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*

Abstract—As the metaverse develops rapidly, 3D facial age transformation is attracting increasing attention, which may bring many potential benefits to a wide variety of users, e.g., 3D aging figures creation, 3D facial data augmentation and editing. Compared with 2D methods, 3D face aging is an underexplored problem. To fill this gap, we propose a new mesh-to-mesh Wasserstein generative adversarial network (MeshWGAN) with a multi-task gradient penalty to model a continuous bi-directional 3D facial geometric aging process. To the best of our knowledge, this is the first architecture to achieve 3D facial geometric age transformation via real 3D scans. As previous image-to-image translation methods cannot be directly applied to the 3D facial mesh, which is totally different from 2D images, we built a mesh encoder, decoder, and multi-task discriminator to facilitate mesh-to-mesh transformations. To mitigate the lack of 3D datasets containing children's faces, we collected scans from 765 subjects aged 5-17 in combination with existing 3D face databases, which provided a large training dataset. Experiments have shown that our architecture can predict 3D facial aging geometries with better identity preservation and age closeness compared to 3D trivial baselines. We also demonstrated the advantages of our approach via various 3D face-related graphics applications. Our project will be publicly available at: https://github.com/Easy-Shu/MeshWGAN.

Index Terms—Age transformation, 3D face geometry, MeshWGAN, mesh generative adversarial networks, multi-task gradient penalty.

1 Introduction

3D age transformation is defined as the process of synthesizing 3D meshes of a person's face across different ages while preserving their identity. Compared with 2D age transformation which focuses on 2D face synthesis [1], [2], [3], [4], [5], [6], the objective of our 3D age transformation is to synthesize the facial shape and albedo with a normalized face pose but no illumination information. This can be applied to many new practical applications, including 3D aging figures creation in animation, film, virtual and augmented reality (VR/AR), age-invariant 3D face recognition, 3D facial data augmentation, and 3D facial attribute editing. With the rapid development of depth cameras, particularly those in mobile phones, it is becoming continually easier to capture 3D facial data, which could make 3D aging transformations more accessible for users. Additionally, the development of the metaverse could make such applications more entertaining and popular.

However, creating lifelong transformations of 3D facial meshes, i.e., synthesizing faces aged 5-70 for any given input age, is a challenging task. The difficulty of capturing and collecting 3D face datasets exacerbates this problem. While the precise facial geometry can be captured using different commercial scanners, the captured textures are usually ill-defined and may include shading, shadowing, specularities, and light source color variation [11]. In this paper, existing 2D facial aging [1] and 3D face reconstruction [7] methods are combined to produce 3D facial aging textures. We aim at creating natural and reliable 3D facial geometric age transformations for facial meshes aged 5-70 years.

Compared to 2D face age transformations, 3D face age transformations have not been fully explored in the literature. Partly, the lack of real 3D face datasets has impeded such studies [9], [12]. While there are many large-scale 2D face images online that can be used for 2D face age transformations [1], [2], [13], [14], there are limited publicly available 3D face datasets [15], [16], [17], and their subjects are mainly adults; child subjects are scarce. Additionally, the structure of a 3D face is mesh (consisting of vertices and triangles), which is entirely different from a 2D image. Hence, although there are many mature 2D face age transformation methods, this difference could potentially cause such methods to struggle with 3D face age transformations.

To solve the above problems, we first established a new dataset of 3D children's faces to expand the existing 3D face datasets. To increase the number of child subjects in the

- *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.)

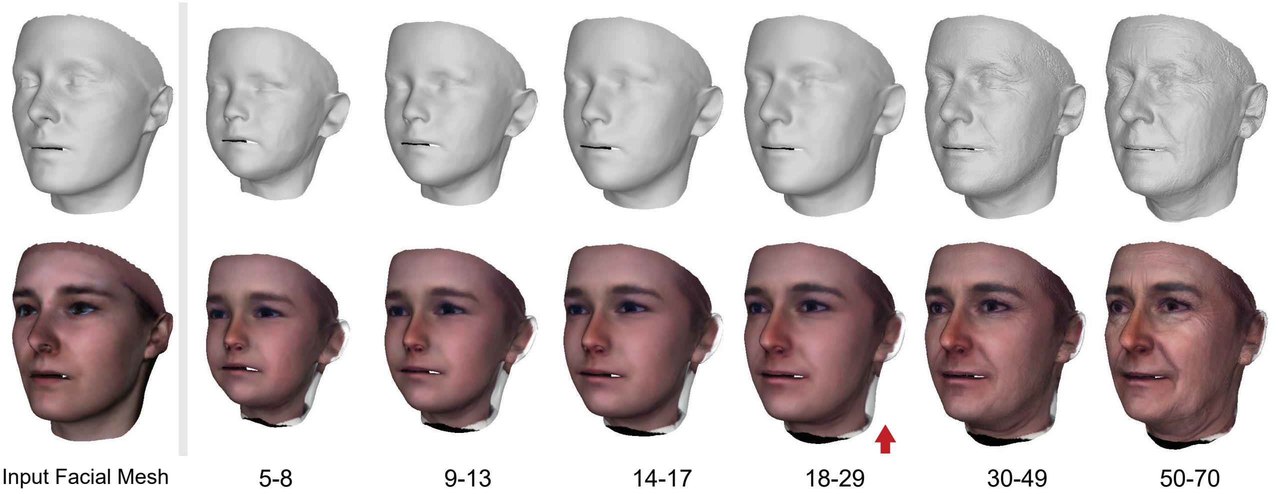


Fig. 1. Bi-directional lifelong age transformation of 3D facial geometries. Given a facial mesh (the red arrow marks the input age group), facial meshes in different age groups are predicted using our MeshWGAN with a multi-task gradient penalty. Rows 1 and 2: 3D facial aging geometries and texture with a 30^o-side pose. To obtain the 3D facial textures, 2D face images at each age were produced from a projected 2D image using SAM [1] and the corresponding 3D facial textures were retrieved using an accurate 3D face reconstruction method [7]. Our method can predict the 3D facial shape and size together, which is consistent with previous anthropometric studies showing that human facial aging is mostly represented by facial growth in children, and by relatively minor shape changes (e.g., skin sagging) and significant texture changes in adults [8], [9], [10].

training dataset, we captured 765 children's faces (ages 5-17 years) and combined them with the publicly available 3D face datasets. We followed previously established methods [2], [18], [19] to approximate the continuous age transformation using a multi-domain age transferring approach and predefine six age groups: three for children (ages 5-8, 9-13, and 14-17), and three for adults (ages 18-29 30-49, and 50-70). After that, we developed a novel mesh-to-mesh conditional Wasserstein generative adversarial network (MeshWGAN) architecture with a multi-task gradient penalty to achieve 3D facial geometric lifelong transformations, as shown in Fig. 1. This method has the ability to represent the desired 3D facial geometric changes across different ages while faithfully preserving the 3D facial geometric identity. Inspired by image-to-image translation architecture (LATS) [2], we designed a novel mesh generator to achieve a mesh-to-mesh transformation. The generator consists of an identity encoder to extract identity features from a facial mesh input, two age mapping networks to produce latent-and-style age spaces from a target age, and a decoder to generate the target facial mesh from the combined latent-and-style age codes and identity features. This differs from that in LATS which only includes style age code, but no latent age code. Additionally, compared with LATS, we also proposed a novel mesh discriminator with different adversarial losses.

Following previous studies [2], [18], [20], we designed a multi-task mesh discriminator (with multiple outputs) to discriminate between multiple age groups [18]. However, compared to Wasserstein GAN (WGAN) without gradient penalty (GP) [21] and WGAN-GP [22] specially designed for a single task, we introduced a novel multi-task gradient penalty to stabilize our multi-task WGAN training. To compute the multi-task gradient penalty, it was assumed that the uniformly sampled facial mesh from two facial meshes from the same age group still belonged to this age group, as facial meshes from the same age group have similar geometric features [16], [23]. Each-task gradient penalty was calculated to ensure the generated face quality in each corresponding age group. Furthermore, in the mesh discriminator, we improved the facial transformation quality by merging the vertex positions and normals as the input, rather than solely the vertex positions.

To the best of our knowledge, this is the first attempt at mesh-to-mesh translation for facial age transformation. Our experimental results demonstrated that our MeshWGAN can predict the 3D facial shape and size together in different ages well, which is consistent with prior anthropometric studies showing that human facial aging is mostly represented by facial growth in children, and by relatively minor shape changes (e.g., skin sagging) in adults [8], [9], [10]. The main contributions of this work are summarized as:

- We propose a novel mesh-to-mesh conditional GAN architecture for 3D facial geometric age transformation. Our generator and discriminator, differing from those in the image-to-image translation architecture (LATS) [2], can produce 3D facial aging meshes with better identity preservation and age closeness.
- We develop a multi-task gradient penalty calculation strategy in the training scheme. It differs from the classic (single-task) WGAN [21], [22], and can more effectively stabilize the multi-task WGAN training.
- We establish a supplementary 3D dataset of real children's faces, effectively addressing the issues of deficiency of child subjects in existing 3D face datasets.

2 Related Work

2.1 3D Face Datasets

The cost and difficulty of capturing 3D face scans are much higher than those for capturing 3D images, resulting in a comparatively small number of 3D face datasets. However, there are a few large-scale publicly available 3D head/face datasets: HeadSpace (1519 subjects aged 1-89 years, predominantly white) [16], FaceScape (938 subjects aged 16-70 years, mainly Asian) [15], FaceWarehouse (150 subjects

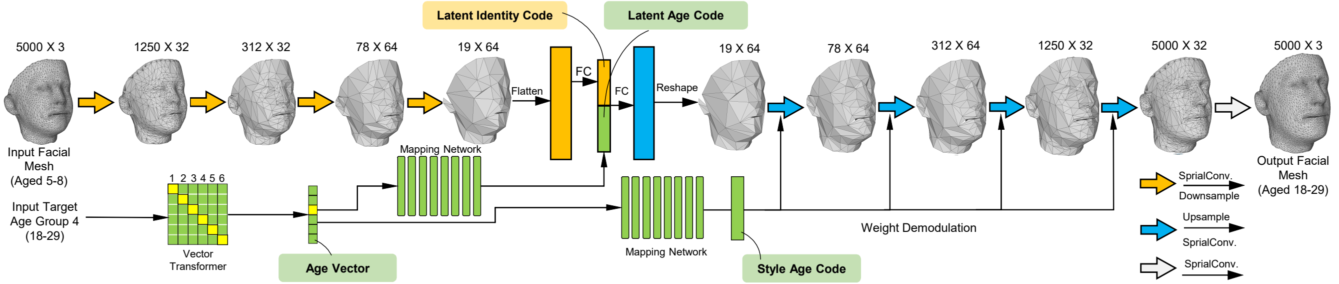


Fig. 2. Model architecture. SpiralConv.: spiral convolution [24] as the mesh convolution. FC: full connection. Weight Demodulation proposed in StyleGAN2 [25] is used in the modulated spiral convolution layers. In the downsampling (upsampling) stages, the new vertices are contracted (recovered) using the predefined barycenteric coordinates of the closest triangle in the decimated mesh [26], where the decimated meshes are generated from the original facial mesh using a surface simplification method based on quadric error metrics [27]. The network receives a source facial mesh and a target age group and outputs a facial mesh of the desired age group. Instead of using image convolution operators, we use the mesh convolution operator as the basic building block to design a novel conditional mesh generator, consisting of three parts, a mesh identity encoder that produces a latent identity code, two mapping networks that produces latent and style age codes from an age vector, and a mesh decoder with modulated convolution layers that produces new meshes from the combination of latent identity and age codes.

aged 7-80 years, various ethnicities) [17], BU-3DFE (100 subjects aged 18-70 years, various ethnicities) [28], Florence (53 white subjects aged 22-61 years) [29]. These datasets are deficient in child subjects, making it challenging to develop methods for lifelong transformations, because facial aging is mostly represented by significant facial geometric growth in children and minor facial geometric changes in adults [12]. To solve this problem, we collected face scans from 765 child subjects aged 5-17 years and combined them with the available face datasets to use for training and validation of our MeshWGAN architecture.

2.2 Face Age Transformation

Previous studies have focused on 2D face age transformations, and many image-to-image translation methods have been developed to translate a given image of a specific age into a new image of a different target age while preserving the individual's identity. The successes of image-to-image translations between two domains (e.g., Pix2Pix [30], CycleGAN [31]) or multiple domains (e.g., StarGAN [19], [32], STGAN [33], FUNIT [18]) provide much inspirations for the task of 2D face aging. In some methods (e.g., SAM [1], HRFAE [34], DAAE [5]), facial age transformation is approached as a continuous-time regression problem to predict the face image with a specific age. However, these methods require large datasets of training subjects in each age and an accurate pretrained age classifier, making them difficult to apply to 3D face aging. In contrast, many methods (e.g., Triple-GAN [35], LATS [2], SFA [20]) approximate this continuous age transformation by representing age with multiple predefined age groups and use a multi-domain image-to-image translation to achieve face aging. One benefit of these methods is reducing the quantity requirement for training datasets. A comprehensive review of 2D face aging research was presented in a recent survey [36]. However, compared to 2D face age transformation, which is a well researched and understood problem, 3D face age transformation is still underexplored. Thus, we propose a mesh-to-mesh transformation to achieve 3D facial mesh aging.

2.3 Mesh Convolution Operators

A 3D facial mesh is composed of vertices and triangles, which is different from 2D images only containing pixels. As a result, image convolution operators cannot be applied directly to the task of 3D face age transformation. Therefore, some studies [37], [38], [39] have converted the 3D facial meshes into 2D domain maps first and then applied standard 2D convolution operators to process them. Many studies [26], [40] have attempted spectral graph convolutions [41] on representations of 3D facial meshes. To process 3D mesh efficiently, a refined spiral convolution (Spiral++) [24], [42] has been developed for 3D facial mesh generation and representation [43], [44]. This mesh convolution operator sufficiently leverages the local geometric features of the mesh surface, their superiority over previous state-of-the-art methods has been demonstrated via experimental results [24], [44]. Therefore, in this study, we adopted Spiral++ as the mesh convolution operator and further enhance our network using weight demodulation in the generator and a multi-task gradient penalty in the discriminator to create 3D facial aging models.

3 Approach

3.1 Overview

Motivated by unpaired image-to-image GAN architectures [2], [18], [19], [31], [32], we propose a multi-domain mesh-to-mesh Wasserstein GAN architecture to achieve lifelong 3D facial geometric age transformations. Since there are no existing large-scale face datasets, we follow previously established methods [2], [18], [19] to approximate continuous age transformation using a multi-domain age transferring approach and predefine six age groups: three for children (ages 5-8, 9-13, and 14-17), three for adults (ages 18-29, 30-49, and 50-70). Furthermore, to achieve the continuous age transformation, an age latent code interpolation is used in our mesh generator. Instead of image convolution, we use a mesh convolution (spiral convolution [24], [44]) as a basic building block to design a novel generative adversarial network architecture consisting of a single conditional mesh generator (see Fig. 2) and a single multi-task discriminator (see Fig. 3). The conditional mesh generator receives a source facial mesh and a target age group and outputs a facial mesh of the desired age group, consisting of three parts: a mesh identity encoder to produce a latent identity code, two mapping networks to produce a latent/style age code

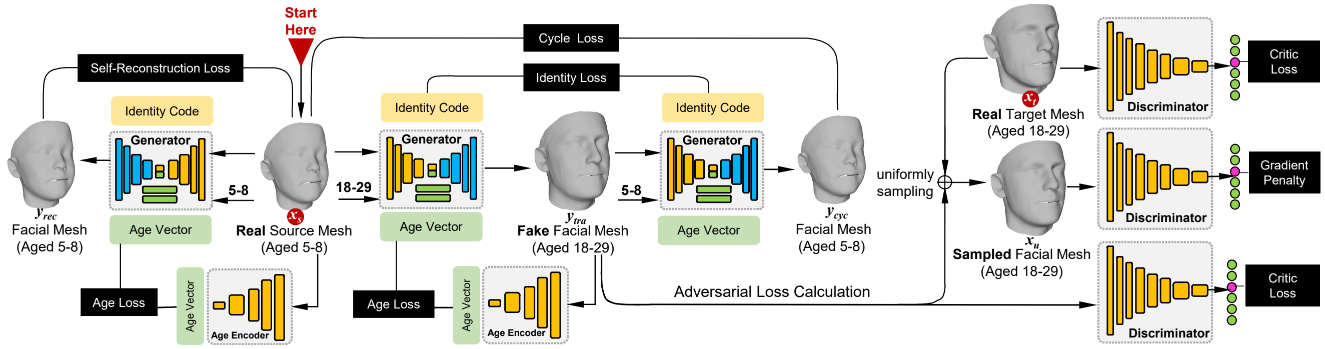


Fig. 3. Training scheme. Two facial meshes $\hat{x}_s$ and $\hat{x}_t$ are sampled from a source age group s and a target age group t ($t \neq s$), respectively. Note that a multi-task gradient penalty was developed to stabilize the Wasserstein GAN training.

from an age vector, and a mesh decoder to produce new meshes from the combination of the identity and age codes. In the multi-task discriminator in Fig. 3, we combine the vertex positions and normals as input, instead of solely the vertex positions. Benefiting from WGAN [21] and WGAN-GP [22], we introduced a multi-task gradient penalty to stabilize our WGAN training. To compute the multi-task gradient penalty, we uniformly sample a new facial mesh between two facial meshes from the same age group.

3.2 Foundations

In our architecture, the spiral convolution is used as a core building block, which sufficiently leverages the local geometric features of the mesh surface. The spiral convolution operator for the ith vertex v_i based on its features x_i is defined as follows [24]:

$$x_i^{(t)} = \coprod^{(t)} \left(\biguplus_{j \in S(i,l)} x_j^{(t-1)} \right), \tag{1}$$

where $\coprod^{(t)}$ and $\biguplus$ donate multi-layer perceptrons and concatenation operation, respectively, and $S(i,l)$ is a dataset in a predefined spiral sequence consisting of l vertices from a concatenation of k-rings. An example of Spiral++ on a facial mesh is shown in Fig. 4. Before training model with Spiral++, the spiral length needs to be determined firstly and then the ordered set of each vertex is confirmed using spiral pattern-based encoding approach [42].

3.3 Architecture

Our model architecture of inference is shown in Fig. 2 and the age encoder and multi-task discriminator are shown in Fig. 3. The conditional generator receives an input facial mesh and a target age group and outputs a facial mesh of the desired age group. In our pre-processing steps, the input facial meshes $\hat{x}$ are parameterized and aligned to a facial mesh template $\bar{m}$.

Conditional Generator We use a predefined vector transformer (proposed in LATS [2]) to convert the input target age group i into a vector α_i with $l \times n$ elements: $\alpha_i = \nu_i + s$, $s \sim N(0, 0.2^2 \cdot \mathbf{I})$, where n is the number of age groups, ν_i is an $l \times n$ element vector that contains ones from $l \times i$ to $l \times (i+1) - 1$ and zeros elsewhere, and $\mathbf{I}$ is the identity matrix. Then, two mapping networks M_z and M_t with an eight-layers MLP network embed an age vector α into a latent code z_{age} and a style code t_{age} with unified N_e elements: $z_{age} = M_z(\alpha)$ and $t_{age} = M_t(\alpha)$, respectively.

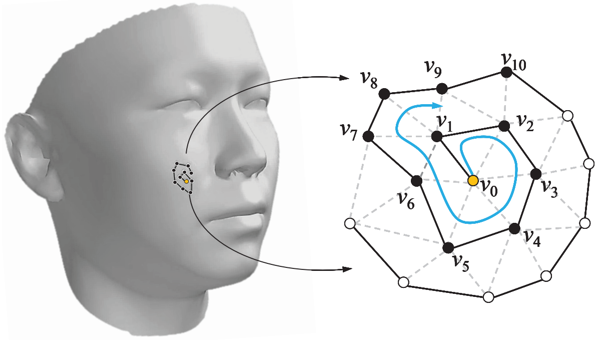


Fig. 4. Example of Spiral++ [42] on a facial mesh. There is one parameter (spiral length) to determine the number of vertices in ordered set.

The identity encoder E_{ide} contains four downsampling layers followed by one fully connected layer. It takes the mesh difference x between an input facial mesh $\hat{x}$ and the facial template $\bar{m}$: $x = \hat{x} - \bar{m}$, and extracts its geometric structure features with N_e elements as a latent identity code: $z_{ide} = E_{ide}(x)$. The benefit of inputting vertex positions differences rather than vertex positions is a reduction in the network learning difficulty, because the facial template has already provided the basic facial structure information in the network. The decoder F contains one fully connected layer and four upsampling layers. To control the facial geometric detail changes, we apply a weight demodulation with the style code t_{age} (proposed in StyleGAN2 [25]) in the modulated spiral convolution layers as shown in Fig. 2. It receives the combination code z with $N_e \times 2$ elements (concatenated from the latent age z_{age} and identity z_{ide} code), and outputs a new mesh difference y at the original size: $y = F(z) = F([z_{ide}, z_{age}], t_{age})$. The new facial mesh $\hat{y}$ can be calculated using the mesh difference y and the facial template $\bar{m}$: $\hat{y} = y + \bar{m}$. The output mesh difference y of our overall generator G from an input facial mesh difference x and an input target age vector α is:

$$y = G(x, \alpha) = F\left([E_{ide}(x), M_z(\alpha)], M_t(\alpha)\right). \tag{2}$$

Multi-Task Discriminator To distinguish between real and fake meshes from multiple age groups, we develop a multi-task mesh discriminator (see Fig. 3). For a real or fake facial mesh from age group i, we penalize only the i-th output. The discriminator D contains eight downsampling layers and four fully connected layers with a minibatch standard deviation to output n values. In each downsampling stage, the number of vertices is reduced by half. In the discriminator, we merge the mesh difference x and normals $\hat{n}$ as the input $[x, \hat{n}]$, rather than solely using the vertex positions. Combined with vertex positions, the vertex normals, as another critical 3D facial features, are used to enhancing the discriminator's distinguishing ability, thereby improving facial transformation quality.

Age Encoder An age encoder of the facial mesh is not used for inference, but only in training (see Fig. 3). The age encoder E_{age} enforces a mapping of the input facial mesh difference x into its corresponding age vector α: $\alpha = E_{age}(x)$. The age encoder contains four downsampling layers (the same ones as that of the identity encoder) and four fully connected layers to output a vector with $l \times n$ elements. In the generator, discriminator and age encoder, the spiral convolution [24] is used as a basic building block. In the downsampling (upsampling) stage of Fig. 2, the new vertices are contracted (recovered) using predefined barycenteric coordinates of the closest triangle in the decimated mesh [26] and the number of vertices is decreased (increased) fourfold.

3.4 Training

To calculate the multi-task gradient penalty and mitigate the imbalance influences between age groups, two facial meshes ($\hat{x}_s$ and $\hat{x}_t$) are sampled from a source age group s and a target age group t ($t \neq s$), respectively, in each training iteration. Fig. 3 shows an overview of our training scheme. Then, three new facial mesh differences are produced from the conditional generator, using

$$y_{rec} = G(x_s, \alpha_s),\ y_{tra} = G(x_s, \alpha_t),\ y_{cyc} = G(y_{tra}, \alpha_s), \tag{3}$$

where, y_{rec} is the self-reconstructed mesh difference at source age group s, y_{tra} is the transformed mesh difference at target age group t and y_{cyc} is the cyclic transformed mesh difference at source age group s from the mesh difference y_{tra}. The corresponding facial meshes are retrieved based on the facial mesh template $\bar{m}$, via:

$$\hat{y}_{res} = y_{rec} + \bar{m},\ \hat{y}_{tra} = y_{tra} + \bar{m},\ \hat{y}_{cyc} = y_{cyc} + \bar{m}. \tag{4}$$

These reconstructed meshes are used to compute their vertices normals $\hat{n}_{res}, \hat{n}_{tra}$ and $\hat{n}_{cyc}$. For each output in multi-task discriminator, the critic losses for real and fake facial meshes, and their gradient penalties are calculated. Particularly, to calculate their gradient penalties, the real and fake facial meshes should come from the same age group, and their uniformly sampled facial mesh is also needed. Hence, a new facial mesh difference x_u is uniformly sampled along straight lines between the target x_t and transformed y_{tra} mesh difference as:

$$x_u = \epsilon \times y_{tra} + (1 - \epsilon) \times x_t, \tag{5}$$

where ϵ is a random number: $\epsilon \sim U[0, 1]$ and the corresponding facial normals are $\hat{n}_u$. These inputs and outputs are used to calculate the objective functions of the model: adversarial loss, self-reconstruction loss, cycle consistency loss, identity preservation loss, and age consistency loss.

Adversarial Loss An adversarial loss $L_{adv}(G, D)$ of WGAN is used to criticize the fake y_{tra} and real x_t facial meshes' differences from the same age group t, and enforce the Lipschitz constraint by calculating a gradient penalty for the random samples x_u, as shown in Fig. 3:

$$\begin{aligned} L_{adv}(G, D) =& \mathbb{E}_{x_t,t}[D([x_t, \hat{n}_t])] - \mathbb{E}_{y_{tra},t}[D([y_{tra}, \hat{n}_{tra}])] + \\ & \lambda \mathbb{E}_{x_u,t}[(\| \bigtriangledown_{[x_u, \hat{n}_u]} D([x_u, \hat{n}_u]) \|_2 - 1)^2], \end{aligned} \tag{6}$$

where λ is usually set as 10 [22]. For a real, fake or sampled facial mesh from age group i, only the i-th element in the output vector of discriminator D is used as the final discriminating result. Unlike WGAN-GP [22] for single-task discriminator, our gradient penalty calculation strategy is especially designed for multi-task discriminator. Additionally, The objective of considering the combination of vertex positions and normals is to improve the discriminator distinguish ability.

Self-Reconstruction Loss A self-reconstruction loss $L_{rec}(G)$, consisting of facial vertex-wise positions' and normals' consistency losses between x_s and y_{rec}, is employed to force the conditional generator to learn the facial mesh identity translation as:

$$L_{rec}(G) = \frac{1}{N} \sum \| x_s - y_{rec} \|_2^2 + \frac{1}{N} \sum (1 - \cos(\hat{n}_s, \hat{n}_{rec})), \tag{7}$$

where, N is the number of vertices in the facial mesh. The normals' consistency loss is calculated by measuring the cosine similarity $\cos(\cdot)$ of each per-vertex normal to guarantee the mesh surface smoothness.

Cycle Consistency Loss A cycle consistency loss $L_{cyc}(G)$ [31] is enforced to help maintain the facial mesh identity, which also consists of facial vertex-wise positions' and normals' consistency losses between x_s and y_{cyc} as:

$$L_{cyc}(G) = \frac{1}{N} \sum \| x_s - y_{cyc} \|_2^2 + \frac{1}{N} \sum (1 - \cos(\hat{n}_s, \hat{n}_{cyc})). \tag{8}$$

Identity Preservation Loss An identity preservation loss $L_{ide}(G)$ is used to enforce the generator to maintain the input facial mesh identity by minimizing the $L1$ distance between the latent identity code of the input x_s and transformed y_{tra} facial meshes as:

$$L_{ide}(G) = \| E_{ide}(x_s) - E_{ide}(y_{tra}) \|_1. \tag{9}$$

Age Consistency Loss An age preservation loss $L_{age}(G)$ is used to enforce the generator to represent the facial geometric features in the target age group by minimizing the the $L1$ distance between the input α_s, α_t and output age vector from the age encoder E_{age} as:

$$L_{age}(G) = \| E_{age}(x_s) - \alpha_s \|_1 + \| E_{age}(y_{tra}) - \alpha_t \|_1. \tag{10}$$

According to the above, G and D are trained to minimize the following optimization loss functions as:

$$\begin{aligned} \min_G \max_D L_{adv}(G, D) + \omega_{rec} L_{rec}(G) + \omega_{cyc} L_{cyc}(G) + \\ \omega_{ide} L_{ide}(G) + \omega_{age} L_{age}(G), \end{aligned} \tag{11}$$

where $\omega_{rec}, \omega_{cyc}, \omega_{ide}$, and ω_{age} are the hyper-parameters for the respective loss terms.

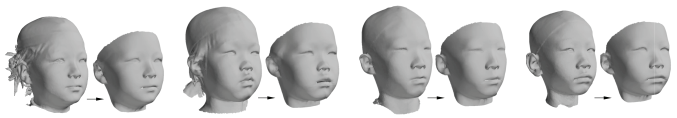


Fig. 5. Examples of our child face scans (on the left arrow) and the resulting parameterized faces (to the right of each arrow) created using the NICP algorithm. We captured and registered the faces of 765 children ages 5-17 years, around 30 female and 30 male children of each age, to compensate for the lack of child subjects in the existing datasets.

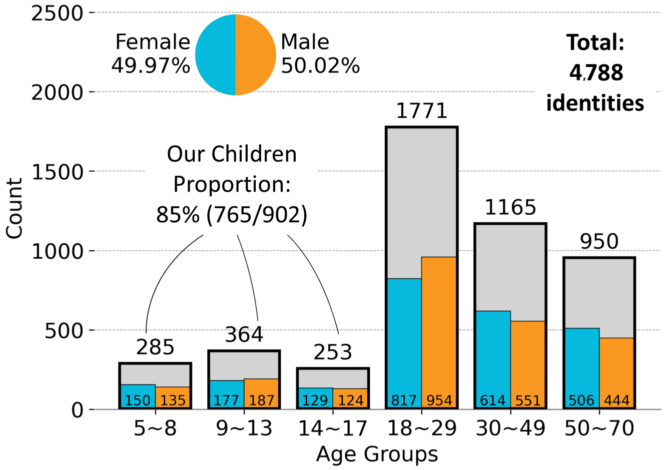


Fig. 6. Age and sex information of subjects in all collected face datasets. Our newly created child dataset includes 765 subjects ages 5-17 years, which accounts for 85% of the total children subjects in training dataset.

4 Experiments

4.1 3D Face Datasets

To establish a large 3D face dataset with sufficient age information, we collected existing 3D head/face datasets, including HeadSpace [16], FaceScape [15], FaceWarehouse [17], BU-3DFE [28], Florence [29], and Adult-Heads [45]. To compensate for this lack of children subjects, we captured 765 children faces ages 5-17 years, as shown in Fig. 5. More information about our children's head dataset is provided in our supplementary file. All facial scans were parameterized using an optimal step nonrigid iterative closest point (NICP) algorithm [46] and aligned to a facial template using procrustes analysis (PA) [47]. All facial meshes have 5,000 vertices and 9,449 triangles. Finally, these datasets - Adult-Heads, HeadSpace, FaceScape, FaceWarehouse, BU-3DFE and Florence, our newly created Children-Faces - provide 1,763, 1,242, 823, 150, 96, 51, and 765 faces, respectively, for a total of 4,890 subjects aged 2-90 years.

These faces were classified into six age groups: three for children (ages 5-8, 9-13, and 14-17) and three for adults (ages 18-29, 30-49, and 50-70). The statistical information of all subjects is shown in Fig. 6. There are around 70.1% Chinese and 29.9% Caucasian. In each group, there are 285, 364, 253, 1,771, 1,165 and 950 identities in sequence, respectively. In total, there are 4,788 identities (2,393 female and 2,395 male), where 75/75 subjects from various age groups of HeadSpace dataset were used as the validation/testing data and the remaining 4,628 as the training data.

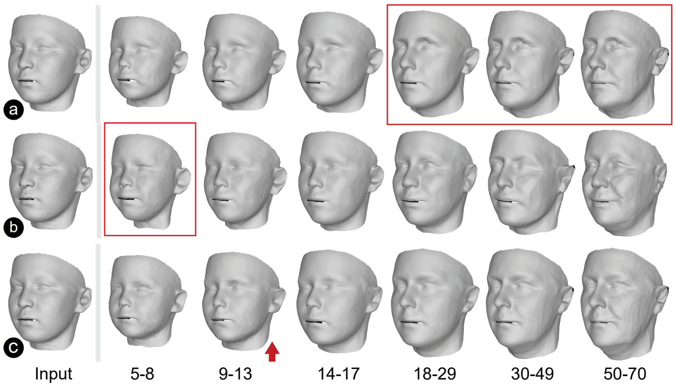


Fig. 7. Qualitative results of ablation study for input age codes in generator (the red arrow marks input age group). (a) With a single latent age code. (b) With a single style age code. (c) With a combination of latent and style age codes. Note that generator (c) has the advantages of both generators (a) and (b), and achieve facial geometric aging while maintaining consistent identity features. Generator (a) is almost identical to Neural3DMM [44], and generator (b) is similar to that of LATS [2] which applies weight demodulation [25] to all layers of the decoder.

4.2 Implementation and Training Details

To train our networks, the Adam optimizer [48] was used with an initial learning rate of 10^{-3}, total iterations of roughly 200k, and a batch size of 16. The learning rate was decayed by 0.5 after 150, 300, and 450 epochs, and the learning rate of the mapping network was decreased by a scale of 0.1. Our model was implemented via Pytorch [49] and PyTorch Geometric [50]. All hyper-parameters - ω_{rec}, ω_{cyc}, ω_{ide}, and ω_{age} - were set as 1. In the generator, the length $l \times n$ in the age vector was set as 50×6 and the lengths N_e of latent age/identity code and style age code were set as 256. Since the facial mesh parameterization is a well-researched and -understood problem, the input facial mesh is produced from face/head scans using the NICP algorithm [46] and aligned to a facial template using the PA method [47]. The output facial meshes are upsampled to a new high-resolution mesh (847,900 vertices and 1,693,440 triangles) for generating high-resolution texture and high-detail geometry (see our supplementary file).

4.3 Ablation Study

We performed four qualitative and quantitative ablation studies in order to prove our main claims, including the combination of latent and style age codes in the generator, the adversarial loss with a multi-task gradient penalty in the discriminator, the input combination of vertex positions and normals in the adversarial loss, and cycle consistency loss.

4.3.1 Qualitative Analysis

In the first study, we compared the single and combination usages of latent and style age codes to demonstrate the superiority of our generator architecture, as shown in Fig. 7. In Fig. 7(a), the generator with a single latent age code is almost identical to Neural3DMM [44] also based on Spiral++ (their main difference is that the input of generator in Fig. 7(a) is facial mesh with 5,000 points, while that of Neural3DMM is head mesh with 5,023 points). The generator in Fig. 7(b) is similar to LATS [2], which applies the weight demodulation [25] with a single style age code to all layers of the decoder,

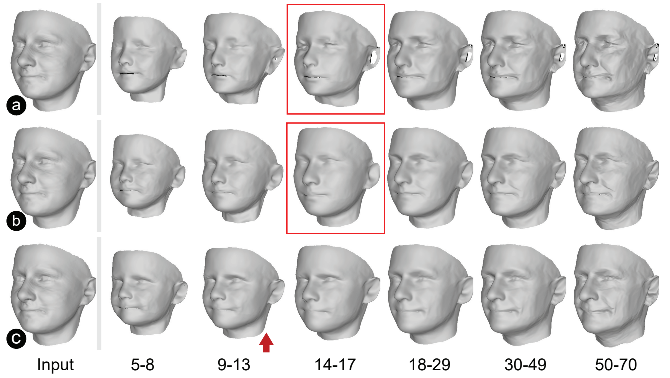


Fig. 8. Qualitative results of ablation study for adversarial loss and input vertex information in multi-task discriminator (the red arrow marks input age group). (a) Using classic multi-task adversarial loss in LATS [2]. (b) Inputting a single vertex position. (c) Using our adversarial loss with multi-task gradient penalty and inputting combination of vertex positions and normals. Note that compared with discriminator (a), our discriminator (c) has better discriminating ability, thereby making the generator produce better facial aging geometries; compared with discriminator (b), our discriminator (c) also has better distinguishing ability, thereby making the generator produce more natural facial aging geometries.

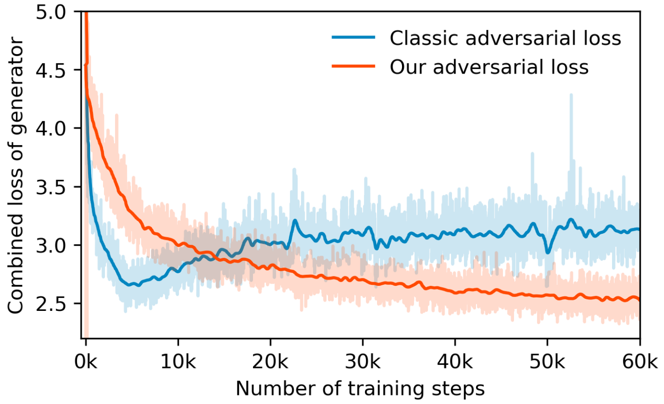


Fig. 9. Trends of generator's combined loss for using two different adversarial losses in multi-task discriminator. While classic GAN suffers from overfitting, ours shows consistent stability in training GAN.

including the full connection and the spiral convolution layers. For the generator in Fig. 7(a), the facial geometries have consistent identity features, but nearly remain unchanged in the three adult groups (indicated by the red rectangles). Particularly, the chin regions of the elderly (50-70) are nearly same as those of the young (18-29), which are inconsistent with the actual situation where the elderly have loose and sagging skin on the chin. In comparison, for the generator in Fig. 7(b), the facial geometries have distinct aging features for children and adult groups, but inconsistent identity features. Specifically, compared to the input face, the child face (5-8, indicated by the red rectangles) has a long and pointed chin. Fortunately, our generator in Fig. 7(c) using both latent and style age codes, can overcome the limitations of both generators and obtain their respective advantages to achieve facial geometric aging with consistent identity features, which proves the superiority of our combination usage of latent and style age codes.

In the second study, we compared our adversarial loss with a multi-task gradient penalty to a classic multi-task adversarial loss (which has been widely used in previous studies, e.g., FUNIT [18], SAF [20] and LATS [2]). This classic loss was also fed by vertex positions and normals of facial mesh in this study, using $L_{adv}(G, D) = \mathbb{E}_{x_t,t}[\log D([x_t, \hat{n}_t])] + \mathbb{E}_{y_{tra},t}[\log(1 - D([y_{tra}, \hat{n}_{tra}]))]$. Here, a non-saturating adversarial loss with R1 regularization, proposed in LATS [2], was used to replace our adversarial loss to train our network. Their comparison results are illustrated in Fig. 8(a) and Fig. 8(c), clearly showing that a generator using classic adversarial loss in Fig. 8(a) can easily lead to distorted facial geometries (indicated by the red rectangles), including eye and ear regions. In comparison, with our adversarial loss in Fig. 8(c), our generator can produce high-quality natural facial geometries. To further demonstrate the ability of our adversarial loss to stabilize training, we compared the training loss $L_{com}(G)$ curves with and without our gradient penalty in multi-task discriminator (see Fig. 9), where $L_{com}(G) = \omega_{rec}L_{rec}(G) + \omega_{cyc}L_{cyc}(G) + \omega_{ide}L_{ide}(G) + \omega_{age}L_{age}(G)$. While classic GAN suffers from overfitting, our proposed approach shows consistent stability in training GAN. It further demonstrates the advantage of using our adversarial loss with a multi-task gradient penalty.

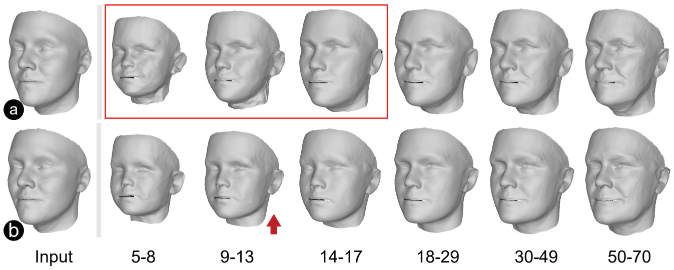
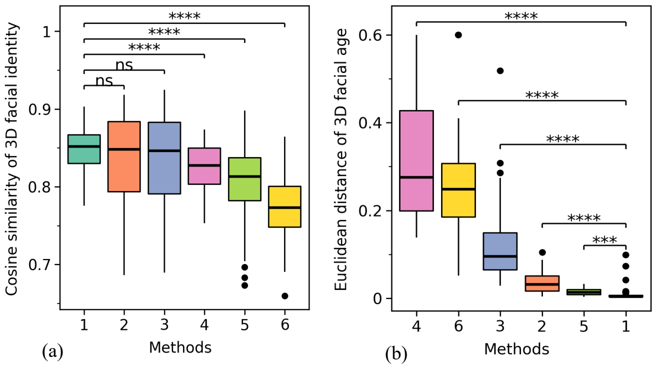


Fig. 10. Qualitative results of ablation study for cycle consistency loss in the training scheme (red arrow marks the input age group). (a) Without cycle consistency loss. (b) With cycle consistency loss. Note that our training scheme (b) can produce more natural facial aging geometries.



Methods	**Identity Preservation**	**Age Closeness**
🟩 1-MeshWGAN	0.846 ± 0.028	0.008 ± 0.014
🟧 2-MeshWGAN with a single vertex position	0.837 ± 0.055	0.036 ± 0.024
🟦 3-MeshWGAN without cycle consistency loss	0.836 ± 0.055	0.118 ± 0.080
🟪 4-MeshWGAN with a single latent age code	0.826 ± 0.029	0.317 ± 0.142
🟩 5-MeshWGAN with a single style age code	0.804 ± 0.047	0.014 ± 0.008
🟨 6-MeshWGAN with classic adversarial loss	0.771 ± 0.037	0.261 ± 0.128

Fig. 11. Quantitative results of ablation study (using the paired-samples t-test). (a) Identity preservation between aging and original 3D facial meshes: average (± standard deviation) cosine similarity of latent identity codes. (b) Age closeness of aging 3D facial meshes: average (± standard deviation) Euclidean distance of age vectors. ns (no significance) =p>0.05; $***$ =p $\leq$ 0.001; $****$ =p $\leq$0.0001.

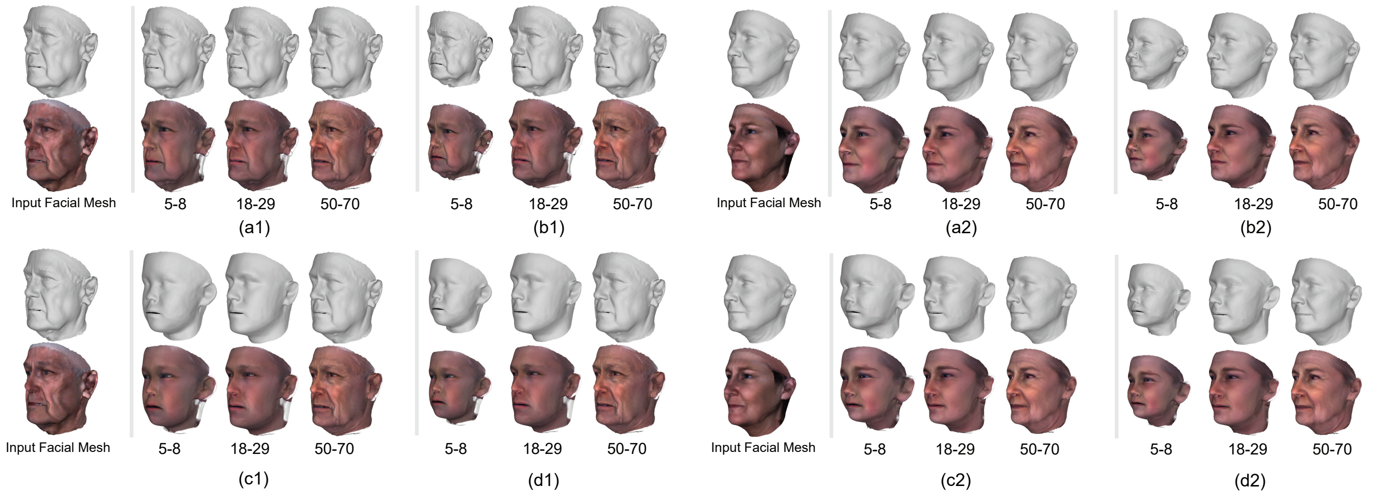


Fig. 12. Qualitative comparisons with 3D trivial baselines using the same texture transformation from SAM [1]. (a) Baseline 1 without geometric transformation. (b) Baseline 2 with the same geometric transformation among age groups. (c) MeshWGAN with unified geometric sizes. (d) MeshWGAN. Note that the unmatching of the textures and geometries in trivial baselines (a) and (b) is noticeable, where the kid's and elder's faces resemble similar even though with different textures; in comparison, our geometric changes between different age groups in our MeshWGAN (c) and (d) are significant (including facial sizes and shapes), especially our kid's facial geometries have fatter cheek and receding chin, which is more consistent with the objective anthropometric features.

In the third study, we showed the importance of using combined vertex positions and normals in the adversarial loss to improve facial transformation quality. Fig. 8(b) and Fig. 8(c) show two sets of facial geometries using a discriminator with combined vertex positions and normals. For the discriminator in Fig. 8(b), that received only a single vertex position, it is evident that the facial geometries of the children's age groups do not resemble the input facial mesh, especially the age groups 5-8 and 14-17. As indicated by the red rectangles (teenager, 14-17), the shapes of cheek, eye, and mouth regions clearly do not look like those of input face; and the overall size is also nearly same as that of the adult (18-29). However, in reality, the teenager's face size should be less than that of the adult. This could be because the single facial vertex position limit the discriminator's distinguishing ability. In comparison, our discriminator in Fig. 8(c) with the combination of vertex positions and normals was able to make our generator produce facial aging geometries with better identity preservation for the Caucasian, even when most of the child subjects in the training dataset are East Asian.

In the fourth study, we demonstrated the importance of using cycle consistency loss to learn age codes in the 3D facial age transformations. Fig. 10 shows two sets of facial geometries using a training scheme without/with cycle consistency loss. For the training scheme without cycle consistency loss in Fig. 10(a), it can be found that the facial geometries of the younger age groups (5-8 and 9-13) (indicated by the red rectangles) have many folds and there are minor but not noticeable changes in shape between the 14-17 and 18-29 groups (indicated by the red rectangles), which is inconsistent with previous anthropometric studies showing that the 3D facial shape grows from adolescence to adulthood [10]. By contrast, our training scheme in Fig. 10(b), which includes cycle consistency loss, can produce more natural aging of facial geometries, which indicates that cycle consistency loss should be included in the training scheme to produce high-quality facial aging.

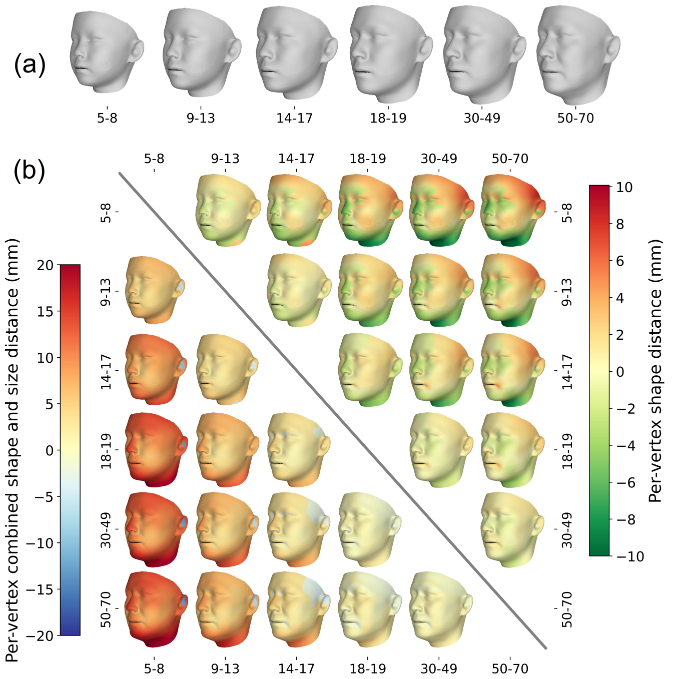


Fig. 13. Visualization of average facial mesh and their differences. (a) Average facial meshes in each age group. (b) Per-vertex distance of combined shape and size (lower left) and shape with unified size (upper right) between two age groups. In the (i, j) per-vertex distance map, the red/blue(green) indicate that the vertex in the ith face mesh is more outwards/inwards on the jth face mesh surface, where i/j is the location of row/column, i/j=1,2,...,6.

4.3.2 Quantitative Analysis

To further support our main claims, in addition to the qualitative analysis, quantitative analysis of the ablation study was also conducted as shown in Fig. 11, including Fig. 11(a) identity preservation, and Fig. 11(b) age closeness. To compare their abilities of identity preservation, the cosine similarity of latent identity codes between aging and original 3D facial meshes was computed. To compare their abilities of

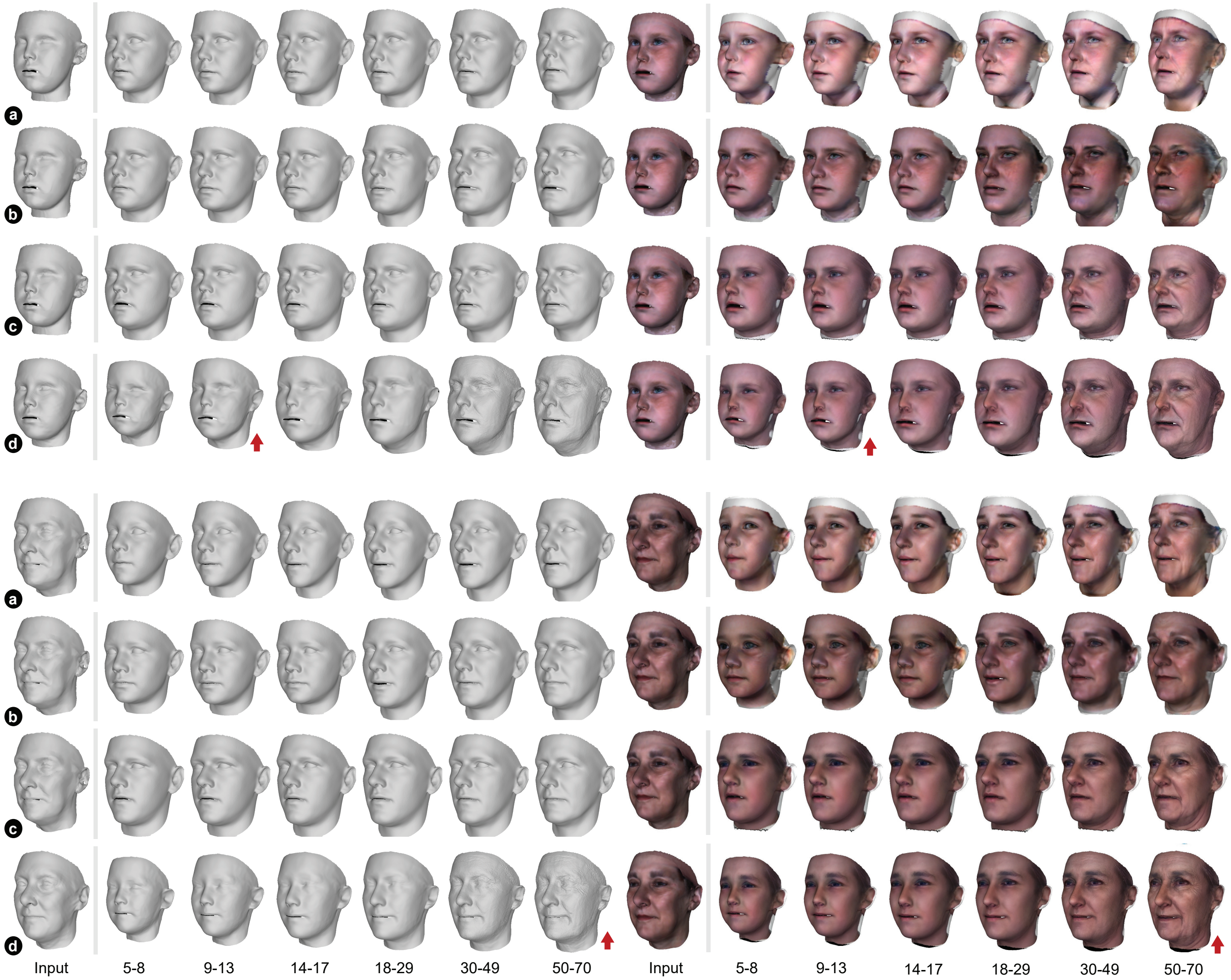


Fig. 14. Qualitative comparisons with state-of-the-art facial aging methods (the red arrow marks input age group). (a) LATS [2] with Deep3DFace [7]. (b) DLFS [51] with Deep3DFace [7]. (c) SAM [1] with Deep3DFace [7]. (d) Our MeshWGAN with 3D facial textures reconstructed using 2D aging images from SAM [1].

age closeness, the Euclidean distance of age vectors for each aging 3D facial mesh was also calculated. The quantitative results (using the paired-samples t-test) in Fig. 11 show our MeshWGAN has the best global performance, our proposed adversarial loss with multi-task gradient penalty in discriminator can most significantly improve the quality of identity preservation, and our proposed combination of latent and style age codes in generator can most significantly increase the accuracy of age closeness. In particular, there are two major differences between MeshWGAN and LATS [2]: the generator and discriminator. In the first ablation study, the qualitative and quantitative comparison of our generator in Fig. 7(c) and LATS's generator [2] in Fig. 7(b) demonstrated that the architecture of our generator is more reasonable and competitive, as our generator can produce the facial meshes with more consistent identity and more accurate age. In the second ablation study, the qualitative and quantitative comparison of adversarial loss in our discriminator in Fig. 8(c) and LATS's discriminator [2] in Fig. 8(a) further demonstrated that our adversarial loss can stabilize training and produce much better results.

4.4 Method Evaluations

Since there are no objective metrics to evaluate the quality of 3D facial age transformation, qualitative comparisons and evaluations were performed to demonstrate our superiority.

4.4.1 Comparisons With 3D Trivial Baselines

To demonstrate the superiority of our method, we compared ours with two 3D trivial baselines: 1. 3D aging faces without geometric transformation, and 2. 3D aging faces with same geometric transformation among age groups. 3D facial textures of SAM were retrieved using a Deep3DFace-based [7] 3D facial texture mapping method (its details are provided in our supplementary file) and applied into all facial aging meshes, as shown in Fig. 12.

In the first baseline without geometric transformation of Fig. 12(a), the unmatching of the textures and geometries is noticeable. Particularly, the faces of the kids aged 5-8 years still resemble the elder, even though it has the rejuvenated textures. This is because that the original facial geometry of the elderly has sagging skins. In comparison, our facial shapes of the younger with unified geometric sizes in Fig.

TABLE 1
Quality evaluation of facial geometries without textures, including Identity Preservation (IP) and Age Closeness (AC). Compared to LATS [2] and SAM [1] with Deep3DFace [7], our MeshWGAN can generate 3D facial geometries with better age closeness and identity preservation.

Item	Method	5-8	9-13	14-17	18-29	30-49	50-70	Overall
IP	LATS [2]	0.11	0.11	0.13	0.08	0.09	0.10	0.10 ± 0.08
	SAM [1]	0.18	0.18	0.16	0.13	0.11	0.09	0.14 ± 0.11
	Ours	**0.71**	**0.71**	**0.71**	**0.79**	**0.79**	**0.81**	**0.75 ± 0.17**
AC	LATS [2]	0.55	0.37	0.31	0.41	0.45	0.41	0.42 ± 0.09
	SAM [1]	0.49	0.33	0.13	0.28	0.39	0.19	0.30 ± 0.09
	Ours	**0.77**	**0.76**	**0.75**	**0.74**	**0.74**	**0.70**	**0.74 ± 0.05**

12(c) are more natural and harmonious when the smooth facial geometries and rejuvenated textures are integrated. Especially, our kid's facial geometries has fatter cheek and receding chin, which is more consistent with the objective anthropometric features.

In the second baseline, average facial meshes $\bar{m}_i$ in each age group were calculated (see Fig. 13(a)) and their mesh difference $d_{i,j}$ of two age groups were computed ($d_{i,j} = \bar{m}_i - \bar{m}_j$, see the lower left of Fig. 13(b)) and applied into the input facial mesh $\hat{x}_j$ (in jth age group) to produce the facial aging meshes: $\hat{x}_i = \hat{x}_j + d_{i,j}$, as shown in Fig. 12(b). It can be clearly found that facial meshes of the younger still remain the facial features of the elderly. By contrast, our rejuvenated facial meshes in Fig. 12(d) have more smooth surfaces, which totally differs from these in Fig. 12(b). This comparison also shows such single facial size changes cannot achieve the 3D facial geometric aging.

To further demonstrate the necessity of facial geometric changes for 3D facial aging, we calculated the per-vertex shape or size distance of pairwise average facial meshes in different age group, as shown Fig. 13(b). The results show that the pairwise facial shapes and dimensions are significantly different and the faces change with the increasing age, which is consistent with previous anthropometric studies [8], [9], [10]. It indicates that the objective facial geometric changes should be achieved for 3D facial aging, including facial shapes and sizes.

4.4.2 Comparisons With 2D Aging Methods

Compared to 2D face aging, 3D facial geometric age transformation is still an underexplored problem. Therefore, we compared our facial geometric aging results with the 3D facial shapes reconstructed from 2D facial aging results produced by three state-of-the-art methods (LATS [2], DLFS [51] and SAM [1]). Both LATS and DLFS are multi-domain translation methods using different age groups, but we were able to leverage its age latent space to produce face images in our age groups. Based on the generated face images, 3D facial shapes were reconstructed using a state-of-the-art 3D face reconstruction method (Deep3DFace) [7]. For our MeshWGAN, because of our 3D facial geometric aging without texture changes, the 3D aging face meshes were rendered and projected using the reconstructed textures from SAM [1] for better visual perception and comparison. Particularly, to reduce the mesh-and-texture mismatches in the elders' facial meshes, we leveraged a pix2pixHD [52] to predict displacement UV map to express and produce 3D facial detailed geometry. Its details are provided in our supplementary file.

TABLE 2
Quality evaluation of facial geometries with textures, including Identity Preservation (IP), and Age Closeness (AC). Our aging facial geometries can significantly improve the original aging visual quality of SAM [1].

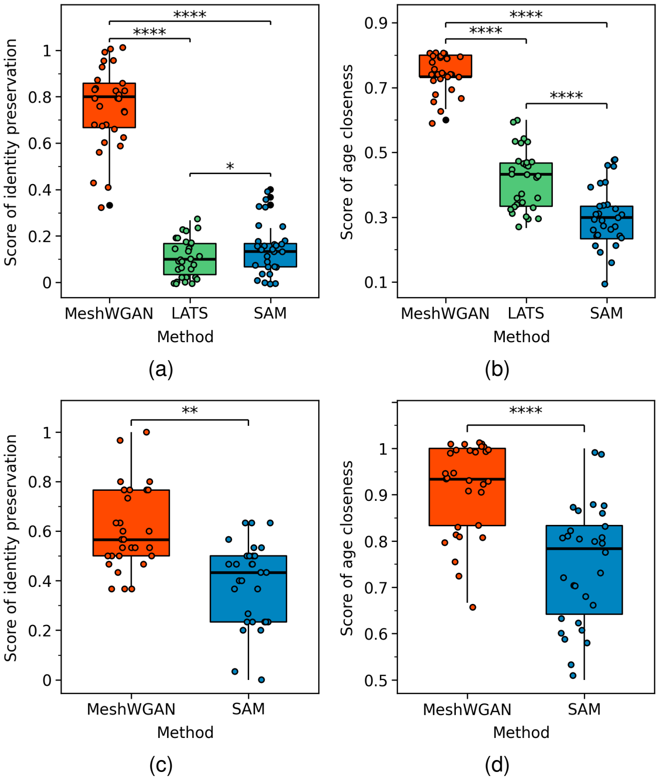
Item	Method	5-8	9-13	14-17	18-29	30-49	50-70	Overall
IP	SAM [1]	0.64	0.66	0.65	0.85	0.79	0.89	0.74 ± 0.13
	Ours	**0.92**	**0.93**	**0.91**	**0.92**	**0.87**	**0.93**	**0.91 ± 0.10**
AC	SAM [1]	0.46	0.42	0.41	0.37	0.36	0.32	0.39 ± 0.04
	Ours	**0.54**	**0.58**	**0.59**	**0.63**	**0.64**	**0.68**	**0.61 ± 0.16**

The qualitative comparisons of different age transformation methods are shown in Fig. 14. The facial shapes reconstructed using SAM do not contain size information and are nearly unchanged with increasing age. The facial shapes created using LATS/DLFS change with age for children ages 5-17, but they are similar for adults ages 18-70. Furthermore, the reconstructed facial shapes do not visually resemble the input facial mesh. In comparison, the aging pattern of our facial geometries is consistent with previous anthropometric studies showing that the human facial aging is mostly represented by facial growth in children, and by relatively large texture changes and minor shape changes in adults [8], [9], [10]. In particular, our elderly facial geometries have marked skin sagging on the cheek and chin regions, and wrinkles on the forehead and the corners of eyes. In addition, our facial geometries visually resemble the input facial mesh, which indicates that they have consistent identity features. Thereby, it is clear that our facial geometries are an improvement over the visual quality of 3D aging via SAM when the reconstructed 3D facial textures from SAM are used.

4.4.3 Human Evaluations

To demonstrate the superiority of our MeshWGAN further, we conducted a user perceptual study to evaluate our results and those created using LATS [2] and SAM [1]. We recruited thirty respondents with experience in 3D graphics/animation design. They were asked to evaluated each generated facial mesh in terms of identity preservation and age closeness. In the experiment, ten facial meshes were input into each method, and the corresponding facial meshes for each of our six age groups are rendered with solid colors using the same direct lighting environment, as well as projected into two color images (512×512 pixels) showing front and 30^o-side views respectively.

For measuring identity preservation, we mixed and showed the rendered images (with the same identity and age group) generated from three methods side-by-side, and asked participants to select which image best portrayed an individual with a consistent identity. The evaluated metric was defined as the percentage of respondents who preferred each method. To measure age closeness for each method, we adopted another approach in which the participants were asked to assign the mixed rendered images of the same identity in each of our six age groups to the estimated age



(a)

(b)

(c)

(d)

Fig. 15. Overall human evaluation results from 30 respondents. (a) Identity preservation evaluation of facial geometries without textures. (b) Age closeness evaluation of facial geometries without textures. (c) Identity preservation evaluation of facial geometries with the same textures from SAM [1]. (d) Age closeness evaluation of facial geometries with the same textures from SAM [1]. $*$ =p $\leq$ 0.05; $**$ =p $\leq$ 0.01; $* * **$ =p $\leq$0.0001. Note that the paired-samples t-test results show that the p-value between our MeshWGAN and LATS/SAM in each subfigure is less than 0.05, which indicates that our method is more competitive.

group. The evaluated metric was defined as the percentage of images correctly assigned by respondents. Their evaluation results are shown in Tables 1 and overall evaluation comparisons are shown in Fig. 15(a) and Fig. 15(b). The paired-samples t-test results show the p-values between our MeshWGAN and LATS/SAM in terms of identity preservation and age closeness are less than 0.05, which indicates that our method can produce aging facial geometries with more consistent identity and age features.

To further demonstrate that our facial geometries can improve the visual quality of 3D aging in images rendered using SAM, we used the same method to measure identity preservation and age closeness of images generated via two methods - our MeshWGAN aging facial geometries with SAM textures, and SAM aging facial geometries with SAM textures - as shown in Table 2, Fig. 15(c) and Fig. 15(d). The paired-samples t-test result had a p-value of less than 0.05, which indicates our aging facial geometries can significantly improve the original visual quality of facial aging images generated by SAM. This also demonstrates the superiority to achieve 3D facial texture and geometry aging together.

4.4.4 Age Interpolation

Although we only used facial meshes from different age groups to train our model, there is one approach to achieve continuous age transformations using age code interpolation, as shown in Fig. 16. Our model possesses the ability to generate continuous age transformations by interpolating a new latent age code $\ddot{z}_{age}$ between two other latent age codes z^{i}_{age} and z^{i+1}_{age} generated from two neighboring age groups i and $i+1$: $\ddot{z}_{age} = \epsilon \times z^{i}_{age} + (1-\epsilon) \times z^{i+1}_{age}$, as well as inserting a new style age code $\ddot{t}_{age}$ between the two other style age codes t^{i}_{age} and t^{i+1}_{age} generated from two neighboring age groups i and $i+1$: $\ddot{t}_{age} = \epsilon \times t^{i}_{age} + (1-\epsilon) \times t^{i+1}_{age}$ where ϵ is an interpolation parameter within [0, 1]. Then, a new facial mesh $\ddot{y}$ is produced using the decoder F from a concatenation of the input latent identity code z_{ide}, the new latent age code $\ddot{z}_{age}$ and the new style age code $\ddot{t}_{age}$: $\ddot{y} = F([z_{ide}, \ddot{z}_{age}], \ddot{t}_{age})$. From Fig. 16, it can be seen that there is obvious continuous facial geometric growth with preserved identity, which indicates our method successfully achieved continuous age transformations using the interpolation of latent and style age codes.

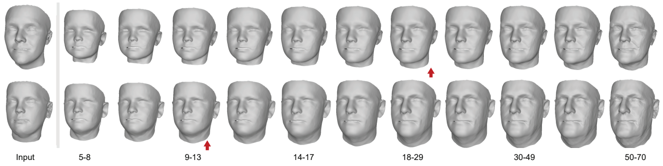


Fig. 16. Continuous age transformations using interpolation of latent and style age codes (the red arrow marks input age group).

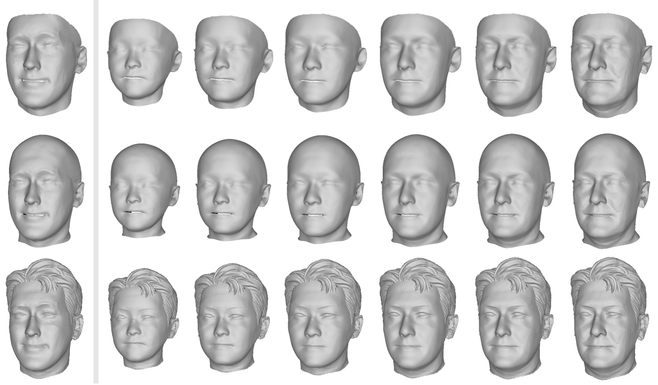


Fig. 17. Head completion (second row) and avatar creation (third row) from 3D aging faces (first row). With our faces, full heads can be predicted using a face-to-head model regression method [54]. In the third row, changes of hair and hairstyle are not represented.

4.4.5 Runtime Analysis

Runtime performance was tested on a computer with an NVIDIA GeForce RTX 3090 GPU (24GB of memory). After a face scan is received, there are only two main steps: facial mesh parameterization and aging face generation. Mesh parameterization from a face scan using AMSGrad-based [53] NICP algorithm, as an optimization method, takes around 77.5 seconds. Then, the generation of facial meshes in our six age groups using our MeshWGAN took around 2.4 seconds.

4.5 Application Scenarios

Our proposed 3D face age transformation method can be applied to many face-related 3D graphics applications, e.g., 3D aging figures/avatar creation, 3D age-invariant face recognition, 3D facial data augmentation, 3D facial attribute editing. In this section, we show two typical applications.

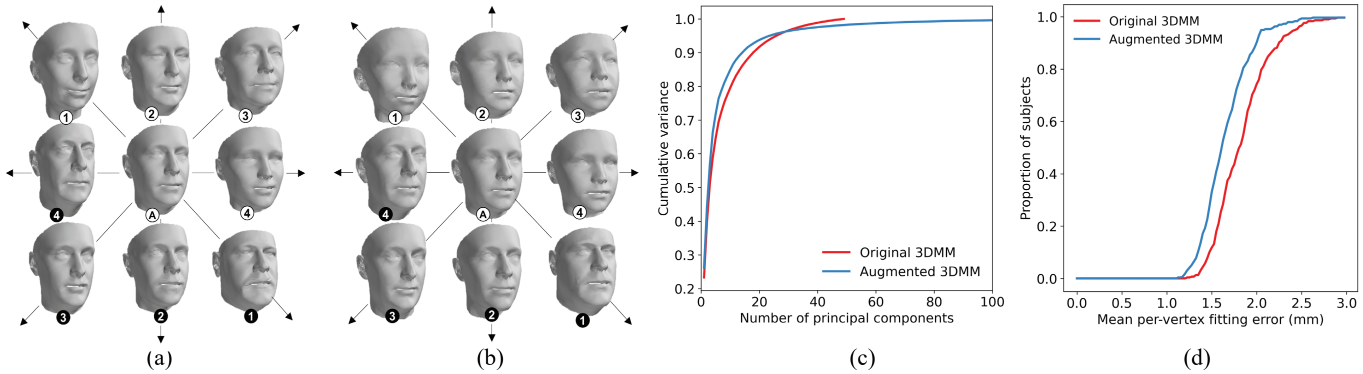


Fig. 18. Generalization ability comparison of two 3DMM. (a) Original 3DMM. (b) Augmented 3DMM, with average shape (shown in center) and first four principal components (PCs) with a weight of 3σ_i (white circle labels) and -3σ_i (black circle labels), where σ_i is the standard deviation of the corresponding PC. (c) Cumulative explained variations. Respectively, 44 PCs in original/augmented 3DMM explain 99.50%/97.94% of their training dataset variance. (d) Mean reconstructed mesh errors.

4.5.1 3D Aging Avatar Creation

Creating realistic digital humans is an increasingly important task in various immersive applications [55] and 3D aging avatar design can bring much fun to this task, especially in the metaverse. Furthermore, 3D aging figures in animation are still created manually by professional designers, a manual workload that can be significantly reduced with the assistance of our proposed method. Fig. 17 shows examples of head completed and avatar created from our 3D aging faces. A face scan can be easily captured using 3D scanners, e.g., Artec Eva3D scanner, smartphone 3D scanner, and then parameterized and aligned with a facial template using using the NICP algorithm [46] and PA method [47]. With aging faces derived from the parameterized face, the full heads can be predicted using a face-to-head model regression method [54], [56] and then transferred to the avatar, with face preservation based on a digital human face template with hair and other accessories.

4.5.2 3D Face Data Augmentation

3D face datasets are critically important for achieving the 3D age-invariant face recognition and building a powerful 3D morphable model (3DMM). The 3DMM is widely used in the task of 3D face reconstructions from 2D images and 3D scans [57]. With our 3D face aging method, one input face can generate five faces of different ages, which can augment the original dataset effectively. To demonstrate this, we used original 50 faces and their augmented 300+50 faces to establish two 3DMMs using a principal components analysis (PCA)-based method [23] with unifying their face sizes using general procrustes analysis (GPA) [47] and comparing their generalization ability with an additional 300 subjects, as shown in Fig. 18. For a fair comparison, we selected the same number (44) of principal components (PCs) that can explain 99.50%/97.94% in original/augmented 3DMM of the training dataset variance. The average (± standard deviation) distance between the reconstructed and original facial meshes are 1.83 (± 0.31) mm and 1.65 (± 0.27) mm for the original and augmented 3DMMs, respectively, and their p-value is less than 0.005 using a paired-samples t-test. This demonstrates that our method can significantly augment the generalization ability of a 3DMM.

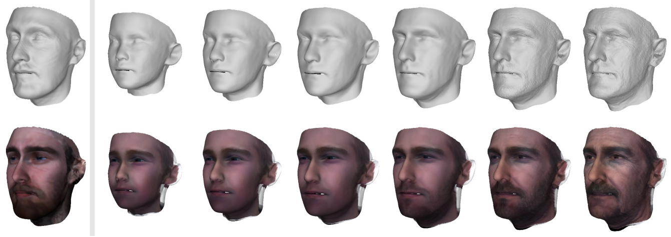


Fig. 19. Limitation of the proposed method. For example, facial hair (e.g., long beards) may lead to inaccurate generation of the chin region.

5 Conclusion and Future Work

We devised a new MeshWGAN architecture with a multi-task gradient penalty to model a continuous bi-directional 3D facial geometric aging. Various experiments showed that our method can predict better facial geometry across different age groups and produce facial geometries more consistent with the input face in the same age group, and the geometric aging process of our method is consistent with previous anthropometric studies. Our model successfully achieved continuous age transformations via age codes' interpolation. However, our method has its limitation. One is that the facial aging geometries are affected by facial hair (e.g., long beards, see Fig. 19), which may result in inaccurate generation of facial regions. This is because the training dataset contains few such uncommon subjects. Besides, to ensure enough training data in each age group, the age ranges within these defined groups (especially 30-49) may be broad. Developing a way to solve these limitations and extending our work to broad 3D face-related translations are directions well worth exploring in future studies.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions.

References

[1] 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.

[2] 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.

[3] P. Li, B. Sheng, and C. L. P. Chen, "Face sketch synthesis using regularized broad learning system," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5346–5360, 2022.

[4] B. Sheng, P. Li, C. Gao, and K.-L. Ma, "Deep neural representation guided face sketch synthesis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 12, pp. 3216–3230, 2019.

[5] P. Li, H. Huang, Y. HuXiang, W. He, and Z. Sun, "Hierarchical face aging through disentangled latent characteristics," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 86–101.

[6] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proceedings of the IEEE International Conference on Image Processing*, 2017, pp. 2089–2093.

[7] 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.

[8] A. M. Albert, K. Ricanek, and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Science International*, vol. 172, no. 1, pp. 1–9, 2007.

[9] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010.

[10] H. S. Matthews, R. L. Palmer, G. S. Baynam, O. W. Quarrell, O. D. Klein, R. A. Spritz, R. C. Hennekam, S. Walsh, M. Shriver, S. M. Weinberg, B. Hallgrimsson, P. Hammond, A. J. Penington, H. Peeters, and P. D. Claes, "Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism," *Scientific reports*, vol. 11, no. 1, pp. 12 175:1–12 175:12, 2021.

[11] W. A. P. Smith, A. Seck, H. Dee, B. Tiddeman, J. Tenenbaum, and B. Egger, "A morphable face albedo model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5010–5019.

[12] Y. Wu, R. Wang, M. Gong, J. Cheng, Z. Yu, and D. Tao, "Adversarial UV-transformation texture estimation for 3D face aging," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4338–4350, 2022.

[13] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial attribute editing by only changing what you want," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5464–5478, 2019.

[14] Z. He, M. Kan, S. Shan, and X. Chen, "S2GAN: Share aging factors across ages and share aging trends among individuals," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9439–9448.

[15] 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.

[16] H. Dai, N. Pears, W. Smith, and C. Duncan, "Statistical modeling of craniofacial shape and texture," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 547–571, 2020.

[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] M.-Y. Liu, X. Huang, A. Mallya, T. Karras, T. Aila, J. Lehtinen, and J. Kautz, "Few-shot unsupervised image-to-image translation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 10 550–10 559.

[19] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8185–8194.

[20] M. Georgopoulos, J. Oldfield, M. A. Nicolaou, Y. Panagakis, and M. Pantic, "Enhancing facial data diversity with style-based face aging," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 66–74.

[21] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the International Conference on Machine Learning*, vol. 70, 2017, pp. 214–223.

[22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5769–5779.

[23] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3D morphable models," *International Journal of Computer Vision*, vol. 126, no. 2, pp. 233–254, 2018.

[24] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, "SpiralNet++: A fast and highly efficient mesh convolution operator," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 4141–4148.

[25] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8107–8116.

[26] 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.

[27] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proceedings of the ACM SIGGRAPH*, 1997, pp. 209–216.

[28] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 211–216.

[29] A. D. Bagdanov, A. Del Bimbo, and I. Masi, "The florence 2D/3D hybrid face dataset," in *Proceedings of the Joint ACM Workshop on Human Gesture and Behavior Understanding*, 2011, pp. 79–80.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5967–5976.

[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2242–2251.

[32] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

[33] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo, and S. Wen, "STGAN: A unified selective transfer network for arbitrary image attribute editing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3668–3677.

[34] X. Yao, G. Puy, A. Newson, Y. Gousseau, and P. Hellier, "High resolution face age editing," in *Proceedings of the International Conference on Pattern Recognition*, 2021, pp. 8624–8631.

[35] H. Fang, W. Deng, Y. Zhong, and J. Hu, "Triple-GAN: Progressive face aging with triple translation loss," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 3500–3509.

[36] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, "Longitudinal face aging in the wild - Recent deep learning approaches," *arXiv:1802.08726*, pp. 1–8, 2018.

[37] S. Moschoglou, S. Ploumpis, M. A. Nicolaou, A. Papaioannou, and S. Zafeiriou, "3DFaceGAN: Adversarial nets for 3D face representation, generation, and translation," *International Journal of Computer Vision*, vol. 128, no. 10, pp. 2534–2551, 2020.

[38] B. Gecer, A. Lattas, S. Ploumpis, J. Deng, A. Papaioannou, S. Moschoglou, and S. Zafeiriou, "Synthesizing coupled 3D face modalities by trunk-branch generative adversarial networks," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 415–433.

[39] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "GANFIT: Generative adversarial network fitting for high fidelity 3D face reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1155–1164.

[40] Z.-H. Jiang, Q. Wu, K. Chen, and J. Zhang, "Disentangled representation learning for 3D face shape," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 957–11 966.

[41] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proceedings of the International Conference on Learning Representations*, 2014, pp. 1–14.

[42] I. Lim, A. Dielen, M. Campen, and L. Kobbelt, "A simple approach to intrinsic correspondence learning on unstructured 3D meshes,"

in *Proceedings of the European Conference on Computer Vision Workshops*, 2018, pp. 349–362.

[43] N. Olivier, K. Baert, F. Danieau, F. Multon, and Q. Avril, "FaceTuneGAN: Face autoencoder for convolutional expression transfer using neural generative adversarial networks," *Computers & Graphics*, vol. 110, pp. 69–85, 2023.

[44] G. Bouritsas, S. Bokhnyak, S. Ploumpis, S. Zafeiriou, and M. Bronstein, "Neural 3D morphable models: Spiral convolutional networks for 3D shape representation learning and generation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7212–7221.

[45] 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.

[46] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid ICP algorithms for surface registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[47] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, 2015, pp. 1–15.

[49] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 721:1–721:12.

[50] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019, pp. 1–9.

[51] 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.

[52] 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.

[53] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," in *Proceedings of the International Conference on Learning Representations*, 2018, pp. 1–23.

[54] S. Ploumpis, E. Ververas, E. O' Sullivan, S. Moschoglou, H. Wang, N. Pears, W. A. P. Smith, B. Gecer, and S. Zafeiriou, "Towards a complete 3D morphable model of the human head," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4142–4160, 2021.

[55] L. Bao, X. Lin, Y. Chen, H. Zhang, S. Wang, X. Zhe, D. Kang, H. Huang, X. Jiang, J. Wang, D. Yu, and Z. Zhang, "High-fidelity 3D digital human head creation from RGB-D selfies," *ACM Transactions on Graphics*, vol. 41, no. 1, pp. 3:1–3:21, 2021.

[56] 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.

[57] B. Egger, W. A. P. Smith, A. Tewari, S. Wuhrer, M. Zollhoefer, T. Beeler, F. Bernard, T. Bolkart, A. Kortylewski, S. Romdhani, C. Theobalt, V. Blanz, and T. Vetter, "3D morphable face models–Past, present, and future," *ACM Transactions on Graphics*, vol. 39, no. 5, pp. 157:1–157:38, 2020.

Jie Zhang received the B.Sc. degree in textile engineering from the Jiangnan University, Wuxi, China, in 2014. He is currently pursuing the Ph.D. degree in design and a Research Associate with The Hong Kong Polytechnic University, Hong Kong. He has published over 30 peer-reviewed journal articles, 3 textbooks, patents and conference papers. His current research interests include AI for design and graphics, 3D face modeling/editing, 3D generative models, color science, and 3D digital human modeling.

Kangneng Zhou received the B.Sc. degree in internet of things from the University of Science and Technology Beijing, Beijing, China, in 2020. He is currently pursuing the M.Eng. degree in computer science and technology with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China. His current research interests include generative models, face synthesis, computer graphics, and deep learning.

Yan Luximon received the Ph.D. degree in ergonomics from The Hong Kong University of Science and Technology, Hong Kong, in 2006. She is currently an Associate Professor with the School of Design, The Hong Kong Polytechnic University, Hong Kong. She is also the Leader for Asian Ergonomics Design Lab and the Deputy Discipline Leader for B.A. Product Design. She has published over 100 peer-reviewed journal articles, book chapters, patents and international conference papers. Her current research interests include computer graphics, 3D digital human modeling and CAD, AI in design and visualization, 3D head and face reconstruction, human-computer interaction, deep learning, and virtual reality.

Tong-Yee Lee (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Washington State University, Pullman, in 1995. He is currently a Chair Professor with the Department of Computer Science and Information Engineering, National Cheng-Kung University (NCKU), Tainan, Taiwan. He leads the Computer Graphics Group, Visual System Laboratory, NCKU (http://graphics.csie.ncku.edu.tw). His current research interests include computer graphics, non-photorealistic rendering, medical visualization, virtual reality, and media resizing. He is a Senior Member of the IEEE and a Member of the ACM. He is an Associate Editor of the *IEEE Transactions on Visualization and Computer Graphics*.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2013. He is currently an Assistant Professor with the Department of Computing and an Assistant Professor with the School of Design, The Hong Kong Polytechnic University, Hong Kong. He has published over 200 top-tier scholarly research articles (e.g., TVCG, TIP, TNNLS, TMI, TMM, TCSVT, TCYB, TBME, TSMC, TII, AAAI, CVPR, NeurIPS), pioneered several new research directions, and made a series of landmark contributions in his areas. He has an excellent research project reported by the *ACM TechNews*, which only reports the top breakthrough news in computer science worldwide. More importantly, however, many of his research outcomes have strong impacts to research fields, addressing societal needs and contributed tremendously to the people concerned. His current research interests include image/video stylization, colorization, artistic rendering and synthesis, realism in non-photorealistic rendering, computational art, and creative media.