OrthoMorphNet: Alignment-Guided Hierarchical Morphological Synthesis of Post-Orthodontic Dentition and Gingiva

Ji Yong Han[1], Woncheol Shin[2], Su Yang[3], Sujeong Kim[1], Hyunbin Yun[1], Sang-Heon Lim[1], Ho Yoon Choi[4], Mobin Ahmadi[1], Jun-Min Kim[5], Sang-Jeong Lee[6], and Won-Jin Yi[1,2,7]

[1]Interdisciplinary Program in Bioengineering, Graduate School of Engineering, Seoul National University, Republic of Korea
[2]School of Medicine, CHA University, Republic of Korea
[3]Department of Medical Informatics, School of Medicine, Kyungpook National University, Republic of Korea
[4]Applied Bioengineering, Graduate School of Convergence Science and Technology, Seoul National University, Republic of Korea
[5]Electronics and Information Engineering, Hansung University, Republic of Korea
[6]Department of Artificial Intelligence, Tech University of Korea, Republic of Korea
[7]Oral and Maxillofacial Radiology and Dental Research Institute, School of Dentistry, Seoul National University, Republic of Korea
{jiyong, wjyi}@snu.ac.kr

Abstract. Digital orthodontic workflows increasingly rely on virtual setups to predict treatment outcomes from pre-treatment intraoral scans. Although learning-based methods can automatically align teeth by estimating per-tooth rigid motions, they often neglect soft-tissue deformation, leaving post-treatment gingival morphology unknown. We propose OrthoMorphNet, a hierarchical framework for joint synthesis of post-orthodontic dentition and gingiva. The model first aligns individual teeth by fusing per-tooth geometry with global spatial context through cross-attention to regress translation and unit-quaternion rotation, supervised by a composite objective enforcing geometric regularization and physical plausibility, including collision avoidance and arch-continuity. Subsequently, the predicted dentition guides gingiva synthesis in a volumetric indicator-grid space using a Swin-transformer-based 3D encoder–decoder, followed by Marching Cubes surface extraction with a grid reconstruction loss. Experiments on a public dataset demonstrate high alignment accuracy and state-of-the-art gingiva reconstruction compared to existing volumetric baselines, supporting clinically faithful end-to-end 3D orthodontic simulation.

Keywords: Digital orthodontics · tooth alignment · gingiva synthesis · transformer · volumetric reconstruction.

1 Introduction

The rapid evolution of digital orthodontic treatments and advances in intraoral 3D scanning have accelerated the shift toward a fully digital workflow [16]. In this paradigm, virtual treatment planning and digital setups are central to clinical decision-making, appliance fabrication, and patient communication. Accurately predicting an optimal post-treatment tooth arrangement from a pre-treatment scan is essential for efficient planning, yet, manual setup remains time-consuming and highly dependent on clinical experience, motivating the development of robust AI-driven automatic tooth alignment.

Recent learning-based approaches have advanced automatic tooth alignment by regressing per-tooth rigid motions through MLP-based predictors [8, 14, 15], graph-based reasoning [20, 19], and diffusion models [6] for improved realism [3]. While diffusion-based approaches improve generative realism, they rely on iterative denoising during inference and produce stochastic outputs, whereas orthodontic treatment planning requires deterministic and reproducible predictions for clinical validation and interactive setup refinement [7, 10, 13, 1]. Nevertheless, these pipelines predominantly focus on rigid tooth repositioning and overlook the non-linear deformation of the surrounding gingiva, which is critical for achieving clinically faithful simulation. Accurate gingival morphology is essential for smile esthetics, periodontal risk assessment like black triangles, and the precise fit of clear aligners, creating a need for unified prediction of both post-treatment dentition and gingival geometry [17]. In this paper, we introduce OrthoMorphNet, a hierarchical framework for synthesizing post-orthodontic dentition and gingiva from pre-treatment 3D data. First, a tooth alignment network predicts per-tooth rigid transformations by fusing local geometry with global spatial context through cross-attention and regressing both translation vectors and unit quaternions for rotation. Subsequently, a gingiva generation network synthesizes the post-treatment gingiva conditioned on the pre-treatment context and the predicted alignment. By learning deformations in a volumetric indicator-grid space, our framework enables high-resolution, noise-robust surface synthesis (Fig. 1). Our contributions can be summarized as follows:

- To the best of our knowledge, OrthoMorphNet is the first framework to jointly predict post-treatment tooth alignment and the corresponding gingival morphology from pre-treatment 3D scans.
- We introduce an alignment-guided hierarchical framework in which the predicted post-treatment dentition serves as an explicit deformation condition for gingiva synthesis, coupling hard-tissue alignment with soft-tissue morphology prediction.
- We incorporate geometric and physical constraints for anatomically plausible alignment and achieve state-of-the-art performance for both tooth alignment and gingiva generation on a public pre/post-treatment dataset.

2 Method

2.1 Problem Formulation and Overview

Let $\mathcal{D}^{pre} = (\mathcal{T}^{pre}, \mathcal{G}^{pre})$ denote the pre-treatment 3D dental model, where $\mathcal{T}^{pre} = \{\mathbf{P}_i^{pre}\}_{i=1}^{N}$ and $\mathcal{G}^{pre}$ are the segmented teeth point sets and gingiva surface, respectively. Our goal is to synthesize the post-treatment outputs $\hat{\mathcal{T}}^{post} = \{\hat{\mathbf{P}}_i^{post}\}_{i=1}^{N}$ and $\hat{\mathcal{G}}^{post}$, supervised by ground-truth $\mathcal{T}^{post}$ and $\mathcal{G}^{post}$. OrthoMorphNet follows a hierarchical design: (i) a tooth alignment module predicts per-tooth rigid transformations to obtain $\hat{\mathcal{T}}^{post}$, and (ii) a gingiva generation module predicts $\hat{\mathcal{G}}^{post}$ conditioned on $\mathcal{T}^{pre}$, $\mathcal{G}^{pre}$, and $\hat{\mathcal{T}}^{post}$.

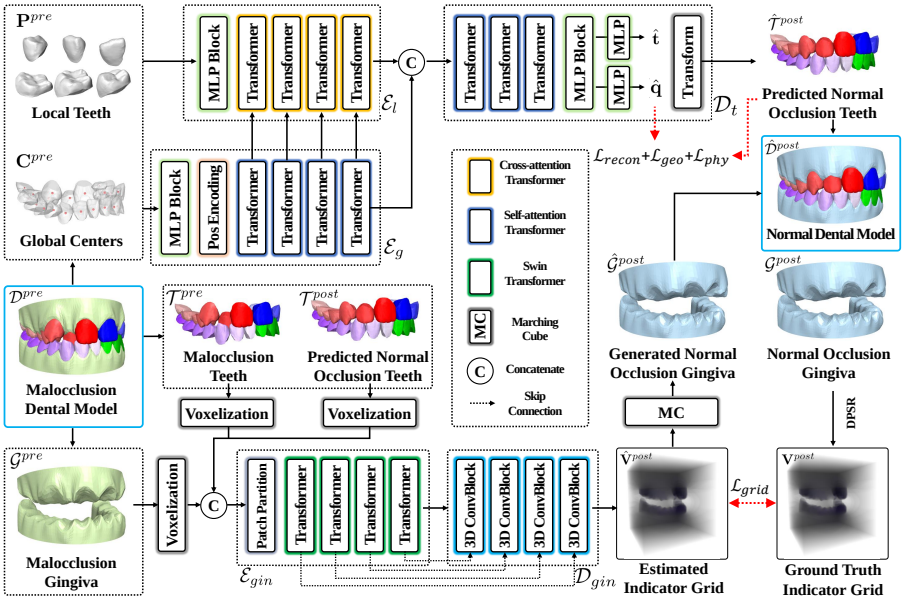

Fig. 1. Entire flow of OrthoMorphNet. OrthoMorphNet regresses per-tooth rigid motions for alignment and synthesizes the corresponding gingival surface.

2.2 Tooth Alignment Network

Local-to-global representations. To capture detailed tooth morphology and global arch context, we encode each tooth point set $\mathbf{P}_i^{pre}$ into a local token $\mathbf{f}_i^l = \mathcal{E}_l(\mathbf{P}_i^{pre})$ and encode the set of tooth centers $\mathbf{C}^{pre} = \{\mathbf{c}_1^{pre}, \ldots, \mathbf{c}_N^{pre}\}$ into a global context $\mathbf{F}^g = \mathcal{E}_g(\mathbf{C}^{pre})$. A dual-branch transformer fuses local and global features with cross-attention, and a regression decoder $\mathcal{D}_t$ predicts a translation vector $\hat{\mathbf{t}}_i \in \mathbb{R}^3$ and a quaternion $\hat{\mathbf{q}}_i \in \mathbb{R}^4$ for each tooth. We normalize $\hat{\mathbf{q}}_i$ to a

unit quaternion and convert it to a rotation matrix $\mathbf{R}(\hat{\mathbf{q}}_i)$ to obtain the aligned tooth:

$$\hat{\mathbf{P}}_i^{post} = \mathbf{R}(\hat{\mathbf{q}}_i)\,\mathbf{P}_i^{pre} + \hat{\mathbf{t}}_i. \tag{1}$$

We supervise the alignment network with $\mathcal{L}_{align} = \mathcal{L}_{recon} + \lambda_{geo}\mathcal{L}_{geo} + \lambda_{phy}\mathcal{L}_{phy}$. As point correspondence is preserved under rigid perturbations, we use point-wise MSE:

$$\mathcal{L}_{recon} = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{P}}_i^{post} - \mathbf{P}_i^{post}\right\|_2^2. \tag{2}$$

where N denotes the number of teeth. This objective ensures that each individual tooth is accurately placed in its target post-treatment position. The geometric term $\mathcal{L}_{geo}$ directly regularizes the predicted per-tooth rigid motion parameters, comprising translation and rotation, by combining an MSE loss on translations and a quaternion cosine distance:

$$\mathcal{L}_{geo} = \frac{1}{N}\sum_{i=1}^{N}\left(\|\hat{\mathbf{t}}_i - \mathbf{t}_i^{gt}\|_2^2 + \left(1 - \left|\langle\hat{\mathbf{q}}_i, \mathbf{q}_i^{gt}\rangle\right|\right)\right), \tag{3}$$

where $\mathbf{t}_i^{gt}$ and $\mathbf{q}_i^{gt}$ represent the ground-truth parameters. The absolute inner product accounts for the antipodal symmetry of quaternion representations. To enforce anatomical validity, we define the physical plausibility term as

$$\mathcal{L}_{phy} = \mathcal{L}_{coll} + \lambda_{adj}\mathcal{L}_{adj}. \tag{4}$$

For each point $\mathbf{p} \in \hat{\mathbf{P}}_i^{post}$, we compute its nearest-neighbor distance to tooth j:

$$d_{ij}(\mathbf{p}) = \min_{\mathbf{r}\in\hat{\mathbf{P}}_j^{post}} \|\mathbf{p} - \mathbf{r}\|_2. \tag{5}$$

We penalize inter-tooth distances smaller than a safety margin ϵ using a hinge penalty $\sigma(x) = \max(0, x)$:

$$\mathcal{L}_{coll} = \frac{1}{|\mathcal{S}|}\sum_{(i,j)\in\mathcal{S}}\left(\frac{1}{|\hat{\mathbf{P}}_i^{post}|}\sum_{\mathbf{p}\in\hat{\mathbf{P}}_i^{post}}\sigma\big(\epsilon - d_{ij}(\mathbf{p})\big)^2\right). \tag{6}$$

Here, $\mathcal{S}$ denotes the set of tooth pairs considered for collision checking. We define the centroid of each predicted and ground-truth post-treatment tooth as

$$\hat{\mathbf{c}}_i = \frac{1}{|\hat{\mathbf{P}}_i^{post}|}\sum_{\mathbf{p}\in\hat{\mathbf{P}}_i^{post}}\mathbf{p}, \quad \mathbf{c}_i^{gt} = \frac{1}{|\mathbf{P}_i^{post}|}\sum_{\mathbf{p}\in\mathbf{P}_i^{post}}\mathbf{p}. \tag{7}$$

Let $\mathcal{A}$ be the set of adjacent tooth pairs along the dental arch. We match the inter-centroid distances of neighboring teeth to those of the ground-truth arch:

$$\mathcal{L}_{adj} = \frac{1}{|\mathcal{A}|}\sum_{(i,j)\in\mathcal{A}}\left(\|\hat{\mathbf{c}}_i - \hat{\mathbf{c}}_j\|_2 - \|\mathbf{c}_i^{gt} - \mathbf{c}_j^{gt}\|_2\right)^2. \tag{8}$$

Together, $\mathcal{L}_{coll}$ and $\mathcal{L}_{adj}$ guide the model toward collision-free and anatomically coherent post-treatment arrangements.

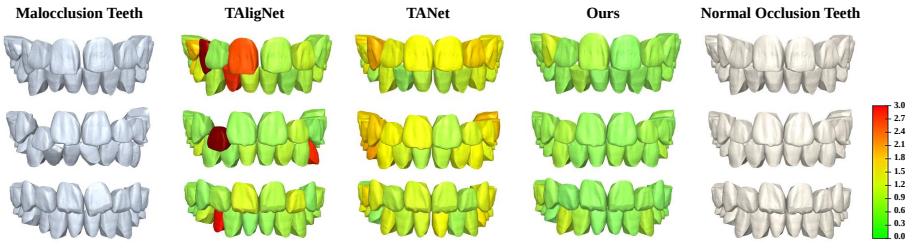

Fig. 2. Qualitative comparison for tooth alignment. We visualize per-surface deviation (mm) from the ground-truth post-treatment (normal occlusion) teeth.

2.3 Gingiva Generation Network

Indicator-grid representation. Direct mesh regression is challenging due to the gingiva's smooth yet complex topology and sensitivity to scan noise. Instead, we represent gingiva surfaces as volumetric indicator grids. Given a surface, we compute an indicator field on a fixed-resolution 3D grid using differentiable Poisson surface reconstruction (DPSR) [12], which provides a high-resolution, noise-robust representation suitable for learning and supervision.

Gingiva Generation. To synthesize the post-treatment gingiva, the network is conditioned on three volumetric cues including the voxelized pre-treatment gingiva $\mathrm{Vox}(\mathcal{G}^{pre})$, the voxelized pre-treatment dentition $\mathrm{Vox}(\mathcal{T}^{pre})$, and the voxelized predicted post-treatment dentition $\mathrm{Vox}(\hat{\mathcal{T}}^{post})$ obtained from the preceding tooth alignment stage. These volumetric representations are concatenated channel-wise to form the joint representation:

$$\mathbf{X} = \mathrm{Concat}(\mathrm{Vox}(\mathcal{G}^{pre}), \mathrm{Vox}(\mathcal{T}^{pre}), \mathrm{Vox}(\hat{\mathcal{T}}^{post})), \tag{9}$$

where the estimated post-treatment dentition $\hat{\mathcal{T}}^{post}$ provides an explicit tooth-movement condition for learning the non-linear deformation of gingival soft tissue. For volumetric synthesis, we employ a 3D encoder-decoder architecture featuring a Swin-transformer backbone for long-range spatial context and convolutional decoding blocks for local detail refinement [9, 21]. The network outputs an estimated post-treatment gingiva indicator grid $\hat{\mathbf{V}}^{post}$, where $\hat{\mathbf{V}}^{post} = \mathcal{D}_{gin}(\mathcal{E}_{gin}(X))$, from which the final surface mesh $\hat{\mathcal{G}}^{post}$ is extracted using the Marching Cubes algorithm denoted as $\hat{\mathcal{G}}^{post} = \mathrm{MC}(\hat{\mathbf{V}}^{post})$. We supervise the predicted indicator grid using a hybrid objective consisting of a voxel-wise reconstruction loss in the continuous indicator space combined with a Dice loss computed on binarized occupancy masks:

$$\mathcal{L}_{grid} = \underbrace{\left\| \hat{\mathbf{V}}^{post} - \mathbf{V}^{post} \right\|_1}_{\mathcal{L}_{L1}} + \lambda_{dice} \underbrace{\left(1 - \mathrm{DSC}(\hat{\mathbf{B}}^{post}, \mathbf{B}^{post})\right)}_{\mathcal{L}_{dice}}. \tag{10}$$

Here, $\mathbf{V}^{post} = \mathrm{DPSR}(\mathcal{G}^{post})$ denotes the ground-truth post-treatment gingiva indicator grid, while $\hat{\mathbf{V}}^{post}$ represents the predicted grid. For the Dice term, we

obtain binary masks by thresholding:

$$\hat{\mathbf{B}}^{post} = \mathbb{I}\Big(\hat{\mathbf{V}}^{post} > \tau\Big), \quad \mathbf{B}^{post} = \mathbb{I}\big(\mathbf{V}^{post} > \tau\big), \tag{11}$$

where $\mathbb{I}(\cdot)$ is the indicator function and τ is a binarization threshold. We compute the Dice similarity coefficient (DSC) as

$$\mathrm{DSC}(\hat{\mathbf{B}}, \mathbf{B}) = \frac{2\sum \hat{\mathbf{B}} \odot \mathbf{B} + \delta}{\sum \hat{\mathbf{B}} + \sum \mathbf{B} + \delta}, \tag{12}$$

with $\odot$ denoting element-wise multiplication and δ a small constant for numerical stability. This combined supervision stabilizes learning, mitigates class imbalance in volumetric occupancy, and better preserves gingival topology during synthesis.

Table 1. Quantitative comparison of tooth alignment performance. Our method achieves superior rotation accuracy ($\Delta\theta_{avg}$) and competitive translation and ADD metrics compared to TAligNet and TANet.

Method	ΔT_{avg} (mm) ↓	$\Delta\theta_{avg}$ (°) ↓	ADD (mm) ↓
TAligNet	1.178 ± 0.602	12.717 ± 4.535	1.264 ± 0.615
TANet	$\mathbf{0.754 \pm 0.174}$	14.452 ± 2.737	0.664 ± 0.242
Ours	0.868 ± 0.380	$\mathbf{10.803 \pm 3.591}$	$\mathbf{0.628 \pm 0.371}$

3 Experiments

3.1 Datasets and Implementation Details

Dataset. We use the publicly available dataset [18] providing paired 3D dental models before/after orthodontic treatment, including segmented teeth and gingiva. For tooth alignment, we use 425 post-treatment dentitions and synthesize malocclusion inputs by per-tooth rigid perturbations (rotation in $[-40°, 40°]$ and translation in $[-3, 3]$ mm per axis, clipped to ≤ 3 mm) [7]. For each subject, we generate four perturbations, yielding 1700 pairs,, split into 1356/172/172 for train/val/test (8:1:1). For gingiva generation, we use 420 paired cases, randomly split into 336/84 for train/test (80/20).

Training setup. For tooth alignment, each tooth point cloud is downsampled to 512 points and normalized to the range $[-1, 1]$. All alignment models are trained using AdamW with a learning rate of 1×10^{-4}, batch size 8, for 300 epochs. We adopt a ReduceLROnPlateau scheduler with patience 20 based on the validation loss. Training is early-terminated if the validation loss does not decrease for more than 40 consecutive epochs. For gingiva generation, the gingiva synthesis network is trained with batch size 1 using the Adam optimizer with a learning rate of 1×10^{-4} for 100 epochs. All models are implemented in Python3 using the PyTorch framework and trained on a single NVIDIA RTX A6000

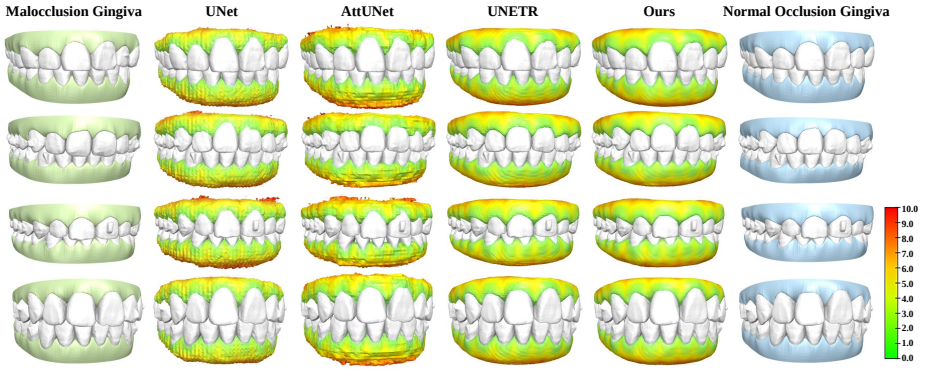

Fig. 3. Qualitative comparison for post-treatment gingiva synthesis. Distance maps (mm) are shown against the ground-truth normal gingiva. Our method demonstrates more anatomically consistent gingival morphology and reduces large-error regions compared with 3D UNet, AttUNet, and UNETR.

GPU (48 GB RAM). All baseline methods are run under the same computing environment for fair comparison.

Evaluation metrics. We evaluate tooth alignment and gingiva generation using task-specific metrics. For tooth alignment, we report the mean translation error ΔT_{avg} (mm) and mean rotation error $\Delta\theta_{avg}$ (°). We further evaluate pose accuracy using the average distance of model points (ADD) [5], defined as the mean point-wise distance between the predicted and ground-truth tooth models. For gingiva generation, we report the mean absolute error (MAE) and the coefficient of determination R^2 on continuous indicator grids, IoU on binarized volumes (threshold at zero), and surface metrics using normal consistency (NC), surface distance error (SDE), and Hausdorff distance (HD). For reporting clarity, MAE/SDE/HD are scaled by 10^3.

Table 2. Quantitative evaluation of gingiva generation. We compare our approach against baseline 3D synthesis networks using various volumetric and surface metrics.

Method	MAE ↓	R^2 ↑	IoU ↑	Dice ↑	SDE ↓	HD ↓	NC ↑
U-Net	36.917 ± 18.389	0.839 ± 0.056	0.777 ± 0.103	0.869 ± 0.102	0.748 ± 0.102	1.937 ± 0.319	0.876 ± 0.022
AttUNet	63.067 ± 18.910	0.720 ± 0.073	0.705 ± 0.105	0.821 ± 0.102	0.867 ± 0.214	2.151 ± 0.534	0.875 ± 0.020
UNETR	19.581 ± 19.538	0.931 ± 0.083	0.873 ± 0.114	0.932 ± 0.108	0.540 ± 0.100	1.506 ± 0.343	0.951 ± 0.018
Ours	$\mathbf{12.486 \pm 19.237}$	$\mathbf{0.947 \pm 0.100}$	$\mathbf{0.883 \pm 0.115}$	$\mathbf{0.934 \pm 0.107}$	$\mathbf{0.511 \pm 0.086}$	$\mathbf{1.446 \pm 0.292}$	$\mathbf{0.954 \pm 0.017}$

3.2 Experimental Results

Quantitative comparison with other methods. We compare OrthoMorphNet with representative deterministic tooth alignment methods [8, 20]. Although diffusion-based models improve generative realism, their iterative and

stochastic inference produces multiple plausible outputs, whereas orthodontic treatment planning requires a single deterministic and reproducible final configuration for clinical decision-making.

Therefore, we restrict our quantitative comparison to deterministic alignment approaches and benchmark gingiva synthesis against established 3D volumetric baselines [2, 4, 11]. For tooth alignment (Table 1), OrthoMorphNet achieves the best overall accuracy with a rotation error ($\Delta\theta_{avg}$) of 10.803±3.591° and an ADD of 0.628 ± 0.371 mm. In terms of translation error (ΔT_{avg}), our method obtains 0.868 ± 0.380 mm, demonstrating competitive performance nearly equivalent to TANet (0.754 ± 0.174 mm), which showed the highest accuracy in this specific metric. For gingiva generation (Table 2), OrthoMorphNet consistently delivers the best performance across all evaluation metrics. Specifically, it surpasses UNETR, the second-best performing baseline, by achieving the top scores in MAE (12.486 ± 19.237), R^2 (0.947 ± 0.100), IoU (0.883 ± 0.115), Dice (0.934 ± 0.107), SDE (0.511 ± 0.086), HD (1.446 ± 0.292), and NC (0.954 ± 0.017). Overall, these results suggest that OrthoMorphNet can generate post-treatment gingiva surfaces with improved geometric fidelity relative to the evaluated baselines.

Qualitative comparison with other methods. Fig. 2 shows qualitative tooth alignment results. OrthoMorphNet more accurately restores anatomically plausible tooth poses with fewer local misalignments than competing methods, especially for teeth with large initial perturbations. Fig. 3 presents qualitative gingiva generation results and distance maps against the ground truth. Our method produces gingiva surfaces that more closely follow the ground-truth contours, particularly around the gingival margin and interproximal regions, while reducing large-error regions where competing baselines tend to exhibit oversmoothing or localized distortions.

Effectiveness of the proposed design. Table 3 validates the contributions of key components in OrthoMorphNet. In (a), combining geometric and physical regularizers with the reconstruction loss achieves the best overall tooth alignment, attaining the lowest translation error ($\Delta T_{avg} = 0.868 \pm 0.380$ mm) and ADD (0.628 ± 0.371 mm). While the rotation error ($\Delta\theta_{avg} = 10.803 \pm 3.591°$) is nearly equivalent to other variants, the full loss configuration provides the most balanced performance. In contrast, removing either regularizer leads to degraded accuracy, as seen in the increased ΔT_{avg} and ADD values. In (b), conditioning the gingiva generator on both aligned post-treatment teeth and pre-treatment context (pre-gingiva and pre-teeth) results in the most accurate synthesis. This strategy improves the MAE to 12.49 ± 19.24 and consistently achieves the top scores across all other metrics, including R^2 (0.947 ± 0.100), IoU (0.883 ± 0.115), Dice (0.934 ± 0.107), SDE (0.511 ± 0.086), HD (1.446 ± 0.292), and NC (0.954 ± 0.017).

4 Conclusion

We proposed OrthoMorphNet, a unified framework for hierarchical orthodontic treatment planning that combines per-tooth rigid alignment with conditional

Table 3. Ablation study results for the proposed framework. Section (a) evaluates the impact of loss functions on tooth alignment, while section (b) examines input conditioning strategies for gingiva generation.

(a) Ablation study on Loss function of Tooth Alignment

Loss Combination	ΔT_{avg} (mm) ↓	$\Delta\theta_{avg}$ (°) ↓	ADD (mm) ↓
$\mathcal{L}_{recon} + \mathcal{L}_{geo}$	0.893±0.417	**9.459±3.397**	0.641±0.387
$\mathcal{L}_{recon} + \mathcal{L}_{phy}$	0.917±0.391	10.309±3.255	0.642±0.420
Full ($\mathcal{L}_{recon} + \mathcal{L}_{geo} + \mathcal{L}_{phy}$)	**0.868±0.380**	10.803±3.591	**0.628±0.371**

(b) Ablation study on Condition of Gingiva Generation

Input Condition	MAE↓	R^2↑	IoU↑	Dice↑	SDE↓	HD↓	NC↑
Post Teeth + Pre Gingiva	13.15±19.01	0.936±0.097	0.870±0.113	0.924±0.107	0.535±0.091	1.476±0.301	0.952±0.017
Post Teeth + Pre Teeth	14.11±19.32	0.926±0.095	0.846±0.111	0.911±0.106	0.585±0.123	1.618±0.394	0.947±0.018
Full (Post + Pre Gin + Pre Teeth)	**12.49±19.24**	**0.947±0.100**	**0.883±0.115**	**0.934±0.107**	**0.511±0.086**	**1.446±0.292**	**0.954±0.017**

gingiva synthesis. Experiments demonstrate competitive tooth translation accuracy, with improved results in rotation and ADD. Furthermore, the framework achieves consistent gains in gingiva reconstruction quality across both quantitative and qualitative evaluations, suggesting its potential for reliable post-treatment dental model generation. These findings highlight the potential of OrthoMorphNet to support clinically reliable and anatomically consistent digital orthodontic treatment planning.

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

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (RS-2023-NR077088 and RS-2026-25504135), and by the Technology Innovation Program (or Industrial Strategic Technology Development Program–Advanced Biomaterials) (RS-2025-14322975), funded by the Ministry of Trade, Industry & Energy (MOTIE).

References

1. Akdeniz, B.S., Aykaç, V., Turgut, M., Çetin, S.: Digital dental models in orthodontics: A review. Deneysel ve Klinik Tıp Dergisi **39**(1), 250–255 (2022)
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
3. Dou, Y., Wu, H., Li, C., Wang, C., Yang, T., Zhu, M., Shen, D., Cui, Z.: Clikdiffusion: Clinical knowledge-informed diffusion model for tooth alignment. Medical Image Analysis **106**, 103746 (2025)
4. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
5. Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N.: Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In: Asian conference on computer vision. pp. 548–562. Springer (2012)

6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Lei, C., Xia, M., Wang, S., Liang, Y., Yi, R., Wen, Y.H., Liu, Y.J.: Automatic tooth arrangement with joint features of point and mesh representations via diffusion probabilistic models. Computer Aided Geometric Design **111**, 102293 (2024)
8. Lingchen, Y., Zefeng, S., Yiqian, W., Xiang, L., Kun, Z., Hongbo, F., et al.: iortho-predictor: model-guided deep prediction of teeth alignment. ACM Transactions on Graphics **39**(6), 216 (2020)
9. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
10. Luo, S., Hu, W.: Diffusion probabilistic models for 3d point cloud generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2837–2845 (2021)
11. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
12. Peng, S., Jiang, C., Liao, Y., Niemeyer, M., Pollefeys, M., Geiger, A.: Shape as points: A differentiable poisson solver. Advances in Neural Information Processing Systems **34**, 13032–13044 (2021)
13. Proffit, W.R., Fields Jr, H.W., Sarver, D.M.: Contemporary orthodontics. Elsevier Health Sciences (2006)
14. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
15. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
16. Rossini, G., Parrini, S., Castroflorio, T., Deregibus, A., Debernardi, C.L.: Efficacy of clear aligners in controlling orthodontic tooth movement: a systematic review. The Angle Orthodontist **85**(5), 881–889 (2015)
17. Tarnow, D.P., Magner, A.W., Fletcher, P.: The effect of the distance from the contact point to the crest of bone on the presence or absence of the interproximal dental papilla. Journal of periodontology **63**(12), 995–996 (1992)
18. Wang, S., Lei, C., Liang, Y., Sun, J., Xie, X., Wang, Y., Zuo, F., Bai, Y., Li, S., Liu, Y.J.: A 3d dental model dataset with pre/post-orthodontic treatment for automatic tooth alignment. Scientific data **11**(1), 1277 (2024)
19. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics (tog) **38**(5), 1–12 (2019)
20. Wei, G., Cui, Z., Liu, Y., Chen, N., Chen, R., Li, G., Wang, W.: Tanet: towards fully automatic tooth arrangement. In: European conference on computer vision. pp. 481–497. Springer (2020)
21. Zhou, H.Y., Guo, J., Zhang, Y., Yu, L., Wang, L., Yu, Y.: nnformer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201 (2021)