Large-Scale 3D Infant Face Model

Till N. Schnabel[1], Yoriko Lill[2,3,4], Benito K. Benitez[2,3,4], Prasad Nalabothu[2,3,4], Philipp Metzler[5,6], Andreas A. Mueller[2,3,4], Markus Gross[1], Baran Gözcü[1], and Barbara Solenthaler[1]

[1] Department of Computer Science, ETH Zurich, Switzerland
till.schnabel@inf.ethz.ch
[2] Oral and Craniomaxillofacial Surgery, University Hospital Basel and University of Basel, Switzerland
[3] Department of Clinical Research, University of Basel, Switzerland
[4] Department of Biomedical Engineering, University of Basel, Switzerland
[5] Clinic for Oral and Maxillofacial Surgery, Cantonal Hospital Aarau, Switzerland
[6] Center for Dental Medicine, University of Zurich, Switzerland

Abstract. Learned 3-dimensional face models have emerged as valuable tools for statistically modeling facial variations, facilitating a wide range of applications in computer graphics, computer vision, and medicine. While these models have been extensively developed for adult faces, research on infant face models remains sparse, limited to a few models trained on small datasets, none of which are publicly available. We propose a novel approach to address this gap by developing a large-scale 3D INfant FACE model (INFACE) using a diverse set of face scans. By harnessing uncontrolled and incomplete data, INFACE surpasses previous efforts in both scale and accessibility. Notably, it represents the first publicly available shape model of its kind, facilitating broader adoption and further advancements in the field. We showcase the versatility of our learned infant face model through multiple potential clinical applications, including shape and appearance completion for mesh cleaning and treatment planning, as well as 3D face reconstruction from images captured in uncontrolled environments. By disentangling expression and identity, we further enable the neutralization of facial features — a crucial capability given the unpredictable nature of infant scanning.

Keywords: 3D Infant Faces · Nonlinear Morphable Model · Unsupervised Disentanglement

1 Introduction

The use of 3D face scans in clinical workflows has been limited, despite their transformative potential in medical applications. Unlike traditional 2D imaging, 3D scans offer a comprehensive representation of facial morphology, enabling the creation of precise digital twins of patients [7,15]. This advancement opens avenues for data-driven models to enhance personalized treatment planning, surgical simulation, and outcome prediction. Although several studies have demonstrated the utility of 3D face scans in medical scenarios, they have predominantly focused on adult populations [10,11]. However, adapting these technologies for pediatric use introduces unique challenges. Infants display spontaneous expressions and frequent occlusions from objects such as pacifiers and hands, complicating face scanning processes. Hence, developing accurate 3D infant face models requires specialized methods to handle such unpredictable and incomplete data.

Face models derived from 3D databases primarily focus on adult faces, often employing linear or nonlinear morphable models to delineate facial shape and appearance variations [2,16,8]. These models can incorporate multiple dimensions to provide semantic control over facial identity and expression [12,6]. Recent advances in developing infant face models [14,9,13] represent significant progress, though they originate from small, curated datasets with limited variability, restricting their scalability and generalizability. Our large-scale 3D INfant FACE model (INFACE) improves upon these shortcomings, proving its versatility through clinical applications, including realistic 3D inpainting for mesh repair and treatment planning, alongside 3D reconstruction from monocular images captured in uncontrolled environments. Furthermore, the disentanglement of expression from identity allows for effective face neutralization, addressing the variable conditions of infant scanning. Our main contributions are:

- The first publicly available 3D infant face model, trained on a large-scale dataset of incomplete face scans using an autoencoder, thereby considering geometric corruptions arising from occlusions during the scanning process.
- The first multi-nonlinear infant face model enabling individual manipulation of identity, facial expression, and age.
- A quantitative evaluation of the face model, compared to its PCA equivalent and to a state-of-the-art adult face model.
- Several applications of the infant face model, including realistic shape and appearance completion (for geometry repair and deformity correction), 3D reconstruction from uncontrolled images, and expression neutralization.

2 Methodology

Our primary objective is to develop a 3D infant face model using a large dataset of uncontrolled and incomplete scans. To this end, we detail the process of learning such a model via masked autoencoder training. Afterwards, we introduce an advanced model specifically developed to disentangle identity, expression, and age in a fully unsupervised manner.

2.1 3D Face Model

INFACE is based on a dataset comprising labeled 3D face scans of healthy individuals. It features uncontrolled facial expressions and corrupted geometry and texture resulting from frequent occlusions during scanning, such as pacifiers and hands. We employ segmentation techniques to isolate these corrupted regions, ensuring robust registration to a template topology through NICP [1] (cf. Section 3). In contrast to the predominant use of PCA models for face modeling

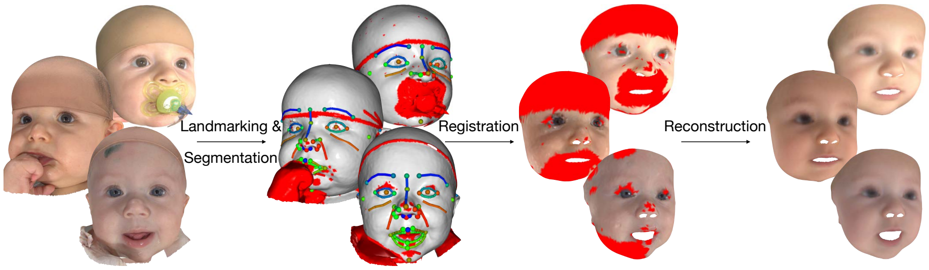

Fig. 1: Raw face scans are manually landmarked and artifacts segmented to be excluded from registration, yielding artifact-free reconstructions from our model.

[2,16], we opt for training a fully-connected autoencoder, enabling the unsupervised disentanglement detailed in Section 2.2 while additionally facilitating seamless integration of partial geometry defined by our segmentation masks [17]. Formally, the masked loss $\mathcal{L}_r$ between an input scan $\mathbf{X}$ and its reconstruction $g(f)(\mathbf{X})$ via our model $g(f)$ is defined as

$$\mathcal{L}_r(g(f)(\mathbf{X}),\mathbf{X}) = \frac{\sum_{i\in\mathcal{M}} \ell(g(f)(\mathbf{X}),\mathbf{X})_i}{|\mathcal{M}|}, \tag{1}$$

where $\mathcal{M}$ defines the mask set containing all indices marked as registered, and ℓ denotes the element-wise loss function between $\mathbf{X}$ and $g(f)(\mathbf{X})$. We train two autoencoders to model shape and appearance separately using the per-vertex 3D position and RGB color information. Figure 1 illustrates the data processing pipeline and model reconstruction.

2.2 Multi-Nonlinear Representation

As infants do not adhere to scanning protocols aimed at capturing a predefined set of facial expressions, we follow Zhou et al. [21] to learn a disentanglement of identity and expression in unsupervised settings, i.e., without requiring any expression labels or a neutral expression common to all identities, via iterative separation during neural network training. We further expand upon this method by additionally separating age from the identity and expression spaces, leveraging scans of the same patients acquired at multiple-month intervals. To this end, we extend the tripled sampling proposed by Zhou et al. [21] to a quadruplet $(\mathbf{X}_1^s, \mathbf{X}_2^s, \mathbf{X}^{s'}, \mathbf{X}^t)$ of scans for each batch element during training. While these four samples may all exhibit unique expressions, identity is shared among $(\mathbf{X}_1^s, \mathbf{X}_2^s, \mathbf{X}^{s'})$, where $(\mathbf{X}_1^s, \mathbf{X}_2^s)$ additionally share age. Weight-independent encoders f_β, f_θ, and f_α map these four samples separately to their respective identity, expression, and age latents, before the decoder g reconstructs an input sample from a mixture of these latents, thus exploiting the samples' shared features to encourage disentanglement during encoding. Specifically, the masked cross consistency loss $\mathcal{L}_C$ reconstructs $\mathbf{X}_2^s$ via

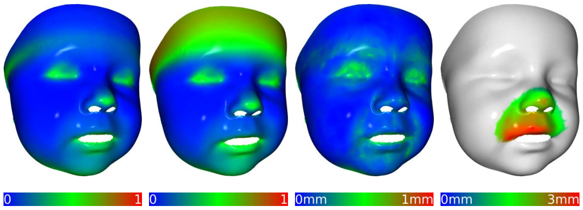

Fig. 2: Left and middle left: Distribution of masked geometry and texture regions, averaging 16 ± 12 % and 27 ± 22 %, respectively, across our dataset. Middle right: Distribution of geometry test reconstruction error, 0.23 ± 0.05 mm. Right: Distribution of nasolabial test reconstruction error, 1.95 ± 1.09 mm.

$$\mathcal{L}_C = \mathcal{L}_r \left(g \left[f_\beta(\mathbf{X}^{s'}), f_\theta(\mathcal{T}(\mathbf{X}^s_2)), f_\alpha(\mathbf{X}^s_1) \right], \mathbf{X}^s_2 \right), \tag{2}$$

whereas the masked self consistency loss $\mathcal{L}_S$ reconstructs $\mathbf{X}^s_1$ as

$$\mathcal{L}_S = \mathcal{L}_r \left(g \left[f_\beta(\mathbf{X}^{s'}), f_\theta(\mathcal{T}(\tilde{\mathbf{X}}^{t'})), f_\alpha(\mathbf{X}^s_2) \right], \mathbf{X}^s_1 \right), \tag{3}$$

using the intermediate $\tilde{\mathbf{X}}^{t'}$ for the expression latent, generated on the fly as

$$\tilde{\mathbf{X}}^{t'} = \mathrm{ARAP} \left(\mathbf{X}^t, g \left[f_\beta(\mathbf{X}^t), f_\theta(\mathcal{T}(\mathbf{X}^s_1)), f_\alpha(\mathbf{X}^t) \right] \right). \tag{4}$$

The As-rigid-as-possible (ARAP) deformation and expression-invariant transformations $\mathcal{T}$ discourage degenerate solution where identity-related features leak into the disentangled expression code [21]. The full training loss is then computed as a uniform average of $\mathcal{L}_C$ and $\mathcal{L}_S$.

3 Data Acquisition and Processing

We compiled a dataset of 2394 3D scans from 816 pediatric patients with normal facial morphology, aged 9±6 months, from the University Hospital Basel and the Cantonal Hospital Aarau. Our dataset maintains equitable gender distribution, with a strong predominance of Caucasian infants. 95 % of the scans were captured using the VECTRA M5 device [7], covering the full head with five cameras, while the remaining scans were obtained with the 3dMDtrio [15] scanner, employing three cameras for rapid facial data capture. Despite comprising approximately 17 000 vertices in the facial area and high-res textures from raw photos, many scans exhibit artifacts due to camera limits and typical infant scanning challenges (e.g., varying head poses, occlusions, movements).

To address artifacts affecting registration and training, we manually marked on average 16 ± 12 % of facial geometry and 27 ± 22 % of textures as corrupted. The distribution of these areas is illustrated in Figure 2 for geometry and texture,

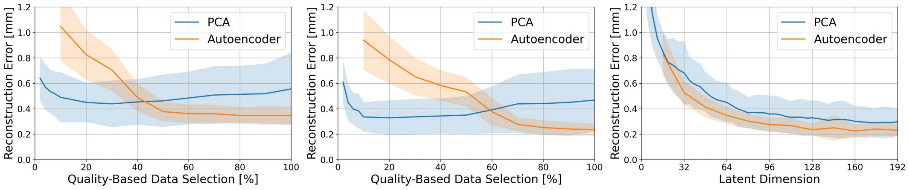

Fig. 3: Average reconstruction error for PCA and our autoencoder on the test set, grouped by data subsets sorted by quality, for latent dimensions 64 (left) and 128 (middle). Model comparison across various latent dimensions (right).

respectively (left and middle left). We implemented an advanced landmarking approach, deploying 112 points to navigate around these corrupted areas, as they sometimes obscure typical landmarks. Beyond pinpointing standard markers like eye corners and glabella, our scheme also traces curvilinear facial features, such as eyebrows and nasolabial folds, through series of interconnected landmarks, allowing the template to align accurately by iteratively matching these curves. This comprehensive landmarking ensures precise correspondence in unoccluded regions, while excluding affected areas from registration. Landmarks and segmented sections are depicted in the second image of Figure 1.

4 Experiments and Results

We first discuss implementation specifics of our model training process, followed by evaluations of the performance of INFACE and its design decisions.

4.1 Training Details

We use distinct training, validation, and test sets with no patient overlap. The test set comprises 50 scans with minimal artifacts, whereas the remaining data were randomly divided into 90 % for training and 10 % for validation. We trained two autoencoders on per-vertex 3D position (standard and multi-nonlinear shape model) and another one on RGB data (appearance model), resulting in input vectors of 15 570 for all models. Each encoder and decoder features a single 256-sized hidden layer with leaky ReLU activation and latent dimensions of 256 for appearance, 128 for the standard shape model, and $(32, 64, 1)$ respectively for identity, expression, and age in the multi-nonlinear model. We chose L1 as per-element reconstruction loss ℓ similar to previous work [6,21,13], and added L1 regularization to counter overfitting and ensure model compactness. Model parameters were optimized in approximately one day on a single NVIDIA RTX 2080 Ti using PyTorch's ADAM for 10 000 epochs and a batch size of 8 at an initial learning rate of 1×10^{-4}, which was progressively lowered upon validation error convergence. The expression-invariant transformations $\mathcal{T}$ employed during training of the multi-nonlinear model consisted of independently rescaling each axis of the mesh within the range $[0.7, 1.3]$ and adding Gaussian noise

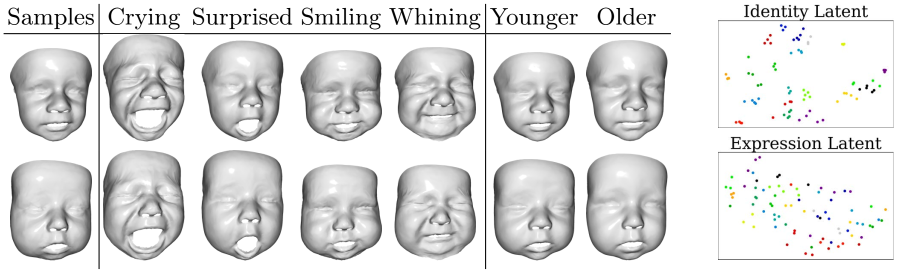

Fig. 4: Left: Expression transfer and age variation demonstrated on two distinct, randomly sampled identities. Right: TSNE-mapped latent distribution of identity versus expression encoded for multiple scans of 20 unique identities, marked in distinct colors.

$\mathcal{N}(0, 0.05\,\text{mm})$ independently to each vertex. The simple, shallow architecture ensures the trained model is easy to adopt and allows for sampling in the order of 100 meshes per second on a modern CPU.

4.2 Model Evaluation

Generalization We report the test reconstruction errors for our shape autoencoder and a classical PCA model [2,3], which were trained on varied data subsets. This analysis considers latent dimensions 64 and 128, as shown in Figure 3 (left and middle). The data subsets were sorted by artifact size, serving as a metric for scan quality. When training the models on small high-quality subsets, PCA performs significantly better than when using larger sets of the data, whereas our autoencoder outperforms PCA's best result when trained on the complete dataset, affirming our masked training approach.

Latent dimension In Figure 3 (right), we show the average reconstruction errors for our autoencoder and PCA across various latent dimensions. The autoencoder was trained on the complete data set and PCA on the 20% of data with the highest quality (i.e., best performing PCA model). Converging at size 128, the autoencoder consistently outperforms PCA ($0.23 \pm 0.05\,\text{mm}$ vs $0.33 \pm 0.14\,\text{mm}$). Figure 2 (middle right) shows the distribution of the reconstruction error, using the 128-dimensional autoencoder trained on the full dataset.

Multi-nonlinear model evaluation Due to the uncontrolled expressions in our dataset, we could not establish a reliable quantitative evaluation of the disentanglement. Instead, Figure 4 provides qualitative results for two sampled identities and additional plots illustrating the separation in the latent encoding, demonstrating meaningful identity-preservation while transferring expressions and manipulating ages (continuous interpolations are shown in the supplementary video). However, it is important to acknowledge that the disentanglement negatively impacts reconstruction accuracy, which is consistent with prior research. Notably, our multi-nonlinear model converged at 2000 epochs with a test

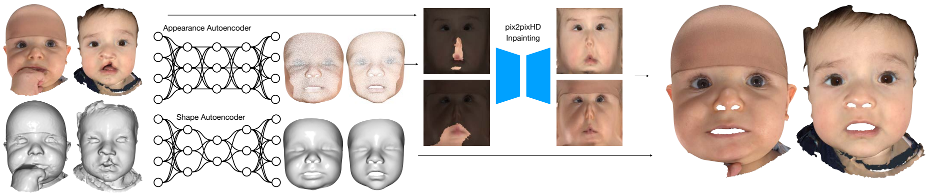

Fig. 5: Our pix2pixHD uses reconstructed shapes and low-res textures to inpaint high-frequency details for realistic facial deformity and occlusion replacement.

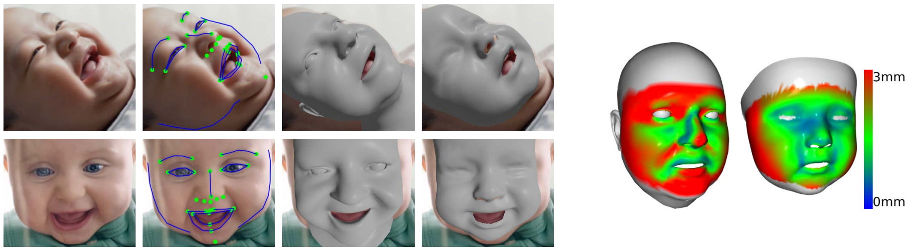

Fig. 6: Using a raw input image [19], we predict 2D landmarks to reconstruct the face in 3D with FLAME and INFACE. On the right, we show the distribution of the average reconstruction error on our test set for both methods.

error of 0.79 ± 0.20 mm, which is significantly higher than the standard model's error of 0.23 ± 0.05 mm after 10 000 epochs.

5 Applications

We illustrate several potential clinical applications of our learned infant face model, highlighting its relevance for incorporation into clinical practice.

5.1 Shape and Appearance Completion

Figure 2 (right) delineates the average nasolabial shape completion test error. To improve texture realism in these reconstructions, we follow Chandran et al. [6] by training a pix2pixHD [20] network for high-detail infusion into low-resolution vertex colors, tailoring the method towards a constrained texture inpainting task. During training, randomized regions [18] of high-frequency textures are replaced with our model's low-frequency reconstructions, achieving realistic shape and appearance completion at test time. Figure 5 showcases two applications: mending occlusion-related gaps and correcting cleft lip deformities in 3D scans, extending traditional 2D GAN methods [4]. These results underline the model's utility in simplifying scan capture and enhancing infant care by providing a valuable tool for surgical planning and improved communication through realistic visuals.

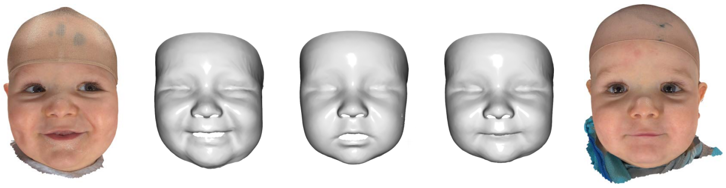

Fig. 7: From an input face scan (left, center left), we compute a neutralized expression (middle) using the multi-nonlinear model, then compare it to the face scan that most closely resembles a neutral expression (center right, right).

5.2 3D Reconstruction from Monocular Images

We use INFACE for the challenging and ill-posed task of 3D reconstruction from single images. Unlike earlier efforts limited to synthetic samples [13], our study applies 3D reconstruction to real infant photos. Given a set of landmarks predicted on the 2D image via [19], we follow previous work [2,5] and optimize camera and latent model parameters to align with the projected landmarks while enforcing regularization to control deviation from the average latent model vector. We utilize two distinct sources for input data. Firstly, selected images from a publicly available dataset [19] are employed to illustrate results visually. Secondly, quantitative analysis is performed on our test dataset, which comprises ground truth image-scan pairs. Figure 6 demonstrates the superior performance of our method in accurately reconstructing infant faces when compared to the adult-oriented FLAME model [12], which fails to capture distinct facial features of infants. Evaluated on our test set, INFACE has an average reconstruction error of 1.69 ± 1.55 mm, while FLAME reaches 2.26 ± 1.72 mm. The error distribution is visually illustrated in Figure 6 (right).

5.3 Neutralization of Facial Expression

Infant scanning often results in unpredictable behavior and uncontrolled facial expressions, adding considerable variability to scan data, thus complicating meaningful information extraction. Neutralizing these factors could enhance data quality and consistency. Figure 7 illustrates the effectiveness of our multi-nonlinear model in neutralizing expressions, highlighting its potential to enable standardized assessment of facial morphology. Using an input face scan (left), we compute the corresponding neutral expression (middle), and compare it to the scan of the same identitiy that is closest to a neutral expression (right).

6 Discussion and Conclusion

We have presented a large-scale, multi-nonlinear 3D infant face model, which is easy to adopt and holds significant promise in enhancing and standardizing the understanding of infant facial morphology. By offering insights into the typical appearance, facial movement during expressions, and developmental changes in infants, alongside making our shape model the initial publicly available resource, we contribute to the progression of pediatric facial analysis and diagnosis. Furthermore, INFACE offers a valuable tool for reconstructing 3D geometry from monocular images taken in unconstrained real-world scenarios and for cleaning erroneous 3D scans, thus enhancing the accuracy and reliability of medical imaging data. Specifically, in the context of cleft lip reconstruction, INFACE serves as a valuable guide for both medical practitioners and parents, facilitating informed decision-making prior to surgery.

Looking ahead, several avenues for future research present themselves. Firstly, automating the labeling process can streamline model training and improve efficiency. Additionally, efforts to eliminate racial bias by diversifying the dataset will further enhance the model's inclusivity and applicability across diverse populations. Moreover, further investigation is required to bridge the gap between accurate model reconstructions and intuitive feature disentanglement. This also encompasses both visual and quantitative evaluations of aging, necessitating suitable datasets to facilitate this analysis. Lastly, exploring novel approaches [22] for 3D reconstruction from monocular images or videos using our learned face model holds promise for expanding its utility in various imaging modalities.

Acknowledgments. Funding was provided by the Swiss National Science Foundation (Grant No. 205321_205008) and by the Botnar Research Centre for Child Health. The collection and further use of the clinical data received the IRB approval by Ethikkommission Nordwest- und Zentralschweiz (EKNZ) on February 25th, 2021 (Protocol No. 2021-00327). We thank the Computer Science students Filip Kalapish, Lukas Nüesch, and Geventh Vignendra for their labeling efforts and Xiaojing Xia and Camille Montemagni for their contributions to this project as part of their master theses. We further extend our gratitude to Samuel Claudio Drack, Maria Fernández Pose, Matias Emanuel Maissen, and José Wittor de Macêdo Santos for the data curation and to Prashanth Chandran for his valuable feedback on the face model development. Lastly, we thank the parents for giving their consent for the use of the image material presented in Figures 1, 5, and 7.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid icp algorithms for surface registration. In: IEEE Conf Comput Vis Pattern Recognit (CVPR). pp. 1–8 (2007). https://doi.org/10.1109/CVPR.2007.383165
2. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. p. 187–194. SIGGRAPH '99, ACM Press/Addison-Wesley Publishing Co., USA (1999). https://doi.org/10.1145/311535.311556
3. Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S.: Large scale 3d morphable models. International Journal of Computer Vision **126**(2), 233–254 (2018). https://doi.org/10.1007/s11263-017-1009-7

4. Boyaci, O., Serpedin, E., Stotland, M.A.: Personalized quantification of facial normality: a machine learning approach. Scientific Reports **10**(1), 21375 (2020). https://doi.org/10.1038/s41598-020-78180-x
5. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics **20**(3), 413–425 (2014). https://doi.org/10.1109/TVCG.2013.249
6. Chandran, P., Bradley, D., Gross, M., Beeler, T.: Semantic deep face models. In: 2020 International Conference on 3D Vision (3DV). pp. 345–354. IEEE Computer Society, Los Alamitos, CA, USA (2020). https://doi.org/10.1109/3DV50981.2020.00044
7. De Stefani, A., Barone, M., Hatami Alamdari, S., Barjami, A., Baciliero, U., Apolloni, F., Gracco, A., Bruno, G.: Validation of vectra 3d imaging systems: A review. International Journal of Environmental Research and Public Health **19**(14) (2022). https://doi.org/10.3390/ijerph19148820
8. Egger, B., Smith, W.A.P., Tewari, A., Wuhrer, S., Zollhoefer, M., Beeler, T., Bernard, F., Bolkart, T., Kortylewski, A., Romdhani, S., Theobalt, C., Blanz, V., Vetter, T.: 3d morphable face models—past, present, and future. ACM Trans. Graph. **39**(5) (2020). https://doi.org/10.1145/3395208
9. Goto, L., Lee, W., Huysmans, T., Molenbroek, J.F.M., Goossens, R.H.M.: The variation in 3d face shapes of dutch children for mask design. Applied Sciences **11**(15) (2021). https://doi.org/10.3390/app11156843
10. Hallgrímsson, B., Aponte, J.D., Katz, D.C., Bannister, J.J., Riccardi, S.L., Mahasuwan, N., McInnes, B.L., Ferrara, T.M., Lipman, D.M., Neves, A.B., Spitzmacher, J.A.J., Larson, J.R., Bellus, G.A., Pham, A.M., Aboujaoude, E., Benke, T.A., Chatfield, K.C., Davis, S.M., Elias, E.R., Enzenauer, R.W., French, B.M., Pickler, L.L., Shieh, J.T.C., Slavotinek, A., Harrop, A.R., Innes, A.M., McCandless, S.E., McCourt, E.A., Meeks, N.J.L., Tartaglia, N.R., Tsai, A.C.H., Wyse, J.P.H., Bernstein, J.A., Sanchez-Lara, P.A., Forkert, N.D., Bernier, F.P., Spritz, R.A., Klein, O.D.: Automated syndrome diagnosis by three-dimensional facial imaging. Genetics in Medicine **22**(10), 1682–1693 (2020). https://doi.org/10.1038/s41436-020-0845-y
11. Knoops, P.G.M., Papaioannou, A., Borghi, A., Breakey, R.W.F., Wilson, A.T., Jeelani, O., Zafeiriou, S., Steinbacher, D., Padwa, B.L., Dunaway, D.J., Schievano, S.: A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. Scientific Reports **9**(1), 13597 (2019). https://doi.org/10.1038/s41598-019-49506-1
12. Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) **36**(6), 194:1–194:17 (2017). https://doi.org/10.1145/3130800.3130813
13. Morales, A., Alomar, A., Porras, A.R., Linguraru, M.G., Piella, G., Sukno, F.M.: Babynet: Reconstructing 3d faces of babies from uncalibrated photographs. Pattern Recognition **139**, 109367 (2023). https://doi.org/https://doi.org/10.1016/j.patcog.2023.109367
14. Morales, A., Porras, A.R., Tu, L., Linguraru, M.G., Piella, G., Sukno, F.M.: Spectral correspondence framework for building a 3d baby face model. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). pp. 708–715 (2020). https://doi.org/10.1109/FG47880.2020.00079
15. Nord, F., Ferjencik, R., Seifert, B., Lanzer, M., Gander, T., Matthews, F., Rücker, M., Lübbers, H.T.: The 3dmd photogrammetric photo system in craniomaxillofacial surgery: Validation of interexaminer variations and perceptions. Jour-

nal of Cranio-Maxillofacial Surgery **43**(9), 1798–1803 (2015). https://doi.org/https://doi.org/10.1016/j.jcms.2015.08.017

16. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3d face model for pose and illumination invariant face recognition. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 296–301 (2009). https://doi.org/10.1109/AVSS.2009.58
17. Pereira, R.C., Santos, M.S., Rodrigues, P.P., Abreu, P.H.: Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes. Journal of Artificial Intelligence Research **69**, 1255–1285 (2020). https://doi.org/10.1613/jair.1.12312
18. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 3172–3182 (2022). https://doi.org/10.1109/WACV51458.2022.00323
19. Wan, M., Zhu, S., Luan, L., Prateek, G., Huang, X., Schwartz-Mette, R., Hayes, M., Zimmerman, E., Ostadabbas, S.: Infanface: Bridging the infant–adult domain gap in facial landmark estimation in the wild. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 4486–4492. IEEE Computer Society, Los Alamitos, CA, USA (2022). https://doi.org/10.1109/ICPR56361.2022.9956647
20. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8798–8807 (2018). https://doi.org/10.1109/CVPR.2018.00917
21. Zhou, K., Bhatnagar, B.L., Pons-Moll, G.: Unsupervised shape and pose disentanglement for 3d meshes. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 341–357. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58542-6_21
22. Zielonka, W., Bolkart, T., Thies, J.: Towards metrical reconstruction of human faces. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 250–269. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-19778-9_15