Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine
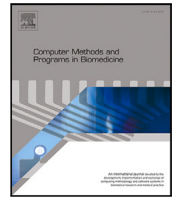
Facial surgery preview based on the orthognathic treatment prediction

Huijun Han [a,e] , Congyi Zhang [b] *, Lifeng Zhu [c], Pradeep Singh [a], Richard Tai-Chiu Hsung [d] ,
Yiu Yan Leung [d], Taku Komura [b], Wenping Wang [e], Min Gu [a,*]

[a] Discipline of Orthodontics, Faculty of Dentistry, the University of Hong Kong, Hong Kong Special Administrative Region of China
[b] Department of Computer Science, Faculty of Engineering, the University of Hong Kong, Hong Kong Special Administrative Region of China
[c] School of Instrument Science and Engineering, Southeast University, Nanjing, China
[d] Discipline of Oral and Maxillofacial Surgery, the University of Hong Kong, Hong Kong Special Administrative Region of China
[e] Department of Computer Science and Engineering, Texas A&M University, TX, USA

ARTICLE INFO

ABSTRACT

Background and Objective: Orthognathic surgery consultations are essential for helping patients understand how their facial appearance may change after surgery. However, current visualization methods are often inefficient and inaccurate due to limited pre- and post-treatment data and the complexity of the treatment. This study aims to develop a fully automated pipeline for generating accurate and efficient 3D previews of postsurgical facial appearances without requiring additional medical images.
Methods: The proposed method incorporates novel aesthetic criteria, such as mouth-convexity and asymmetry, to improve prediction accuracy. To address data limitations, a robust data augmentation scheme is implemented. Performance is evaluated against state-of-the-art methods using Chamfer distance and Hausdorff distance metrics. Additionally, a user study involving medical professionals and engineers was conducted to evaluate the effectiveness of the predicted models. Participants performed blinded comparisons of machine learning-generated faces and real surgical outcomes, with McNemar's test used to analyze the robustness of their differentiation.
Results: Quantitative evaluations showed high prediction accuracy for our method, with a Hausdorff Distance of 9.00 millimeters and Chamfer Distance of 2.50 millimeters, outperforming the state of the art. Even without additional synthesized data, our method achieved competitive results (Hausdorff Distance: 9.43 millimeters, Chamfer Distance: 2.94 millimeters). Qualitative results demonstrated accurate facial predictions. The analysis revealed slightly higher sensitivity (54.20% compared to 53.30%) and precision (50.20% compared to 49.40%) for engineers compared to medical professionals, though both groups had low specificity, approximately 46%. Statistical tests showed no significant difference in distinguishing Machine Learning-Generated faces from Real Surgical Outcomes, with p-values of 0.567 and 0.256, respectively. Ablation tests demonstrated the contribution of our loss functions and data augmentation in enhancing prediction accuracy.
Conclusion: This study provides a practical and effective solution for orthognathic surgery consultations, benefiting both doctors and patients by improving the efficiency and accuracy of 3D postsurgical facial appearance previews. The proposed method has the potential for practical application in pre-surgical visualization and aiding in decision-making.

1. Introduction

Orthognathic surgery addresses facial asymmetry and abnormalities, significantly improving aesthetics, oral function, and psychosocial well-being. Despite these benefits, uncertainty about the appearance of the postoperative period often leads to anxiety among patients, affecting their decision-making process and communication with doctors. Visualizing expected surgical outcomes has emerged as a crucial step in mitigating presurgical anxiety, setting realistic expectations, and improving overall patient satisfaction, particularly when CBCT data is not available.

Data-driven approaches have shown promise in assessing surgical necessity and complexity. For example, neural networks have been used to predict the need for orthognathic surgery from facial photographs [1] and to estimate the difficulty of tooth extraction from radiographic images [2]. However, these studies focus primarily on probability predictions rather than visual outcomes.
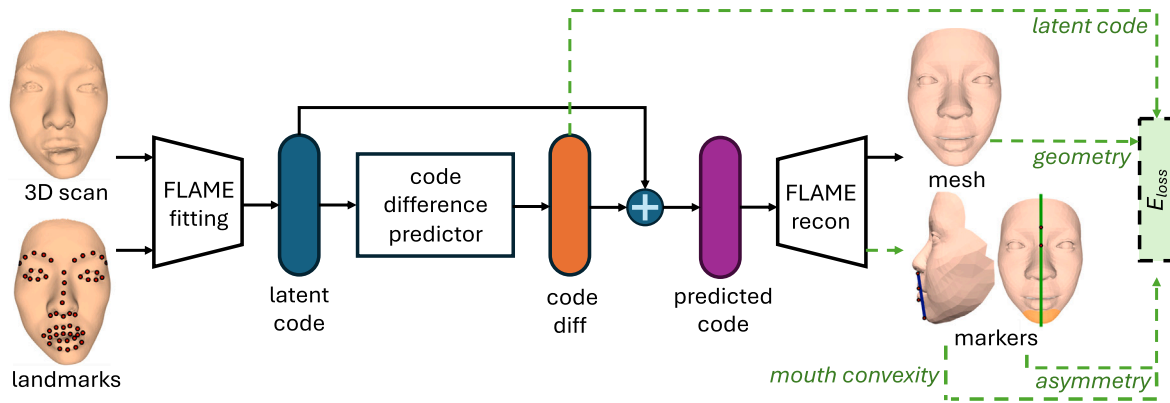
---

* Corresponding author.
E-mail addresses: cyzhang@cs.hku.hk (C. Zhang), drgumin@hku.hk (M. Gu).

Fig. 1. Net architecture for predicting the postoperative appearance from a captured 3D scan. During the training phase, the captured mesh and auto-annotated landmarks are first passed through the FLAME fitting procedure, where they are transformed into a compressed latent code. It then passes through a predictor and has the code difference added to it. The FLAME reconstructing procedure helps to calculate well-designed losses using markers and mesh. With the help of medical, latent code, and geological types of loss, the parameter of the code difference predictor can be continuously updated. During the testing phase, the data do not flow through the dashed line. The predicted postoperative appearance is generated through reconstruction using FLAME.

To address the need for predictive visualizations, researchers have begun exploring the use of machine learning algorithms to predict facial appearance after surgery [3]. These algorithms employ various advanced methods, including dense multilayer perceptrons (MLP) [4–10], conditional generative adversarial networks (cGAN) [11,12], transformers [13], and convolutional neural networks (CNN) [14]. These deep learning models excel at capturing spatial hierarchies in images, enabling refined feature extraction and accurate prediction of 3D coordinate changes. Despite these advancements, existing methods remain clinician-oriented, heavily reliant on X-ray or CT data and anatomical landmarks provided by medical professionals [15]. This reliance limits full automation and increases costs, making these approaches inaccessible for patients seeking quick and intuitive consultations. Moreover, several limitations persist. For example, the model in [6] overlooks key reductions in facial asymmetry after surgery, potentially compromising accuracy for asymmetric patients. Similarly, Tanikawa's network [5], trained on fewer than 100 patients, faces overfitting issues due to high-dimensional input and output layers (18,000), while requiring CBCT data, which are challenging to obtain. Other approaches, such as Park's cGAN [11], are effective for orthodontic surgery but lack robustness for orthognathic procedures due to their complexity. Finally, methods like FC-Net [16] require clinicians to provide precise bone movement data during inference, limiting their usability for direct patient engagement.

To accurately predict 3D post-surgery facial appearance, it is essential to use parametric 3D facial reconstruction methods that capture fine articulation, especially the jaw joint, which is critical for both aesthetic and functional outcomes in orthognathic surgery. Existing methods like Large Scale Facial Model (LSFM) [17] and Structure-Aware Editable Morphable Model (SEMM) [18] fail to model dynamic jaw movement, treating the face as a morphable structure where all features, including the jaw, are considered as a whole. SCULPTOR [19], an articulated model, can capture skull-face joint distribution but is difficult to integrate due to limited data and incomplete code. To overcome this, we integrate FLAME [20–22], a model with nonlinear jaw articulation, enabling vivid facial reconstruction and efficient encoding for accurate post-surgical predictions.

Our algorithm is designed to predict 3D post-surgery facial appearance based on preoperative facial scans of patients without the need for additional medical images (e.g., X-ray or CBCT). To improve the robustness of the model, we develop a data augmentation method that substantially amplifies the available dataset for orthognathic treatment. To explicitly encode the assessment rules in orthognathic treatment, we introduce two medical losses, mouth-convexity loss and asymmetry loss, consistent with surgical definitions that help achieve the optimization goals of surgery.

In summary, our **contributions** are: (1) We integrate two criteria — minimizing mouth-convexity and minimizing asymmetry — into the machine learning procedure, which align with the goals of orthognathic surgery, to enhance prediction accuracy; (2) we devise an augmentation method to expand the available dataset, achieving more robust predictions; (3) we complete a fully automated pipeline that showcases the postoperative changes in facial appearance to patients in 3D.

2. Methods

We present a simple and effective pipeline to predict 3D face models after surgery, as shown in Fig. 1. The reconstructed 3D facial model is initially encoded into a latent code using FLAME [20], a parametric facial model. A face predictor then estimates the difference between post- and pre-surgery facial features, guided by custom-tailored loss functions that specifically incorporate facial asymmetry and mouth-convexity rules relevant to orthognathic surgery (explained in Section 2.1). These rules are directly applied to the predicted facial geometry and back-propagated through the network in a differentiable manner (discussed in Section 2.3). To provide a fully textured post-treatment preview, we leverage the mesh coherence of the FLAME model, transferring the texture from pre-treatment 3D facial scans to the predicted face models (explained in Section 2.4). To address the challenge of limited training data, we introduce a novel data augmentation technique that significantly increases the number of face pairs, enhancing the robustness of the model (described in Section 2.2). The effectiveness of the predicted models is evaluated through a user study involving both medical professionals and engineers, with the robustness of the predictions assessed using McNemar's test (outlined in Section 2.5).

2.1. Prediction network

We adopt a predictor and its associated losses for the prediction of facial appearance of orthognathic surgery, as shown in Fig. 1. As a parameterized and differentiable facial model, FLAME can serve as both a pre-trained encoder, compressing geometric information into a latent code, and also a decoder layer, providing the necessary geometric information in a differentiable manner for explicit supervision, which helps update the parameters in the code difference predictor.

The overall loss function used in our algorithm is composed of four components: the mouth convexity loss, the asymmetry loss, the latent code loss, and the geometry loss. The main challenge in training the predictor comes from the imperfect training data. Our training data consist of real orthognathic surgery cases. However, due to some practical limitations or customized considerations, the post-treatment facial
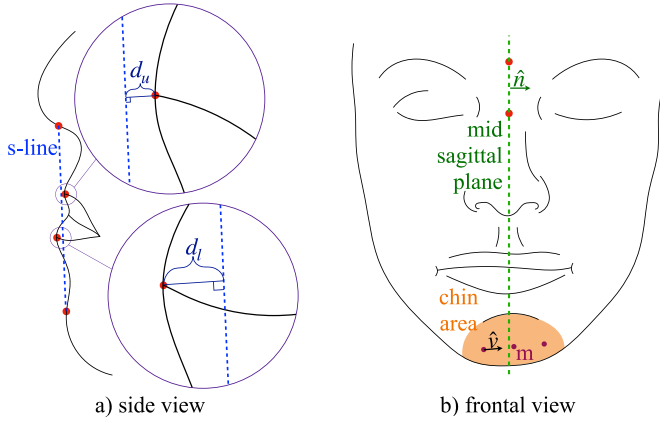
Fig. 2. Side view (left) and frontal view (right) of a orthognathic patient for calculating mouth-convexity loss and asymmetry loss respectively.

appearance may not be ideal from a clinical perspective, e.g. some of them are still asymmetric or protruding to a certain extent. Adopting a pure data-driven loss in supervision would yield the same artifacts as observed in the training data. To this end, we formulate two explicit clinical rules as our novel losses, the mouth-convexity loss and the asymmetry loss, to enhance the functional and aesthetic aspects.

Mouth-convexity loss: *Mouth convexity* refers to a measurement used in orthognathic surgery to describe the relative position of the mouth, nose, and chin. Mouth convexity helps categorize facial profiles into two main types: convex profiles (where the lower jaw protrudes outward, creating a rounded appearance) and concave profiles (where the lower jaw appears retruded or inwardly curved, resulting in a flatter facial contour).

Orthognathic surgery is a highly effective approach to correcting mouth convexity. By carefully realigning the jawbones, this surgical procedure can significantly improve both the functional and aesthetic aspects of the patient's facial structure. Therefore, we tailor a specific loss function that penalizes the protruding mouth issues, so as to enable the network to generate more pleased facial geometries.

According to [23], a reference line known as *Steiner-line* (s-line, blue dashed line in Fig. 2a) can be drawn from the middle of the nose base to *pogonion* (the extreme anterior point of the chin) to serve as a basis for assessing the protrusion of the mouth. We denote the distances from the s-line to the midpoints of the upper and lower lips as d_u and d_l respectively. Medical standards suggest that lip midpoints within a range of 3 millimeters deviation from the s-line are acceptable. To this end, we devise a mouth convexity loss function $L(d)$, where the squared distance is used to penalize cases with mouth convexity.

$$L(d) = \begin{cases} 0, & \text{if } d < 3 \text{ mm} \\ (d-3)^2, & \text{if } d \geq 3 \text{ mm} \end{cases} \tag{1}$$

Mouth-convexity loss L_p is the sum of $L(d_u)$ and $L(d_l)$.

Asymmetry loss: The degree of chin asymmetry in orthognathic surgery is a major concern, and can be quantified by measuring the symmetry of the chin with respect to the *mid sagittal plane* [24] shown in Fig. 2b, which is the focus of our designed asymmetry loss.

According to the medical definition, we determine the mid-sagittal plane S (green dashed line in Fig. 2b) by solving a plane that passes through points at the midpoint of the eyebrows and the midpoint of the inner corners of the eyes (red points in Fig. 2b). Because all 3D reconstructed head models of patients are captured with the same natural head position, the ideal unit normal vector $\hat{n}$ of the mid-sagittal plane is always along the x-axis. In practice, for each case, we solve the least squares system to determine the mid-sagittal plane S.

As FLAME uses a topologically symmetric template mesh, we first pair all vertices in the chin area, denoted as $\{p_i, q_i\}_{i=1,\dots,k}$ (orange

region in Fig. 2b). Then we compute the unit direction vectors $\hat{v}_i$ of the line segments connecting the paired points p_i and q_i and their midpoints $m_i = \frac{1}{2}(p_i + q_i)$. Following that, we can measure the overall asymmetry of paired points with respect to the mid-sagittal plane using the asymmetry loss

$$L_a = \sum_{i=1}^{k} d(m_i, S) + (1 - \hat{n} \cdot \hat{v}_i) \tag{2}$$

where $d(m, S)$ denotes the distance function from point m to mid-sagittal plane S and $\cdot$ is the dot product.

Latent code loss: With the aid of a parameterized model serving as an encoder, each pair of preoperative and postoperative captured scans can be encoded as a pair of latent codes. To improve the approximation between the predicted and ground-truth values in the latent space during the training phase, we calculate the squared error of the latent codes as our latent code loss L_f:

$$L_f = \|\vec{\beta}_{pred} - \vec{\beta}_{gt}\|_2^2 \tag{3}$$

where $\|\cdot\|_2$ denotes the l^2 norm, $\vec{\beta}_{pred}$ and $\vec{\beta}_{gt}$ are the latent codes of predicted and postoperative GT faces.

Geometry loss: The geometry loss L_g consists of two parts: the point-to-point distance and surface normal errors between the predicted and true meshes:

$$L_g = \frac{1}{N} \sum_{i=1}^{N} \|p_i^{pred} - p_i^{gt}\|_2^2 + w \frac{1}{M} \sum_{j=1}^{M} 1 - cos(\theta_j), \tag{4}$$

where N and M are the numbers of points and triangles in the face region, p^{pred} and p^{gt} are the points on the predicted and ground-truth mesh, θ is the angle between the predicted and ground-truth surface normal, and w is the balancing weight.

The geometry part is designed to encourage all the points within our facial mask to be close to their ground-truth values, preventing solely focusing on the chin and mouth areas due to the mouth-convexity loss and asymmetry loss. In addition, the normal part helps the network distinguish between the upper and lower lips, aiding in a better understanding of the facial appearance. We conducted thorough ablation studies in Section 4.1 to verify that geometry loss and latent code loss are not redundant.

Overall loss function: we compute the weighted sum of the four losses to form the total loss:

$$L = \alpha_p L_p + \alpha_a L_a + \alpha_f L_f + \alpha_g L_g. \tag{5}$$

2.2. Data augmentation

A sufficient amount of diverse data is crucial for enhancing the robustness of machine learning models [25]. In orthognathic surgery, data augmentation offers significant advantages by improving the algorithm's ability to generalize across a wide range of cases. By simulating diverse populations, data augmentation enhances the model's predictive accuracy and ensures reliable 3D facial outcome predictions.

A common assumption in orthognathic surgery is that there is a horizontal plane dividing the unchanged upper part of the face (above the plane) from the key modified region (below the plane), where the surgical alterations occur. The goal of our data augmentation method is to generate synthetic data where the upper face remains intact, preserving its natural characteristics above the horizontal plane, while the lower part, which is affected by the surgery, is modified according to the surgical requirements. Traditional 3D data augmentation techniques, such as translation, rotation, and scaling, may not be effective in our framework, as they do not generate meaningful geometry that meets the requirements for preserving the natural structure and characteristics of the orthognathic surgery.

To create the synthetic data, we first identify the horizontal plane P_s that separates the unaltered upper face from the altered lower region.

(a) segmentation   (b) conditional generation   (c) stitching
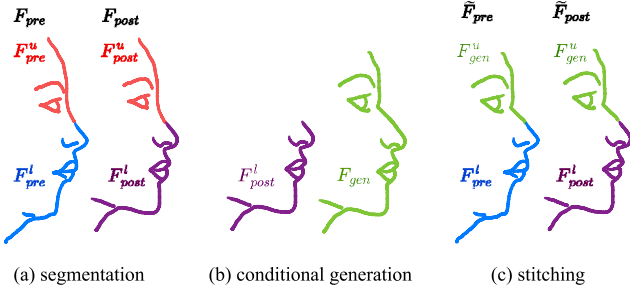
Fig. 3. An illustration of synthetic data generation process. The process involves three stages: face segmentation, conditionally generating a synthetic face F_{gen}, and stitching the upper and lower regions to create plausible pre- and post-surgery pairs for training.

Table 1
Comparison of Sensitivity, Specificity, and Precision values for Engineers and Medical Professionals in distinguishing Machine Learning-Generated faces (MLG) and Real surgical outcomes (RSO).

	Medical Professionals	Engineers
Sensitivity	53.30%	54.20%
Specificity	45.30%	46.20%
Precision	49.40%	50.20%

As illustrated in Fig. 3, for the pre-surgery scan, we obtain F_{pre}^u (upper part) and F_{pre}^l (lower part). For the post-surgery scan, we obtain F_{post}^u (upper part) and F_{post}^l (lower part). Note that F_{pre}^u is the same as F_{post}^u. Next, we conditionally generate a synthetic face F_{gen} based on the lower part of the postoperative scan F_{post}^l with a random variable . Then, we use the horizontal plane P_s to separate the upper part of F_{gen}, denoted as F_{gen}^u. Finally, we create the pair of generated faces $\{\tilde{F}_{pre}, \tilde{F}_{post}\}$ by stitching F_{gen}^u with both F_{pre}^l and F_{post}^l, respectively. We use a random variant ξ to alter the latent code of the upper face, and then we use the decoder to reconstruct the facial geometry. We randomize the ξ to synthesize multiple plausible pairs of $\{\tilde{F}_{pre}, \tilde{F}_{post}\}$ to train the prediction network. This approach ensures that the synthetic data reflects the surgical changes below the horizontal plane while keeping the upper part of the face unchanged.

In summary, our goal is to generate paired facial scans that reflect the pre- and post-surgery appearance. We modify the upper part of the face while directly copying the lower part from the dataset, following the paired changes between the pre- and post-surgery scans. As a result, we obtain a pair of faces where the upper part remains consistent, and the lower part exhibits the surgical changes.

Additionally, we include a data cleaning step to remove outliers introduced by the fitting process. By calculating the fitting error at key facial landmarks and applying a threshold, we ensure that only meshes with an acceptable error are retained, thereby guaranteeing the accuracy and reliability of the synthetic meshes. To evaluate the impact of data augmentation, we conducted an ablation study, with the results presented in the last row of Table 6 and the last row of Fig. 5

2.3. Integrating FLAME as an encoder–decoder

In our pipeline, FLAME is integrated not only as an encoder to obtain the latent code of captured scans, but also as a decoder to reconstruct facial meshes. As a full-head model with an underlying face skeleton tree composed of neck, jaw, and eyeball joints, FLAME model [20] is defined as:

$$M(\vec{\beta}, \vec{\theta}, \vec{\psi}) = W(T_P(\vec{\beta}, \vec{\theta}, \vec{\psi}), \mathbf{J}(\vec{\beta}), \vec{\theta}, \mathcal{W}), \qquad (6)$$

where $\vec{\beta}$, $\vec{\theta}$ and $\vec{\psi}$ denote shape, pose and expression vector respectively. During our fitting and reconstruction procedure, the expression vector $\vec{\psi}$ is not involved, because the patient scanning data were obtained when they were relaxed and in neutral facial expressions and poses.

2.4. Visualization of prediction as a textured mesh

We aim to provide individuals seeking surgical consultations with a textured 3D mesh that can be viewed from multiple angles, offering a comprehensive and intuitive visualization of their postoperative appearance.

To achieve this, we deform the textured scan to align with the geometry of the predicted parametric model. Specifically, using barycentric coordinates, we deform the vertices to match the geometry of the predicted model. For each vertex in the scanned mesh, the closest point in the parametric model is identified as its correspondence, and the displacement for each vertex is computed based on the barycentric coordinates ($\lambda_1, \lambda_2, \lambda_3$) and their predicted displacements d_1, d_2, and d_3. The texture is then morphed using the formula $\lambda_1 d_1 + \lambda_2 d_2 + \lambda_3 d_3$ to approach the geometry of the predicted parametric model.

2.5. User study

The aim of this study was to assess the participants' ability to differentiate between machine learning-generated faces (MLG) and real surgical outcomes (RSO). Participants in this study consisted of five medical professionals (comprising two surgeons and three orthodontists) with a minimum of 5 years of experience in their respective fields. Additionally, 15 engineers (including twelve males and three females) were included in the study, each possessing a minimum of 3 years of relevant experience in engineering and technical disciplines. The selection criteria for medical professionals were based on their expertise in surgical procedures and orthodontic treatments, ensuring a minimum threshold of 5 years of clinical experience. Engineers were chosen for their technical background and familiarity with machine learning concepts, with a minimum of 3 years of professional experience in their respective fields. Preoperative and post-surgery 3D facial data were collected for each patient, standardized using FLAME parameters, and textures were applied consistently across all data. To facilitate participants' observation of facial symmetry and feature positioning, both frontal and side views of the faces were provided, along with a 180-degree rotating animation from left to right. The preoperative images were clearly labeled, while the MLG and RSO were presented to the participants in a blinded manner. Each participant was simultaneously shown two images (labeled as A and B) and asked to distinguish between MLG and RSO without knowing the group to which each image belonged. A total of 30 randomly selected images (including A and B) were presented to each participant. Participants were required to digitally record their responses. The responses provided by the participants were then compared with the correct answers separately for engineers and medical professionals, and specificity, sensitivity, and precision values were calculated for each group. To further analyze the participants' ability to differentiate between MLG and RSO, a McNemar's test was conducted for each group.

3. Results

3.1. Datasets

A total of 163 pairs of pre- and post-operative 3D facial scans were collected using the 3dMD facial scan acquisition system from patients undergoing orthognathic surgery at the University of Hong Kong School of Dentistry. Static three-dimensional (3D) images of each participant were taken using a 3dMDface System (3dMD Inc., Atlanta, GA, USA) by professional photographers. The accuracy of the system had been previously published and was reported to be lower than 0.2 mm root mean square (RMS) [26]. The system was calibrated according to the manufacturer's instructions before each image capture. During scanning, patients were relaxed and instructed to look straight into a

Table 2
Comparison of Engineers' and Medical Professionals' ability to distinguish between Machine Learning-Generated faces and Real surgical outcomes.

	Medical Professionals (N_{med} = 5) n = 150	Engineers (N_{eng} = 15) n = 450
MLG	34 (23%)	104 (23%)
MLG identified as RSO	41 (27%)	121 (27%)
RSO identified as MLG	35 (23%)	103 (23%)
RSO	40 (27%)	122 (27%)
95% CI	12.58 to 17.02	14.01 to 16.12
P-value	0.567	0.256

N_{med}, Number of medical professionals; N_{eng}, Number of engineers; n, number of responses;
MLG, Machine learning generated faces; RSO, Real surgical outcomes; CI, Confidence Interval;
McNemar chi-square test was performed unless otherwise mentioned.
$p < 0.05$ (in **bold** *italics*), considered statistically significant.

Table 3
Confusion matrix presenting True Positives, False Positives, True Negatives, and False Negatives.

		True Responses	
		RSO	MLG
Predicted Responses	Predicted RSO_med	53.30%	54.70%
	Predicted MLG_med	46.70%	45.30%
	Predicted RSO_eng	54.20%	53.80%
	Predicted MLG_eng	45.80%	46.20%

RSO: Real surgical outcomes;
MLG: Machine learning-generated faces.

mirror at their own eyes, ensuring the capture of their forehead, chin, and ears. Immediately prior to image capture, participants were seated 100 cm away from the system, looking forward with the Frankfort plane parallel to the floor, and any glasses and jewelry were removed. The camera system captured six 2D images—four black-and-white pictures to depict facial structures and spatial relationships, and two colored images to project texture information onto the mesh framework. The scan process took 1.5 ms and the 3D facial surfaces were exported as Wavefront OBJ files for further processing. Postoperative scans were recorded at least three months after the surgery to ensure that facial swelling had subsided, providing an accurate reflection of the patient's post-recovery appearance. The male-to-female ratio was 59:104, with the majority of patients being of Asian descent.

3.2. Implementation details

3.2.1. Exclusion of non-facial elements

Data cleaning is a crucial step in our prediction algorithm, as it ensures that the focus remains solely on the relevant facial regions by deliberately excluding non-facial elements such as hair and disposable surgical caps. This process is essential to eliminate extraneous data that could introduce noise, allowing the algorithm to concentrate on critical facial features like the chin and nose. By removing these non-facial elements, we not only enhance the fitting accuracy but also improve the overall performance of the model. The careful segmentation and exclusion of non-facial data ensure that the algorithm's analysis is based on high-quality, relevant information, leading to more precise and reliable predictions.

To isolate the facial regions, we render the 3D scans from three distinct viewpoints: frontal and both side views with a 45-degree rotation. We then apply BiSeNet [27], a bilateral segmentation network, to segment the facial regions from these images, retaining only the pertinent areas for subsequent analysis. Once the 2D segmentation maps are obtained, we reproject them back into the 3D space to ensure that the segmented regions correspond accurately to the facial areas in the original 3D scans. The non-facial elements, identified in the segmentation, are then removed from the 3D scans, leaving only the relevant facial regions for further analysis.

3.2.2. Landmarks annotation

We utilized Supervision by Registration [28] in conjunction with a facial landmark detector to automatically annotate facial landmarks. A point light was positioned in front of the patient's face to render the 3D mesh into a frontal view image. A pre-trained landmark detector was then applied to extract 2D coordinates for 68 facial landmarks. These 2D coordinates were subsequently transformed back to 3D using the vertex-to-pixel mapping. For challenging preoperative cases, any discrepancies or drift in the landmark positions were manually corrected by a professional to ensure the dataset's accuracy and reliability.

3.2.3. Training settings

Our neural network comprises two fully connected modules with a hidden layer of 100 dimensions and input and output layers of 300 dimensions. Additionally, the modules are connected by a batch normalization layer, a nonlinear layer activated by ReLU, and a dropout layer with a 50% probability. The balance parameters α_p, α_a, α_f, and α_g are set to 5000, 5000, 1, and 1 respectively.

During the training process, we set the batch size to 150, and trained our model for a total of 500 epochs. The original learning rate was set to 10^{-3}, and we employed a learning rate decay strategy. Specifically, we decayed the learning rate by 50% every 100 epochs, which helped to prevent over-fitting and improve the generalization ability of our model. We conducted our training on a NVIDIA RTX 3090 GPU.

The training process took approximately 25 min to complete. To ensure the robustness and reliability of our results, we employed a 5-fold cross-validation strategy. The 163 valid pairs of facial scans were subjected to random shuffle and then split into 5 consecutive folds for cross-validation. Each fold was used once as the validation set, while the remaining four folds formed the training set. Data splitting was performed prior to data augmentation, which was applied exclusively to the training dataset. We report the average score in Table 4.

3.3. Analysis of user study results

The sensitivity, specificity, and precision values are presented in Table 1. The results indicated that engineers had a slightly higher sensitivity percentage (54.20%) compared to medical professionals (53.30%), suggesting that engineers were slightly better at identifying Machine Learning-Generated faces (MLG) and Real Surgical Outcomes (RSO) accurately. However, both groups exhibited low specificity percentages, with engineers at 46.20% and medical professionals at 45.30%, indicating challenges in distinguishing between MLG faces and RSO. In terms of precision, engineers had a slightly higher percentage (50.20%) compared to medical professionals (49.40%), suggesting a slightly higher accuracy for engineers in identifying MLG faces. Nevertheless, there was no statistically significant difference in the ability to differentiate between MLG faces and RSO within each group, with p-values of 0.567 for engineers and 0.256 for medical professionals (Table 2). Additionally, Table 3 presents the confusion matrix, illustrating the classification performance for both groups. Noteworthily, although the subjective measure employed in this study holds value in assessing participants' ability to differentiate between MLG faces and RSO, it is important to acknowledge that subjective evaluation alone may not entirely validate clinical accuracy.

Table 4
Performance comparison.

Algorithms	HD* (mm) ↓			CD* (mm) ↓			Data Amount
	mean	min	max	mean	min	max	
OUR's	**9.00**	7.63	**11.30**	**2.50**	**1.24**	**3.60**	**1330**
LARS	9.68	7.50	15.41	2.77	1.72	4.77	1330
OUR's w/o synthesized data	9.43	8.00	13.53	2.94	1.91	6.38	133
LARS w/o synthesized data	9.67	**7.46**	16.13	2.98	1.57	5.21	133

HD and CD represent Hausdorff and Chamfer Distance respectively.

Table 5
Statistical analysis of Hausdorff and Chamfer distances.

Comparison	Metric	t-statistic	p-value	Significance
Our model vs LARS model	Hausdorff	2.113	0.039	Significant
Our model vs LARS model	Chamfer	2.143	0.036	Significant

3.4. Quantitative metrics

We introduce two key metrics to quantitatively evaluate the accuracy of the predictions: the Hausdorff distance and the Chamfer distance. These metrics are commonly used to assess the similarity between two point clouds or surfaces, providing insights into the geometric alignment of the predicted and ground truth meshes.

The **Hausdorff distance** measures the greatest of all the distances from a point in one set to the closest point in the other set [29]. It is defined as:

$$d_H(A, B) = \max \left(\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|b - a\| \right) \qquad (7)$$

where A and B are two point sets, and $\|a - b\|$ denotes the Euclidean distance between points a and b. The Hausdorff distance is sensitive to outliers and emphasizes the maximum deviation between the two sets.

The **Chamfer distance** provides a more balanced measure of the overall geometric difference between two point clouds [30]. It is computed as:

$$d_C(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \|a - b\|^2 + \frac{1}{|B|} \sum_{b \in B} \min_{a \in A} \|b - a\|^2 \qquad (8)$$

where $|A|$ and $|B|$ are the number of points in sets A and B, respectively, and $\|a-b\|$ is the Euclidean distance between points a and b. The Chamfer distance measures the average distance between the points in both sets, offering a less sensitive alternative to the Hausdorff distance for assessing surface matching.

Both of these metrics allow for a precise comparison of the predicted and actual 3D facial geometries, providing valuable insights into the performance of our model.

3.5. Quantitative comparison

These examples compare the real postoperative appearances of patients with the prediction results from our algorithm and LARS, the state-of-the-art model from [4]. We quantitatively compared the prediction errors of our network with those of LARS using the Chamfer distance (CD) and Hausdorff distance (HD) metrics. Table 4 summarizes the results across different settings.

With the full dataset (1330 samples), our network achieved a mean HD of 9.00 mm and a mean CD of 2.50 mm, outperforming LARS, which had a mean HD of 9.68 mm and a mean CD of 2.77 mm. This demonstrates that our network not only reduces overall errors but also minimizes large deviations, as indicated by lower maximum values for both metrics.

When trained without synthesized data (133 samples), our network still performed better than LARS, with a mean HD of 9.43 mm and a mean CD of 2.94 mm, compared to LARS's mean HD of 9.67 mm and mean CD of 2.98 mm. This highlights the effectiveness of our approach even in limited data scenarios, though the inclusion of synthesized data

further improves performance by enhancing the network's generalization capability. Additionally, we observed that data augmentation was very helpful in improving our performance, but had little effect on the performance of LARS.

Table 5 presents the results of the statistical analysis using a t-test to compare our model with the LARS model based on two distance metrics: Hausdorff and Chamfer distances. The t-statistics and corresponding p-values indicate that both metrics show statistically significant differences between the models. Specifically, for the Hausdorff distance, our model demonstrated a t-statistic of 2.113 with a *p*-value of 0.039, which is below the 0.05 threshold for significance. Similarly, for the Chamfer distance, the t-statistic was 2.143 with a *p*-value of 0.036, also demonstrating a significant difference. These results suggest that our model outperforms the LARS model in terms of both distance metrics.

3.6. Qualitative comparison

To qualitatively evaluate our algorithm, we selected representative cases for visualization, as shown in Fig. 4. These examples compare the real postoperative appearances of patients with the prediction results from our algorithm and LARS [4]. To ensure a fair assessment of accuracy, experiments were also conducted using the Basel network and compared against LARS. The showcased patients underwent bimaxillary orthognathic surgery, with or without genioplasty.

For the first patient, our approach effectively predicted the correction of mouth alignment, showing only minor discrepancies in the lower face when compared to LARS. Notably, in the lower nasal region, the linear model predicted a longer, wider, and lower nasal base, deviating significantly from the surgical plan. For the second patient, our algorithm successfully predicted the repositioning of the jaw, whereas LARS struggled to address the protruding chin accurately. In the cases of the third and fourth patients, both of whom underwent bimaxillary orthognathic surgery combined with genioplasty to shorten facial length by adjusting the chin's tilt angle, our algorithm precisely captured the jaw's angle changes, resulting in a more aesthetically balanced facial shape. In contrast, LARS failed to accurately predict these adjustments, underscoring the superior predictive capability of our method.

4. Discussion

4.1. Ablation experiment

We conducted an ablation study to examine the effectiveness of the four losses and data augmentation techniques introduced in our model. Both quantitative (as shown in Table 6) and qualitative (as shown in Fig. 5) results demonstrate their effectiveness.

The ablation study highlights the contributions of various components in our model, as summarized in Table 6. The full model (OUR's)
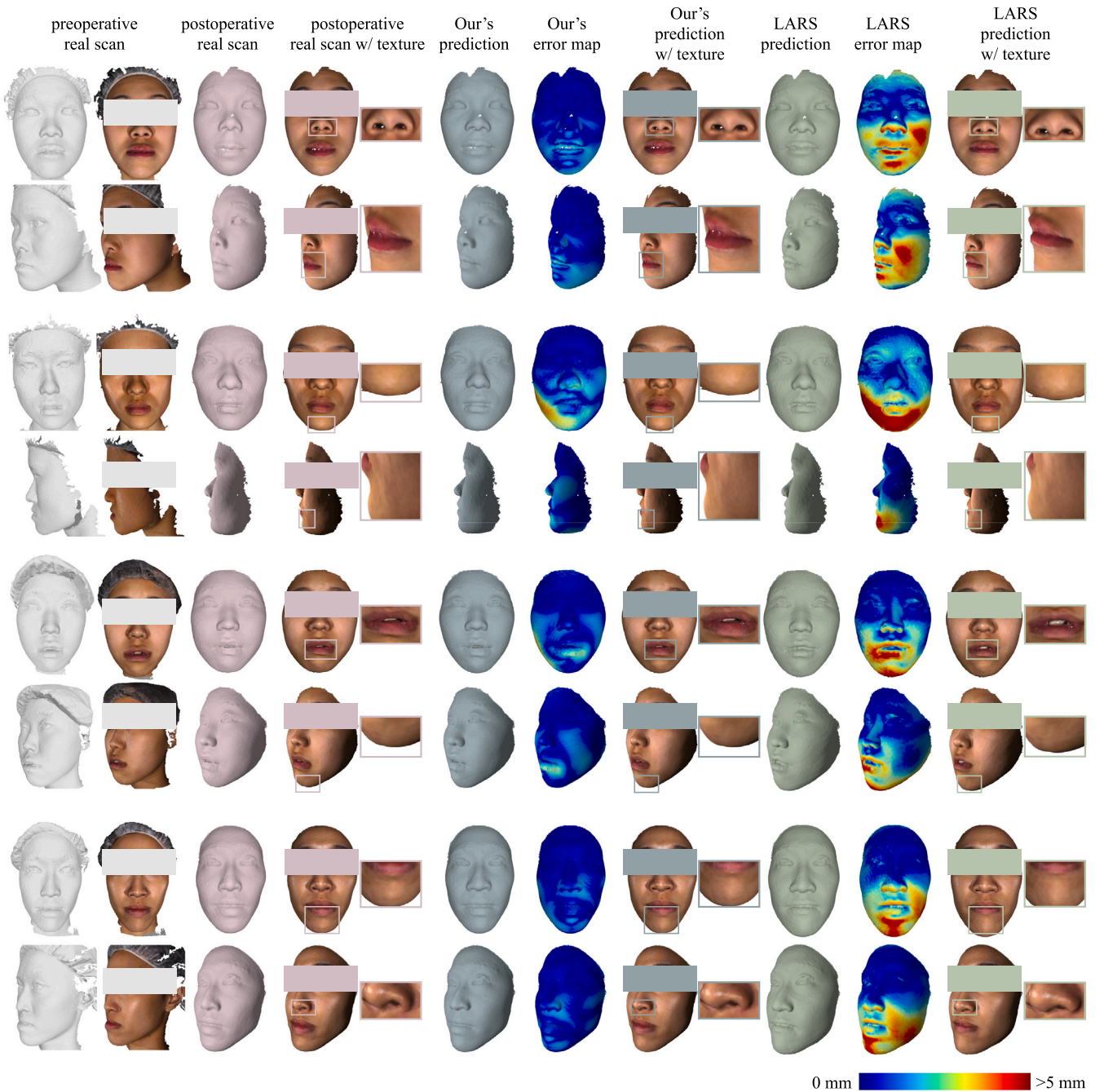
Fig. 4. Comparison of our method with LARS across four patient cases. On the left, the ground truth (GT) pre- and post-surgical scans are shown for reference. The middle columns display our predicted results, and the right columns show the LARS model's predictions. To compare the prediction errors with the real outcomes, heatmaps are provided showing the error distribution across facial regions, with the error bars located in the bottom-right corner.

Table 6
Ablation study of our model.

Prediction Network	Hausdorff Distance (mm)	Chamfer Distance (mm)	Training Data Amount
OUR's	**8.999**	**2.503**	**1330**
- Mouth-convexity loss	9.009	2.561	1330
- Asymmetry loss	9.142	2.571	1330
- Latent code loss	9.047	2.635	1330
- Geometry loss	9.237	2.644	1330
- Augmented Data	9.517	2.940	133

achieves the best performance with a Hausdorff Distance of 8.999 mm and a Chamfer Distance of 2.503 mm. Removing the mouth-convexity loss slightly degrades performance (0.11% in Hausdorff, 2.32% in Chamfer), while the asymmetry loss has a more noticeable impact (1.59% in Hausdorff, 2.72% in Chamfer). The latent code loss affects face consistency (0.53% in Hausdorff, 5.27% in Chamfer), and the
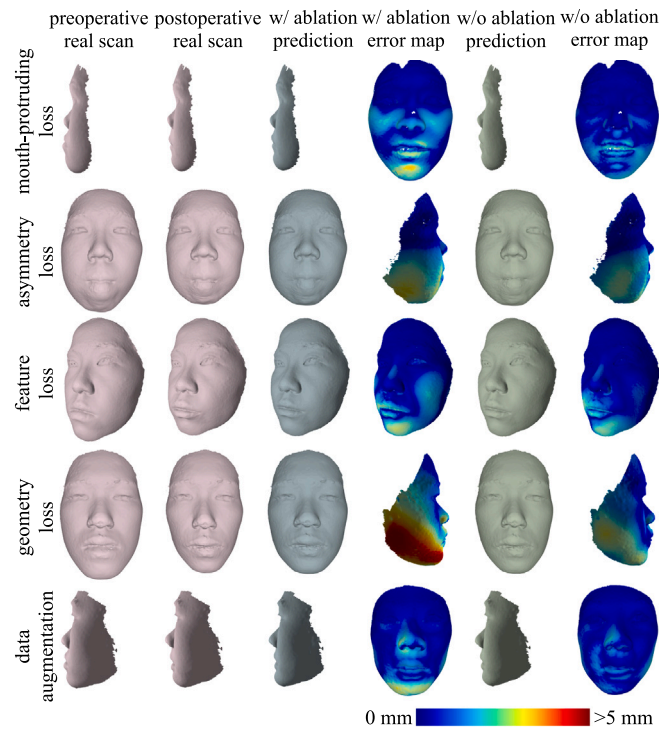
Fig. 5. Impact of losses and data augmentation on predictions. The left section illustrates actual surgical data, the middle section presents results after removing specific modules, and the right section showcases the full-model predictions.

geometry loss is critical, causing the largest increases among the four losses (2.64% in Hausdorff, 5.63% in Chamfer).

Data augmentation plays a pivotal role in model training, as its removal causes the most pronounced effect, increasing Hausdorff Distance by 5.75% and Chamfer Distance by 17.46%, while reducing the training dataset by 90%. The addition of augmented data appears to help reduce prediction error.

Fig. 5 further underscores these findings, vividly illustrating the qualitative impact of each component. For instance, removing the mouth-convexity loss leads to an exaggerated chin, while the absence of the asymmetry loss results in pronounced facial asymmetry. Similarly, excluding the latent code loss causes notable drift in the cheek contours, and removing the geometry loss significantly exacerbates overall structural errors. Additionally, data augmentation seems to play an important role, as its absence introduces noticeable inconsistencies in facial predictions.

4.2. Evaluation of methodology robustness

The results of our user study revealed that both engineers and medical professionals encountered similar challenges in accurately distinguishing between MLG faces and RSO. The minimal differences in sensitivity, specificity, and precision values between the two groups indicate that both groups faced similar difficulties, highlighting the reliability and accuracy of our novel method. The consistency in the performance of both groups confirms the effectiveness and robustness of our methodology and algorithm in predicting facial appearance following orthognathic treatment using only 3D face geometry.

4.3. Comparison with commercial VSP tools

Although Virtual Surgical Planning (VSP) tools (e.g., Dolphin, Morpheus 3D FaceMaker, and 3dMD Vultus) provide clinically accurate results for surgical planning, they require Computed Tomography (CT) or
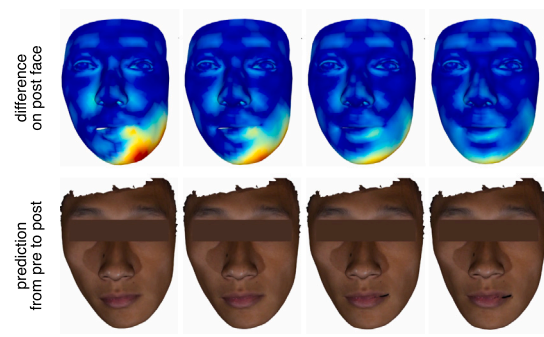
Fig. 6. Visualization of the latent space interpolation results. The faces generated from interpolated latent codes are shown in the second row. The face on the far left is the pre-surgery face, and the face on the far right is the prediction of the post-surgery outcome. The first row represents the distances between these faces and the post-surgery face. Use the same color map as in Fig. 5.

Cone Beam Computed Tomography (CBCT) imaging data as additional inputs [31].

A patient-oriented solution for visualizing surgical outcomes should be more accessible than a professional doctor-oriented one. In the early consultation stage, patients are not obligated to take CBCT or X-ray imaging, which would expose users to radiation. Instead, 3D facial reconstruction (e.g., 3dMD (3dMD Inc., Atlanta, GA, USA), Bellus3D (Bellus3D Inc., Los Gatos, CA, USA), and Ein-Scan3D (Shining 3D Technology, Hangzhou, China)) is much safer for post-treatment preview purpose.

Our fully automated and highly efficient method delivers accurate visualizations, offering an automatic preview of facial surgery outcomes that serves as an effective educational and motivational tool for patients, particularly during the consultation phase. By functioning as a visual aid, it enhances patients' understanding and acceptance of treatment plans. To further engage patients, we generated an animation by interpolating the latent vectors, illustrating the transformation from pre-surgery to post-surgery. This visualization highlights the movement of the chin and the overall changes in facial appearance, as depicted in Fig. 6. For a more comprehensive demonstration, please refer to our supplementary video.

5. Limitation

Despite the comprehensive analysis conducted in this study, several limitations need to be acknowledged. The current training dataset is limited to Asian patients, which may impact the generalizability of the findings. Therefore, future research endeavors can focus on expanding the dataset to encompass greater diversity, including various ethnicities for improved applicability across different populations. Additionally, the study may benefit from addressing long-term skeletal changes, as the dataset primarily consists of adults with post-surgical scans taken at least six months after surgery to capture stabilized facial structures. While results are deemed reliable over an extended period, ongoing monitoring and consideration of potential long-term changes will be crucial for enhancing the study's robustness. Furthermore, the model's inability to offer multimodal-based predictions for individual patient attributes, such as age, gender, and skin condition, underscores the need for additional data with multi-modal labels to enable more targeted predictions without compromising data integrity. The current focus on providing automatic visualization services for patients rather than adjustable parameters for orthognathic professionals highlights a potential limitation for medical professionals seeking customizable features for surgical planning. In future work, considerations may include incorporating user-friendly interfaces with adjustable parameters to cater to professional needs effectively. Moreover, while the model primarily supports facial prediction for Asian facial types, efforts to

expand its applicability to other ethnic groups through further research and validation are essential. Finally, while the subjective evaluation provided insights into the perceptual differences between MLG faces and RSO, it is crucial to recognize the limitations of relying solely on subjective assessments. Subjective measures, by nature, may introduce bias and subjective interpretation, potentially affecting the overall validity and reliability of the study findings. To mitigate these limitations, future studies can integrate quantitative clinical metrics and objective measures to complement subjective assessments, enhancing the study outcomes' validity and reliability. Incorporating quantitative measures such as cephalometric analyses, facial landmark tracking, and patient-reported outcomes will facilitate a more comprehensive assessment and validation of machine learning-generated facial predictions in the context of orthognathic surgery.

6. Conclusions

In this paper, we introduce a novel method for predicting facial appearance following orthognathic treatment using only multi view images. During the training phase, our approach utilizes customized mouth-convexity and asymmetry losses, combined with latent code, geometric losses, and data augmentation, to enhance robustness and outperform existing methods in terms of accuracy.

CRediT authorship contribution statement

Huijun Han: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Congyi Zhang:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Lifeng Zhu:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Pradeep Singh:** Writing – review & editing, Validation, Investigation, Data curation. **Richard Tai-Chiu Hsung:** Supervision, Investigation, Formal analysis. **Yiu Yan Leung:** Supervision, Investigation. **Taku Komura:** Funding acquisition. **Wenping Wang:** Funding acquisition. **Min Gu:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Funding acquisition, Conceptualization.

Ethics statement

This study was conducted with the approval of the Institutional Review Board (IRB) under the joint agreement of the University of Hong Kong and the Hospital Authority Hong Kong West Cluster. The IRB reference number for this study is UW 21-140, with approval granted on February 17, 2021. All procedures adhered to the Declaration of Helsinki and local regulatory requirements, ensuring ethical conduct throughout the research process.

Participants provided informed consent, and their confidentiality and privacy were safeguarded. The study followed all necessary ethical guidelines to minimize risks and ensure the well-being of all individuals involved.

Declaration of competing interest

The authors declare that they have no financial or personal relationships that could have influenced the work presented in this paper. This work is supported by the GRF grants (17107321) from the Research Grants Council (RGC) of Hong Kong. There are no other conflicts of interest related to the study design, data collection, analysis, or publication process.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cmpb.2025.108781.

References

[1] S.H. Jeong, J.P. Yun, H.-G. Yeom, H.J. Lim, J. Lee, B.C. Kim, Deep learning based discrimination of soft tissue profiles requiring orthognathic surgery by facial photographs, Sci. Rep. 10 (1) (2020) 16235.

[2] J.-H. Yoo, H.-G. Yeom, W. Shin, J.P. Yun, J.H. Lee, S.H. Jeong, H.J. Lim, J. Lee, B.C. Kim, Deep learning based prediction of extraction difficulty for mandibular third molars, Sci. Rep. 11 (1) (2021) 1954.

[3] E.D. Rekow, Digital dentistry: The new state of the art—Is it disruptive or destructive? Dent. Mater. 36 (1) (2020) 9–24.

[4] P.G. Knoops, A. Papaioannou, A. Borghi, R.W. Breakey, A.T. Wilson, O. Jeelani, S. Zafeiriou, D. Steinbacher, B.L. Padwa, D.J. Dunaway, et al., A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery, Sci. Rep. 9 (1) (2019) 1–12.

[5] C. Tanikawa, T. Yamashiro, Development of novel artificial intelligence systems to predict facial morphology after orthognathic surgery and orthodontic treatment in Japanese patients, Sci. Rep. 11 (1) (2021) 1–11.

[6] R. Ter Horst, H. van Weert, T. Loonen, S. Bergé, S. Vinayahalingam, F. Baan, T. Maal, G. de Jong, T. Xi, Three-dimensional virtual planning in mandibular advancement surgery: Soft tissue prediction based on deep learning, J. Cranio-Maxillofacial Surg. 49 (9) (2021) 775–782.

[7] N. Chaiprasittikul, B. Thanathornwong, S. Pornprasertsuk-Damrongsri, S. Raocharernporn, S. Maponthong, S. Manopatanakul, Application of a multi-layer perceptron in preoperative screening for orthognathic surgery, Heal. Inform. Res. 29 (1) (2023) 16–22.

[8] J.N. Saeed, A.M. Abdulazeez, D.A. Ibrahim, Automatic facial aesthetic prediction based on deep learning with loss ensembles, Appl. Sci. 13 (17) (2023) 9728.

[9] J.-A. Park, J.-H. Moon, J.-M. Lee, S.J. Cho, B.-M. Seo, R.E. Donatelli, S.-J. Lee, Does artificial intelligence predict orthognathic surgical outcomes better than conventional linear regression methods? Angle Orthod. (2024).

[10] I.-H. Kim, J.-S. Kim, J. Jeong, J.-W. Park, K. Park, J.-H. Cho, M. Hong, K.-H. Kang, M. Kim, S.-J. Kim, Y.-J. Kim, S.-J. Sung, Y.H. Kim, S.-H. Lim, S.-H. Baek, N. Kim, Orthognathic surgical planning using graph CNN with dual embedding module: External validations with multi-hospital datasets, Comput. Methods Programs Biomed. 242 (2023) 107853, http://dx.doi.org/10.1016/j.cmpb.2023.107853.

[11] Y. Park, J. Choi, Y. Kim, S. Choi, J. Lee, K. Kim, C. Chung, Deep learning–based prediction of the 3D postorthodontic facial changes, J. Dent. Res. 101 (11) (2022) 1372–1379, PMID: 35774018.

[12] D. Laurinavičius, R. Maskeliūnas, R. Damaševičius, Improvement of facial beauty prediction using artificial human faces generated by generative adversarial network, Cogn. Comput. 15 (3) (2023) 998–1015.

[13] M. Cheng, X. Zhang, J. Wang, Y. Yang, M. Li, H. Zhao, J. Huang, C. Zhang, D. Qian, H. Yu, Prediction of orthognathic surgery plan from 3D cephalometric analysis via deep learning, BMC Oral. Health 23 (1) (2023) 161.

[14] Q. Ma, E. Kobayashi, B. Fan, K. Hara, K. Nakagawa, K. Masamune, I. Sakuma, H. Suenaga, Machine-learning-based approach for predicting postoperative skeletal changes for orthognathic surgical planning, Int. J. Med. Robot. Comput. Assist. Surg. 18 (3) (2022) e2379.

[15] H. Sankar, R. Alagarsamy, B. Lal, S.S. Rana, A. Roychoudhury, A. Agrawal, S. Wankhar, Role of artificial intelligence in treatment planning and outcome prediction of jaw corrective surgeries by using 3-D imaging- A systematic review, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. (2024) http://dx.doi.org/10.1016/j.oooo.2024.09.010, URL https://www.sciencedirect.com/science/article/pii/S2212440324005078.

[16] L. Ma, D. Kim, C. Lian, D. Xiao, T. Kuang, Q. Liu, Y. Lang, H.H. Deng, J. Gateno, Y. Wu, et al., Deep simulation of facial appearance changes following craniomaxillofacial bony movements in orthognathic surgical planning, in: MICCAI, Springer, 2021, pp. 459–468.

[17] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, D. Dunaway, A 3d morphable model learnt from 10,000 faces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5543–5552.

[18] J. Ling, Z. Wang, M. Lu, Q. Wang, C. Qian, F. Xu, Structure-aware editable morphable model for 3D facial detail animation and manipulation, in: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III, Springer, 2022, pp. 249–267.

[19] Z. Qiu, Y. Li, D. He, Q. Zhang, L. Zhang, Y. Zhang, J. Wang, L. Xu, X. Wang, Y. Zhang, J. Yu, SCULPTOR: Skeleton-consistent face creation using a learned parametric generator, ACM Trans. Graph. 41 (6) (2022).

[20] T. Li, T. Bolkart, M.J. Black, H. Li, J. Romero, Learning a model of facial shape and expression from 4D scans, in: ACM Proc. SIGGRAPH Asia, vol. 36, 2017, pp. 194:1–194:17.

[21] W. Zheng, J. Zhao, X. Liu, Y. Pan, Z. Gan, H. Han, N. Liu, FLAME-based multi-view 3D face reconstruction, in: Advances in Computer Graphics: 40th Computer Graphics International Conference, CGI 2023, Shanghai, China, August 28 – September 1, 2023, Proceedings, Part IV, Springer-Verlag, Berlin, Heidelberg, 2023, pp. 327–339, http://dx.doi.org/10.1007/978-3-031-50078-7_26.

[22] Y. Liang, C. Zhang, J. Zhao, W. Wang, X. Li, Skull-to-face: Anatomy-guided 3D facial reconstruction and editing , IEEE Trans. Vis. Comput. Graphics (01) (2025) 1–13, http://dx.doi.org/10.1109/TVCG.2024.3515093, 5555, URL https://doi.ieeecomputersociety.org/10.1109/TVCG.2024.3515093.

[23] C.C. Steiner, The use of cephalometrics as an aid to planning and assessing orthodontic treatment: report of a case, Am. J. Orthod. 46 (10) (1960) 721–735.

[24] A. Dobai, Z. Markella, T. Vízkelety, C. Fouquet, A. Rosta, J. Barabás, Landmark-based midsagittal plane analysis in patients with facial symmetry and asymmetry based on CBCT analysis tomography, J. Orofac. Orthopedics/ Fortschritte der Kieferorthopädie 79 (6) (2018) 371–379.

[25] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning (still) requires rethinking generalization, Commun. ACM 64 (3) (2021) 107–115.

[26] Z. Shan, R.T.-C. Hsung, C. Zhang, J. Ji, W.S. Choi, W. Wang, Y. Yang, M. Gu, B.S. Khambay, Anthropometric accuracy of three-dimensional average faces compared to conventional facial measurements, Sci. Rep. 11 (1) (2021) 12254.

[27] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: Bilateral segmentation network for real-time semantic segmentation, in: ECCV, 2018, pp. 325–341.

[28] X. Dong, S.-I. Yu, X. Weng, S.-E. Wei, Y. Yang, Y. Sheikh, Supervision-by-Registration: An unsupervised approach to improve the precision of facial landmark detectors, in: CVPR, 2018, pp. 360–368.

[29] M.-P. Dubuisson, A. Jain, A modified hausdorff distance for object matching, in: Proceedings of 12th International Conference on Pattern Recognition, vol. 1, 1994, pp. 566–568, http://dx.doi.org/10.1109/ICPR.1994.576361.

[30] Y. Yang, C. Feng, Y. Shen, D. Tian, FoldingNet: Point cloud auto-encoder via deep grid deformation, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 206–215, http://dx.doi.org/10.1109/CVPR.2018.00029.

[31] T. Starch-Jensen, F. Hernández-Alfaro, Ö. Kesmez, R. Gorgis, A. Valls-Ontañón, Accuracy of orthognathic surgical planning using three-dimensional virtual techniques compared with conventional two-dimensional techniques: a systematic review, J. Oral Maxillofacial Res. 14 (1) (2023).