Electronic Letters on Computer Vision and Image Analysis 25(3):1-22, 2026

Landmark-Guided Geometric Warping and GAN Refinement for Controllable and Explainable Cephalometric Profile Transformation

Gadha Lekshmi P[1] , Govind Arun[2], Ajo Babu George[3] and Arif Ahmed Sekh[4]

[1] *IGNOU: Indira Gandhi National Open University*
[2] *College of Computer, Mathematical, and Natural Sciences, University of Maryland College Park, USA*
[3] *Dicemed, Cuttak, India*
[4] *Senior Researcher, UiT The Arctic University of Norway, Norway*

Received 15th of November 2025; accepted 15th of July 2026

Abstract

Orthodontic treatment planning requires visualizing skeletal profile modifications, with CONVEX and CONCAVE morphologies requiring geometric corrections toward balanced configurations. However, existing GAN-based latent editing methods achieve insufficient skeletal control for clinically significant Pogonion displacement. To address this limitation, a hybrid geometric–generative framework is introduced that decouples deterministic skeletal modification from photorealistic synthesis. Radial basis function warping with adaptive amplification (1.8–$4.5\times$) produces controlled mandibular advancement or setback. StyleGAN2 refinement with $50\times$ landmark weighting then restores anatomical realism without compromising structural accuracy. Multi-task regularization combining diversity loss with landmark-consistency constraints mitigates dataset imbalance and enables balanced generation across profile classes (approximately 31% each), countering the predominance of CONVEX cases (63%). The system achieves 5–15 pixel Pogonion displacement with $<$3-pixel landmark drift (mean: 0.9 mm), demonstrating 7–10$\times$ greater skeletal control than latent editing while maintaining anatomical coherence. Interactive profile morphing animations provide explainable visualization of transformation trajectories through transparent, frame-by-frame display, enabling clinically interpretable exploration of treatment outcomes across the full CONVEX-to-CONCAVE spectrum. Integrated Vision Transformer classifiers and regressors automate the Visual Treatment Objective workflow, reaching 89% profile-classification accuracy and 95.6% landmark-regression accuracy within a 2-pixel tolerance. Ablation analysis reveals that standard latent editing achieves only 1.4 pixels of Pogonion movement, confirming that skeletal landmarks encode patient identity rather than transformable attributes in StyleGAN's learned manifold. The framework is designed as a pre-clinical foundational research stage visualization aid for patient communication and treatment outcome exploration, intended to support informed consent discussions in preventive, interceptive, and corrective orthodontic and orthognathic treatment planning.

Key Words: Computer Vision, Image Analysis, Latent Space Editing, Cephalometric Landmarking, Explainability, Facial Geometry Transformation, Thin-Plate Spline Warping, Vision Transformer

Correspondence to: gadhalekshmip@gmail.com

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2469
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

1 Introduction

Assessment of cephalometric landmarks has long guided diagnosis and surgical planning in orthodontics and maxillofacial orthopedics, linking cast-based simulations with anticipated skeletal and soft-tissue changes on lateral cephalograms [2]. Large cohort analyses of surgical and orthodontic treatment have further characterized patterns of correction across the mandible and bimaxillary procedures and quantified the likelihood of achieving target occlusal and soft-tissue ranges in severe skeletal discrepancies [3].

Data-driven prediction has recently advanced beyond manual tracing and rule-based forecasting. Learning from three-dimensional cephalometric analyses, transformer-based regression has been used to predict landmark-driven reposition vectors with clinically small errors and interpretable feature importance [4]. In parallel, diffusion-based generative modeling has been used to synthesize postoperative lateral cephalograms from preoperative inputs by combining predicted landmarks with profile line segmentations to improve planning fidelity and communication [5]. While these advances improve prediction accuracy, they do not address the visualization of treatment outcomes through controlled geometric transformation.

Aesthetic treatment objectives vary systematically across profile morphologies, with mandibular position, incisor inclination, nasolabial angle, and lip posture differing significantly between straight, concave, and convex profiles. These variations motivate profile-specific correction targets rather than universal norms [1]. However, existing methods face opposing limitations: classical landmark-based warping (e.g., thin-plate splines) provides deterministic geometric control but produces unrealistic texture distortions and compromises anatomical coherence, whereas generative adversarial networks (GANs) preserve photorealism but lack precise steerability at diagnostically critical anatomical points.

This limitation is particularly significant for controlling Pogonion horizontal position (chin) relative to the Nasion-Point A (N-A) reference line, the primary determinant of profile classification (Figure 1). Latent space editing in StyleGAN typically achieves only 0.6–1 px Pogonion displacement under stable editing conditions, insufficient for convex-to-straight correction requiring 5–10 pixels, while over-amplification causes identity collapse or unrealistic distortions. The fundamental limitation is architectural: StyleGAN is trained to model the distribution of existing facial morphologies, not the transformation trajectories between profile classes, as distinct individuals rather than as pre- and post-treatment states of the same patient. The proposed framework is therefore designed as a pre-clinical visualization aid for patient communication and treatment outcome exploration, and is not intended as a substitute for clinical surgical planning or biomechanical treatment simulation.

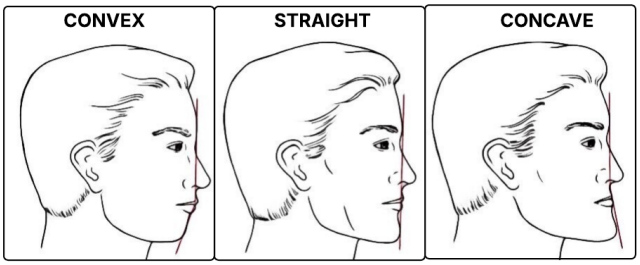


Figure 1: Facial profile classification based on Facial Angle of Convexity defined by chin position relative to the Nasion-Point A reference line. CONVEX: receding chin (Skeletal Class-II); STRAIGHT: orthognathic (Skeletal Class-I); CONCAVE: protruding chin (Skeletal Class-III).

This is addressed through a hybrid geometric–generative framework that separates transformation into two stages. First, deterministic radial basis function (RBF) warping enforces target Pogonion displacement with mathematical precision. Second, a StyleGAN2-based refinement step restores photorealism while preserving geometric accuracy through heavily weighted landmark constraints, maintaining geometric fidelity via a $50\times$ penalty on N, A, and Pog consistency. This design parallels the five-phase Visual Treatment Objective (VTO) [26] workflow, replacing subjective manual steps with quantified computational modules. Three core innova-

tions define the proposed approach:

- **Hybrid geometric–generative framework:** Deterministic RBF warping coupled with StyleGAN2 refinement achieves controlled 5–15 px skeletal adjustments while maintaining photorealism, exceeding the 0.6–1 px displacement attainable through latent editing alone.

- **Geometry-preserving optimization:** A multi-task regularization scheme integrates landmark, diversity, and classification losses, ensuring balanced training and stable transformations. Landmark constraints are explicitly enforced through high-magnitude weighting ($50\times$ overall, $3\times$ Pogonion emphasis), maintaining under 3 px drift for 15 px transformations.

- **End-to-end clinical pipeline:** The framework unifies ViT-based classification, landmark regression, geometric warping, and StyleGAN2 refinement into an interpretable, profile-specific visualization tool for patient communication and treatment outcome exploration.

2 Related Works

2.1 Cephalometric Landmark Detection and Classification

Automated cephalometric analysis has transitioned from manual tracing to fully learned approaches, with systematic reviews [6] reporting mean errors near 2 mm and approximately 80% of landmarks within a 2 mm clinical tolerance across nineteen studies. However, substantial risk of bias and limited generalizability remain. Broader technical surveys such as Juneja [7] discuss knowledge-based, machine learning, and deep learning-based techniques, outlining common imaging modalities, preprocessing strategies, and persistent challenges including noise, anatomical overlap, and limited dataset size. Methodological advances target robustness through cascaded refinement [9] and anatomical context learning [8]. However, most systems focus exclusively on detection accuracy rather than enabling downstream editing applications, which require both localization precision and transformation control. Our work extends this lineage by integrating landmark regression with profile classification (convex/concave/straight), explicitly linking anatomical localization to transformation objectives.

2.2 GAN-Based Facial Editing and Latent Space Manipulation

Recent developments in GAN-based face editing demonstrate substantial progress in controllable manipulation of facial attributes. The latent-search framework in [10] introduced a two-stage design that first aligns a latent code with an input image and then applies attribute-specific shifts within the latent space, allowing high-resolution modifications while preserving global facial structure. Unsupervised discovery methods [13, 12] enable semantic manipulations without attribute labels, reducing entanglement between correlated features. Extensions to diffusion models [14] reveal semantic structure in Jacobian mappings between noise states and U-Net representations, yielding globally consistent edits across timesteps.

However, comprehensive surveys [11] highlight persistent challenges like maintaining identity consistency during large edits, disentangling correlated attributes, and ensuring stable behavior across demographic variation.

Critically, these methods excel at perceptual attribute manipulation (age, expression, lighting) but are fundamentally limited for large geometric transformations. When skeletal landmarks must be displaced by clinically significant distances (5–15 pixels for Pogonion correction), latent editing either produces imperceptible changes or collapses identity through over-amplification. As established in Section 1, this architectural limitation, that generative models are optimized for sampling plausible faces rather than enforcing geometric constraints, motivates our hybrid approach separating deterministic warping from learned refinement.

2.3 Geometric Priors in Facial Manipulation

Recent work in facial manipulation detection has highlighted the importance of geometric priors for distinguishing real from synthetic faces [15]. These geometric constraints, particularly landmark consistency and anatomical ordering, inform our refinement module's regularization strategy. The mask-guided manipulation strategy in [16] demonstrated that explicit geometric control signals enable more precise structural edits while preserving global fidelity. Our framework builds on these insights by enforcing hard geometric constraints during both warping and refinement stages.

3 Dataset

3.1 Data Sources and Composition

A unified dataset of 2,061 lateral cephalometric radiographs was assembled from three independent collections: one private clinical dataset and two publicly available datasets. The private proprietary dataset consists of 682 clinical radiographs obtained from routine maxillofacial diagnostic practice. The images were obtained using PaX-i SC Scan Type Ceph system, manufactured by VATECH Company Limited, Korea that uses a Complementary Metal-Oxide-Semiconductor (CMOS) sensor with a 100 μm pixel size. These images represent a secondary analysis of data originally collected under a protocol approved by the Institutional Ethics Committee (IEC) of SCB Dental College & Hospital on 17.09.2019. To ensure participant privacy and compliance for this subsequent work, all original patient records and identifying information were irreversibly dissociated from the image data. Prior to inclusion in this dataset, all direct identifiers (names, IDs, dates) were removed, and a new, arbitrary project-specific identifier was assigned to each image. The irreversible nature of this anonymization procedure meets the criteria for using non-identifiable data for secondary analysis. Two additional public datasets were incorporated: the Aariz dataset [17], a benchmark for automatic cephalometric landmark detection, contributes 1,000 radiographs, and the Kaggle [18] Cephalometric Dataset provides an additional 379 images.



Figure 2: Profile class distribution across three data sources.

The dataset exhibits severe class imbalance (Figure 2): 1,298 CONVEX (63.0%), 423 CONCAVE (20.5%), and 340 STRAIGHT (16.5%). This distribution reflects clinical prevalence, as convex profiles dominate orthodontic and surgical case loads, but introduces challenges for balanced generative modeling, addressed

through multi-task regularization during training. A notable limitation of the assembled dataset is its lack of demographic diversity: the collection predominantly represents a single clinical cohort without systematic representation across ethnic backgrounds, age groups, or skeletal variation beyond the three profile classes. Furthermore, the dataset contains lateral cephalograms from different individuals at single time points rather than longitudinal pre- and post-treatment records of the same patient. This absence of paired intra-patient data precludes direct supervised learning of treatment transformation trajectories and limits assessment of prediction accuracy against true post-treatment outcomes. These constraints on generalizability are addressed through multi-task regularization but motivate future multicentric data collection efforts. The data is partitioned into training (70%, 1,443 images), validation (20%, 412 images), and test (10%, 206 images) subsets using stratified sampling to preserve class ratios.

3.2 Annotation Protocol

All images receive two types of annotations. Profile classification is based on the horizontal projection of Pogonion relative to the Nasion–Point A (N–A) reference line, which is a standard anteroposterior skeletal assessment in cephalometric analysis. Profile classification labels (CONVEX, CONCAVE, STRAIGHT) are assigned according to this relationship (Figure 1). Profiles in which Pogonion is positioned posterior to the N–A line are classified as CONVEX. Profiles in which Pogonion is positioned anterior to the line are classified as CONCAVE. Profiles in which Pogonion is aligned with the line are classified as STRAIGHT. These assignments follow established cephalometric classification criteria.

Cephalometric landmark coordinates are provided for three diagnostically critical points (Figure 3). Nasion (N) is the most anterior point of the frontonasal suture at the nasal bridge. Point A (A) is the deepest midline concavity on the maxilla between the anterior nasal spine and supradentale. Pogonion (Pog) is the most anterior point of the bony chin. All coordinate values are normalized to $[0, 1]$ relative to the image dimensions.

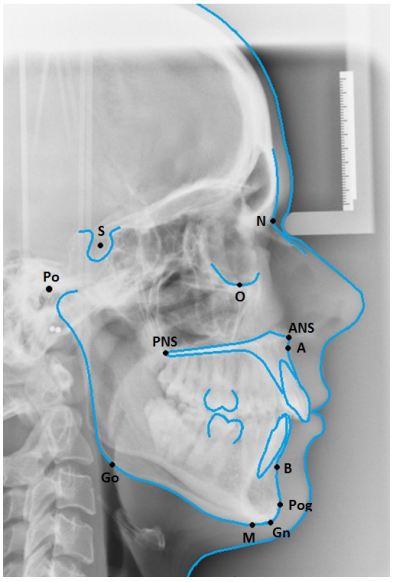


Figure 3: Cephalometric landmark annotation on a lateral skull radiograph. The three diagnostically critical points used in this study are Nasion (N, nasal bridge), Point A (A, maxillary apex), and Pogonion (Pog, chin) [19].

3.3 Profile Classification Rationale

The three-class taxonomy reflects clinical orthodontic practice, wherein treatment goals systematically vary by profile type [1]. CONVEX profiles, characterized by mandibular retrognathism or maxillary prognathism, con-

stitute the majority of orthodontic cases and require anterior repositioning of Pogonion toward the N–A line. STRAIGHT profiles represent balanced anteroposterior skeletal relationship. CONCAVE profiles, associated with Class III skeletal malocclusion, exhibit mandibular prognathism. The proposed framework is designed to enable transformations toward STRAIGHT profiles from either extreme, necessitating profile-specific displacement vectors and balanced generative capabilities.

4 Methodology

4.1 Overview of the hybrid pipeline

Our hybrid geometric-generative framework decouples profile transformation into three sequential stages (Figure 4): landmark-guided analysis, deterministic geometric warping, and GAN-based photorealistic refinement. This architecture addresses the fundamental limitation that StyleGAN latent editing achieves insufficient displacement for skeletal correction by separating geometric transformation from texture synthesis.

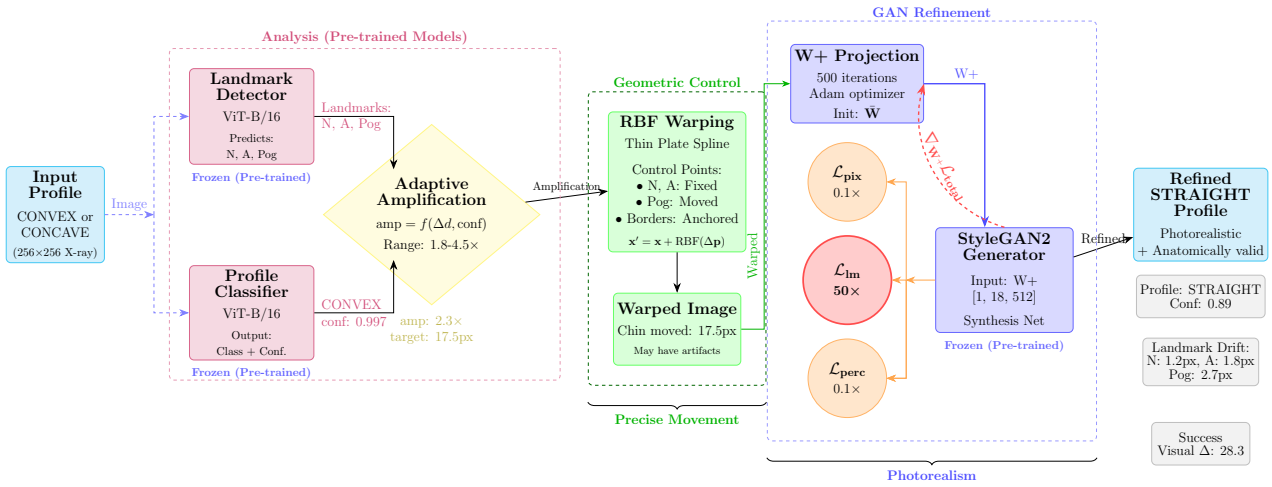


Figure 4: Hybrid geometric-generative framework for cephalometric profile transformation.

Stage 1: Landmark-Guided Analysis. A frozen Vision Transformer classifier assigns profile labels (CONVEX, CONCAVE, STRAIGHT) and confidence scores. Simultaneously, a frozen ViT landmark regressor predicts normalized coordinates for Nasion (N), Point A (A), and Pogonion (Pog). These outputs feed an adaptive amplification module that computes profile-specific displacement magnitude based on the distance from Pog to the N-A reference line and classifier confidence.

Stage 2: Geometric Warping. Radial basis function interpolation with thin-plate spline kernels computes smooth deformations mapping source landmarks to target positions. Control points include the three cephalometric landmarks (N and A fixed, Pog displaced) plus dense border anchors. The warping module provides deterministic geometric control but introduces texture distortions requiring refinement.

Stage 3: GAN Refinement. A pre-trained StyleGAN2 generator projects the warped image into $\mathcal{W}^+$ latent space via iterative optimization. The objective combines pixel-level reconstruction, perceptual similarity, and landmark-weighted consistency losses with high emphasis on geometric preservation. This formulation prevents geometric collapse during photorealistic synthesis while restoring radiographic texture consistent with the learned distribution of cephalometric images. The refinement stage exploits StyleGAN's learned manifold of valid cephalometric appearances without retraining, enabling artifact removal while preserving the precise geometric transformation from Stage 2.

All modules except the refinement optimizer are deterministic, ensuring reproducible transformations for clinical documentation.

4.2 Vision Transformer Modules for Profile Analysis

Both profile classification and landmark localization employ Vision Transformer (ViT) architectures [20] fine-tuned from ImageNet pre-trained base models. ViT partitions input images into fixed-size non-overlapping patches, embeds each patch via linear projection, and processes the sequence through Transformer encoder layers with multi-head self-attention. This architecture adapts to cephalometric analysis through task-specific fine-tuning on lateral cephalograms.

The profile classifier appends a learnable classification token to the patch sequence and projects its final representation to three logits corresponding to CONVEX, CONCAVE, and STRAIGHT profiles. During inference, both the predicted class and softmax confidence score jointly determine the amplification factor applied during geometric warping. High-confidence predictions permit stronger displacement, as classification certainty reduces the risk of transforming across incorrect profile boundaries. The landmark regressor uses an identical backbone but projects the classification token representation to six continuous outputs representing normalized coordinates for Nasion (N), Point A (A), and Pogonion (Pog). These predictions define control points for geometric warping and serve as hard constraints during GAN refinement.

Both models remain frozen after initial fine-tuning and serve as auxiliary modules during StyleGAN2 training to provide gradient-based guidance without degrading learned cephalometric representations. Freezing enables the use of pre-trained discriminative knowledge to regularize generative training while maintaining computational efficiency during inference.

4.3 StyleGAN2 Training with Multi-Task Regularization

StyleGAN2 [21] is trained to generate diverse cephalometric profiles across all three classes despite severe class imbalance in the dataset (63% CONVEX, Figure 5). The generator architecture follows the standard design: an 8-layer mapping network transforms Gaussian latent codes $z \sim \mathcal{N}(0, I_{512})$ into the intermediate style space $\mathcal{W}$, which is extended to $\mathcal{W}^+ \in \mathbb{R}^{18\times512}$ by allowing layer-specific style vectors. The synthesis network progressively produces 256×256 images from a learned constant input through convolutional blocks modulated by adaptive instance normalization (AdaIN), which applies learned affine transformations $(\text{scale}, \text{bias})$ derived from $\mathcal{W}^+$ to normalize feature maps at each resolution.

Beyond the standard adversarial objective, two auxiliary losses computed by frozen ViT modules explicitly regularize generation toward balanced profile distributions and anatomical plausibility. The total generator loss combines three terms:

$$\mathcal{L}_{\text{G}} = \mathcal{L}_{\text{adv}} + \lambda_{\text{sx}}\mathcal{L}_{\text{sx}} + \lambda_{\text{lm}}\mathcal{L}_{\text{lm}} \tag{1}$$

where $\lambda_{\text{sx}} = 0.5$ and $\lambda_{\text{lm}} = 2.0$ amplify auxiliary gradients to compete with adversarial loss magnitude.

StyleX Diversity Loss. The frozen ViT classifier predicts profile class probabilities $\mathbf{p}_i \in \mathbb{R}^3$ for each generated image $G(z_i)$ in a batch. Two penalties encourage both prediction confidence (low entropy) and batch-level diversity (uniform class distribution):

$$\mathcal{L}_{\text{sx}} = \underbrace{\frac{1}{B}\sum_{i=1}^{B} H(\mathbf{p}_i)}_{\text{confidence}} + \underbrace{D_{\text{KL}}\left(\bar{\mathbf{p}} \,\|\, \mathbf{u}\right)}_{\text{diversity}} \tag{2}$$

where $H(\mathbf{p}_i) = -\sum_c p_{i,c} \log p_{i,c}$ penalizes ambiguous classifications, with $c \in \{\text{CONVEX}, \text{STRAIGHT}, \text{CONCAVE}\}$. The batch-averaged distribution is $\bar{\mathbf{p}} = \frac{1}{B}\sum_i \mathbf{p}_i$, and $\mathbf{u} = [1/3, 1/3, 1/3]$ is the uniform target. This formulation is designed to discourage mode collapse to the dominant CONVEX class by explicitly rewarding generators that produce varied profiles within each training batch.

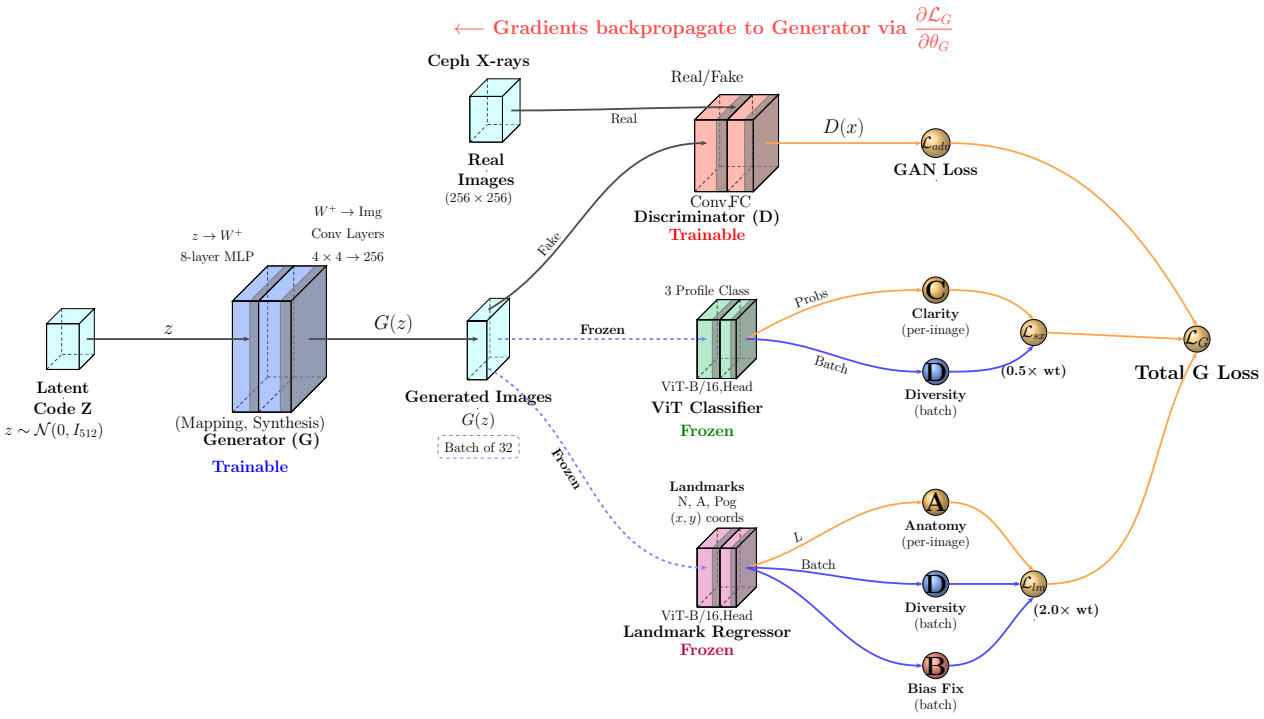


Figure 5: Multi-task StyleGAN2 training architecture with auxiliary losses for balanced profile generation.

Landmark Consistency Loss. The frozen landmark regressor predicts normalized (x, y) coordinates for the three landmarks $\mathbf{L}_i = (N, A, \text{Pog})$ in each generated image. We enforce anatomical constraints through four terms:

$$\mathcal{L}_{\text{lm}} = \mathcal{L}_{\text{vert}} + \mathcal{L}_{\text{dist}} + \mathcal{L}_{\text{div}} + \mathcal{L}_{\text{bias}} \tag{3}$$

The vertical ordering constraint $\mathcal{L}_{\text{vert}}$ penalizes inversions where N falls below A or A below Pog using ReLU-gated differences. The distance constraint $\mathcal{L}_{\text{dist}}$ regularizes $\|N - A\|$ and $\|A - \text{Pog}\|$ toward typical cephalometric proportions (0.25 and 0.22 in normalized coordinates) via mean squared error penalties.

The **diversity term** $\mathcal{L}_{\text{div}}$ penalizes low variance in Pogonion horizontal position across batch samples, preventing all generated faces from having identical chin projection. The bias correction term $\mathcal{L}_{\text{bias}}$ penalizes batch-mean Pogonion horizontal position when it deviates significantly from the N-A line, directly countering the 63% CONVEX training bias.

The discriminator follows standard StyleGAN2 design with R1 gradient penalty for training stability. This multi-task design enables balanced profile generation from imbalanced training data while maintaining anatomical validity, providing the diverse synthesis network required for subsequent $\mathcal{W}^+$ projection during inference-time refinement.

4.4 Geometric Warping Module

Geometric warping provides deterministic control over Pogonion displacement through smooth, landmark-constrained deformation. The approach employs radial basis function interpolation with thin-plate spline kernels [22, 23], a standard technique in medical image registration and morphometric analysis. Given detected landmarks (N, A, Pog) from the regression module and profile classification, a target Pogonion position Pog' is computed aligned to the N-A reference line, then RBF interpolation generates a spatially smooth warp field that repositions Pog while preserving surrounding anatomical structure.

Target Landmark Calculation. For CONVEX profiles requiring anterior chin movement, the perpendicular projection of Pog onto the N-A line is first computed, where landmarks are represented as 2D coordinate vectors (x, y):

$$\mathrm{Pog}_{\mathrm{proj}} = N + \frac{(\mathrm{Pog} - N) \cdot (A - N)}{\|A - N\|^2}(A - N) \tag{4}$$

The geometric displacement vector $\Delta = \mathrm{Pog}_{\mathrm{proj}} - \mathrm{Pog}$ defines the movement required to achieve a STRAIGHT profile. However, a known limitation of $\mathcal{W}^+$ latent projection methods is the tendency for optimization to pull generated images back toward the training distribution [25, 21]. To compensate for this pull-back effect, which empirically reduces effective Pogonion displacement during GAN refinement, adaptive amplification is applied:

$$\mathrm{Pog}' = \mathrm{Pog} + \alpha \cdot \Delta \tag{5}$$

where $\alpha \in [1.8, 4.5]$ is determined by a piecewise function of classifier confidence and displacement magnitude. High-confidence CONVEX predictions with small geometric distances receive stronger amplification (up to $4.5\times$) to overcome GAN manifold constraints, while low-confidence predictions with large distances use conservative factors ($1.8\times$) to prevent overshoot. For CONCAVE profiles, the displacement direction reverses (Pog moves posteriorly), with amplification factors in the range $[2.0\times, 4.0\times]$ empirically determined to account for asymmetric warping behavior in posterior versus anterior transformations.

Radial Basis Function Warping. A thin-plate spline deformation is constructed using RBF interpolation to smoothly propagate landmark displacement throughout the image. Control points include the three cephalometric landmarks plus dense border anchors:

$$\mathcal{C}_{\mathrm{src}} = \{N, A, \mathrm{Pog}\} \cup \mathcal{B}, \quad \mathcal{C}_{\mathrm{dst}} = \{N, A, \mathrm{Pog}'\} \cup \mathcal{B} \tag{6}$$

where $\mathcal{B}$ denotes a grid of boundary points along image borders that anchor edges to prevent global drift. Nasion and Point A remain fixed ($N' = N, A' = A$) to preserve upper facial structure, while only Pogonion undergoes prescribed displacement. The RBF interpolators for horizontal and vertical displacement fields are:

$$f_x(x, y) = \sum_i w_{i,x}\phi(\|\mathbf{r}_i - \mathbf{r}\|), \quad f_y(x, y) = \sum_i w_{i,y}\phi(\|\mathbf{r}_i - \mathbf{r}\|) \tag{7}$$

where $\phi(r) = r^2 \log(r)$ is the thin-plate spline basis function [24], chosen for its smoothness properties and minimal bending energy characteristics. The vectors $\mathbf{r}_i$ are control point coordinates, and weights $\{w_{i,x}, w_{i,y}\}$ are solved via least-squares to satisfy boundary constraints $f(\mathbf{r}_i) = \mathbf{r}'_i - \mathbf{r}_i$. The final warped image is obtained through bilinear sampling:

$$I_{\mathrm{warp}}(x, y) = I_{\mathrm{input}}(x + f_x(x, y), y + f_y(x, y)) \tag{8}$$

This geometric stage produces controlled Pogonion displacements with deterministic reproducibility, but introduces texture artifacts (stretched bone features, discontinuous soft-tissue boundaries) that require subsequent GAN refinement to restore anatomical realism.

4.5 GAN Refinement Module

The warped image achieves precise geometric correction but exhibits texture stretching, discontinuous bone boundaries, and unrealistic trabecular patterns. Photorealism is restored through $\mathcal{W}^+$ latent space projection [25], optimizing a latent code such that the pre-trained StyleGAN2 generator synthesizes an image matching the warped geometry while respecting the learned manifold of valid cephalometric appearances.

$\mathcal{W}^+$ Projection Framework. Given warped image I_{warp} and its detected landmarks $\mathbf{L}_{\mathrm{warp}} = (N, A, \mathrm{Pog})$, a latent code $\mathbf{w}^+ \in \mathbb{R}^{18\times512}$ is initialized from the mean latent vector $\bar{\mathbf{w}}$ computed over the training data and optimized via gradient descent to minimize:

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{pix}}\mathcal{L}_{\mathrm{pix}} + \lambda_{\mathrm{lm}}\mathcal{L}_{\mathrm{lm}} + \lambda_{\mathrm{perc}}\mathcal{L}_{\mathrm{perc}} \tag{9}$$

with penalty weights $\lambda_{\text{pix}} = 0.1$, $\lambda_{\text{lm}} = 50.0$, and $\lambda_{\text{perc}} = 0.1$. The dominant landmark weight ($\lambda_{\text{lm}} = 50.0$) is substantially larger than reconstruction losses, preventing the generator from reverting to nearby training samples and ensuring geometric transformation is preserved during refinement. The optimization performs 500 Adam steps with learning rate 0.01, balancing convergence speed against refinement quality.

Loss Function Design. The pixel reconstruction loss $\mathcal{L}_{\text{pix}} = \|G(\mathbf{w}^+) - I_{\text{warp}}\|_2^2$ provides pixel-level guidance but requires perceptual loss to capture structural similarity. The perceptual loss $\mathcal{L}_{\text{perc}} = \|G_{64}(\mathbf{w}^+) - I_{\text{warp},64}\|_2^2$ computed on downsampled 64×64 images captures global structure. Critically, the landmark consistency loss penalizes deviation of generated landmarks $\mathbf{L}_{\text{gen}} = \text{Regressor}(G(\mathbf{w}^+))$ from warped targets:

$$\mathcal{L}_{\text{lm}} = \frac{1}{5}\left(\|N_{\text{gen}} - N_{\text{warp}}\|_2^2 \cdot 256^2 + \|A_{\text{gen}} - A_{\text{warp}}\|_2^2 \cdot 256^2 + 3 \cdot \|\text{Pog}_{\text{gen}} - \text{Pog}_{\text{warp}}\|_2^2 \cdot 256^2\right) \quad (10)$$

where the 256^2 scaling converts normalized coordinates to pixel space, preventing gradient vanishing relative to pixel loss. The division by 5 normalizes across coordinate components (2 per landmark) plus the $3\times$ Pogonion emphasis term, which prioritizes chin stability as the primary transformation target. This formulation enables precise geometric control while allowing the generator to synthesize realistic bone texture and soft-tissue boundaries consistent with the learned distribution.

The refinement module exploits the pre-trained generator's learned prior over valid cephalometric structures without requiring retraining or dataset expansion. By formulating refinement as constrained optimization rather than latent space editing, photorealistic synthesis is achieved while enforcing hard geometric constraints incompatible with StyleGAN's original training objective.

5 Experiments and Results

All experiments were conducted on a single NVIDIA Tesla A100 GPU (32GB) using PyTorch 1.12.1. The dataset was split into 70% training (1,443 images), 20% validation (412 images), and 10% test (206 images) with stratified sampling.

Both ViT-based auxiliary models were fine-tuned from `google/vit-base-patch16-224-in21k` using identical hyperparameters: AdamW optimizer with learning rate 2×10^{-5}, batch size 16, and 30 epochs. Images were resized to 224×224 pixels with ImageNet normalization. The complete transformation pipeline requires approximately 15–20 seconds per image: 0.1 seconds for ViT inference, 2–3 seconds for RBF warping (CPU), and 12–15 seconds for GAN refinement (GPU optimization with 500 Adam steps).

5.1 Profile Classification and Landmark Detection

Table 1 summarizes the performance of both auxiliary models.

Table 1: Auxiliary model architectures and performance metrics.

Model	Architecture	Task	Test Performance
Classifier	ViT-Base/16	Profile classification	89.0% (3-class)
Landmark Regressor	ViT-Base/16	N, A, Pog detection	95.6% (<2px tolerance)

The landmark regressor demonstrated mean radial error of 1.8 pixels (0.7% of image diagonal), providing sub-pixel precision that enables deterministic control point specification for geometric warping and serves as hard constraints during GAN refinement.

5.2 StyleGAN2 Training

StyleGAN2 was trained for 20,000 iterations (batch size 32) on a single NVIDIA A100 GPU, requiring 8 hours and 4 minutes. The generator synthesizes 256×256 images through 18 style-modulated layers, while

the discriminator uses residual blocks with progressive downsampling. Both networks used Adam optimizers (learning rate 0.002) with the R1 gradient penalty ($\lambda_R = 10.0$) applied every 16 iterations. Beyond standard adversarial training, auxiliary losses computed by frozen ViT modules enforce profile diversity and anatomical validity. The total generator loss follows Equation (1), with penalty weights $\lambda_{\text{sx}} = 0.5$ and $\lambda_{\text{lm}} = 2.0$. The StyleX diversity loss ($\mathcal{L}_{\text{sx}}$) penalizes batch-level class imbalance via KL divergence from the uniform distribution, while the landmark consistency loss ($\mathcal{L}_{\text{lm}}$) enforces anatomical constraints

Table 2: StyleGAN2 training outcomes and profile distribution.

Component	**Configuration**	**Final Loss**
Generator	256×256, 18 layers	1.447
Discriminator	Residual blocks	1.058
StyleX Diversity ($\lambda_{\text{cls}} = 0.5$)	KL divergence	0.166
Landmark Consistency ($\lambda_{\text{lm}} = 2.0$)	Anatomical constraints	0.056
Profile Distribution (CONVEX / STRAIGHT / CONCAVE):		
Training Data	63% / 16% / 20%	
Generated (n=100)	31% / 31% / 31%	

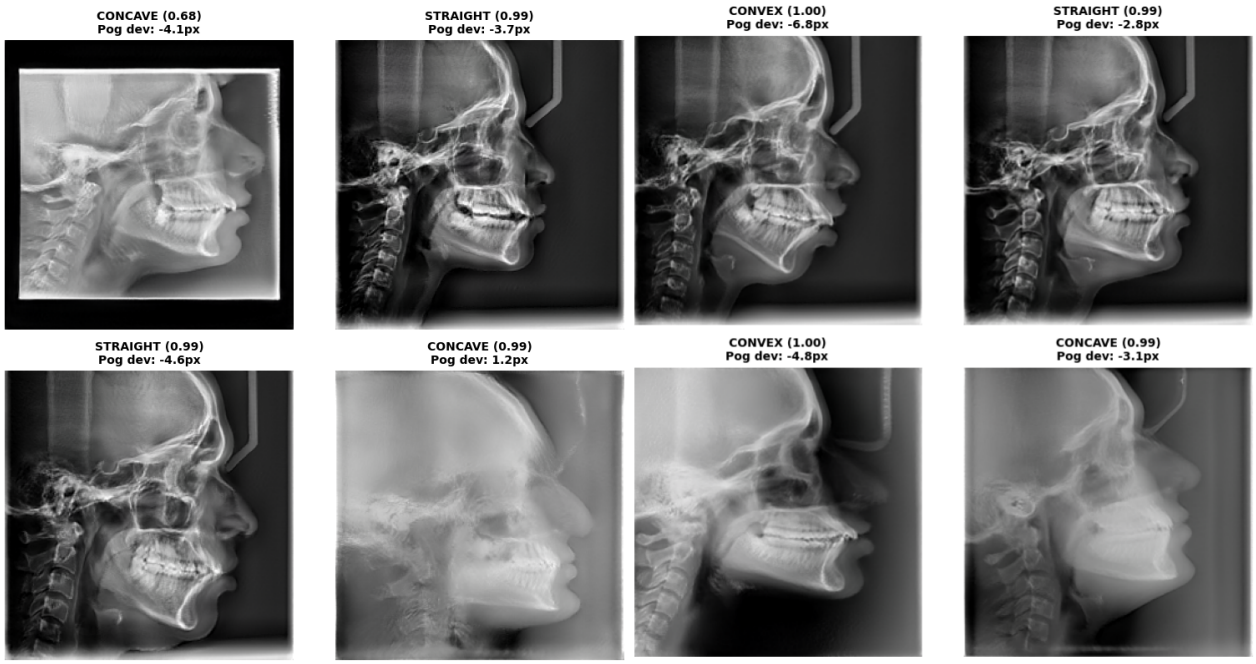


Figure 6: Representative generated samples demonstrating diverse profile coverage across all three classes.

The generative model produces diverse, anatomically consistent outputs (Figure 6). During optimization (Figure 7), the discriminator loss stabilized at 1.06, while the generator loss converged to 1.45, indicating stable training without mode collapse. The landmark consistency loss remained in the 0.05–0.06 range throughout, confirming that anatomical constraints were continuously enforced. The StyleX diversity loss declined steadily from 1.03 to 0.17, reflecting increasing balance in generated profile distributions.

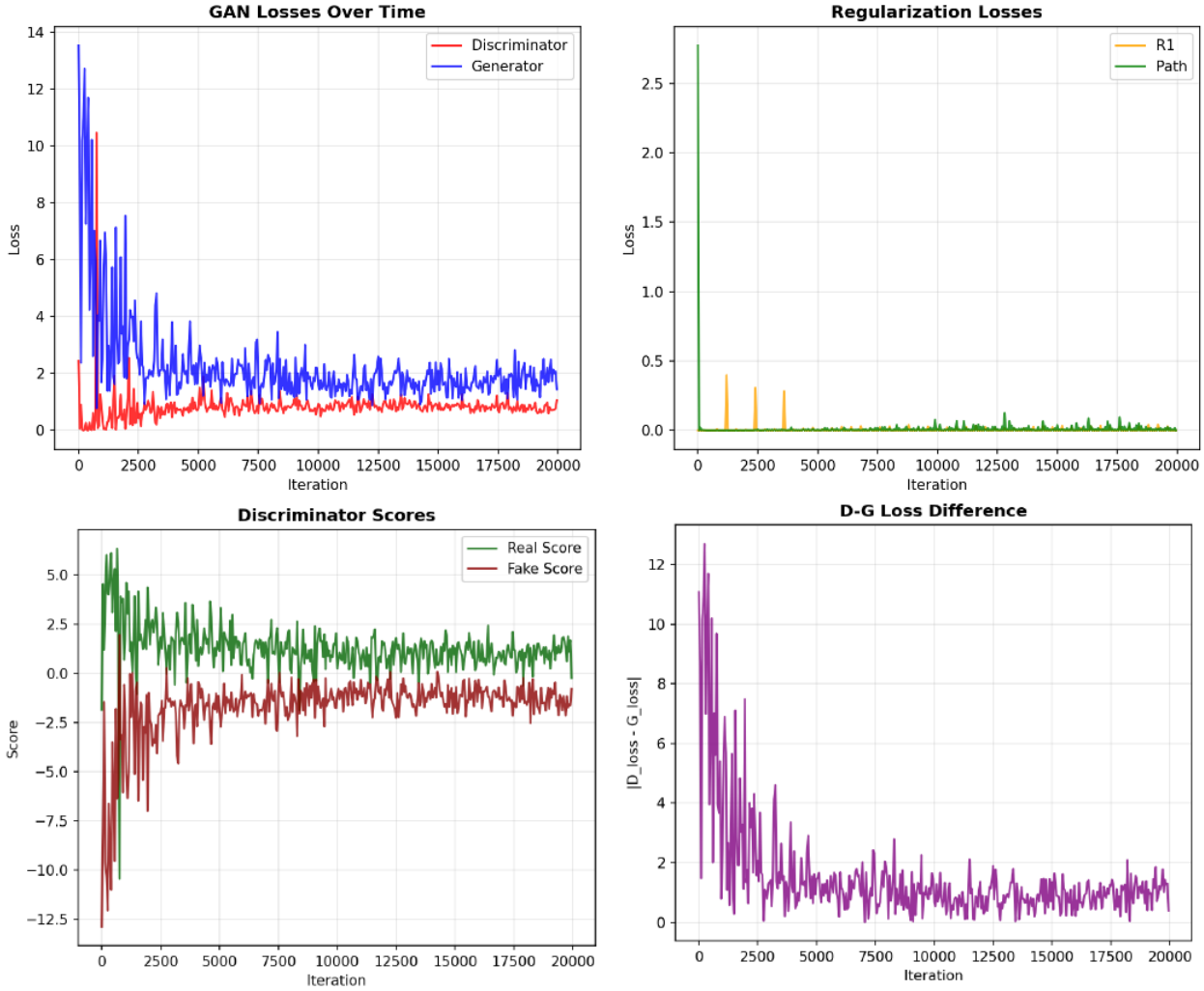


Figure 7: Training loss curves over 20,000 iterations.

5.3 Hybrid pipeline results

The complete transformation pipeline yields anatomically coherent and visually natural adjustments, as illustrated in Figure 8. In the case of a CONVEX profile (classifier confidence: 0.997), anterior repositioning of the chin is achieved through a controlled warping step, producing a 9.5 px Pogonion displacement. The red–green overlay makes this shift explicit, with the refinement stage subsequently reinstating realistic cortical boundaries and soft-tissue gradients. A parallel workflow is applied to CONCAVE morphology (classifier confidence: 0.943). Here, posterior displacement requires $2\times$ amplification to overcome warping resistance and achieve the target correction. The resulting shift ($\Delta x = -5.5$ px) is preserved in the refined output, which eliminates texture artifacts while maintaining the intended geometric change.

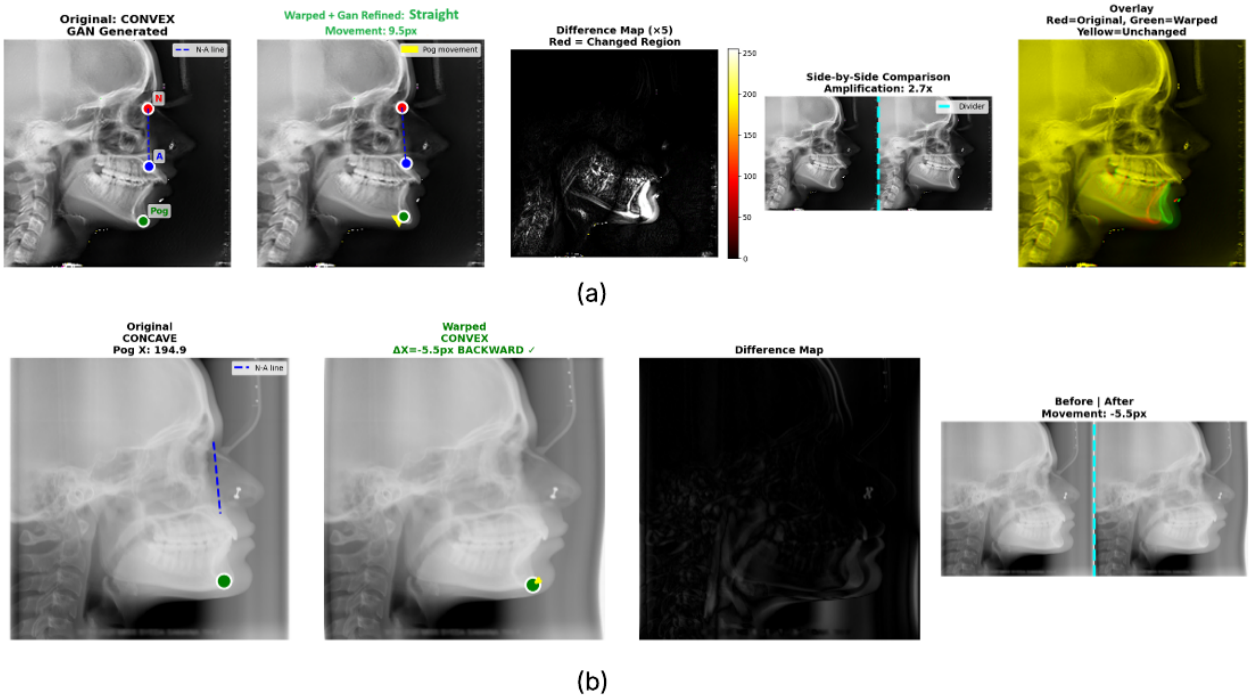


Figure 8: Profile transformation examples demonstrating CONVEX→STRAIGHT (top) and CONCAVE→STRAIGHT (bottom) corrections with geometric displacement visualization.

5.3.1 Importance of GAN Refinement

The refinement stage plays a central role in restoring photorealistic structure following deformation.

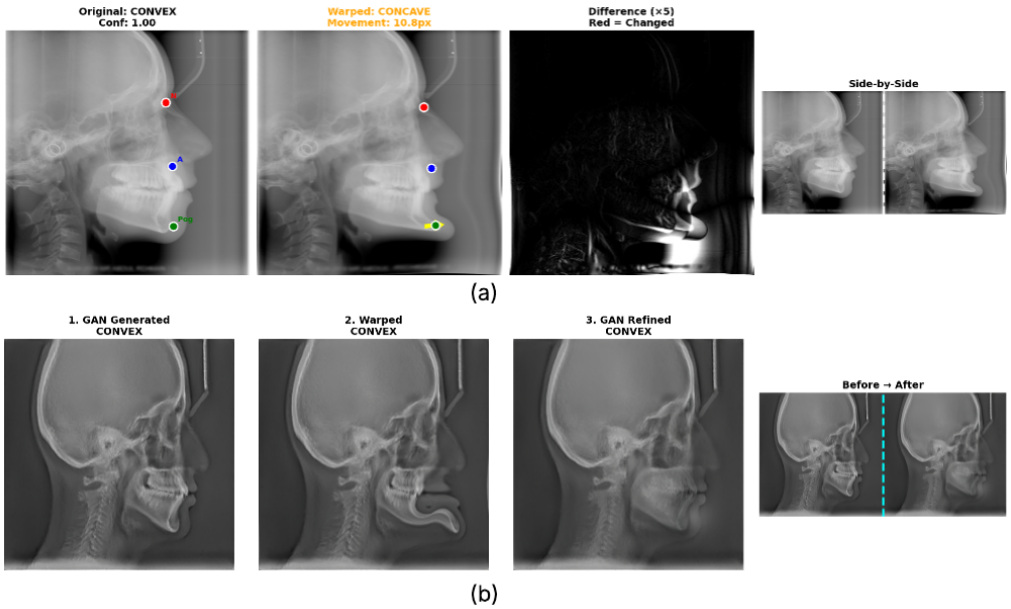


Figure 9: Comparison of warped output before and after GAN refinement demonstrating artifact removal

Prior to refinement (Figure 9(a)), warped images exhibit irregular trabecular texture, breaks along cortical boundaries, and soft-tissue patterns inconsistent with valid cephalometric appearance. After refinement (Figure 9(b)), these artifacts are resolved, yielding profiles that preserve the intended geometric adjustment while remaining anchored to the learned distribution of anatomically credible structures.

5.3.2 Landmark Preservation During Refinement

GAN projection without proper landmark constraints causes geometric collapse, where the generator reverts to the nearest training manifold sample and undoes the warping transformation. Table 3 demonstrates the effect of landmark loss weighting during $\mathcal{W}^+$ projection. Standard projection with $\lambda_{\text{lm}} = 2.0$ results in geometric collapse (5–15 px Pogonion drift), as the generator prioritizes photorealism over geometric fidelity. Increasing the weight to $\lambda_{\text{lm}} = 50.0$ while reducing the pixel loss to $\lambda_{\text{pix}} = 0.1$ ensures geometric preservation (< 3 px drift) while maintaining photorealistic output. The 256^2 coordinate scaling (converting normalized landmarks to pixel space) prevents gradient vanishing relative to pixel-level losses.

Table 3: GAN projection loss weight ablation study.

$\lambda_{\text{landmark}}$	λ_{pixel}	**Pogonion Drift (px)**	**Outcome**
2.0	1.0	>10	Geometric collapse
10.0	0.5	5–8	Partial preservation
50.0	**0.1**	**<3**	**Geometric fidelity**

5.3.3 Interactive Profile Morphing

A continuous spectrum of profile shapes can be explored through incremental geometric control parameters, providing explicit geometric explainability where each intermediate frame corresponds to a quantified skeletal displacement. Figure 10 presents a seven-frame sequence demonstrating bidirectional transformation across the complete profile range. The sequence begins at the CONVEX extreme (slider: -3, 15 px posterior Pogonion displacement), progresses through intermediate states (-2, -1) toward the STRAIGHT baseline (slider: 0), then continues to CONCAVE configurations ($+1$, $+2$, $+3$, 15 px anterior displacement). Each transition maintains anatomical integrity through RBF interpolation with fixed superior landmarks (N, A) and mobile Pogonion, ensuring that transformation trajectories are anatomically valid rather than arbitrary latent-space interpolations. The resulting trajectory offers clinically interpretable visualization of potential treatment outcomes across the entire profile range, with each intermediate frame remaining structurally coherent and directly mappable to skeletal geometry.

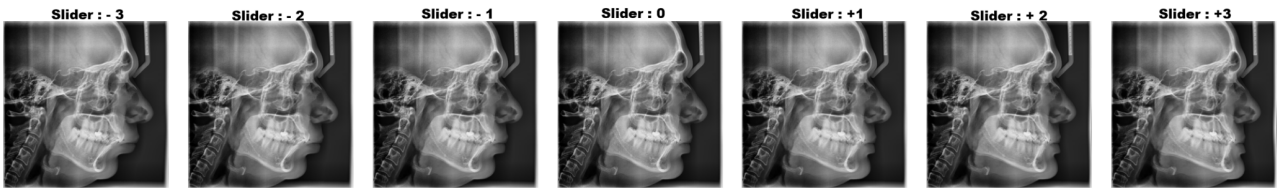


Figure 10: Seven-frame morphing sequence showing profile transformation from CONVEX (-3) through STRAIGHT (0) to CONCAVE ($+3$).

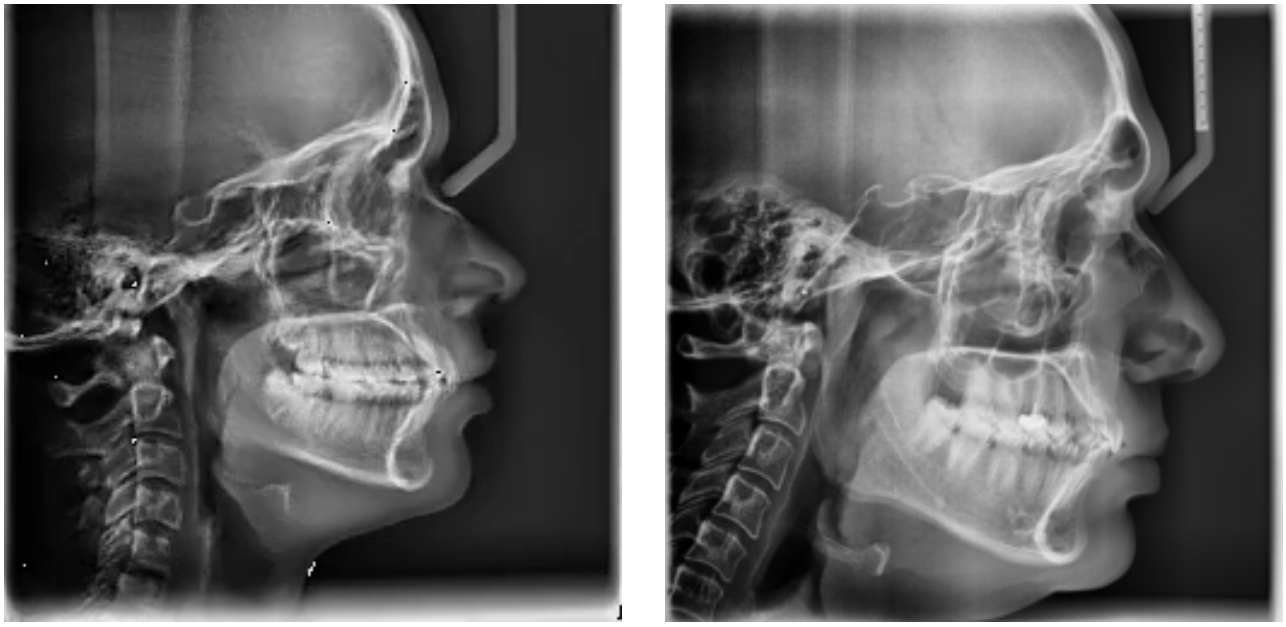


Figure 11: Animated cephalometric profile morphing demonstrating continuous transformation from STRAIGHT through CONVEX to CONCAVE. *Note: Interactive animation requires Adobe Acrobat Reader DC for playback.*

5.4 Latent Space Editing Ablation

To validate the necessity of the hybrid approach, three latent editing strategies were systematically evaluated. $\mathcal{W}$-space editing applies uniform style codes across all 18 synthesis layers, enabling global geometric changes but lacking fine control. $\mathcal{W}^+$-space extends this by allowing independent style vectors per layer ($\mathbf{w}^+ \in \mathbb{R}^{18\times512}$), with editing focused on mid-level layers (4–11) responsible for facial geometry, while preserving global structure (early layers) and texture (late layers).

StyleSpace (S-space) offers the highest level of control by exposing intermediate feature activations across all synthesis layers, comprising approximately 9,088 channels. Three complementary S-space strategies were examined: (1) single-channel search through gradient-based sensitivity analysis; (2) multi-channel bundles, where groups of 3–5 related channels are adjusted jointly to minimize cross-layer interference; and (3) gradient-based direction extraction identifying coherent manipulation directions:

$$\nabla_{\mathbf{s}} \left\| \mathrm{Pog}_{\mathrm{gen}} - \mathrm{Pog}_{\mathrm{target}} \right\| \tag{11}$$

Despite this comprehensive exploration, all S-space approaches converged to texture variations rather than skeletal shifts.

Table 4: Latent space editing effectiveness compared to hybrid approach.

Method	**Pogonion Movement (px)**	**Outcome**
$\mathcal{W}^+$ Optimization	1.4	Insufficient displacement
$\mathcal{W}$-space Editing	9.4	Uncontrolled overshoot
S-space (StyleSpace)	0.8–1.1	Insufficient displacement
Pure RBF Warping	$>$25	Unrealistic artifacts
Hybrid (Proposed)	**5–15**	**Controlled transformation**

These results confirm that StyleGAN encodes skeletal morphology as patient identity rather than as a trans-

formable attribute. Latent editing approaches, which operate within the learned distribution of existing faces, cannot achieve clinically significant geometric displacement without identity collapse. The hybrid framework circumvents this limitation by performing explicit geometric transformation in image space before refinement, achieving 7–10$\times$ greater Pogonion displacement than latent editing while maintaining anatomical coherence.

5.5 Hyperparameter Sensitivity Analysis

Beyond the component necessity ablations (Tables 3, 4), two critical hyperparameters governing geometric transformation were validated.

Geometric Amplification Factors. The adaptive amplification range $\alpha \in [1.8, 4.5]$ for CONVEX profiles was validated against alternative ranges $[1.0, 3.0]$ and $[2.5, 6.0]$ on 50 validation images. Values below 1.5 produced insufficient final Pogonion displacement after GAN pull-back ($<$5 px), while values above 5.0 introduced anatomical distortions (excessive bone stretching, discontinuous cortical boundaries) that could not be resolved during refinement. The selected range compensates for the empirically observed $\sim$50% displacement reduction during $\mathcal{W}^+$-space projection while avoiding overshoot artifacts. CONCAVE profiles use $\alpha \in [2.0, 4.0]$, reflecting asymmetric warping behavior in posterior transformations.

ViT Fine-Tuning Duration. Both auxiliary models were evaluated at 20, 30, and 50 training epochs. Classification accuracy improved from 87.2% (20 epochs) to 89.0% (30 epochs), with minimal change to 89.3% at 50 epochs. Landmark regression accuracy plateaued similarly at 30 epochs. The 30-epoch configuration balances convergence against computational cost.

6 DISCUSSION

6.1 Latent Editing Failure and Hybrid Necessity

The ablation study revealed that all latent space methods produced negligible skeletal changes, as shown in Table 4. These results demonstrate that StyleGAN learned skeletal structure as patient identity rather than as a malleable attribute. StyleGAN encodes skeletal morphology as a fundamental identity characteristic, analogous to how it treats ethnicity or facial structure, rather than as a transformable attribute like age or expression. This occurs because the training data contained different individuals with different skeletal profiles but never presented before-and-after pairs of the same individual undergoing treatment. Consequently, attempts to move Pogonion in latent space trigger identity changes rather than geometric corrections. Mathematically, this manifests as orthogonal editing directions: profile-altering latent shifts necessarily traverse dimensions that StyleGAN has learned to associate with inter-individual variation, causing collapse to different training samples rather than controlled transformation.

This finding has direct implications for medical image editing. Unlike texture attributes (age, expression) that vary continuously within individuals in natural face datasets, skeletal geometry defines the individual in cephalometric imaging. GANs cannot edit what they were never trained to transform. The hybrid framework circumvents this limitation by performing geometric modification in image space where control points and displacement vectors are explicit. RBF warping achieves deterministic 5–15 pixel movements that would require amplification factors ($> 20\times$) causing identity collapse in latent space.

6.2 Clinical Interpretation and VTO Workflow

Traditional Visual Treatment Objective methodology progresses through five manual phases [26]. Profile assessment by experienced clinicians establishes baseline skeletal relationships. Cephalometric landmark tracing identifies Nasion, Point A, and Pogonion alongside other parameters for planning maxillofacial orthodontic, orthopaedic and surgical procedures. Displacement estimation projects required skeletal movements based on treatment goals. Skeletal tissue based soft tissue prediction applies empirical ratios derived from population

studies to forecast lip and chin response. Treatment communication uses cephalometric tracings or gnathostatic models to visualize outcomes [38].

This computational framework automates key steps of the clinical workflow, yielding measurable and reproducible outputs that enhance patient education and compliance with treatment procedures. Automated profile classification provides explicit confidence estimates, with performance validated through clinical review. Landmark localization reaches sub-2-pixel precision, meeting the threshold commonly cited for surgical planning [35]. The geometric warping stage applies prescribed displacements directly, avoiding the subjective variability inherent to manual estimation. Subsequent StyleGAN-based refinement restores photorealistic soft-tissue structure without the need for hand-drawn adjustments.

A critical distinction must be emphasized: the framework is designed as a pre-clinical foundational research stage visualization aid to support patient education, informed consent discussions, and treatment outcome exploration. By providing realistic previews of potential outcomes across preventive, interceptive, and corrective orthodontic and orthognathic treatment planning, it supports clinician-patient communication and ensures transparency in shared decision-making. The current framework models geometric response of skeletal structures and synthesizes soft-tissue appearance through learned statistical patterns, and does not account for biomechanical tissue dynamics, remaining cephalometric landmarks beyond N, A, and Pog, or 3D volumetric changes. It should at present be utilized as a communicative adjunct rather than a primary tool for determining clinical skeletal or tooth movements. The GAN refinement generates visually plausible soft-tissue draping based on training data distributions, representing a data-driven approach to outcome visualization. This differs from physics-based biomechanical modeling of tissue elasticity, muscular dynamics, or volumetric changes following skeletal repositioning [28]. Traditional VTO relies on empirical ratios (e.g., 1:1 for upper lip response to maxillary advancement [38]) derived from population studies, whereas the proposed approach learns these relationships implicitly through StyleGAN's exposure to existing cephalometric profiles. Recent systematic reviews [31, 32, 33] indicate that soft-tissue prediction remains an active research area across both learning-based and physics-based approaches, with ongoing work to improve accuracy and generalizability.

The framework is designed as a visualization tool for treatment communication, enabling orthodontists and maxillofacial surgeons to generate realistic previews aligned to specific profile classifications for informed consent discussions [27]. Clinical dosing decisions (millimeters of skeletal movement, surgical approach selection) continue to require traditional cephalometric analysis and biomechanical consideration [30]. The reported landmark drift (mean: 0.9 mm) quantifies internal geometric consistency—the system's ability to preserve intended transformations during refinement—demonstrating technical feasibility for controlled geometric manipulation with photorealistic synthesis.

6.3 Limitations and Clinical Validation Requirements

A primary limitation is the dataset's lack of demographic diversity, owing to limited multicentric data collection and insufficient representation across ethnic backgrounds and age groups. The demographic constraints described in Section 3 have direct implications for clinical generalizability. The 63% CONVEX predominance may reduce model reliability for CONCAVE correction, and the absence of multicentric data limits applicability across diverse ethnic populations. Multicentric validation is required before broader deployment [34].

The framework operates on 2D lateral cephalograms without access to the transverse (anteroposterior width) dimension. The system analyzes only three landmarks (N, A, Pog), while comprehensive cephalometric analysis requires additional parameters such as ANB angle, mandibular plane angle, and facial height ratios.

The system should be interpreted as a research-stage framework demonstrating computational feasibility rather than a clinically deployable tool. Translation to clinical practice requires: (1) prospective studies comparing predicted profiles to actual post-treatment outcomes, quantifying prediction error in millimeters across diverse patient populations; (2) regulatory consideration for software as a medical device classification; (3) integration with DICOM workflows and clinical information systems; (4) validation by independent orthodontic practitioners assessing clinical utility in treatment planning discussions.

7 Conclusion and Future Scope

Visualizing skeletal profile transformations for patient communication requires balancing geometric precision with photorealistic synthesis, a challenge existing generative models cannot address through latent editing alone. This work demonstrates that profile-specific cephalometric transformation with quantified geometric control is achievable through hybrid geometric-generative methods, successfully addressing the fundamental limitation of latent-space editing for skeletal modification.

The framework establishes computational parallels to Visual Treatment Objective methodology: automated profile classification replaces subjective assessment, landmark regression provides sub-2-pixel precision for control point specification, geometric warping enforces deterministic skeletal displacement, and generative refinement synthesizes photorealistic soft-tissue appearance. The system achieves controlled Pogonion displacement (5–15 px) with preserved geometric fidelity (mean landmark drift: 0.9 mm), demonstrating 7–10$\times$ greater skeletal control than latent editing approaches while maintaining anatomical coherence.

Although validated on pre-clinical anonymized data, the framework is designed to serve as a high-fidelity visualization aid to enhance patient education and informed consent discussions. By providing realistic previews of potential outcomes across preventive, interceptive, and corrective orthodontic and orthognathic treatment planning, it supports clinician-patient communication and ensures transparency in shared decision-making. The system does not account for biomechanical tissue dynamics, remaining cephalometric landmarks beyond N, A, and Pog, or 3D volumetric changes, and requires prospective clinical validation before deployment. It should be utilized as a communicative adjunct rather than a primary tool for surgical planning or determining clinical skeletal and tooth movements.

The hybrid architecture opens several directions for technical extension. The ViT classifier could be made more robust to geometric transformations by training on warp-augmented datasets incorporating synthetic geometric variations. The GAN pull-back effect could be reduced by fine-tuning the generator on warped samples, teaching it to treat geometric modifications as valid configurations and expanding the learned distribution to encompass transformation trajectories. Adversarial training strategies where the discriminator explicitly learns to distinguish real radiographs from both poorly warped samples with artifacts and well-warped samples with correct geometry could guide the generator to preserve transformations. Conditional generation architectures that directly encode profile class as input would enable end-to-end transformation without requiring two-stage processing.

Potential applications span multiple deployment contexts. Automated profile classification on smartphone photographs enables screening in primary care and in low-resource settings. Geometric warping provides transparent surgical simulation for secondary care orthodontic planning. Photorealistic visualization tools support informed-consent discussions in tertiary orthognathic centers [31].

Extension to 3D CBCT data would enable comprehensive volumetric analysis, requiring architectures with 3D convolutions and multi-view consistency enforcement. Volumetric warping methods and 3D StyleGAN architectures conditioned on skeletal geometry would synthesize view-consistent soft-tissue predictions [36]. Paired pre-treatment and post-treatment CBCT datasets from longitudinal studies provide ground truth for training conditional transformation models. Integrating finite element modeling of soft-tissue biomechanics [37] would provide anatomically principled constraints on deformation, combining learned appearance models with physics-based tissue response. Hybrid approaches integrating data-driven synthesis with biomechanical simulation could improve prediction accuracy while reducing dependence on training data size. Also in future this can extend the possibility of a classifier free text-to-image diffusion model [40] .

Clinical translation requires prospective longitudinal studies comparing generated profiles to actual post-treatment imaging data with quantified error metrics. Regulatory pathways for software medical device classification must address liability when predictions inform treatment decisions. Integration with DICOM workflows requires interoperability standards and independent validation by orthodontic practitioners.

The research provides a methodological foundation for treatment visualization systems, demonstrating that controlled geometric transformation with photorealistic synthesis is computationally feasible through explicit

separation of geometric manipulation from texture generation.

8 Acknowledgment

The authors appreciate Dr. Pooja Bohidar, SCB Dental College & Hospital, Cuttack, Odisha, for her guidance and technical support during the protocol development and ethical approval process. This study was conducted under the Institutional Ethics Committee clearance IEC/SCBDCH/048/2019, granted by the Institutional Ethics Committee of SCB Dental College & Hospital.

References

[1] V. Pornsirianand, J. Chavanavesh, P. Kanpittaya, "Aesthetically acceptable cephalometric treatment goal differences for straight, concave, and convex facial profiles between genders", *Clinical and Investigative Orthodontics* 83(1):33–43, 2024. https://doi.org/10.1080/27705781.2023.2298077

[2] R. W. McNeill, W. R. Proffit, R. P. White, "Cephalometric prediction for orthodontic surgery", *The Angle Orthodontist* 42(2):154–164, 1972. https://api.semanticscholar.org/CorpusID:76707554

[3] C. Johnston, D. Burden, D. Kennedy, N. Harradine, M. Stevenson, "Class III surgical-orthodontic treatment: a cephalometric study", *American Journal of Orthodontics and Dentofacial Orthopedics* 130(3):300–309, 2006. https://doi.org/10.1016/j.ajodo.2006.05.027

[4] 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):161, 2023. https://doi.org/10.1186/s12903-023-02844-z

[5] I.-H. Kim, J. Jeong, J.-S. Kim, J. Lim, J.-H. Cho, M. Hong, K.-H. Kang, M. Kim, S.-J. Kim, Y.-J. Kim, et al., "Predicting orthognathic surgery results as postoperative lateral cephalograms using graph neural networks and diffusion models", *Nature Communications* 16(1):2586, 2025. https://doi.org/10.1038/s41467-025-57669-x

[6] F. Schwendicke, A. Chaurasia, L. Arsiwala, J.-H. Lee, K. Elhennawy, P.-G. Jost-Brinkmann, F. Demarco, J. Krois, "Deep learning for cephalometric landmark detection: systematic review and meta-analysis", *Clinical Oral Investigations* 25(7):4299–4309, 2021. https://doi.org/10.1007/s00784-021-03990-w

[7] M. Juneja, P. Garg, R. Kaur, P. Manocha, S. Batra, P. Singh, S. Singh, P. Jindal, et al., "A review on cephalometric landmark detection techniques", *Biomedical Signal Processing and Control* 66:102486, 2021. https://doi.org/10.1016/j.bspc.2021.102486

[8] K. Oh, I.-S. Oh, D.-W. Lee, et al., "Deep anatomical context feature learning for cephalometric landmark detection", *IEEE Journal of Biomedical and Health Informatics* 25(3):806–817, 2020. https://doi.org/10.1109/JBHI.2020.3002582

[9] M. Zeng, Z. Yan, S. Liu, Y. Zhou, L. Qiu, "Cascaded convolutional networks for automatic cephalometric landmark detection", *Medical Image Analysis* 68:101904, 2021. https://doi.org/10.1016/j.media.2020.101904

[10] S. Liu, D. Li, T. Cao, Y. Sun, Y. Hu, J. Ji, "GAN-based face attribute editing", *IEEE Access* 8:34854–34867, 2020. https://doi.org/10.1109/ACCESS.2020.2974043

[11] Y. Liu, Q. Li, Q. Deng, Z. Sun, M.-H. Yang, "GAN-based facial attribute manipulation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(12):14590–14610, 2023. https://doi.org/10.1109/TPAMI.2023.3298868

[12] H. Yang, L. Chai, Q. Wen, S. Zhao, Z. Sun, S. He, "Discovering interpretable latent space directions of GANs beyond binary attributes", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12177–12185, 2021. https://doi.org/10.1109/CVPR46437.2021.01200

[13] A. Voynov, A. Babenko, "Unsupervised discovery of interpretable directions in the GAN latent space", *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 9786–9796, 2020. https://proceedings.mlr.press/v119/voynov20a.html

[14] Y.-H. Park, M. Kwon, J. Jo, Y. Uh, "Unsupervised discovery of semantic latent directions in diffusion models", 2023. https://doi.org/10.48550/arXiv.2302.12469

[15] P. Liang, G. Liu, Z. Xiong, H. Fan, H. Zhu, X. Zhang, "A facial geometry based detection model for face manipulation using CNN-LSTM architecture", *Information Sciences* 633:370–383, 2023. https://doi.org/10.1016/j.ins.2023.03.079

[16] C.-H. Lee, Z. Liu, L. Wu, P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5549–5558, 2020. https://doi.org/10.1109/CVPR42600.2020.00559

[17] M. A. Khalid, K. Zulfiqar, U. Bashir, A. Shaheen, R. Iqbal, Z. Rizwan, G. Rizwan, and M. M. Fraz, "Aariz: A Benchmark Dataset for Automatic Cephalometric Landmark Detection and CVM Stage Classification," 2023. https://doi.org/10.1038/s41597-025-05542-3

[18] KambingBersayapHitam, "Cephalometric Profile Dataset," Kaggle, 2023. https://www.kaggle.com/datasets/kambingbersayaphitam/cephalometric-profile-dataset

[19] L. Harris, "A Beginners Guide to Lateral Cephalometric Radiographs," *Dental Protection*, March 3 2015; last updated November 14 2018. https://www.dentalprotection.org/uk/articles/a-beginners-guide-to-lateral-cephalometric-radiographs

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations (ICLR)*, 2021. https://doi.org/10.48550/arXiv.2010.11929

[21] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, "Analyzing and improving the image quality of StyleGAN," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, 2020. https://doi.org/10.1109/CVPR42600.2020.00813

[22] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989. https://api.semanticscholar.org/CorpusID:47302

[23] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," *Constructive Theory of Functions of Several Variables*, vol. 571, pp. 85–100, Springer, 1977. https://api.semanticscholar.org/CorpusID:123055447

[24] A. Sotiras, C. Davatzikos, N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013. https://doi.org/10.1109/TMI.2013.2265603

[25] R. Abdal, Y. Qin, P. Wonka, "Image2StyleGAN: How to embed images into the StyleGAN latent space?" *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4432–4441, 2019. https://doi.org/10.1109/ICCV.2019.00453

[26] C. Toepel-Sievers, H. Fischer-Brandies, "Validity of the computer-assisted cephalometric growth prognosis VTO (Visual Treatment Objective) according to Ricketts", *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie* 60(3):185–194, 1999. https://doi.org/10.1007/BF01365265

[27] K.J. Nirayath, A. Husain, "Assessment of Visualized Treatment Objective and Post-Treatment Outcome in Skeletal Class II Functional Cases: A Comparative Study", *Asian Journal of Dental Sciences* 6(1):211–219, 2023. https://journalajds.com/index.php/AJDS/article/view/149

[28] R.J. Gupta, S. Schendel, "Soft Tissue Changes and Prediction with Orthognathic Surgery", In: M. Miloro, G.E. Ghali, P.E. Larsen, P. Waite (eds), *Peterson's Principles of Oral and Maxillofacial Surgery*, Springer, Cham, 2022. https://doi.org/10.1007/978-3-030-91920-7_67

[29] A.K. Dall'Magro, et al., "Orthognathic surgery and orthodontics associated with orofacial harmonization: Case report", *International Journal of Surgery Case Reports* 83:106013, 2021. https://doi.org/10.1016/j.ijscr.2021.106013

[30] G. Van Hemelen, et al., "Three-dimensional virtual planning in orthognathic surgery enhances the accuracy of soft tissue prediction", *Journal of Cranio-Maxillofacial Surgery* 43(6):918–925, 2015. https://doi.org/10.1016/j.jcms.2015.04.006

[31] A.A. Almarhoumi, "Accuracy of artificial intelligence in predicting facial changes post-orthognathic surgery: a comprehensive scoping review", *Journal of Clinical and Experimental Dentistry* 16(5):e624, 2024. https://doi.org/10.4317/jced.61500

[32] D. Salazar, P.E. Rossouw, F. Javed, D. Michelogiannakis, "Artificial intelligence for treatment planning and soft tissue outcome prediction of orthognathic treatment: A systematic review", *Journal of Orthodontics* 51(2):107–119, 2024. https://doi.org/10.1177/14653125231203743

[33] B. Gong, et al., "Research of orthodontic soft tissue profile prediction based on conditional generative adversarial networks", *Journal of Dentistry* 154:105570, 2025. https://doi.org/10.1016/j.jdent.2025.105570

[34] P. Smołka, K. Nelke, N. Struzik, K. Wiśniewska, S. Kiryk, J. Kensy, W. Dobrzyński, J. Kiryk, J. Matys, M. Dobrzyński, "Discrepancies in cephalometric analysis results between orthodontists and radiologists and artificial intelligence: A systematic review", *Applied Sciences* 14(12):4972, 2024. https://doi.org/10.3390/app14124972

[35] B. Liu, C. Liu, Y. Xiong, et al., "Accuracy and reliability of 3D cephalometric landmark detection with deep learning", *European Journal of Medical Research* 30(1):1000, 2025. https://doi.org/10.1186/s40001-025-03198-8

[36] H. Tang, S. Liu, Y. Shi, J. Wei, J. Peng, H. Feng, "Automatic segmentation and landmark detection of 3D CBCT images using semi supervised learning for assisting orthognathic surgery planning", *Scientific Reports* 15:8814, 2025. https://doi.org/10.1038/s41598-025-93317-6

[37] J. Bao, X. Zhang, S. Xiang, H. Liu, M. Cheng, Y. Yang, X. Huang, W. Xiang, W. Cui, H.C. Lai, S. Huang, Y. Wang, D. Qian, H. Yu, "Deep Learning-Based Facial and Skeletal Transformations for Surgical Planning", *Journal of Dental Research* 103(8):809–819, 2024. https://doi.org/10.1177/00220345241253186

[38] W.R. Proffit, H.W. Fields, D.M. Sarver, "Contemporary Orthodontics 4th ed", *Mosby*, Philadelphia, 2007.

[39] A. B. George, S. Bathini and G. A, "Grad-CAM & Grad-CAM++ for Explainable Oral Squamous Cell Carcinoma Detection using Deep Learning on Orthopantomograms," 2025 *International Conference on Sensors and Related Networks (SENNET) Special Focus on Digital Healthcare(64220)*, Vellore, India, 2025, pp. 1–6. https://doi.org/10.1109/SENNET64220.2025.11136014

[40] K. Song, L. Han, B. Liu, D. Metaxas and A. Elgammal, "StyleGAN-Fusion: Diffusion Guided Domain Adaptation of Image Generators," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024. https://doi.org/10.1109/WACV57701.2024.00537