Tooth Alignment via Virtual Trajectories with Clinical Constraints

SeungKwan Jo[1*], Taejin Moon[1*], So Hyun Kim[1], Hyensoo Jang[2], and Minyoung Chung[3**]

[1] Soongsil University, Seoul, Republic of Korea
[2] SNUAILAB, Seoul, Republic of Korea
[3] Osstem Implant, Seoul, Republic of Korea
chungmy.contact@gmail.com

Abstract. Tooth alignment is a fundamental procedure in orthodontic treatment, requiring prolonged stage-wise clinical evaluation to precisely manage complex tooth movements. Conventional automatic methods regress a direct rigid transformation in a one-shot manner and impose clinical constraints only on the final configuration, limiting the ability to reflect the progressive nature of actual orthodontic treatment and yielding unreliable outcomes in complex tooth movement cases. To address this limitation, we propose a framework that models a virtual progressive tooth movement trajectory in which clinical constraints can be evaluated at each stage. The framework decomposes alignment into K incremental rigid-transformation steps and generates a stable and continuous virtual tooth trajectory from pre-/post-treatment data alone. By enforcing clinical constraints at every intermediate stage, the framework models progressive tooth movements in orthodontic treatment, demonstrating more precise and reliable tooth alignment outcomes. Our code is publicly available at https://github.com/moontaijin/ProMoT-Tooth-Alignment.

Keywords: Intraoral scan · Orthodontics · Tooth alignment · Virtual tooth-movement trajectory · Clinical constraints

1 Introduction

Tooth alignment is a core procedure in orthodontic treatment, directly determining the accuracy of the final occlusion and the overall treatment outcome. However, clinical alignment is not merely a problem of predicting a target state. In practice, it involves managing complex tooth displacements over an extended period and performing repeated fine-grained adjustments, and the outcome can vary substantially depending on patient-specific biomechanical conditions in clinical workflows. Due to these characteristics, orthodontic treatment entails considerable time and cost burdens, and there has been a sustained demand for reliable automated tooth alignment technologies [1–5].

* These authors contributed equally.
** Corresponding author.

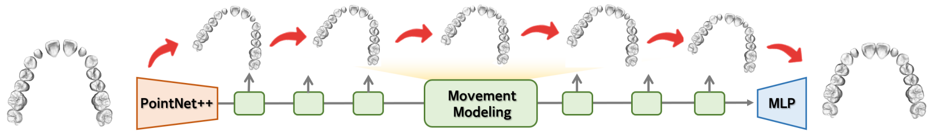

Fig. 1. Overview of our proposed framework. The model learns a virtual staged trajectory of progressive tooth movement from only pre-/post-treatment observations.

Most practical automated tooth alignment pipelines follow a one-shot regression paradigm: the task is reduced to estimating per-tooth 6-DoF rigid motions, and a transformation matrix to the final state is directly regressed in a single step [6–13]. However, this single-step formulation often fails to capture irregular and complex tooth displacements. In clinical practice, tooth alignment is refined through repeated stage-wise evaluations (i.e., *stage-wise clinical constraints*) and fine-grained adjustments rather than being finalized from the initial diagnosis alone [14–16]. Intermediate states are iteratively assessed under clinical criteria such as the occlusal plane, interproximal spacing, and collision risk, and the plan is updated accordingly [17, 18]. Therefore, precise tooth alignment is inherently a process-centric problem (i.e., *progressive tooth movement modeling*) in which inter-tooth interactions and clinical constraints are cumulatively reflected, requiring stage-wise evaluation and multiple rounds of adjustment.

Implementing such process-centric alignment in learning-based systems is challenging. Most public and real-world datasets provide only pre-/post-treatment pairs [6, 19], making it difficult to learn realistic movement trajectories and to supervise intermediate-stage assessments. Recent approaches augment the one-shot objective by adding clinically motivated terms to the final prediction [6–13]. Despite this, they still provide no learning signal for intermediate stages—no stage-wise evaluation, no micro-adjustment, and no accumulation of clinically valid refinements. This limitation can degrade reliability and consistency, particularly in cases that demand complex multi-tooth motions.

To address this issue, we propose a process-centric tooth alignment framework that models a virtual progressive tooth movement trajectory and performs clinical evaluation and step-wise refinement at each stage. The key idea is to enable the generation of stable and continuous virtual trajectories even from pre/post data, by regressing and accumulating stage-wise transformations using a state-space model (SSM)-based sequential predictor. In addition, to ensure physical plausibility, we introduce (i) a trajectory stabilization strategy that suppresses excessive increases in stage-wise displacement and prevents trajectory collapse, and (ii) an intermediate-state re-training mechanism that provides richer and more fine-grained supervision along the trajectory. It reconstructs the generated intermediate states and reuses them as new training inputs. As a result, virtual tooth trajectory modeling becomes feasible, and process-centric tooth alignment is achieved by reflecting the "stage-wise evaluation $\rightarrow$ micro-adjustment $\rightarrow$ accumulation" loop.

Our contributions are threefold:

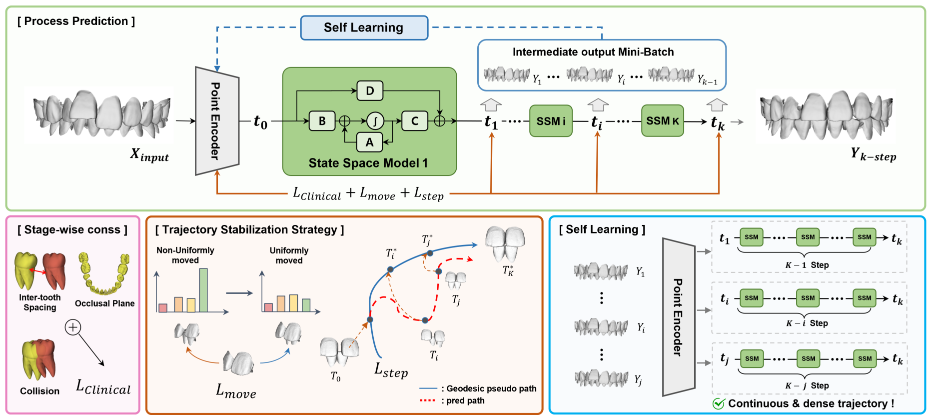

Fig. 2. Overview of progressive tooth-movement modeling with stage-wise clinical evaluation. Given the input tooth point clouds $X_{\text{input}} = \{X_i\}_{i=1}^{N}$, PointNet^{++} [20] encodes stage-wise features and an SSM-based sequential model predicts incremental SE(3) updates $\Delta\xi_i^{(k)}$, which are accumulated as $T_i^{(k)}$ with the final prediction $\hat{T}_i = T_i^{(K)}$ (Sec. 2.1). The trajectory stabilization strategy (TSS) and progressive self-learning support stable and dense intermediate trajectories (Secs. 2.2, 2.3). Clinical constraints are enforced at every stage to support progressive refinement (Sec. 2.4).

– We reformulate tooth alignment as a process-centric K-step incremental 6-DoF trajectory prediction problem, enabling intermediate-stage evaluation using only pre-/post-treatment data.
– We propose an SSM-based sequential regressor with a trajectory stabilization strategy to generate stable and continuous virtual movement trajectories.
– We introduce intermediate-state re-learning and stage-wise clinical evaluation to cumulatively refine incremental steps, consistent with clear-aligner micro-adjustments.

2 Method

Overview We learn a virtual staged trajectory of tooth movements that models orthodontic staged refinement using only pre-/post-treatment observations. Starting from the identity, the model predicts K incremental rigid updates in $SE(3)$ (K=21), composes them cumulatively, and enforces stage-wise clinical constraints at every intermediate stage. Stage-wise updates are generated by an SSM-based sequential model [21] for stable long-range dependency modeling across refinement stages.

2.1 Progressive Tooth Movement Modeling

Given a per-tooth point cloud $X_i \in \mathbb{R}^{P\times 3}$, we predict a rigid transformation $\hat{T}_i \in SE(3)$ for each tooth ($i = 1, \ldots, N;\ N$=32) by composing K=21 incremental twist updates $\Delta\xi_i^{(k)} \in se(3)$ via $\exp : se(3) \to SE(3)$, where K is selected with reference to reported stage settings in clear aligner-based orthodontic treatment [22]. Points are extended to homogeneous coordinates and transformed by left multiplication. We initialize $T_i^{(0)} = I$ and accumulate updates as

$$T_i^{(k+1)} = \exp\big(\Delta\xi_i^{(k)}\big) \cdot T_i^{(k)}, \qquad \hat{T}_i = T_i^{(K)}. \tag{1}$$

Stage-wise Encoding and Sequential Modeling At each stage k, the transformed point cloud $\tilde{X}_i^{(k)}$ is obtained by applying $T_i^{(k)}$ to X_i in homogeneous coordinates, and a PointNet^{++} [20] encoder $E(\cdot)$ extracts a per-tooth feature:

$$f_i^{(k)} = E\big(\tilde{X}_i^{(k)}\big). \tag{2}$$

An SSM-based sequential model [21] maintains a per-tooth hidden state $h_i^{(k)} \in \mathbb{R}^d$ to capture long-range dependencies and regresses the incremental twist:

$$h_i^{(k+1)} = \Psi\big(h_i^{(k)}, f_i^{(k)}\big), \qquad \Delta\xi_i^{(k)} = G\big(h_i^{(k+1)}\big), \tag{3}$$

where $\Psi(\cdot)$ is the SSM [21] transition and $G(\cdot)$ is an MLP regressing the twist $\Delta\xi = [\omega^\top, v^\top]^\top$. Stage-wise clinical constraints are enforced over the adjacent tooth-pair set $\mathcal{E}_{\text{adj}}$ (Section 2.4).

2.2 Trajectory Stabilization Strategy

Training a virtual trajectory from only pre-/post-treatment data can lead to trajectory collapse—manifesting as excessively large per-step movements or failure to converge toward the target. In real orthodontic treatment, all teeth move progressively across stages—magnitudes may vary, but no single tooth should dominate. We address both through per-stage movement distribution regularization and geodesic pseudo-target supervision.

Per-Stage Movement Distribution Regularization We normalize per-tooth movement magnitudes $m_i^{(k)} = \|\Delta\xi_i^{(k)}\|$ into a probability distribution via softmax and minimize its cross-entropy against a uniform target $u_i = 1/N$, encouraging equal contribution from all teeth at each stage:

$$L_{\text{move}} = \sum_{k=1}^{K} \left(-\frac{1}{N} \sum_{i=1}^{N} \log p_i^{(k)} \right), \qquad p_i^{(k)} = \frac{\exp(m_i^{(k)})}{\sum_{j=1}^{N} \exp(m_j^{(k)})}. \tag{4}$$

Geodesic Pseudo-Target Supervision Let $\log : SE(3) \to se(3)$ be the matrix logarithm with principal branch in $[0, \pi)$. Stage-wise pseudo-targets interpolate the $SE(3)$ geodesic from I to T_i^*:

$$T_i^{*(k)} = \exp\left(\tfrac{k}{K} \log T_i^*\right), \tag{5}$$

with supervision via the left-invariant error averaged over all teeth:

$$L_{\text{step}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\rho_k \left\|\log\!\left((T_i^{(k)})^{-1}T_i^{*(k)}\right)\right\|_1, \qquad L_{\text{TSS}} = L_{\text{move}} + L_{\text{step}}, \tag{6}$$

where ρ_k is set to 1 throughout training, and rotation/translation components are normalized to comparable scales. Both supervisions are applied only in the early training phase via a cosine schedule.

2.3 Progressive Self-learning from Intermediate States

Clinical tooth movement is effectively continuous [14–16], whereas our predictor models it as a discrete K-stage process. To embed denser movement dynamics, we reuse intermediate configurations as new starting states and retrain the model to predict the remaining motion with another K-step rollout. After the main rollout, intermediate transforms $\{T_i^{(k)}\}_{k=0}^{K}$ are recorded via a stop-gradient path:

$$\tilde{T}_i^{(k)} = \mathrm{sg}\left(T_i^{(k)}\right), \qquad \tilde{X}_i^{(k)} = \tilde{T}_i^{(k)} \cdot X_i^{\mathrm{h}}, \tag{7}$$

where X_i^{h} denotes X_i in homogeneous coordinates. Each restart defines the remaining motion as

$$T_{i,\text{rem}}^{(k)} = \left(\tilde{T}_i^{(k)}\right)^{-1} T_i^{*}. \tag{8}$$

The model is trained from $\tilde{X}_i^{(k)}$ to predict $T_{i,\text{rem}}^{(k)}$ using the same objective as the main pass, enforcing consistent refinement from a dense set of partially progressed states.

2.4 Stage-wise Clinical Constraints and Training

A key advantage of progressive tooth movement modeling is that it enables direct evaluation at intermediate stages, allowing clinical constraints—such as preventing inter-tooth collision, maintaining the occlusal plane, and avoiding excessive interproximal gaps—to be enforced at every intermediate stage. At each stage k, we compute a clinical energy on $\{\tilde{X}_i^{(k)}\}$ over the adjacent tooth-pair set $\mathcal{E}_{\text{adj}}$:

$$C_T^{(k)} = \lambda_{\text{inter}} L_{\text{inter}}^{(k)} + \lambda_{\text{occ}} L_{\text{occ}}^{(k)} + \lambda_{\text{spacing}} L_{\text{spacing}}^{(k)}, \tag{9}$$

where L_{inter} [11] penalizes inter-tooth collision and interpenetration; L_{occ} [6] penalizes deviation from the occlusal plane; and L_{spacing} [7] penalizes excessive interproximal gaps between adjacent teeth. $\mathcal{E}_{\text{adj}}$ consists of consecutive tooth pairs in both jaws and is fixed across all stages.

Training Objective The final-stage alignment loss combines the ADD point-distance metric and AAE geodesic rotation error, following [7]:

$$L^{(K)}_{\text{align}} = \frac{1}{N}\sum_{i=1}^{N}\Big[\lambda_{\text{ADD}}\, d_{\text{ADD}}(\hat{T}_i, T_i^*) + \lambda_{\text{AAE}}\, d_{\text{AAE}}(\hat{T}_i, T_i^*)\Big]. \tag{10}$$

The full training objective is then defined as

$$L = L^{(K)}_{\text{align}} + \lambda_{\text{TSS}} L_{\text{TSS}} + \lambda_{\text{clin}} \sum_{k=1}^{K} w_k\, C_T^{(k)}, \tag{11}$$

with $w_k = 1$ throughout training; λ_{TSS} and λ_{clin} control the contribution of each term.

3 Results

3.1 Experimental Setup and Comparisons

We conduct experiments on a public IOS dataset [6] of 837 paired pre-/post-treatment IOS samples with per-tooth segmentation, using a 670/167 train/test split and reporting mean ± std over 5-fold cross-validation. We train for 500 epochs (batch size 32) with N=512 sampled points and normalized inputs, optimized by Adam (lr 1.5×10^{-4}, weight decay 10^{-4}). Our sequential refinement uses an S4-style SSM [21] with rollout length K=21; intermediate-state re-training uses a micro-batch of three and is enabled from epoch 200, with TSS applied only in early epochs to provide an initial scaffold without constraining later fine-grained optimization. We evaluate using ADD/AUC, ADD (mm), AAE (degree), ME_{Rot} (degree), and ME_{Trans} (mm). We compare against TANet [7], TAlignNet [8], PSTN [9], and STTAlign [6]; these baselines perform one-shot transformation regression, whereas ours performs progressive movement modeling (ProMoT) with stage-wise constraints and TSS for intermediate corrections.

3.2 Tooth Alignment Performance

Table 1 summarizes quantitative results. Our method achieves the best overall performance, reaching an ADD/AUC of 0.863 ± 0.049 and an ADD of 0.794 ± 0.058 mm, with consistently low AAE, ME_{Rot}, and ME_{Trans}. Relative to the strongest baseline STTAlign [6], our approach improves ADD/AUC by 2.3% and reduces ADD by 3.2%. Notably, the gains are not limited to point-wise proximity: we also reduce AAE by 13.8% and ME_{Trans} by 4.7%, indicating improved arch-level consistency and motion accuracy under clinically relevant constraints.

Fig. 3 shows qualitative comparisons. One-shot methods often yield misalignment and inter-tooth collision in challenging cases. STTAlign [6] produces globally more coherent arrangements but leaves localized errors, consistent with the absence of intermediate constraint evaluation. In contrast, our method more closely matches the ground truth, maintaining a geometrically consistent arch form with minimal collision, qualitatively corroborating the effectiveness of stage-wise clinical constraint enforcement along a progressive movement trajectory.

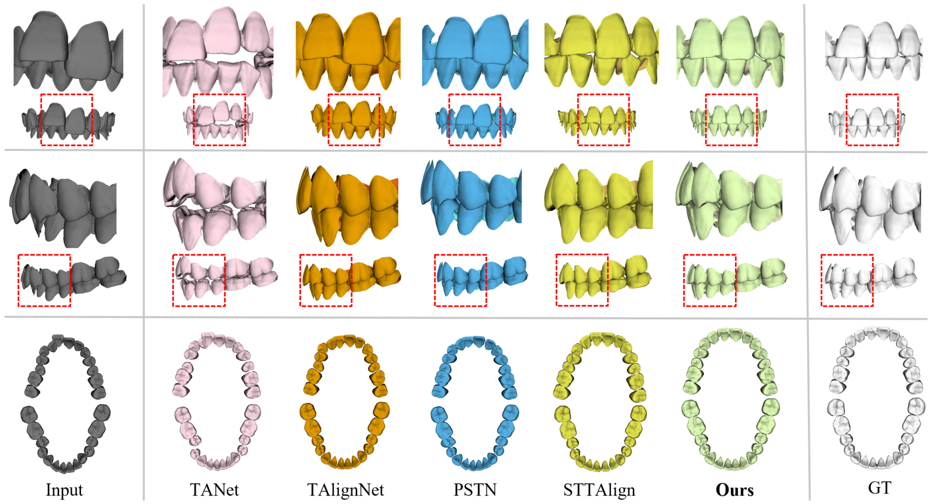

Fig. 3. Qualitative comparison of alignment results produced by different methods along with the ground truth. The red boxes indicate regions where common orthodontic issues typically occur (e.g., protrusion and inter-tooth spacing).

Table 1. Evaluation of tooth alignment performance (mean ± std). Best results are highlighted in **bold**.

Method	ADD/AUC ↑	ADD ↓	AAE ↓	ME_{Rot} ↓	ME_{Trans} ↓
TANet [7]	0.681±0.082	1.476±0.121	1.894±0.213	8.894±1.053	3.400±0.418
TAlignNet [8]	0.704±0.078	1.426±0.139	1.773±0.217	7.776±1.286	2.892±0.381
PSTN [9]	0.754±0.061	1.231±0.155	1.221±0.149	11.922±1.511	3.586±0.263
STTAlign [6]	0.844±0.064	0.820±0.103	1.136±0.117	3.136±0.730	1.836±0.169
Ours	**0.863±0.049**	**0.794±0.058**	**0.980±0.109**	**3.121±0.661**	**1.750±0.163**

Virtual Trajectory Visualization Fig. 4 visualizes virtual progressive trajectories generated by our framework on representative cases. In spacing cases, interproximal gaps are reduced gradually across stages while avoiding collision, reflecting direct regulation of tooth-to-tooth interactions by stage-wise constraints during refinement. In protrusion cases, the trajectories exhibit smooth curved paths rather than piecewise-linear displacements typical of one-shot or coarse updates, suggesting that TSS supports smoother intermediate dynamics. Intermediate-state re-training further supports denser trajectory refinement, resulting in more continuous trajectories that approximate smooth tooth movement.

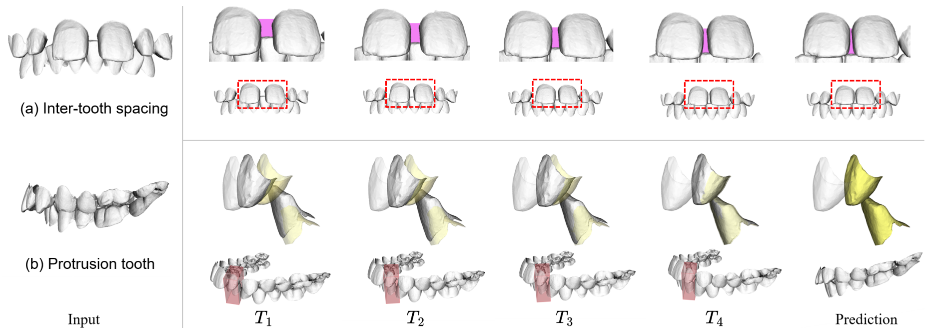

Fig. 4. Progressive tooth alignment over steps. As the step index increases, the dentition is gradually rearranged toward a plausible final configuration. In the upper case, teeth come into closer contact while the posterior gap is reduced. In the lower case, the teeth form a clear movement trajectory and are iteratively aligned across steps. Here, $T^{(k)}$ denotes the predicted tooth transformations at the k-th step produced by the SSM, where $k \in \{1, 2, 3, 4\}$ corresponds to steps $\{5, 10, 15, 20\}$, respectively.

Table 2. Ablation study with sequential component additions.

Method	**ADD/AUC** ↑	**ADD** ↓	**AAE** ↓	$\mathbf{ME}_{\text{Rot}}$ ↓	$\mathbf{ME}_{\text{Trans}}$ ↓
Baseline (PointNet^{++} [20])	0.6421	1.6206	2.1203	12.3007	3.9505
+ Progressive modeling	0.7650	1.0806	1.4208	5.2001	2.1504
+ Stage-wise constraints	0.8107	0.9501	1.2803	4.3002	1.9801
+ Trajectory module	0.8267	0.9141	1.1303	3.7662	1.9241
+ Self-learning	**0.8631**	**0.7938**	**0.9802**	**3.1209**	**1.7502**

3.3 Ablation Study

Table 2 validates the role of each component by incrementally extending a one-shot PointNet^{++} [20] baseline. The dominant improvement comes from progressive modeling: decomposing large displacements into bounded stage-wise SE(3) updates increases ADD/AUC by about 19% and reduces point-wise error by roughly one third, while cutting rotation error by more than half, indicating that step decomposition is crucial for nonlinear rotations. Adding stage-wise clinical constraints yields a further, consistent gain (about +6% in ADD/AUC) by injecting intermediate feedback that suppresses rotational overshoot and prevents collision/spacing violations from propagating to the final configuration. The TSS module primarily stabilizes early trajectory learning and provides a reliable scaffold for later refinement; its standalone effect is modest once progressive modeling and constraints are present, but it complements progressive self-learning. Finally, self-learning densifies intermediate states to produce smoother and more continuous refinement, delivering an additional ∼6% ADD/AUC gain and the lowest motion errors overall.

4 Conclusion

We address the core limitation of one-shot tooth alignment—its inability to reflect the progressive nature of clinical orthodontic treatment—and propose a framework that models a virtual progressive tooth movement trajectory from pre-/post-treatment data alone. By enforcing clinical constraints at every intermediate stage, the framework models progressive tooth movements in orthodontic treatment, demonstrating reliable and precise alignment even in complex tooth movement cases. However, the generated virtual trajectories remain limited by the absence of actual intermediate scans. Future work will focus on validating these trajectories with intermediate clinical data and extending the framework to recent diffusion- and Transformer-based models and more diverse orthodontic cases.

Acknowledgments. This work was supported by the Materials and Components Technology Development Program funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and Korea Evaluation Institute of Industrial Technology (KEIT) (Project Number: RS-2023-KT224280).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Gkantidis, N., Christou, P., Topouzelis, N.: The orthodontic–periodontic interrelationship in integrated treatment challenges: a systematic review. Journal of oral rehabilitation **37**(5), 377–390 (2010)
2. Dawson, P.E.: Functional occlusion: from TMJ to smile design. Elsevier Health Sciences (2006)
3. Grauer, D., Proffit, W.R.: Accuracy in tooth positioning with a fully customized lingual orthodontic appliance. American Journal of Orthodontics and Dentofacial Orthopedics **140**(3), 433–443 (2011)
4. Cheng, C., Cheng, X., Dai, N., Liu, Y., Fan, Q., Hou, Y., Jiang, X.: Personalized orthodontic accurate tooth arrangement system with complete teeth model. Journal of medical systems **39**(9), 84 (2015)
5. Kazimierczak, N., Kazimierczak, W., Serafin, Z., Nowicki, P., Nożewski, J., Janiszewska-Olszowska, J.: Ai in orthodontics: Revolutionizing diagnostics and treatment planning—a comprehensive review. Journal of clinical medicine **13**(2), 344 (2024)
6. Dong, Z., Chen, J.: Transformer-based tooth alignment prediction with occlusion and collision constraints. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 25145–25154 (2025)
7. 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)
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. Li, X., Bi, L., Kim, J., Li, T., Li, P., Tian, Y., Sheng, B., Feng, D.: Malocclusion treatment planning via pointnet based spatial transformation network. In: International conference on medical image computing and computer-assisted intervention. pp. 105–114. Springer (2020)
10. Deng, Q., Yang, X., Huang, M., Jiang, L., Zhang, D.: Taposenet: Teeth alignment based on pose estimation via multi-scale graph convolutional network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 314–323. Springer (2024)
11. 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)
12. Wang, C., Wei, G., Wei, G., Wang, W., Zhou, Y.: Tooth alignment network based on landmark constraints and hierarchical graph structure. IEEE Transactions on Visualization and Computer Graphics **30**(2), 1457–1469 (2022)
13. 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)
14. Jaber, S.T., Hajeer, M.Y., Burhan, A.S., Latifeh, Y.: The effect of treatment with clear aligners versus fixed appliances on oral health-related quality of life in patients with severe crowding: a one-year follow-up randomized controlled clinical trial. Cureus **14**(5) (2022)
15. D'Antò, V., Bucci, R., De Simone, V., Huanca Ghislanzoni, L., Michelotti, A., Rongo, R.: Evaluation of tooth movement accuracy with aligners: a prospective study. Materials **15**(7), 2646 (2022)
16. Kang, J., Jeon, H.H., Shahabuddin, N.: Does aligner refinement have the same efficiency in deep bite correction?: A retrospective study. BMC Oral Health **24**(1), 338 (2024)
17. Proffit, W.R., Fields, H., Larson, B., Sarver, D.M.: Contemporary orthodontics-e-book: contemporary orthodontics-E-Book. Elsevier Health Sciences (2018)
18. Fleming, P., Marinho, V., Johal, A.: Orthodontic measurements on digital study models compared with plaster models: a systematic review. Orthodontics & craniofacial research **14**(1), 1–16 (2011)
19. 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)
20. 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)
21. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
22. Haouili, N., Kravitz, N.D., Vaid, N.R., Ferguson, D.J., Makki, L.: Has invisalign improved? a prospective follow-up study on the efficacy of tooth movement with invisalign. American Journal of Orthodontics and Dentofacial Orthopedics **158**(3), 420–425 (2020)