3DGrowthNet: A Deep Learning Model for Synthetic Aging and Conditional Shape Generation Using 3D Facial Meshes

Nina CLAESSENS [*] [1,2], Susan WALSH [3], Mark SHRIVER [4], Seth M WEINBERG [5], Paolo M CATTANEO [6], Anthony J PENINGTON [7,8,9], Peter CLAES [1,2,10]
[1] UZ Leuven, Medical Imaging Research Center, Leuven, Belgium;
[2] Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium;
[3] Department of Biology, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA;
[4] Department of Anthropology, Pennsylvania State University, State College, PA, USA;
[5] Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA;
[6] Melbourne Dental School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia;
[7] Facial Sciences Research Group, Murdoch Children's Research Institute, Parkville, Australia;
[8] Department of Plastic and Maxillofacial Surgery, Royal Children's Hospital, Melbourne, Australia;
[9] Department of Pediatrics, University of Melbourne, Melbourne, Australia;
[10] KU Leuven, Human Genetics Department, Leuven, Belgium

https://doi.org/10.15221/25.27

Abstract

Accurately modeling facial growth is essential for applications in forensic science, clinical genetics, and developmental biology. We present 3DGrowthNet, a multi-task geometric deep learning framework that performs continuous synthetic aging, age and sex estimation, and conditional shape generation from 3D facial meshes. We introduce a multi-task training strategy that unifies existing synthetic aging frameworks with conditional shape generation and a continuous label embedding mechanism into a single CVAE-GAN architecture. This integration enables the network to disentangle age from identity while learning to generate anatomically plausible faces across the full age range of 0-88 years.

Trained on over 5,000 scans and validated using a smaller longitudinal dataset of 60 children, the model achieves a mean prediction error of ~2 mm, improving age-invariant identification performance by nearly 20%. It also generates realistic, demographically consistent synthetic faces with high coverage (98.7%) and low Minimum Matching Distance, supporting robust data augmentation. Biomedical relevance is demonstrated through simulations of sexual dimorphism across age, revealing expected developmental trends even in sparsely sampled age ranges. Experiments show that the model is capable of generating a wide variety of realistic and demographically consistent 3D faces and supports robust data augmentation across the age spectrum. In addition to its generative capabilities, 3DGrowthNet performs age and sex estimation resulting in a median absolute age estimation error of 2.0 years and an overall sex classification accuracy of 87.7%, with performance varying across age groups. These results confirm that the model effectively encodes biologically relevant demographic information.

3DGrowthNet sets a new benchmark for realistic, demographically informed mesh synthesis and provides a foundation for advancing personalized growth modeling, forensic identification, and clinical assessment of facial dysmorphism.

Keywords: facial growth, geometric deep learning, synthetic aging, conditional shape generation, age prediction, sex classification, 3D surface scans, mesh synthesis

1. Introduction

Accurately predicting growth and estimating age are central to various forensic domains, including cases of missing children and disaster victim identification. At the same time, understanding how normal facial shape variation evolves over the human lifespan is crucial in biomedical fields such as clinical genetics, pediatric surgery, and orthodontics. Because facial growth is a complex process influenced by genetic, biological, and environmental factors, its modeling requires approaches that capture detailed 3D morphology and treat age as a continuous, nonlinear variable.

* Corresponding author: Nina Claessens, email: nina.claessens@kuleuven.be

Despite recent progress, the field still faces several limitations. Firstly, most existing facial growth models rely on 2D images [1-5], which lack the rich structural information present in 3D data. This is partly due to the limited availability of large-scale 3D datasets and the relative scarcity of general-purpose deep learning frameworks for 3D data. Additionally, when large-scale datasets are not available—as is usually the case in 3D studies—age is often discretized into bins [1,3,5], yet such encoding sacrifices precision and ignores correlations between neighboring age bins. Lastly, existing 3D growth models are often based on linear approximations or sparse landmarks [6-10], leaving many discriminative aspects of the face underexplored.

Zhang et al. recently introduced MeshWGAN [11], the first geometric deep learning model to predict facial growth from 3D surface scans. The MeshWGAN architecture consists of a conditional autoencoder (CAE) in combination with a generative adversarial network (GAN) and was based on the two-dimensional LATS framework [12]. The model successfully captured age progression across six discrete bins spanning 5 to 70 years. However, the use of such broad age intervals limits precision, particularly at younger ages where the face changes rapidly. Furthermore, the accuracy of the aging simulations was assessed only through visual inspection, as no longitudinal data were available for quantitative evaluation.

To overcome these limitations, we propose 3DGrowthNet, a unified framework trained on 3D facial meshes that enables synthetic aging, age and sex estimation, and conditional face generation through continuous age conditioning. Adapting the model architecture of LATS and MeshWGAN, we included a variational autoencoder (VAE) to enable shape generation and a label input mechanism based on the approach of Ding et al. [13] to encode age as a continuous variable (Fig. 1). This label input mechanism allows the model to predict growth and generate shapes at any continuous age. The growth predictions were validated using a longitudinal dataset of 60 children and the added value in real-world scenarios was tested in a biometric identification experiment.

In addition to predicting individual growth trajectories, we further propose a multi-task training scheme that integrates the LATS training scheme with existing generative learning frameworks, which makes it possible to generate an unlimited number of shapes for any specified age and sex. This is valuable for augmenting existing 3D datasets, particularly in age ranges where data are typically scarce or in situations where privacy constraints prevent the use of real scans. Beyond its role in data augmentation, the model provides valuable insight into how normal facial morphology changes over time, which is essential for studying biological phenomena like sexual dimorphism. Moreover, this characterization of normal facial development opens the possibility to detect and quantify facial dysmorphism, which is crucial in clinical fields such as syndrome classification. Finally, our model can estimate age and sex by minimizing the reconstruction error relative to the conditional label. This approach can be used to impute missing data or as a validation method to assess how well the model encodes demographic information.

2. Methods

2.1. General overview

We propose a multi-task CVAE-GAN that performs (1) continuous synthetic aging, (2) age and sex estimation, and (3) conditional shape generation on dense 3D facial meshes. The architecture consists of four primary components, as depicted in Fig. 1:

1. A conditional variational autoencoder (CVAE) with an Encoder (E) and a Decoder/Generator (G).
2. An Age/Sex Classifier (C) that provides auxiliary supervision to make sure the generated shapes match the specified label.
3. A Discriminator (D) to enforce realistic and sharp shape generation.
4. A label Embedder (E_y) that learns a continuous embedding of the age label.

Standard 2D convolutions are replaced with SpiralNet++ [14] layers for efficient feature learning on fixed-topology facial meshes. To train the model, we propose a multi-task learning strategy that integrates the LATS [12] training scheme for synthetic aging with VAE [15] and GAN [16] training schemes for conditional shape generation.

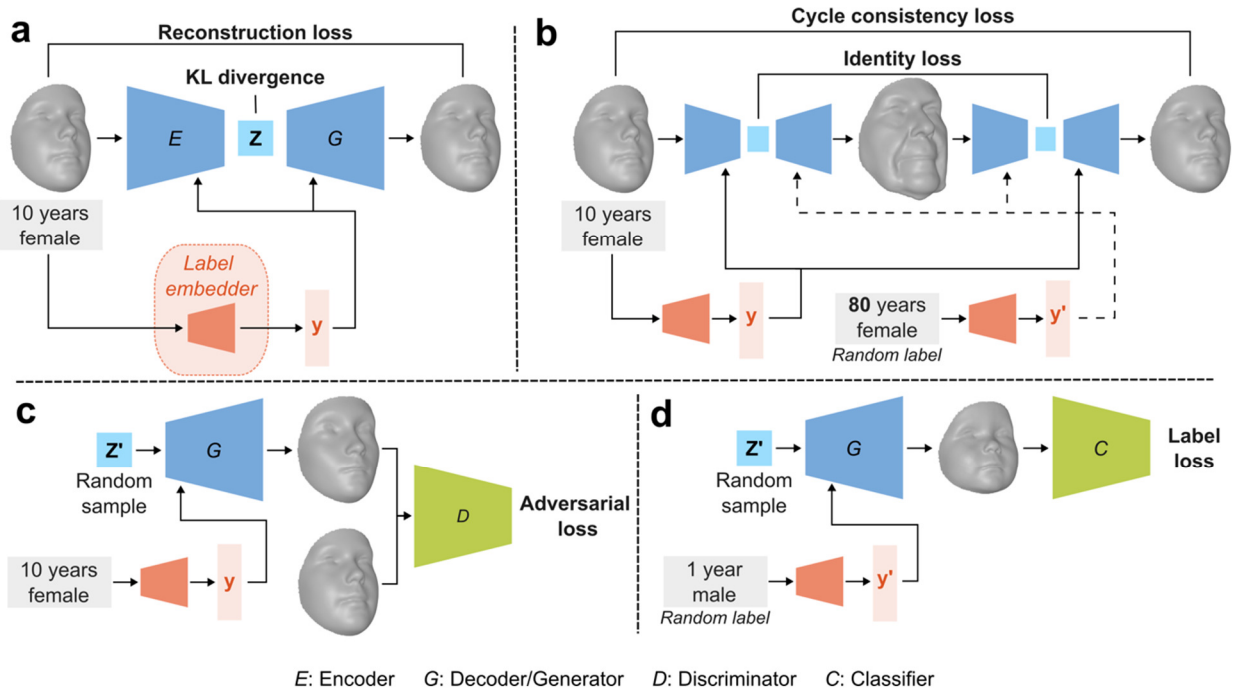

Fig. 1: 3DGrowthNet training overview. (a) Standard VAE losses. (b) Cycle consistency and identity loss. (c) Adversarial loss. (d) Classification loss.

2.2. Ethical approval

This research received ethical approval from the KU Leuven Ethics Committee and University Hospitals Gasthuisberg, Leuven (S56392). Approval for data acquisition across multiple institutions was also obtained, including: Royal Children's Hospital, Australia (IRB 29008I); University of Pittsburgh (IRB PRO09060553 and RB0405013); Seattle Children's Hospital (IRB 12107); University of Texas Health Science Center (HSC-DB-09-0508); University of Iowa (IRB 200912764 and 200710721); Pennsylvania State University (IRB 13103, 45727, 2503, 44929, 4320, and 1278); University of Cincinnati (IRB 2015-3073); Indiana University–Purdue University Indianapolis (IRB 1409306349); University College London Hospital (IRB JREC00/E042); and Sheffield Children's Hospital (IRB MREC/03/4/022).

2.3. Dataset

Our dataset, sourced from four different centers and previously described by Matthews et al. [9], consists of 5443 3D surface scans spanning an age range from 0-88 years (Fig. 2). Longitudinal data are available for 60 children, each with a follow-up scan taken approximately six years after the initial visit. Only individuals of self-reported European, Australian or "white" ancestry were retained due to limited representation of other groups. All shapes were represented with 7160 dense quasi-landmarks obtained with MeshMonk [17] and rigidly aligned via Generalised Procrustes Analysis (GPA) [18] to remove pose variation.

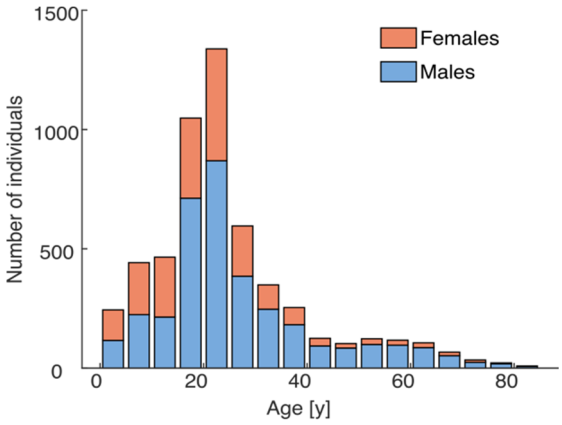

Fig. 2: Demographics of the dataset used in this study.

2.4. Conditional variational autoencoder (CVAE)

The CVAE consists of an encoder (E) and a decoder/generator (G). The encoder learns a distribution $P(z|x,y)$ that transforms an input x to a latent representation z given the age and sex label y. The decoder learns $P(x|z,y)$, generating a sample x' from the latent code z conditioned on a label y.
The encoder architecture includes four spiral convolutional layers followed by two fully connected layers (Fig. 3). The final linear layer branches into two outputs to produce the mean μ and log variance σ of the latent vector. The decoder mirrors this structure.

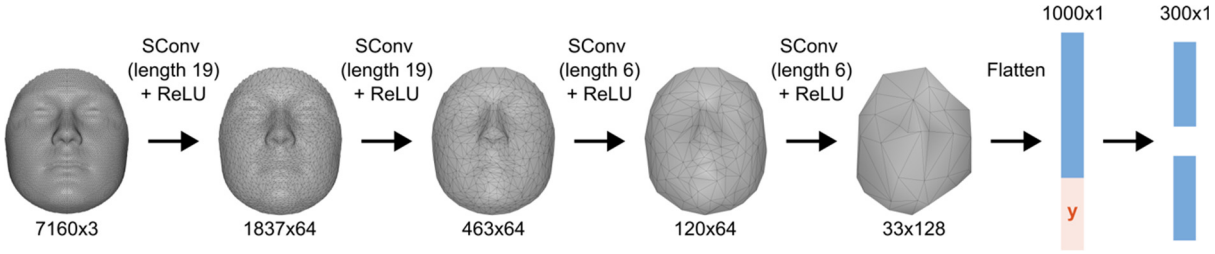

Fig. 3: Architecture details of the encoder. It consists of four SpiralNet++ [14] layers followed by two fully connected layers.

Latent sampling is performed via:

$$z_{x,y} = \mu_{x,y} + \varepsilon \cdot exp(\sigma_{x,y}) \text{ with } \varepsilon \sim N(0,1).$$

The VAE loss consists of the reconstruction loss and the KL-divergence [15]. The reconstruction loss is calculated as:

$$\mathcal{L}_{rec} = \frac{1}{V} \cdot \sum_{i=1}^{V} \left\| x_i - x_{rec,i} \right\|^2$$

With x_{rec} the reconstructed mesh with V vertices. The KL-divergence is calculated over a batch of samples together.

$$\mathcal{L}_{KL} = \frac{1}{2} \cdot [\boldsymbol{\mu}^T \boldsymbol{\mu} + sum(exp(\boldsymbol{\sigma}) - \boldsymbol{\sigma} - 1)]$$

Following the LATS training strategy [12], two additional objectives are introduced: the cycle consistency loss and the identity loss. For the cycle consistency loss, every image x is encoded with its real demographic label y and decoded with a random label y'. This transformed image x_{tra} is then encoded again with the same random label and decoded with its original label. This leads to the reconstructed image x_{cyc}. The cycle consistency loss encourages identity preservation by enforcing that x_{cyc} closely resembles x.

$$\mathcal{L}_{cyc} = \frac{1}{V} \cdot \sum_{i=1}^{V} \left\| x_i - x_{cyc,i} \right\|^2$$

The identity loss promotes the separation of demographic attributes from individual identity within the latent space. It is computed by comparing the latent representation $z_{x,y}$ obtained from the original image x with its true label y, to the latent representation $z_{tra,y'}$ obtained from the transformed image x_{tra} with the randomly assigned label y'. The model is encouraged to keep identity-specific features consistent, regardless of changes in the conditioning labels, by minimizing the difference between them.

$$\mathcal{L}_{id} = \left\| z_{x,y} - z_{tra,y'} \right\|^2$$

2.5. Label classifier

To ensure that the generated meshes correspond closely to the specified age and sex label, we pretrained an age and sex classifier (C_{age} and C_{sex}), which consists of the same type of convolutional layers as the encoder followed by two separate multilayer heads (Lin(512), Lin(256), Lin(64)) with ReLU activation function. The classifier was trained with L2 loss for the age head and binary cross-entropy loss for the sex head. The same loss functions were used to train the CVAE in a later stage.

$$\mathcal{L}_{age} = \left| C_{age}(x_{a,s}) - a \right|^2$$
$$\mathcal{L}_{sex} = -[s \cdot log(C_{sex}(x_{a,s})) + (1-s) \cdot log(1 - C_{sex}(x_{a,s}))]$$

where $x_{a,s}$ is a randomly generated sample with age a and sex s. Rather than evaluating the classifier on reconstructed or transformed meshes (as in LATS [12] and MeshWGAN [11]), it was applied to purely

synthetic meshes drawn from the latent space with randomly assigned demographic labels. By leveraging our variational autoencoder, we can train directly on these fully generated samples, enabling the model to learn conditional shape generation as well as synthetic aging.

2.6. Discriminator

The discriminator follows the same architecture as the classifier, with an extra linear layer to produce a single scalar output. It is trained to tell real samples apart from those generated by the model. We adopt the WGAN-GP [19] approach, which employs a Wasserstein loss with gradient penalty to improve training stability compared to conventional GANs. The WGAN-GP discriminator produces a continuous validity score for each image and employs a gradient penalty that enforces Lipschitz continuity, preventing the model from diverging during training. Similar to the classifier loss, the generator loss $\mathcal{L}_G$ and the discriminator loss $\mathcal{L}_D$ are calculated with respect to newly generated samples instead of reconstructed images to train for conditional shape generation.

$$\mathcal{L}_G = E_{x_G,y}[-D(x_G, y)]$$
$$\mathcal{L}_D = E_{x_G,y}[D(x_G, y)] - E_{x,y}[D(x, y)] + \lambda \cdot E_{\hat{x}}[(\|\nabla_{\hat{x}} \mathrm{D}(\hat{x}, y)\|_2 - 1)^2]$$

with $x_G = G(z, y)$ and $\hat{x} = \alpha \cdot x + (1 - \alpha) \cdot x_G$ with $\alpha \sim U(0,1)$.

2.7. Continuous label embeddings

The continuous age label was embedded following the label input mechanism proposed by Ding et al. [13], allowing the model to generalize across the full continuous domain, including regions with sparse training data. Following their strategy, a learned embedding function transforms the scalar labels into high-dimensional representations. This function is parameterized by a fully connected neural network E_y that transforms the age a to a representation of the same size as the features extracted by the age classifier right before the last linear layer. We split the pretrained classifier into two neural networks C_1 and C_2, where C_1 contains all the layers of C except for the last and C_2 consists only of the last layer. The objective of E_y is then to minimize

$$\mathcal{L}_{E_y} = \left| C_2(E_y(a + \varepsilon)) - (a + \varepsilon) \right|^2$$

with $\varepsilon \sim N(0, \sigma^2)$, which is a small amount of noise that is added to the original age to make sure the model learns a smooth embedding. The embedding network E_y consists of three linear layers (16, 32, 64), each followed by a ReLU activation function.

To enable conditional shape generation on sex as well as age, the sex label is concatenated to the embedded age label vectors, resulting in the final conditional label y. The label is then concatenated into the encoder, decoder and discriminator at distinct layers. This happens right before the first linear layer in the encoder and discriminator and right after the last linear layer in the decoder.

2.8. Training details

The dataset was split up in a train, test and validation set (85% train, 10% test, 5% validation). To ensure an even age distribution in the train and test sets, the data was divided into age categories with a width of 1 year between ages 1-25 and 5 years for 25y+ and the split was then applied stratified for each category. All individuals from the longitudinal dataset were put in the test set. Before training the main model, the age and sex classifiers and age label embedder were pretrained for 100 and 50 epochs and with a learning rate of 5e-5 and 1e-4 respectively. The noise variation σ in the label embedder was set to 0.5 years.

The CVAE-GAN was then trained in multiples steps. It was first pretrained for 50 epochs without the age- and sex-classifier, the cycle consistency loss and the identity alignment loss. After 50 epochs, the age- and sex-loss was added and after 100 epochs, the cycle consistency loss and identity alignment loss was added. The VAE was trained for 500 epochs in total and was updated with a learning rate of 1e-3 using the following loss function:

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \beta \cdot \mathcal{L}_{KL} + c_1 \cdot \mathcal{L}_{cyc} + c_2 \cdot \mathcal{L}_{id} + c_3 \cdot \mathcal{L}_{age} + c_4 \cdot \mathcal{L}_{sex} + c_5 \cdot \mathcal{L}_G$$

with β = 1e-4, c_1 = 1e-2, c_2 = .15, c_3 = c_4 = 5e-4 and c_5 = 1e-4. The discriminator was updated after each batch with $\mathcal{L}_D$ and a learning rate of 1e-6. All components were trained with the Adam optimizer, a learning rate decay of .99 after every epoch, and a batch size of 32 on an ASUS Turbo GeForce RTX 3090, 24G RAM, with PyTorch 2.1.0. Each epoch required approximately 36 seconds, and inference time per mesh was <1 second.

3. Experiments

To evaluate its performance comprehensively, we tested the model on four tasks. (1) The reconstruction was measured to see how well the network preserves detailed information of an individual's face. (2) The synthetic aging capabilities were analyzed visually as well as quantitatively using the longitudinal dataset. (3) The age and sex estimation was used to evaluate how well the network encodes the label information. (4) The generative capabilities were assessed visually and by measuring its ability to capture sexual dimorphism.

3.1. Reconstruction

The reconstruction error was quantified using the average root mean square error (RMSE) across all vertices. On average, the RMSE was 0.74mm with the highest errors in regions with sharp geometric features, such as the nose tip, eyes, and lips (Fig. 4). For comparison, a principal component analysis (PCA) model trained on the same dataset and retaining 97 % of the variance produced a similar RMSE of 0.78 mm. This suggests that the CVAE smooths the input while preserving most of the shape variability. This smoothing effect is a known characteristic of variational autoencoders (VAEs), which must balance the trade-off between generating sharp meshes and maintaining a continuous, well-structured latent space.

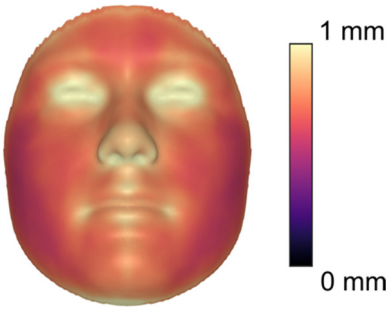

Fig. 4: Average reconstruction error visualized on the average face.

3.2. Synthetic aging

Synthetic aging was achieved by encoding a facial scan with its true age label and decoding it with a target age. This was qualitatively assessed by decoding two real examples across a range of ages (Fig. 5). The transformation shows a pronounced increase in centroid size between ages 0 and 15, followed by a plateau in adulthood, reflecting the nonlinear nature of facial growth. Younger faces exhibit an overall rounder shape and smoother features, while older faces display characteristic signs of aging such as skin folding and sagging around the chin, which is consistent with prior facial growth studies [9]. While age-related features change, the underlying facial identity and expression remain consistent, indicating successful disentanglement of age and identity in the latent space.

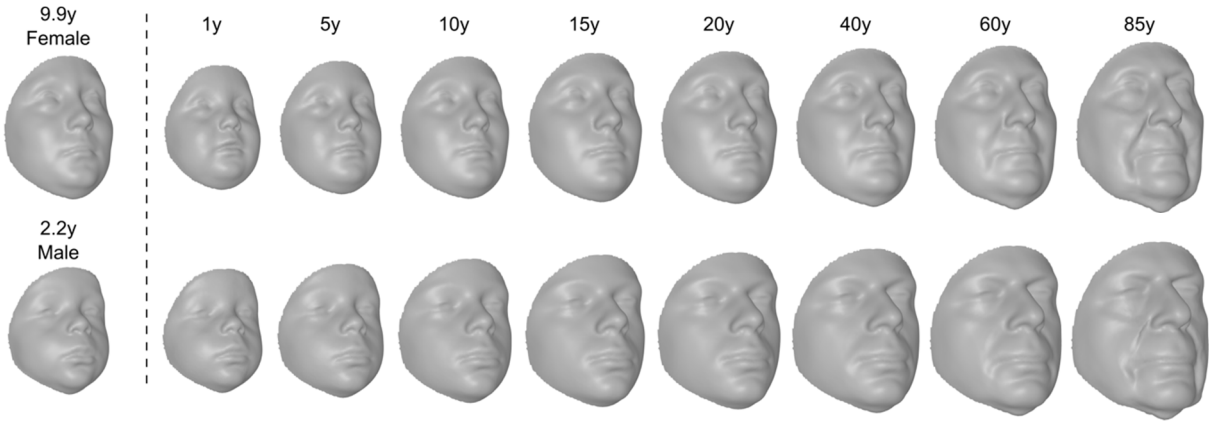

Fig. 5: Two examples of predicted aging trajectories. The original image (left) was encoded using its real age and sex and decoded using a range of different ages (right).

The longitudinal dataset was used to quantitatively assess the performance. The first image of every individual was encoded with its real age and decoded with its follow-up age (Fig. 6a). Although the predicted faces share some key identity traits with the real faces, the model cannot predict environmental factors such as weight fluctuations or scarring. The average prediction error is 2.00mm after size normalization and 2.40mm for unnormalized data, indicating that size has a substantial effect on the prediction accuracy. Prediction errors decrease with increasing age at the first timepoint ($p = 0.005$, Pearson's $r = -0.35$), likely because early facial development exhibits greater variability and more rapid structural change.

To illustrate the practical value of synthetic aging in real-world scenarios, a simple identification experiment was conducted (Fig. 6b). As a baseline, images from the second timepoint were identified by matching unaltered facial surfaces from the first timepoint to the most similar image based on average root mean square error (RMSE) in millimeters. Each lineup included the actual aged image alongside the 99 closest faces in terms of age and sex. To assess the added value of our model, the facial surfaces from the first timepoint were first synthetically aged to correspond to the age at the second timepoint, and the identification procedure was repeated using these aged representations. Performance was evaluated using the Cumulative Matching Curve (CMC), which reports the percentage of correct identifications within the top-K ranked candidates. We further compared results using both unnormalized and normalized faces. As shown in Fig. 6b, size normalization consistently improved identification accuracy. In the normalized condition, the rank-1 identification rate increased from 39% to 57% when using synthetically aged faces, demonstrating that the model enhances age-invariant recognition.

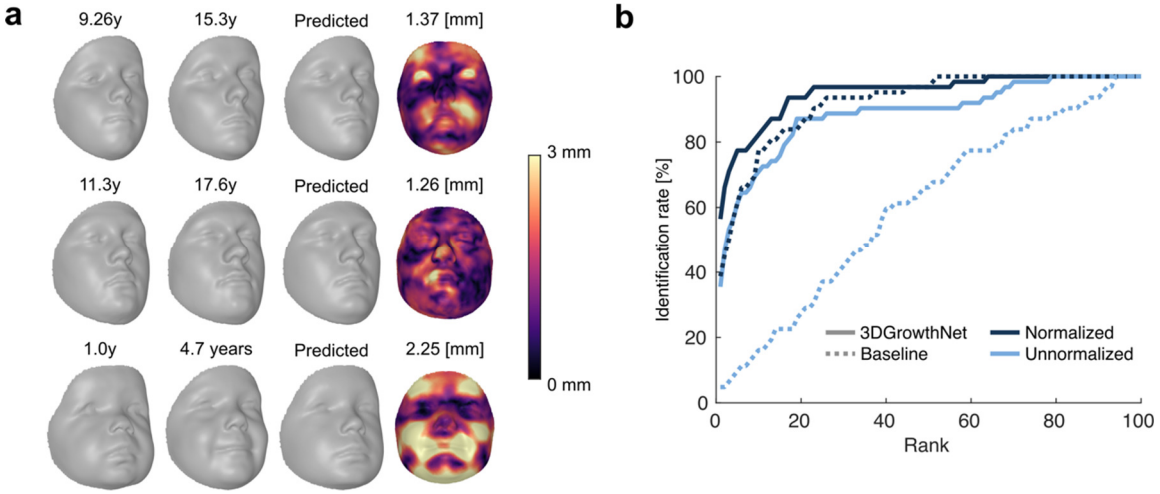

Fig. 6: Evaluation of synthetic aging on the longitudinal dataset. (a) The first, second and third columns contains the original mesh, the real aged mesh, and the predicted face by the model respectively. The numbers in the fourth column shows the average prediction errors over the whole face in mm and the colors indicate the error in mm per vertex. (b) CMC of the identification experiment.

3.3. Age and sex estimation

Although the model was not explicitly trained for age and sex classification, it can infer age and minimizing the reconstruction error with respect to the label. The core assumption is that reconstructing a face with incorrect label information will distort its latent representation and increase the reconstruction error.

For age estimation, we fixed the sex label and evaluated reconstruction error across a finely sampled age range (0 to 88 years, in 0.1-year intervals). The age label corresponding to the lowest reconstruction error was selected as the predicted age. This approach achieved a median absolute error of 2 years, with generally lower errors in younger individuals (Fig. 7a). Remarkably, this result is close to the performance of our pretrained age classifier (median absolute error = 2.30y), confirming that the model captures age-related shape variation effectively.

We applied a similar approach for sex classification. Each facial scan was processed twice—once with a male label and once with a female label—while keeping the true age fixed. The predicted sex was assigned based on which label yielded the lower reconstruction error. This achieved an overall classification accuracy of 87.7%, with lower accuracy in younger age groups (Fig. 7b). Again, this aligns well with the pretrained classifier's performance (accuracy = 89.7%), demonstrating the model's capacity to capture relevant sex-specific features.

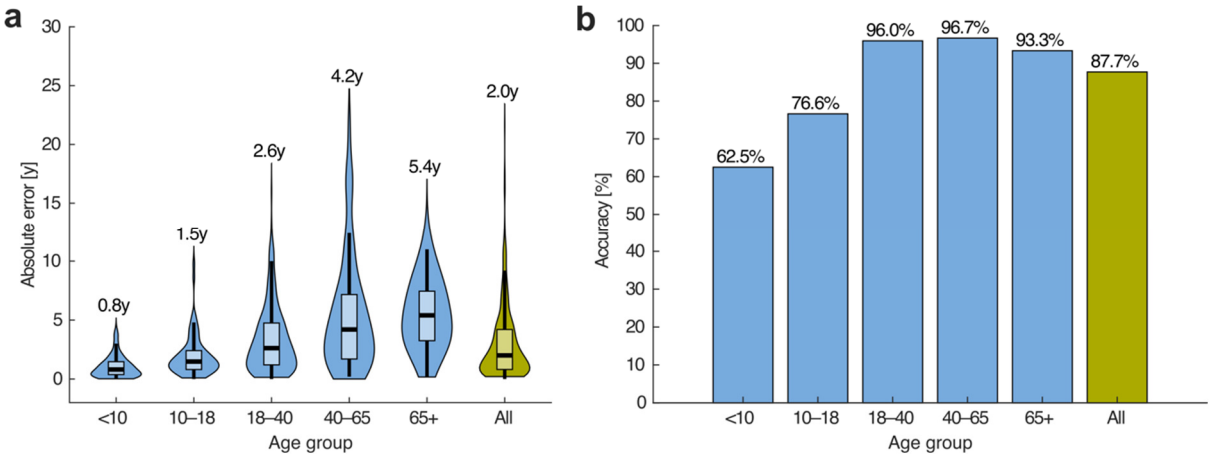

Fig. 7: Results of the age and sex estimation experiments for different age ranges. (a) Absolute age estimation error in years. The numbers on top of the violin plot indicate the median absolute error in that group. (b) Sex prediction accuracy.

3.4. Conditional shape generation

One of the key advantages of using a variational autoencoder (VAE) over a standard autoencoder is the ability to sample an unlimited number of synthetic facial shapes conditioned on age and sex. Fig. 8 illustrates examples of such synthetic aging trajectories, demonstrating the model's capacity to generate a wide variation of anatomically plausible and demographically consistent 3D faces.

Fig. 8: Examples of 3 female (top) and 3 male (bottom) synthetically generated growth trajectories.

To quantitatively assess the performance of our model, we generated up to 50,000 synthetic faces of the same age and sex distribution as the train set. Two metrics were used to evaluate generative quality: Coverage and Minimum Matching Distance (MMD) [20]. All measurements were taken on size

normalized shapes to focus on shape variation rather than size. The Coverage metric assesses the extent to which the set of generated shapes "covers" the distribution of reference shapes in the test set. For each generated shape, the method finds its nearest neighbor in the test set based on the average RMSE. Coverage is then computed as the percentage of unique test shapes that are selected as the closest match to at least one generated shape. This serves as a measure of the diversity of shapes that the network can generate. The coverage was evaluated as a function of the multiplication factor X, where X indicates how many times the size of the test set was sampled from the generated data. This was calculated considering all shapes as well as within predefined age groups, to assess whether the model performance was consistent over all ages. As a ground truth measurement, we also calculated the coverage by considering the train set as the generated set of shapes. As shown in Fig. 9a, coverage increases progressively up to 98.7% as more generated faces are sampled, closely matching the ground truth results. The coverage surpasses that of the original train set as additional shapes are generated, suggesting that the model can sample a broad and diverse range of faces rather than solely replicating existing examples.

The Minimum Matching Distance measures how closely the generated shapes approximate real ones. For each reference shape in the test set, it calculates the distance to its nearest counterpart in the generated set and then averages this over all reference shapes. This metric expresses the quality in addition to the coverage of the generated shapes. The MMD was evaluated as a function of the number of samples, as visualized in Fig. 9b. To get a ground truth measurement, the MMD was also calculated over the full dataset using a leave-one-out approach. The MMD plateaued to around 1.58mm and follows the same trend as the ground truth MMD of the dataset. This indicates that the model is capable of producing a diverse range of realistic samples that closely match the dataset's distribution. The MMD was the highest for the 65+ category, which may reflect the limited availability of training data in this range or the increased shape variation typically seen in older individuals [9].

To assess the generative capabilities of 3DGrowthNet in a data augmentation context, we conducted an experiment comparing reconstruction errors within a PCA space built from original training data versus one built from a combination of training and synthetically generated samples. 97% percent of variance was retained for the PCA space built from the training data and the PCA space of the combined dataset was constructed with the same number of PCs for a fair comparison. The experiment was performed in multiple predefined age groups, only retaining the shapes within the corresponding age limits. The inclusion of augmented data resulted in consistently lower reconstruction errors across all age groups, with the most notable improvements in underrepresented populations such as infants under one year and adults over 75 (Fig. 9c). The suggests that the model with the augmented dataset is capable of modeling more meaningful shape variation with the same number of components. In other words, this means that the quality of the individual components increases, which relates to its ability to generalize to unseen data. These results underscore the growing importance of data augmentation when modeling narrowly defined age intervals or demographic subsets with limited available training data.

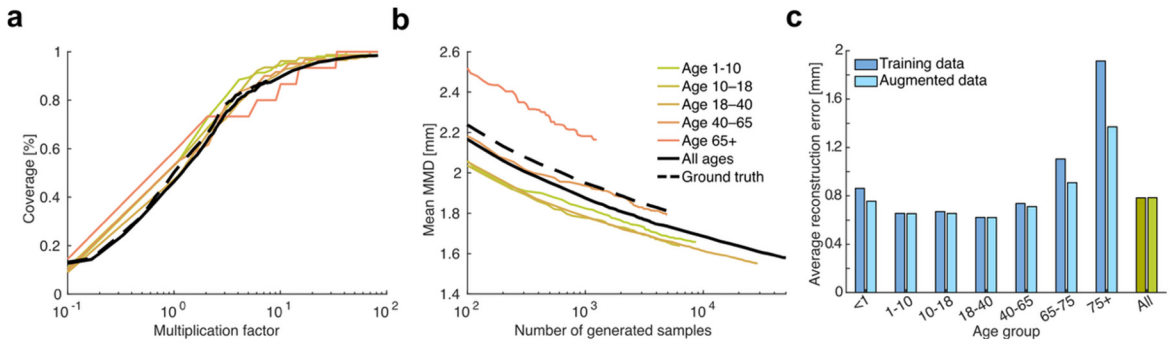

Fig. 9: Results of generative experiments. (a) Coverage. (b) Minimum Matching Distance (MMD). (c) Reconstruction error after PCA with original data and augmented data.

To show the value of the models' generative capabilities in a biomedical context, we examined average sex-based differences in facial shape across age. To examine the differences between male and female facial shapes, we sampled 500 males and 500 females from different ages and subtracted the expected female facial shape from the expected male shape (Fig. 10). As expected, sex-based differences became more pronounced during adolescence, particularly between ages 10 and 15, corresponding with the onset of puberty. This aligns to the sex classification results (Fig. 7b) where children were

generally harder to classify compared to adults. Male faces develop a more prominent nose, brow ridge, and elongated facial contour, while female faces retain rounder, softer features, which is consistent with the findings in previously published studies [21-23]. The estimated dimorphism remains consistent and biologically feasible, even at age 85 while there is no training data available for men for such high ages, indicating that the model is capable of extrapolating meaningful shape information outside the limits of the training data.

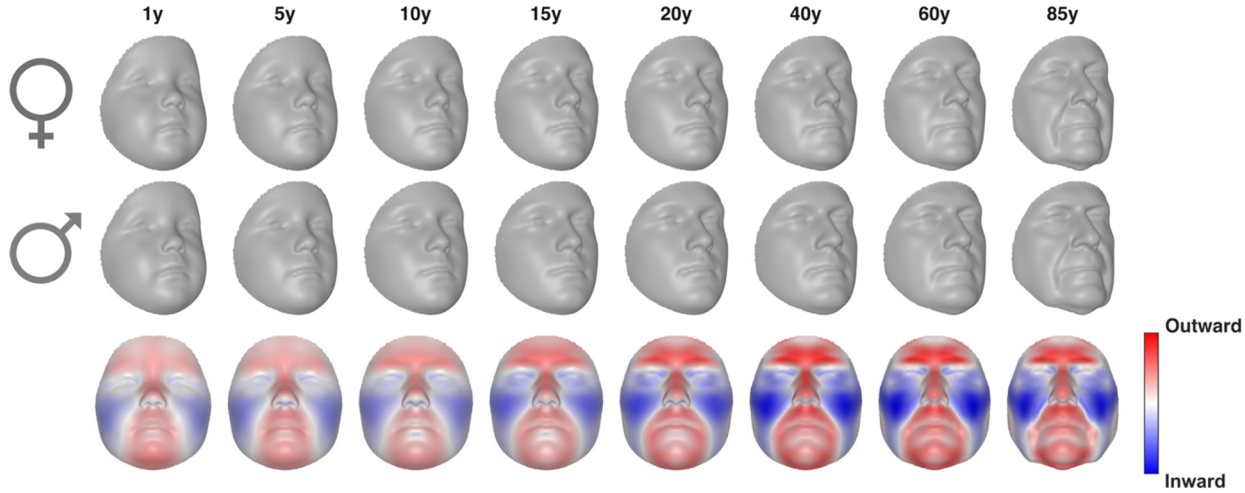

Fig. 10: Sexual dimorphism at different ages. The first and second row show the average faces at different ages for women and men respectively, normalized for centroid size. The third row indicates the different between these average shapes at different ages. Red indicates that the male average face is more outward than the female average face, and blue indicates the opposite.

4. Conclusion and future work

In this work we proposed 3DGrowthNet, a multi-task geometric deep learning framework for modeling facial growth from 3D surface scans. We adopted the continuous age embedding mechanism inspired by Ding et al [13], which allows us to predict growth at any continuous age over the full lifespan. The predicted growth trajectories showed that the model successfully disentangles age from identity, modeling age-related morphological changes while preserving individual characteristics. Quantitative validation using a longitudinal dataset of children yielded an average prediction error of approximately 2 mm, improving age-independent identification with almost 20%.

We also introduced a multi-task training scheme that combines the LATS [12] synthetic aging protocol with VAE and GAN objectives. This enables unlimited sampling of age- and sex–conditioned synthetic faces or even full aging trajectories. The model achieved high coverage (98.7%) and a Minimum Matching Distance that closely matches the ground truth results, indicating that the model is capable of generating diverse and realistic 3D meshes. A data augmentation experiment showed that incorporating synthetically generated shapes from 3DGrowthNet enabled the model to capture more meaningful shape variation, particularly within underrepresented age groups. We further demonstrated biomedical relevance by simulating average sex differences across different ages. The generated aging trajectories revealed expected developmental trends, even at late ages where real data are sparse, further supporting the model's data augmentation capabilities.

Additionally, our model performs age and sex estimation by minimizing the reconstruction error with respect to the demographic label. Estimation accuracy for both age and sex were highly dependent on the subject's age. Age prediction was most precise in younger individuals, whereas sex classification proved less reliable in children due to less sexual dimorphism in children. The ability to accurately predict age and sex shows that the model successfully encodes the label information.

A notable limitation of the VAE architecture is the tendency toward overly smooth mesh outputs. Although the discriminator partially counteracts this effect, future work could explore diffusion-based or other advanced generative models to further enhance sharpness and realism. Another constraint is that, trained on cross-sectional data alone, 3DGrowthNet can only predict population-average growth rather than individual trajectories. Longitudinal scans and the incorporation of additional covariates such as ancestry, weight, height or hormone levels would allow personalized growth predictions and further improve performance. In addition, because 3DGrowthNet was trained only on individuals of European

ancestry, its applicability to other populations remains to be investigated. Craniofacial growth patterns vary across ancestries, and applying the current model outside its training domain may introduce bias or reduce accuracy. Adding data of other populations will be essential to improve fairness, robustness, and biological validity in future studies. Until such data are incorporated, results should be interpreted within the demographic boundaries of the training set.

In conclusion, 3DGrowthNet represents the first unified approach for continuous, conditional aging and shape generation on 3D facial meshes, establishing a new benchmark for realistic, demographically informed mesh synthesis. It enhances age-invariant identification in forensic settings and provides a powerful data augmentation tool for underrepresented demographic groups, which enables more precise clinical assessment of facial dysmorphism in future studies.

Code availability

The code will be made publicly available on Gitlab upon publication:
https://gitlab.kuleuven.be/mirc/public-projects/nina-claessens/3DGrowthNet .

Funding

The KU Leuven research team and analyses were supported by the National Institutes of Health (R01-DE027023), the research fund KU Leuven (BOF-C1, C14/20/081) and the Research Program of the Research Foundation Flanders (Belgium) (FWO, G0D1923N). The data was collected by the research teams of the University of Pittsburgh and the Indiana University Purdue University Indianapolis (IUPUI). The University of Pittsburgh data collection was supported by the National Institute of Dental and Craniofacial research (U01-DE020078, R01-DE016148, R01-DE027023; SMW). The IUPUI data collection was supported by the National Institute of Justice (2014-DN-BX-K031, 2018-DU-BX-0219; SW).

References

[1] Z. Zhang, Y. Song, and H. Qi, “Age Progression/Regression by Conditional Adversarial Autoencoder,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 4352–4360. doi: https://doi.org/10.1109/CVPR.2017.463.

[2] “Hierarchical Face Aging Through Disentangled Latent Characteristics,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2020, pp. 86–101. doi: https://doi.org/10.1007/978-3-030-58580-8_6.

[3] X. Yao, G. Puy, A. Newson, Y. Gousseau, and P. Hellier, “High Resolution Face Age Editing,” in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy: IEEE, Jan. 2021, pp. 8624–8631. doi: https://doi.org/10.1109/ICPR48806.2021.9412383.

[4] Y. Alaluf, O. Patashnik, and D. Cohen-Or, “Only a matter of style: age transformation using a style-based regression model,” *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–12, Aug. 2021, doi: https://doi.org/10.1145/3450626.3459805.

[5] 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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 8789–8797. doi: https://doi.org/10.1109/CVPR.2018.00916.

[6] B. Y. Ainuz, R. R. Hallac, and A. A. Kane, “Longitudinal composite 3D faces and facial growth trends in children 6–11 years of age using 3D cephalometric surface imaging,” *Annals of Human Biology*, vol. 48, no. 7–8, pp. 540–549, Nov. 2021, doi: https://doi.org/10.1080/03014460.2021.2012257.

[7] M. Krimmel *et al.*, “Three-Dimensional Normal Facial Growth from Birth to the Age of 7 Years:,” *Plastic and Reconstructive Surgery*, vol. 136, no. 4, pp. 490e–501e, Oct. 2015, doi: https://doi.org/10.1097/PRS.0000000000001612.

[8] C. Sforza, G. Grandi, M. De Menezes, G. M. Tartaglia, and V. F. Ferrario, “Age- and sex-related changes in the normal human external nose,” *Forensic Science International*, vol. 204, no. 1–3, p. 205.e1-205.e9, Jan. 2011, doi: https://doi.org/10.1016/j.forsciint.2010.07.027.

[9] H. S. Matthews *et al.*, “Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism,” *Sci Rep*, vol. 11, no. 1, p. 12175, Jun. 2021, doi: https://doi.org/10.1038/s41598-021-91465-z.

[10] H. Matthews, A. Penington, J. Clement, N. Kilpatrick, Y. Fan, and P. Claes, “Estimating age and synthesising growth in children and adolescents using 3D facial prototypes,” *Forensic Science International*, vol. 286, pp. 61–69, May 2018, doi: https://doi.org/10.1016/j.forsciint.2018.02.024.

[11] J. Zhang, K. Zhou, Y. Luximon, T.-Y. Lee, and P. Li, “MeshWGAN: Mesh-to-Mesh Wasserstein GAN With Multi-Task Gradient Penalty for 3D Facial Geometric Age Transformation,” *IEEE Trans. Visual. Comput. Graphics*, vol. 30, no. 8, pp. 4927–4940, Aug. 2024, doi: https://doi.org/10.1109/TVCG.2023.3284500.

[12] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, “Lifespan Age Transformation Synthesis,” in *Computer Vision – ECCV 2020*, vol. 12351, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in Lecture Notes in Computer Science, vol. 12351. , Cham: Springer International Publishing, 2020, pp. 739–755. doi: https://doi.org/10.1007/978-3-030-58539-6_44.

[13] X. Ding, Y. Wang, Z. Xu, W. J. Welch, and Z. J. Wang, “Continuous Conditional Generative Adversarial Networks: Novel Empirical Losses and Label Input Mechanisms,” *EEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8143–8158, Jul. 2023, doi: https://doi.org/10.1109/TPAMI.2022.3228915.

[14] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, “SpiralNet++: A Fast and Highly Efficient Mesh Convolution Operator,” in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 4141–4148. doi: https://doi.org/10.1109/ICCVW.2019.00509.

[15] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” 2013, *arXiv*: arXiv:1312.6114. Accessed: Jun. 10, 2024. [Online]. Available: http://arxiv.org/abs/1312.6114

[16] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” *Advances in Neural Information Processing Systems*, vol. 3, Jun. 2014, doi: https://doi.org/10.1145/3422622.

[17] J. D. White *et al.*, “MeshMonk: Open-source large-scale intensive 3D phenotyping,” *Sci Rep*, vol. 9, no. 1, p. 6085, Apr. 2019, doi: https://doi.org/10.1038/s41598-019-42533-y.

[18] J. C. Gower, “Generalized procrustes analysis,” *Psychometrika*, vol. 40, pp. 33–51, 1975, doi: https://doi.org/10.1007/bf02291478.

[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein GANs,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 5769–5779.

[20] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, “Learning Representations and Generative Models for 3D Point Clouds,” Jun. 12, 2018, *arXiv*: arXiv:1707.02392. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/1707.02392

[21] P. Claes *et al.*, “Sexual dimorphism in multiple aspects of 3D facial symmetry and asymmetry defined by spatially dense geometric morphometrics,” *Journal of Anatomy*, vol. 221, no. 2, pp. 97–114, Aug. 2012, doi: https://doi.org/10.1111/j.1469-7580.2012.01528.x.

[22] H. Matthews, T. Penington, I. Saey, J. Halliday, E. Muggli, and P. Claes, “Spatially dense morphometrics of craniofacial sexual dimorphism in 1-year-olds,” *Journal of Anatomy*, vol. 229, no. 4, pp. 549–559, Oct. 2016, doi: https://doi.org/10.1111/joa.12507.

[23] H. S. Matthews *et al.*, “Using data-driven phenotyping to investigate the impact of sex on 3D human facial surface morphology,” *Journal of Anatomy*, vol. 243, no. 2, pp. 274–283, Aug. 2023, doi: https://doi.org/10.1111/joa.13866.