structure-aware editable morphable model for 3D facial detail animation and manipulation," in *Proceedings of the European Conference on Computer Vision*, 2022, pp. 249–267.

[16] J. Ling, Z. Wang, M. Lu, Q. Wang, C. Qian, and F. Xu, "Semantically disentangled variational autoencoder for modeling 3D facial details," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–12, 2022.

[17] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "FaceWarehouse: A 3D facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2014.

[18] J. Zhang, Y. Luximon, L. Zhu, and P. Li, "3DCMM: 3D comprehensive morphable models for accurate head completion," in *Proceedings of the ACM SIGGRAPH VRCAI*, 2022, pp. 13:1–13:8.

[19] T. Baltrusaitis, E. Wood, V. Estellers, C. Hewitt, S. Dziadzio, M. Kowalski, M. Johnson, T. J. Cashman, and J. Shotton, "A high fidelity synthetic face framework for computer vision," *arXiv:2007.08364*, pp. 1–10, 2020.