Comput. Aided Geom. Des. 111 (2024) 102293

Contents lists available at ScienceDirect

Computer Aided Geometric Design

journal homepage: www.elsevier.com/locate/cagd

Automatic tooth arrangement with joint features of point and mesh representations via diffusion probabilistic models

Changsong Lei [a], Mengfei Xia [a], Shaofeng Wang [b], Yaqian Liang [a,*], Ran Yi [c], Yu-Hui Wen [d], Yong-Jin Liu [a,**]

[a] Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, Beijing, China
[b] Beijing Stomatological Hospital, Capital Medical University, Beijing, 100069, Beijing, China
[c] The School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, Shanghai, China
[d] Beijing Key Laboratory of Traffic Data Analysis and Mining, School of Computer Science and Technology, Beijing Jiaotong University, Beijing, 100044, Beijing, China

ARTICLE INFO

Keywords:
Automatic tooth arrangement
Diffusion probabilistic models
Transformation matrices prediction

ABSTRACT

Tooth arrangement is a crucial step in orthodontics treatment, in which aligning teeth could improve overall well-being, enhance facial aesthetics, and boost self-confidence. To improve the efficiency of tooth arrangement and minimize errors associated with unreasonable designs by inexperienced practitioners, some deep learning-based tooth arrangement methods have been proposed. Currently, most existing approaches employ MLPs to model the nonlinear relationship between tooth features and transformation matrices to achieve tooth arrangement automatically. However, the limited datasets (which to our knowledge, have not been made public) collected from clinical practice constrain the applicability of existing methods, making them inadequate for addressing **diverse** malocclusion issues. To address this challenge, we propose a general tooth arrangement neural network based on the diffusion probabilistic model. Conditioned on the features extracted from the dental model, the diffusion probabilistic model can learn the distribution of teeth transformation matrices from malocclusion to normal occlusion by gradually denoising from a random variable, thus more adeptly managing real orthodontic data. To take full advantage of effective features, we exploit both mesh and point cloud representations by designing different encoding networks to extract the tooth (local) and jaw (global) features, respectively. In addition to traditional metrics ADD, PA-ADD, CSA, and ME_{rot}, we propose **a new evaluation metric** based on dental arch curves to judge whether the generated teeth meet the individual normal occlusion. Experimental results demonstrate that our proposed method achieves state-of-the-art tooth alignment results and satisfactory occlusal relationships between dental arches. We will publish the code and dataset.

* Corresponding author.
** Principal corresponding author.
E-mail addresses: leics23@mails.tsinghua.edu.cn (C. Lei), xmf20@mails.tsinghua.edu.cn (M. Xia), 2939108747@ccmu.edu.cn (S. Wang), yaqianliang@tsinghua.edu.cn (Y. Liang), ranyi@sjtu.edu.cn (R. Yi), yhwen1@bjtu.edu.cn (Y.-H. Wen), liuyongjin@tsinghua.edu.cn (Y.-J. Liu).

https://doi.org/10.1016/j.cagd.2024.102293

Available online 22 April 2024

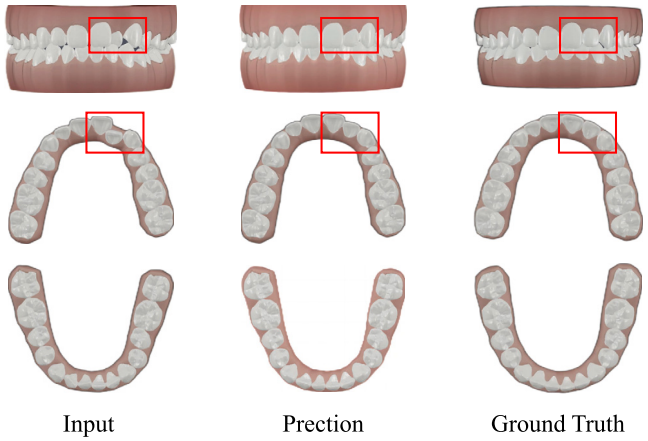

Fig. 1. Comparison between the input dental model, the predicted result of our method, and the ground truth. In this paper, we focus on not only the teeth alignment but also the occlusion between the upper and lower dental arches.

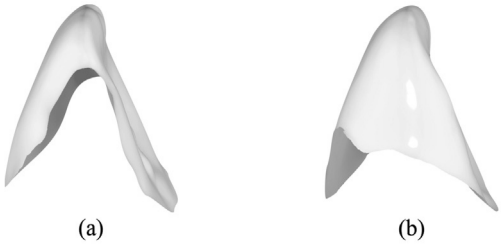

Fig. 2. A single tooth in our dataset, where (a) and (b) are the same tooth before and after orthodontic treatment observed from the same angle.

1. Introduction

Malocclusion refers to the irregular tooth arrangement, the abnormal relationship between the upper and lower dental arches, or the abnormal size/shape/position of jaws (Proffit et al., 2018). Currently, a significant number of people worldwide are plagued by malocclusion, with the probability of malocclusion ranges from 39% to 93% depending on region and ethnicity (Almotairy and Almutairi, 2022; Balachandran and Janakiram, 2021; Lin et al., 2020). Orthodontic treatment is the primary approach to address malocclusion. By applying specific forces in different directions with various types of orthodontic appliances, teeth can be moved towards predetermined positions to achieve an aesthetically pleasing, balanced, and stable dental arch and occlusal relationship (Gkantidis et al., 2010). Fig. 1 illustrates the effect of tooth arrangement.

Traditional orthodontic treatment methods require dentists and aligner design technicians to manually arrange teeth on digital dental models. Although current computer-aided design systems have simplified this task to some extent, the process still involves extensive and time-consuming human-computer interaction. With the growing integration of digital technology in orthodontic practices, digital tooth alignment simulations are becoming more prevalent in clinical settings (Ogodescu et al., 2010). These tools assist dentists in creating treatment plans by simulating teeth movements and predicting their final optimal positions. In recent years, the successful applications of artificial intelligence technology have markedly propelled the advancement of the medical field. Along with it, a variety of data-driven methods have been proposed in dentistry and orthodontics, such as tooth segmentation on oral CBCT data (Li et al., 2022; Cui et al., 2022) and teeth alignment prediction (Wei et al., 2020; Lingchen et al., 2020; Wang et al., 2022).

Existing tooth arrangement network models rely on the post-treatment tooth segmentations as the ground truth to calculate the loss functions, making tooth segmentation an essential step. However, due to the possibility of changes in crown shape during clinical treatment, we found that tooth segmentation results, whether from state-of-the-art intelligent methods or manually annotated segmentations, are not always completely consistent before and after orthodontic treatment. Fig. 2 demonstrates the segmented single tooth from a dental model in our dataset, where (a) and (b) are the same tooth before and after orthodontic treatment. Although Fig. 2(a) and Fig. 2(b) belong to the same tooth, there are significant differences between their shapes. This kind of inconsistency will affect the calculation of the loss functions, e.g., (1) the chamfer distance between the predicted teeth and ground truth teeth, and (2) the difference between the predicted teeth transformation matrix and the ground truth teeth transformation matrix. The aforementioned issue leads to the existing methods yielding inaccurate predictions of transformation matrices.

To tackle this challenge, we design our method with two steps: (1) We construct a simulated pre-orthodontic teeth dataset by randomly shifting the teeth after orthodontic treatment, thereby ensuring that each tooth segmentation is exactly the same *before* and *after* orthodontic treatment. At the same time, we note that due to large disparities between the distribution of the constructed pre-orthodontic dataset and the distribution of the real dataset, directly training existing methods on the constructed dataset still results in

poor generalization for real orthodontic cases. (2) To better cope with real orthodontic data, we further propose to use the diffusion model to learn the real distribution of teeth's transformation matrices when moving from an unreasonable (i.e., malocclusion) position to a reasonable (i.e., individual normal occlusion) position. Diffusion probabilistic models (DPMs) (Ho et al., 2020) have recently received growing attention due to their powerful representation learning ability and effectiveness in modeling complex data distributions. We utilize the features extracted from the dental model as the condition guiding the diffusion model, where the 3D mesh models are utilized to extract the local geometric details of a single tooth, and the point clouds are utilized to extract the global shape of the whole dental model. Compared with existing methods, our two-step framework demonstrates superior generalization to real patient data, as shown in Fig. 1.

To sum up, we make the following contributions:

- We propose TADPM, an automatic Tooth Arrangement neural network via Diffusion Probabilistic Model. By generating the teeth transformation matrices and applying them on the teeth models, our method predicts satisfactory orthodontic effects.
- To extract effective dental model features, we design different encoder networks at local and global levels.
- Integrating professional orthodontic knowledge, we propose a new metric based on dental arch curves to precisely evaluate the occlusal relationship between dental arches. We apply it to validate the alignment performance of our tooth arrangement network.

Experiments based on a real dataset collected from patients show that compared to existing methods, our method can significantly improve the teeth alignment results.

2. Related works

2.1. Learning-based teeth processing methods

The two most critical steps in orthodontic treatment are segmenting the 3D dental model and developing the arrangement plan. Manually performing these two tasks is time-consuming, tedious, and highly dependent on orthodontists' experiences due to the abnormality and large-scale variance of patients' teeth. In recent years, the successful applications of artificial intelligence technology have provided a strong impetus for advancements in various medical fields (Xu et al., 2018). For example, early works propose to use neural networks to conduct tooth segmentation on CBCT images (Wirtz et al., 2018; Cui et al., 2019; Shaheen et al., 2021; Cui et al., 2022; Xie et al., 2023; Li et al., 2022) or X-ray images (Zhao et al., 2020). To address problems caused by the irregular data format of 3D models, some works introduce Graph Convolution Network (GCN) to learn discriminative geometric features for 3D dental model segmentation (Sun et al., 2020; Zhang et al., 2021; Zheng et al., 2023). To alleviate the issue of insufficient annotations on 3D dental model data, Darch (Qiu et al., 2022) proposes a dental arch prior-assisted 3D tooth segmentation method when only weak annotation is provided, and Wu et al. (2022) propose a two-stage framework based on mesh deep learning (called TS-MDL) for joint tooth labeling and landmark identification on raw intraoral scans. To detect the axis direction of teeth, Fan et al. (2022) propose the rotation transformation encoding method based on the feature confidence-aware attention mechanism.

Tooth arrangement methods are usually developed based on segmentation results. TANet (Wei et al., 2020) is the first learning-based tooth arrangement predicting approach, which employs PointNet to extract the features of the crown point cloud and utilizes a graph neural network (GNN) to implement feature propagation between teeth through topological relations. Iorthopredictor (Lingchen et al., 2020) generates the facial image with aligned teeth, mimicking a real orthodontic treatment effect. Furthermore, OrthoGAN (Shen et al., 2022) focuses on patients' frontal facial images and progressively generates the visual results based on transfer learning GAN (Wang et al., 2024). Li et al. (2020) propose to conduct tooth arrangement by leveraging the spatial interrelationship between different teeth. Wang et al. (2022) propose a tooth arrangement network based on tooth landmark constraints and a hierarchical graph structure. All the aforementioned methods employ point clouds as input data. Nevertheless, the discrete points sampled from dental models lose geometric details and topological connections among teeth, thus potentially impacting subsequent transformation prediction processes. In contrast, we propose to study automatic tooth arrangement methods with 3D dental mesh models, which is unexplored. Experiments show that extracting information from mesh models with more geometric details yields better results.

2.2. Diffusion probabilistic models

Diffusion probabilistic models (DPMs) are a class of generative models that transform a Gaussian distribution into the distribution of the given data via an iterative denoising process (Sohl-Dickstein et al., 2015; Song and Ermon, 2019; Song et al., 2021; Ho et al., 2020). Suppose that $\mathbf{x}_0 \in \mathbb{R}^D$ is a D-dimensional random variable with an unknown distribution $q_0(\mathbf{x}_0)$. DPMs define a forward diffuse process by gradually corrupting the information of $\mathbf{x}_0$ with Gaussian noises, such that for any $t \in [0, T]$, the transition distribution is (Song et al., 2021; Sohl-Dickstein et al., 2015):

$$q_{0t}(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}), \tag{1}$$

where α_t and $\sigma_t > 0$ are differentiable functions of t with bounded derivatives.

To date, DPMs have been applied to various applications, including image generation (Sohl-Dickstein et al., 2015; Song et al., 2021; Ho et al., 2020), image editing (Avrahami et al., 2022; Choi et al., 2021), 3D model generation (Luo and Hu, 2021; Lyu et

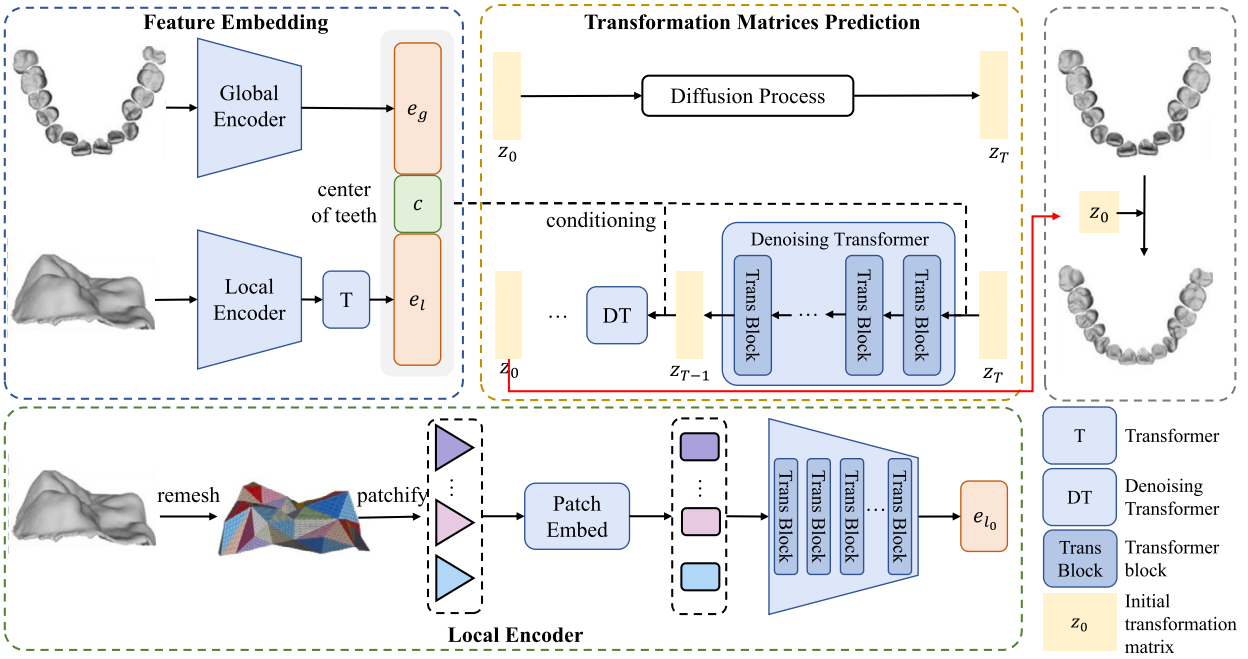

Fig. 3. The overall architecture of the proposed framework. In the feature embedding module, we utilize MeshMAE (Liang et al., 2022) as the local encoder network, where the faces of the single tooth mesh are divided into patches and embedded to compute the feature embedding e_{l_0}, then e_{l_0} is fed into a transformer network to produce the local feature e_l. We utilize PointNet++ as the global encoder network, where the point cloud of the whole jaw is utilized to extract the global feature e_g. In the transformation matrices prediction module, we build upon diffusion models and design the denoising transformer to predict the transformation matrices conditioned on the extracted local and global features. The lower part of the figure presents the local encoder, including the data pre-processing and feature embedding process.

al., 2023; Liu et al., 2023; Vahdat et al., 2022; Nichol et al., 2022), and conditional generation (Choi et al., 2021), demonstrating significant performance in the generative domain. Considering the *diverse* categories of dental malocclusion in clinical settings and the difficulty of data collection, we model the automatic tooth arrangement task as *a generative task*. We introduce DPMs to generate tooth transformation matrices, which are expected to learn the orthodontic data distribution, thereby increasing the generalization of the network and enabling more effective handling of real clinical data.

3. Methods

3.1. Overview

The tooth arrangement task aims at estimating a 6-dimensional transformation parameter for each tooth in the input dental models. As shown in Fig. 3, we propose TADPM to solve this task, comprising two key modules: a Feature Embedding module and a Transformation Matrices Prediction module. Specifically, we denote the input data, a set of segmented dental mesh models (without gingiva), as $M = \{M_k | k \in \mathcal{K}\}$, where $\mathcal{K}$ is the set of teeth labels according to the FDI two-digit notation (Harris et al., 2005), and $M_k = (V_k, F_k)$ denotes the 3D mesh model of a single tooth with label k, where V_k and F_k are the collections of vertices and faces of M_k, respectively. First, the Feature Embedding module uses a local encoder and a global encoder to separately extract the local and global features from the input dental models. Then, the Transformation Matrices Prediction module generates the pose transformation matrix for each tooth conditioned on the feature embeddings with the Denoising Transformer blocks. We describe these two modules in detail as follows.

3.2. Feature embedding module

Existing tooth arrangement methods (Wei et al., 2020; Li et al., 2020; Lingchen et al., 2020; Wang et al., 2022) only use point clouds as input to extract features. Compared with point clouds, 3D mesh models contain more geometric details (especially the topological connectivity) and can provide fine geometric features that are helpful for subsequent transformation prediction; however, mesh processing requires more time and resources. To achieve effective and efficient feature extraction of dental models, we propose to use both 3D meshes and point clouds for feature extraction. This method involves extracting local features from 3D meshes to accurately represent the shape of individual teeth, and global features from point clouds to conserve computational time and memory.

3.2.1. Local feature

In this paper, we conduct tooth arrangement on dental models in the 3D mesh representation. Compared with point clouds, mesh models contain rich connectivity-related information, which inspires us to extract powerful features directly from the mesh models. However, due to the intrinsic irregular nature of the data format of mesh models and complex adjacency between vertices, it is difficult to directly use neural networks to process 3D meshes. Moreover, due to the difficulty of annotating labels for dental models, the number of training samples is actually insufficient for performing supervised learning on dental meshes. To address these limitations, we introduce MeshMAE (Liang et al., 2022) to extract the feature embedding of each tooth, so that the self-supervised pre-training process in MeshMAE would benefit the downstream tooth arrangement tasks.

First, following MeshMAE, we perform the re-mesh operation on a single tooth mesh to construct a hierarchical structure. During the process, the faces F_k in mesh M_k are divided into several non-overlapping patches. Then, we concatenate the features of all faces in a patch into the patch feature according to the order from the re-meshing process. The patch feature is then projected into the patch embedding by a multi-layer perceptron (MLP). Furthermore, we integrate positional information by computing the positional embedding based on the center coordinates of the patches, which is more appropriate for unordered geometric data. Finally, the patch embedding and positional embedding are combined to form the input embedding.

To address the challenge posed by limited data, we adopt the self-supervised pre-training strategy of MeshMAE to learn an effective feature representation from the re-meshed dental datasets. Specifically, a portion of the input embeddings are randomly masked, and only the remaining unmasked embeddings are fed into the encoder for further processing. Then, the masked embeddings are replaced with a shared mask embedding, which are eventually combined with the encoder's output to produce the input of the decoder. Finally, the decoder predicts the face features and vertex coordinates of the masked patches to reconstruct the information of the masked parts. After the pre-training stage, we utilize the encoder of MeshMAE as the local encoder and finetune[1] it to compute the feature embedding $e_{l_{k_0}}$ of the tooth M_k.

3.2.2. Feature propagation

The local features extracted by MeshMAE capture the tooth-level geometric details. However, they lack the interdependent information shared among adjacent teeth. To address this issue, we introduce a Transformer network to transfer geometric information among all teeth. The output sequence $\{e_{l_{k_0}} | k \in \mathcal{K}\}$ obtained from the local encoder is fed into the Transformer, and the geometry center c_k of M_k is utilized as the positional embedding. By leveraging the attention mechanism, the resulting output $\{e_{l_k} | k \in \mathcal{K}\}$ is able to capture inter-teeth mutual information across all teeth. This mechanism enables the model to effectively capture and integrate the context of geometry relationships among all teeth.

3.2.3. Global feature

Instead of using mesh as in the local feature extraction, we propose to utilize point clouds to extract the global feature of the whole dental model. The reasons are two-fold: 1) on the one hand, using MeshMAE to extract global features is memory-consuming since there are too many faces in the whole jaw; 2) on the other hand, local features already contain sufficient fine geometric details, which makes it redundant to re-extract geometric details for global feature extraction. Therefore, we leverage point clouds for the extraction of global features from the entire dental model.

Specifically, we define the input point cloud as $P = \{p_k \in \mathbb{R}^{N_k \times 3} | k \in \mathcal{K}\}$, where p_k denotes the point cloud sampled from the single tooth mesh M_k and N_k denotes the number of points in p_k. Then, we use PointNet++ (Qi et al., 2017) as the global feature extraction module to obtain a global feature e_g from P. Although features extracted by PointNet++ do not contain very fine geometry details, they nevertheless provide sufficient global information for tooth arrangement.

Finally, the local and global features and the center of teeth are integrated through Stack and Concatenation operations to form the overall feature representation of the dental model:

$$e = \mathrm{Stack}\{\mathrm{Concat}(e_g, c_k, e_{l_k}) | k \in \mathcal{K}\}, \tag{2}$$

where Stack and Concat denote corresponding operations in PyTorch, respectively.

3.3. Transformation matrices prediction module

After feature extraction, the tooth arrangement task can be regarded as a generation task conditioned on the input feature e. We adopt a representative diffusion model (Ho et al., 2020) by elaborately reformulating our task as a *conditioned* DPM model to generate a 6-DoF transformation matrix of each tooth.

During the training stage, each tooth is assigned with a 6-DoF ground truth[2] transformation parameter $z_0^k = (m_0^k, r_0^k)$, where $m_0^k \in \mathbb{R}^3$ and $r_0^k \in \mathbb{R}^3$ respectively denote the translation and rotation vectors of the single tooth M_k in the coordinate system. By concatenating all $\{z_0^k | k \in \mathcal{K}\}$, we construct the transformation matrix $z_0 \in \mathbb{R}^{|\mathcal{K}| \times 6}$, which serves as input for the diffusion model, as illustrated in Fig. 3. Then, we introduce perturbations to z_0 to generate a noisy matrix z_t at timestep t:

$$z_t = \alpha_t z_0 + \sigma_t \epsilon, \tag{3}$$

[1] See the training process in Algorithm 1.

[2] Note that we construct the training dataset by shifting dental models after orthodontics, so we can obtain the ground truth transformation matrices directly.

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, α_t and σ_t are as defined in Eq. (1).

Subsequently, we employ the U-ViT (Bao et al., 2023) as the denoising network, denoted as $z_\theta(z_t, t, e)$ (where θ represents the parameters of the network), to predict the original z_0 from the noisy z_t with the feature e extracted from the feature embedding module as the condition. U-ViT is a simple and general ViT-based architecture for the generation with diffusion models, which treats all inputs including the time, condition, and noisy image patches as tokens and employs long skip connections between shallow and deep layers. The network's actual output, $\bar{z}_0 = \{\bar{z}_0^k | k \in \mathcal{K}\} \in \mathbb{R}^{|\mathcal{K}| \times 6}$, is a sequence of 6-DoF transformation parameters. Each $\bar{z}_0^k = (m_0^k, r_0^k) \in \mathbb{R}^6$ defines the transformation parameters of the single tooth mesh M_k, where m_0^k and r_0^k represent the translation and rotation parameters of M_k, respectively. We can obtain a transformation matrix $T_k \in \mathbb{R}^{4\times4}$ from z_0^k through SE(3) exponential mapping, defined as follows:

$$\Omega = \begin{bmatrix} 0 & -r_k^z & r_k^y \\ r_k^z & 0 & -r_k^x \\ -r_k^y & r_k^x & 0 \end{bmatrix}, \tag{4}$$

$$R_k = exp(\Omega) = I_{3\times3} + \frac{sin\phi}{\phi}\Omega + \frac{1 - cos\phi}{\phi^2}\Omega^2, \tag{5}$$

$$T_k = \begin{bmatrix} R_k & m_0^k \\ \mathbf{0} & 1 \end{bmatrix}, \tag{6}$$

where $\phi = \|r_0^k\|$ represents the rotation angle of M_k.

Finally, we apply the transformation matrix T_k on the original dental model M to get an aligned dental model $\bar{M}$. For convenience, we represent the whole aligning process as follows:

$$\bar{M} = aligner(M, \bar{z}_0). \tag{7}$$

Algorithm 1 Training stage of TADPM

Input: Features e, Ground truth of transformation matrix z_0, dental model M before orthodontic treatment and ground truth M^* after treatment.

1: **repeat**
2: $t \sim \text{Uniform}[1, \cdots, T]$
3: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $z_t \leftarrow \alpha_t z_0 + \sigma_t \epsilon$
5: $\bar{z}_0 \leftarrow z_\theta(z_t, t, e)$
6: $\bar{M} \leftarrow aligner(M, \bar{z}_0)$
7: Take gradient descent step on

$$\nabla_\theta[\lambda_1 \mathcal{L}_{CD} + \lambda_2 \mathcal{L}_{diff} + \lambda_3 \mathcal{L}_{pos}]$$

8: **until** converge

The training process is summarized in Algorithm 1, where step 7 refers to the loss function that we describe in Sec. 3.4. In the sampling stage, we first generate a random transformation matrix $z_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and gradually denoise it to predict the desired transformation parameters of each tooth. We further adopt the determinative sampling method introduced in (Song et al., 2020) to accelerate the sampling process:

$$z_{t-\Delta t} = \sqrt{\bar{\alpha}_{t-\Delta t}}\bar{z}_0 + \sqrt{1 - \bar{\alpha}_{t-\Delta t}} \frac{z_t - \sqrt{\bar{\alpha}_t}\bar{z}_0}{\sqrt{1 - \bar{\alpha}_t}}, \tag{8}$$

where Δt denotes the step size. Algorithm 2 summarizes the whole sampling process.

Algorithm 2 Sampling stage of TADPM

Input: Features e.
Output: Prediction of transformation matrix z_0.

1: $z_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, T - \Delta t, \cdots, \Delta t$ **do**
3: $\bar{z}_0 \leftarrow z_\theta(z_t, t, e)$
4: $z_{t-\Delta t} \leftarrow \sqrt{\bar{\alpha}_{t-\Delta t}}\bar{z}_0 + \sqrt{1 - \bar{\alpha}_{t-\Delta t}} \frac{z_t - \sqrt{\bar{\alpha}_t}\bar{z}_0}{\sqrt{1 - \bar{\alpha}_t}}$
5: **end for**
6: **return** z_0

3.4. Loss functions

3.4.1. Reconstruction loss

Chamfer distance is a widely used metric for point cloud reconstruction tasks. We first convert the predicted result $\bar{M}$ and ground truth M^* to point clouds by simply utilizing their vertex sets $\bar{V}$ and V^*. Then, we compute the chamfer distance between corresponding teeth in the prediction result $\bar{M}$ and the ground truth M^*:

$$\mathcal{L}_{CD} = \sum_{k \in \mathcal{K}} \left(\frac{1}{|\bar{V}_k|} \sum_{x \in \bar{V}_k} \min_{y \in V_k^*} \|x - y\|^2 + \frac{1}{|V_k^*|} \sum_{y \in V_k^*} \min_{x \in \bar{V}_k} \|x - y\|^2 \right), \tag{9}$$

where $\|\|$ denotes L2 norm and $\bar{V}_k$ denotes vertices of the predicted tooth with label k and V_k^* denotes vertices of the ground truth with label k. $\mathcal{K}$ is the set of labels for all teeth.

3.4.2. Diffusion loss

We use the predicted transformation matrix output by diffusion model to supervise the training directly. Generally speaking, the squared error between prediction and uncorrupted data is a natural choice for the loss function of diffusion models:

$$\mathcal{L}_{diff} = \|z_0 - z_\theta(z_t, t, e)\|^2, \tag{10}$$

where z_0 and $z_\theta(z_t, t, e)$ denote the ground truth and the predicted transformation matrix, respectively.

3.4.3. Relative position loss

Relative positional relationship among teeth constrains the space between adjacent teeth and is critical to tooth arrangement. We define a distance matrix D to represent the positional structure of a dental model M:

$$D(M)_{ij} := \|c_i - c_j\|_1, \tag{11}$$

where $\|\|_1$ denotes L1 norm, $D(M)_{ij}$ denotes the distance between the i-th and j-th teeth, and c_i and c_j denote the geometry center of the i-th and j-th tooth, respectively. Compared with the aforementioned chamfer distance, calculating our designed distance matrix requires less time.

We then calculate the difference between the distance matrices of the prediction result and the ground truth, providing supervision for the relative positional relationship:

$$\mathcal{L}_{pos} = \|D(\bar{M}) - D(M^*)\|^2, \tag{12}$$

where $D(\bar{M})$ and $D(M^*)$ denote the distance matrices of the prediction result $\bar{M}$ and the ground truth M^*, respectively.

$\mathcal{L}_{pos}$ constrains distances between corresponding teeth in the upper and lower jaws as well as distances between adjacent teeth, thereby achieving a better occlusal relationship and tooth arrangement.

Finally, we combine the above three losses together to obtain the final loss function:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{CD} + \lambda_2 \mathcal{L}_{diff} + \lambda_3 \mathcal{L}_{pos}, \tag{13}$$

where λ_1, λ_2, and λ_3 denote weights for each part of the loss function. Due to the magnitude of $\mathcal{L}_{CD}$ is much larger than those of $\mathcal{L}_{diff}$ and $\mathcal{L}_{pos}$, we use a small weight to constrain $\mathcal{L}_{CD}$ and maximize the influence of each loss function, where λ_1, λ_2, and λ_3 are set as 0.05, 0.5, and 1, respectively.

4. Experimental results

Since there is no public dataset for tooth arrangement available, we evaluated the proposed TADPM on our newly built dataset collected from clinical patients. We compared TADPM with several baselines qualitatively and quantitatively. We also conducted ablation studies to demonstrate the effectiveness of each component in TADPM.

4.1. Experimental setups

4.1.1. Datasets

Our dataset[3] consists of 212 pairs of dental models before and after orthodontic treatment, which are collected from patients treated between June 2016 and April 2023. We invited four orthodontists with over ten years of clinical experience to annotate the tooth segmentation labels and tooth position numbering. The doctors were trained to use the 'Mesh Labeler' software and delineated the boundaries of the teeth according to the FDI international tooth numbering system.

In our experiment, we observed that some of the tooth segmentation labels are different before and after orthodontic treatment in our dataset (see Fig. 2 for an example), which would have a negative impact on the training process. To address this problem, we

[3] The amount of dental models in our dataset is comparable to that of the datasets in existing methods, e.g., the dataset in (Li et al., 2020) has 178 pairs of dental models.

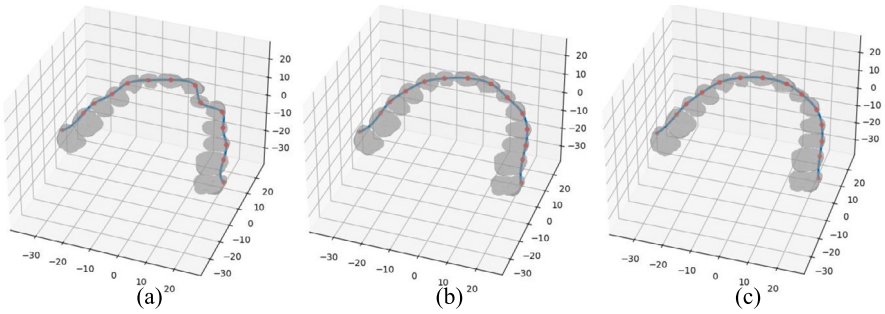

Fig. 4. Visualization of the dental arch curves, where (a), (b), and (c) show the arch curve of input data, prediction, and ground truth, respectively. The coordinate unit is mm. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

added random disturbance to the dental models after orthodontics to simulate "pre-orthodontic" dental models. For each patient, we randomly created 10 pairs, which will expand the dataset size by 10 times (the actual amount of our input data is around 2120 pairs of dental models). The constructed data pairs are used for *training*, while in *testing*, we mix constructed and real data pairs to see whether our model can handle different types of malocclusion. Additionally, the dental models in our dataset are usually distributed far from the origin, leading to challenges in feature extraction. To mitigate this issue, we relocated the geometric centers of all dental models to the origin.

4.1.2. Evaluation metrics

Following (Wei et al., 2020; Li et al., 2020; Wang et al., 2022), we adopt the metrics of ADD (Hinterstoisser et al., 2013), PA-ADD, cosine similarity accuracy (CSA), and ME_{rot} to measure the performance of our method. ADD is the point-wise mean distance between the predicted dental models and ground truth, which provides an intuitive measure of alignment error. PA-ADD is the ADD metric calculated after rigid registration from the prediction to the ground truth. Since the process of tooth arrangement can be described by predicting the transformation parameters of each single tooth, we can calculate the CSA of predicted matrices and the ground truth. ME_{rot} denotes the mean error of rotation.

To measure the teeth alignment effect and occlusion relationship between dental arches, we introduce the Fréchet Distance (Eiter and Mannila, 1994) between dental arch curves of prediction and ground truth as a metric, denoted as FD_{cur}. We leverage the B-spline curve to represent the dental arch curve, which is obtained by interpolating the landmarks (red points in the figure) on the teeth, as shown in Fig. 4.

4.1.3. Implementation details

We train the proposed TADPM on the platform of PyTorch (Paszke et al., 2019), in a Linux environment with 2 NVIDIA Tesla A100 GPUs. In all our experiments, we set the batch size to 8 and the epochs to 500. We use the AdamW optimizer with a learning rate of 1e-4. Because the shape of input z_0 has a great aspect ratio ($|\mathcal{K}| \times 6$), we utilize the transformer as the backbone instead of the U-Net, for the convolutional operations may not perform well on the input data with unbalanced dimension size. Moreover, the previous studies (Peebles and Xie, 2023) show that replacing U-Net backbone with transformer backbone can achieve better results in some tasks. In feature propagation model, we use Vision Transformer (ViT) (Dosovitskiy et al., 2021) to learn mutual information among teeth. In our diffusion model, we use U-ViT as our backbone since it is a common choice for the diffusion model. We use 12 transformer blocks for the pertaining and also for the diffusion model.

The number of parameters in our model is 173M, and the computational complexity is approximately 6,000 GFlops. During the inference stage, we adopt the DDIM sampler for acceleration. The average inference time of a single dental model is 38.72 s, which is acceptable for clinical usage.

4.2. Comparisons with state-of-the-arts

To demonstrate the effectiveness of TADPM, we compare it with three state-of-the-art methods, i.e. TANet (Wei et al., 2020), PSTN (Li et al., 2020), and TAligNet (Lingchen et al., 2020). The experimental results are summarized in Table 1, where our method achieves the best performance on all four metrics. Notably, our method shows significant advantages on ADD and FD_{cur}, demonstrating that our method can not only obtain good tooth alignment effectiveness but also improve the occlusal relationship of the dental arches.

In addition, we show the distribution of mean pointwise distance (the average of distances between each pair of corresponding vertices in predicted tooth mesh and ground truth) in Fig. 5, where the x-axis represents the range of mean pointwise distance and the y-axis represents the accuracy within this range. Compared with baseline methods, the results of our method have a wider distribution within a smaller range, indicating that the tooth arrangement effect of our method is closer to the ground truth.

Table 1
Comparison with state-of-the-art methods. ↓ indicates the lower the better, while ↑ indicates the higher the better. The coordinate unit is degree for ME_{rot}, mm for ADD, PA-ADD and FD_{cur}.

Model	ADD (↓)	PA-ADD (↓)	CSA (↑)	ME_{rot} (↓)	FD_{cur} (↓)
TANet	1.618	1.325	0.813	5.564	2.320
PSTN	1.784	1.440	0.781	6.073	2.552
TAligNet	1.752	1.418	0.785	5.893	2.665
TADPM (Ours)	**1.487**	**1.202**	**0.847**	**5.472**	**1.932**

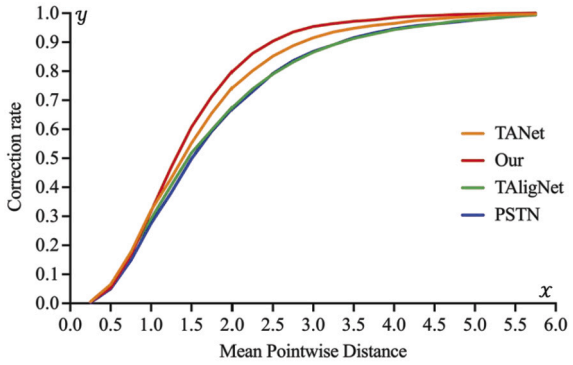

Fig. 5. The distribution of pointwise distance, where the x-axis represents the range of the mean pointwise distance, and the y-axis represents the accuracy within the range.

4.3. Qualitative evaluation

Fig. 6 illustrates some examples of the aligned dental models obtained by our method, which can generate satisfactory dental alignment effects for different kinds of malocclusions, demonstrating the generalization of our method. In detail, our method focuses not only on tooth alignment issues such as teeth crowding or dental diastema, but also on dental occlusion problems such as deep overbite or deep overjet. Tailored to distinct problems, our method ensures even and smooth teeth arrangements.

4.4. Subjective evaluations

In this subsection, we provide subjective evaluations of dental alignment effectiveness from professional clinical orthodontists. The responses were collected from 5 orthodontists, who were asked to rate 10 randomly selected samples by answering three questions for each sample:

- Question I: Does the alignment result of the upper and lower jaw meet the clinical standards?
- Question II: Does the occlusal relationship between the upper and lower jaw meet the clinical standards?
- Question III: Does the result meet your expectation for the post-treatment effect of the patient?

For each question, orthodontists are required to rate between 1 and 5 scores, with 5 representing the best. We list the average and variance of rates for each sample in Table 2. The average rates for three questions are all over 4 (Good), indicating that the results of TADPM can meet clinical standards.

4.5. Ablation study

To further analyze the effectiveness of the components and loss functions of the proposed method, we conduct ablation studies on the proposed modules and loss functions.

4.5.1. Effectiveness of proposed modules

As described in Sec. 3, we adopt the MeshMAE as the local encoder and introduce the diffusion probabilistic module to predict the transformation matrix. To verify the effectiveness of the above two strategies, we conduct ablation studies by replacing the MeshMAE and the diffusion probabilistic module with PointNet++ and MLP, respectively. The experimental results are shown in columns 2 ("w/o MeshMAE") and 5 ("w/o DPM") in Table 3. To promote the feature extraction effect, we introduce the global encoder and the feature propagation module into our model. Here, we conduct the ablation experiments by removing these two parts to verify whether they are necessary for feature extraction. The experimental results are shown in columns 3 ("w/o Global") and 4 ("w/o FP") in Table 3.

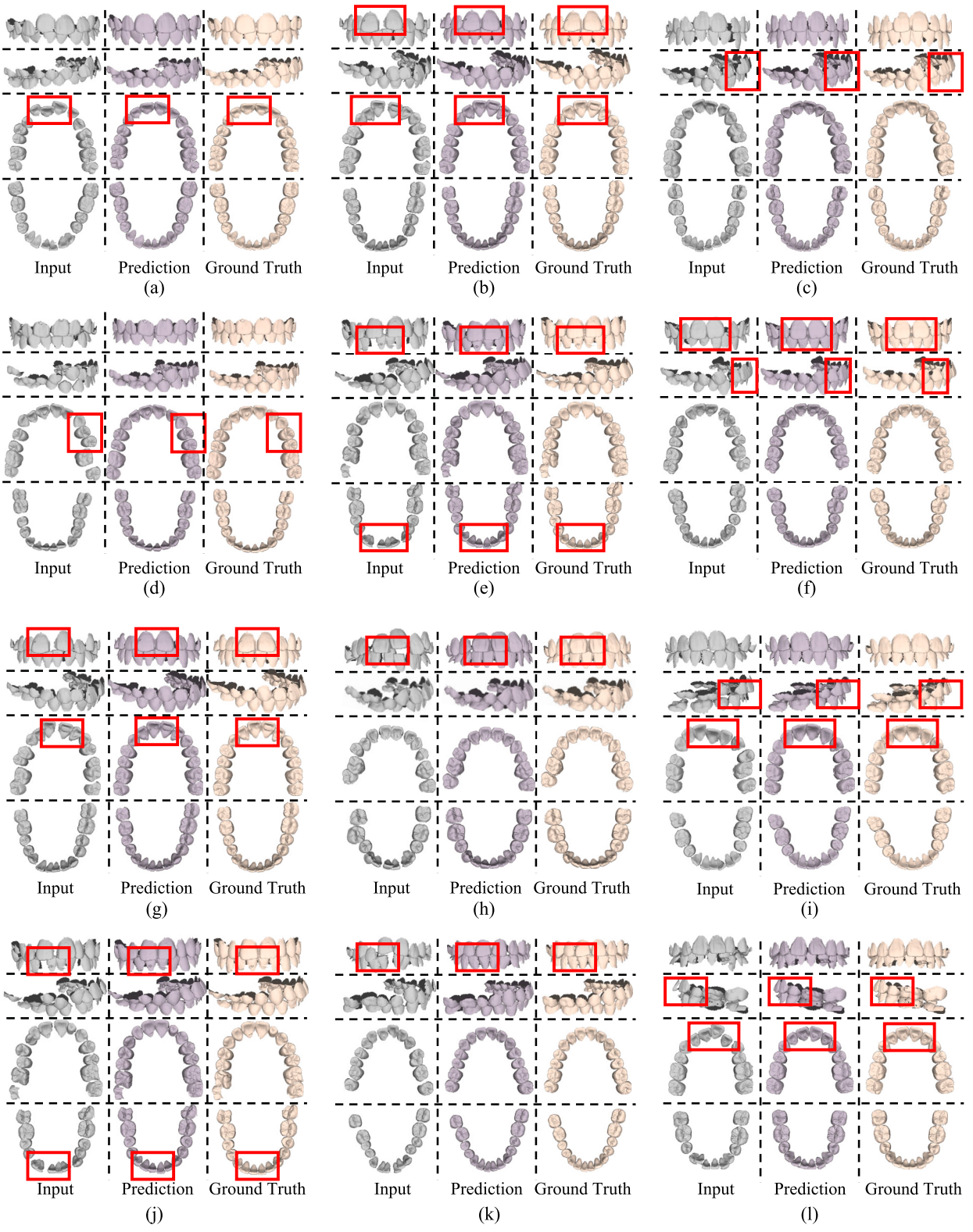

Fig. 6. Visualization of the orthodontic results of our method, where (a) and (d) belong to the malocclusion problem of teeth crowding, (b) and (e) belong to the dental diastema, (c) and (f) belong to the deep overbite, (g) and (j) belong to the large dental gap, (h) and (k) belong to the abnormal tooth occlusion, (i) and (l) belong to the deep overjet.

Table 2
Orthodontists' rates for 10 randomly selected samples.

Model	Question I		Question II		Question III	
	Average	Variance	Average	Variance	Average	Variance
Sample I	4.20	0.16	4.80	0.16	4.40	0.24
Sample II	3.80	0.56	4.20	0.16	4.00	0.40
Sample III	4.20	0.16	4.60	0.24	4.20	0.16
Sample IV	3.80	0.56	4.80	0.16	4.40	0.64
Sample V	4.40	0.24	4.60	0.24	4.60	0.24
Sample VI	4.00	0.00	4.40	0.24	4.20	0.16
Sample VII	4.00	0.00	4.40	0.24	4.40	0.24
Sample VIII	4.40	0.24	4.80	0.16	4.20	0.16
Sample IX	4.20	0.16	4.60	0.24	4.20	0.16
Sample X	3.60	0.24	4.40	0.64	4.20	0.16
Average	4.06	0.23	4.56	0.25	4.28	0.26

Table 3
The effectiveness of different modules. "Local" denotes the local encoder module, "Global" denotes the global encoder module, "FP" denotes the feature propagation module, "DPM" denotes the diffusion probabilistic module, and the last column is the result of our method.

Modules	w/o MeshMAE	w/o Global	w/o FP	w/o DPM	Ours
Local	PointNet++	✓	✓	✓	✓
Global	✓		✓	✓	✓
FP	✓	✓		✓	✓
DPM	✓	✓	✓	MLP	✓
ADD (↓)	1.637	1.793	1.703	1.536	**1.487**
PA-ADD (↓)	1.336	1.507	1.430	1.248	**1.202**
CSA (↑)	0.813	0.770	0.792	0.836	**0.847**
ME_{rot} (↓)	5.652	5.951	5.695	5.492	**5.472**
FD_{cur} (↓)	2.638	3.073	2.607	2.152	**1.932**

Table 4
The performance of different loss functions, where the last row is the result of our method.

$\mathcal{L}$	ADD (↓)	PA-ADD (↓)	CSA (↑)	ME_{rot} (↓)	FD_{cur} (↓)
$\mathcal{L}_{CD}$	1.587	1.319	0.829	5.504	2.034
$\mathcal{L}_{diff}$	1.709	1.531	0.797	5.879	2.891
$\mathcal{L}_{CD} + \mathcal{L}_{diff}$	1.502	1.229	0.832	5.478	**1.879**
$\mathcal{L}_{CD} + \mathcal{L}_{diff} + \mathcal{L}_{pos}$	**1.487**	**1.202**	**0.847**	**5.472**	1.932

The results of our full method are shown in the last column of Table 3. From these results, the appropriate feature embedding modules can significantly improve the performance. Unlike other methods that utilize MLP to directly regress the transformation matrices, we introduce the diffusion model to learn the distribution of transformation matrices. This strategy further enhances the effectiveness of our method on all metrics, especially on FD_{cur}, demonstrating that the introduction of the diffusion model can indeed bring a good occlusal relationship.

4.5.2. Ablation study on loss functions

The loss functions in our full method include three parts, i.e., reconstruction loss $\mathcal{L}_{CD}$, diffusion loss $\mathcal{L}_{diff}$, and relative position loss $\mathcal{L}_{pos}$. We conduct ablation studies to verify the effectiveness of each part. The results are shown in Table 4. We can observe that it is difficult to achieve good results using only $\mathcal{L}_{diff}$ to guide model learning due to the lack of geometric information, while the combination of $\mathcal{L}_{diff}$ and $\mathcal{L}_{CD}$ can boost the performance. To constrain the distance between adjacent teeth, we also add $\mathcal{L}_{pos}$ to the overall loss function, which contributes to achieving better results. Although involving $\mathcal{L}_{pos}$ in the overall loss function may slightly decrease FD_{cur}, it is still deserved. First, the other four metrics are all improved. Second, FD_{cur} is utilized to measure the similarity between the dental arch curves of prediction and ground truth, which can be accepted as long as it meets clinical principles and individual ideal occlusion, without the fixed evaluation indicators. Furthermore, compared with other methods, our method still gets the highest FD_{cur}, indicating the effectiveness of the proposed method.

4.6. Iterative experiment

The output of our method is a set of teeth, which is in the same format as the input data. Therefore, a natural assumption is that the network can achieve better tooth alignment than the current output by reusing the output as input. Based on this assumption,

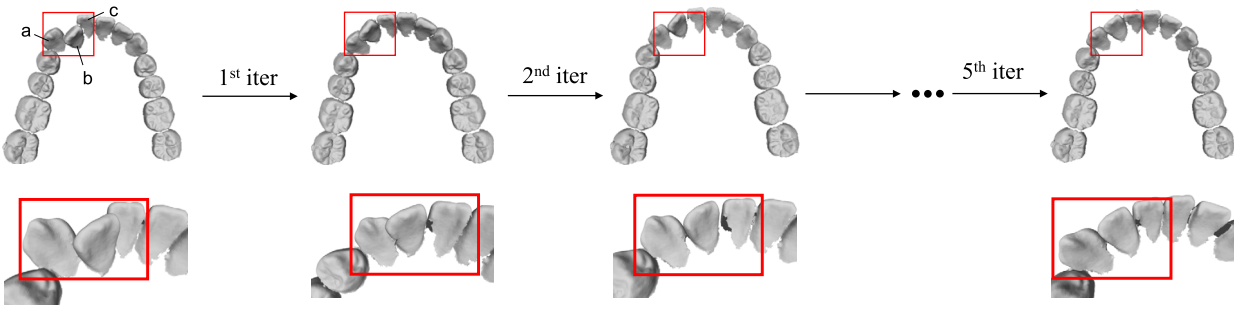

Fig. 7. An example of the iterative experiment result. The second row shows the details of the red boxes in the first row.

we explore the performance of iterative experiments, i.e., using the output of our network as input and conducting teeth alignment iteratively. Fig. 7 shows an example of the iterative experiment. We can observe from the original input that the space between teeth *a* and *c* is too small to let tooth *b* fit in. After the first iteration, *a* and *b* are still touching. In order to squeeze out some space, most teeth need to be moved. Therefore, the network requires more subsequent iterations to gradually adjust the whole dental model towards a global optimization.

Iterative teeth alignment is helpful in extreme cases like Fig. 7, where the network fails to align all teeth well at once. But in most cases, our model can already achieve good performance in one round.

5. Conclusion

In this paper, we present a novel automatic tooth arrangement network to generate tooth transformation matrices for orthodontics treatment. To address the challenges posed by complex real-patient dental data, our method constructs the "pre-orthodontic" teeth dataset by randomly moving the teeth after orthodontic treatment and leverages the diffusion probabilistic models to generate the transformation parameters for each tooth. Additionally, we extract local geometric details of the tooth based on mesh models, which is verified to contribute to the effectiveness of feature embedding extraction in our experiments. Extensive experiments and user studies demonstrate the effectiveness of our method. In the future, we would combine our method with Deepcache (Ma et al., 2023) and Faster Diffusion (Li et al., 2023), which will further improve the inference speed.

CRediT authorship contribution statement

Changsong Lei: Writing – review & editing, Writing – original draft, Methodology. **Mengfei Xia:** Conceptualization, Methodology. **Shaofeng Wang:** Resources, Conceptualization. **Yaqian Liang:** Writing – review & editing, Writing – original draft, Methodology. **Ran Yi:** Writing – original draft, Conceptualization. **Yu-Hui Wen:** Writing – original draft, Conceptualization. **Yong-Jin Liu:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by Beijing Natural Science Foundation (No. L222008), Beijing Hospitals Authority Clinical Medicine Development of Special Funding Support (No. ZLRK202330), National Natural Science Foundation of China (No. 62302297, No. 62202257), Shanghai Sailing Program (No. 22YF1420300), Young Elite Scientists Sponsorship Program by CAST (No. 2022QNRC001), and Talent Fund of Beijing Jiaotong University (No. 2023XKRC045).

References

Almotairy, N., Almutairi, F., 2022. A nation-wide prevalence of malocclusion traits in Saudi Arabia: a systematic review. J. Int. Soc. Prev. Commun. Dent. 12, 1–11.
Avrahami, O., Lischinski, D., Fried, O., 2022. Blended diffusion for text-driven editing of natural images. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 18208–18218.
Balachandran, P., Janakiram, C., 2021. Prevalence of malocclusion among 8–15 years old children, India–a systematic review and meta-analysis. J. Oral Biol. Craniofacial Res. 11, 192–199.

Bao, F., Nie, S., Xue, K., Cao, Y., Li, C., Su, H., Zhu, J., 2023. All are worth words: a vit backbone for diffusion models. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 22669–22679.
Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S., 2021. Ilvr: conditioning method for denoising diffusion probabilistic models. In: 2021 IEEE, Int. Conf. Comput. Vis., vol. 1, p. 2.
Cui, Z., Li, C., Wang, W., 2019. Toothnet: automatic tooth instance segmentation and identification from cone beam ct images. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 6368–6377.
Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., et al., 2022. A fully automatic ai system for tooth and alveolar bone segmentation from cone-beam ct images. Nat. Commun., 2096.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: transformers for image recognition at scale. In: Int. Conf. Learn. Represent., pp. 1–12.
Eiter, T., Mannila, H., 1994. Computing Discrete Fréchet Distance. Technical Report. Christian Doppler Laboratory for Expert System, TU, Vienna, Austria.
Fan, Y., Ma, Q., Wei, G., Cui, Z., Zhou, Y., Wang, W., 2022. Tad-net: tooth axis detection network based on rotation transformation encoding. Graph. Models 121, 101138.
Gkantidis, N., Christou, P., Topouzelis, N., 2010. The orthodontic–periodontic interrelationship in integrated treatment challenges: a systematic review. J. Oral Rehabil., 377–390.
Harris, E.F., et al., 2005. Tooth-coding systems in the clinical dental setting. Dent. Anthropol. J. 18, 43–49.
Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N., 2013. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In: Asian Conf. Comput. Vis. Springer, pp. 548–562.
Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. In: Adv. Neural Inform. Process. Syst., pp. 6840–6851.
Li, P., Liu, Y., Cui, Z., Yang, F., Zhao, Y., Lian, C., Gao, C., 2022. Semantic graph attention with explicit anatomical association modeling for tooth segmentation from cbct images. IEEE Trans. Med. Imaging, 3116–3127.
Li, S., Hu, T., Khan, F.S., Li, L., Yang, S., Wang, Y., Cheng, M.-M., Yang, J., 2023. Faster diffusion: rethinking the role of unet encoder in diffusion models. arXiv preprint. arXiv:2312.09608.
Li, X., Bi, L., Kim, J., Li, T., Li, P., Tian, Y., Sheng, B., Feng, D., 2020. Malocclusion treatment planning via pointnet based spatial transformation network. In: Medical Image Computing and Computer Assisted Intervention, pp. 105–114.
Liang, Y., Zhao, S., Yu, B., Zhang, J., He, F., 2022. Meshmae: masked autoencoders for 3d mesh data analysis. In: Eur. Conf. Comput. Vis. Springer, pp. 37–54.
Lin, M., Xie, C., Yang, H., Wu, C., Ren, A., 2020. Prevalence of malocclusion in Chinese schoolchildren from 1991 to 2018: a systematic review and meta-analysis. Int. J. Paed. Dent. 30, 144–155.
Lingchen, Y., Zefeng, S., Yiqian, W., Xiang, L., Kun, Z., Hongbo, F., Zheng, Y., 2020. iorthopredictor: model-guided deep prediction of teeth alignment. ACM Trans. Graph. 39, 216.
Liu, Z., Feng, Y., Black, M.J., Nowrouzezahrai, D., Paull, L., Liu, W., 2023. Meshdiffusion: score-based generative 3d mesh modeling. In: Int. Conf. Learn. Represent., pp. 1–12.
Luo, S., Hu, W., 2021. Diffusion probabilistic models for 3d point cloud generation. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 2837–2845.
Lyu, Z., Wang, J., An, Y., Zhang, Y., Lin, D., Dai, B., 2023. Controllable mesh generation through sparse latent point diffusion models. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 271–280.
Ma, X., Fang, G., Wang, X., 2023. Deepcache: accelerating diffusion models for free. arXiv preprint. arXiv:2312.00858.
Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., Chen, M., 2022. Point-e: a system for generating 3d point clouds from complex prompts. arXiv preprint. arXiv:2212.08751.
Ogodescu, A.S., Sinescu, C., Ogodescu, E.A., Negrutiu, M., Rominu, R., Bratu, E., 2010. Computer science in the orthodontic treatment of adult patients. In: Proceedings of the European Conference of Systems, and European Conference of Circuits Technology and Devices, and European Conference of Communications, and European Conference on Computer Science, pp. 15–18.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.
Peebles, W., Xie, S., 2023. Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4195–4205.
Proffit, W.R., Fields, H.W., Larson, B., Sarver, D.M., 2018. Contemporary Orthodontics-e-Book. Elsevier Health Sciences.
Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In: Adv. Neural Inform. Process. Syst., pp. 5105–5114.
Qiu, L., Ye, C., Chen, P., Liu, Y., Han, X., Cui, S., 2022. Darch: dental arch prior-assisted 3d tooth instance segmentation with weak annotations. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 20752–20761.
Shaheen, E., Leite, A., Alqahtani, K.A., Smolders, A., Van Gerven, A., Willems, H., Jacobs, R., 2021. A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. a validation study. J. Dent., 103865.
Shen, F., Liu, J., Li, H., Fang, B., Ma, C., Hao, J., Feng, Y., Zheng, Y., 2022. Orthogan: high-precision image generation for teeth orthodontic visualization. arXiv preprint. arXiv:2212.14162.
Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: Int. Conf. Mach. Learn., pp. 2256–2265.
Song, J., Meng, C., Ermon, S., 2020. Denoising diffusion implicit models. In: Int. Conf. Learn. Represent., pp. 1–12.
Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. In: Adv. Neural Inform. Process. Syst., vol. 32, pp. 1–13.
Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2021. Score-based generative modeling through stochastic differential equations. In: Int. Conf. Learn. Represent., pp. 1–12.
Sun, D., Pei, Y., Li, P., Song, G., Guo, Y., Zha, H., Xu, T., 2020. Automatic tooth segmentation and dense correspondence of 3d dental model. In: Medical Image Computing and Computer Assisted Intervention, pp. 703–712.
Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K., et al., 2022. Lion: latent point diffusion models for 3d shape generation. Adv. Neural Inf. Process. Syst. 35, 10021–10039.
Wang, C., Wei, G., Wei, G., Wang, W., Zhou, Y., 2022. Tooth alignment network based on landmark constraints and hierarchical graph structure. IEEE Trans. Vis. Comput. Graph., 1457–1469.
Wang, Y., Gonzalez-Garcia, A., Wu, C., Herranz, L., Khan, F.S., Jui, S., Yang, J., van de Weijer, J., 2024. Minegan++: mining generative models for efficient knowledge transfer to limited data domains. Int. J. Comput. Vis. 132, 490–514.
Wei, G., Cui, Z., Liu, Y., Chen, N., Chen, R., Li, G., Wang, W., 2020. Tanet: towards fully automatic tooth arrangement. In: Eur. Conf. Comput. Vis., pp. 481–497.
Wirtz, A., Mirashi, S.G., Wesarg, S., 2018. Automatic teeth segmentation in panoramic x-ray images using a coupled shape model in combination with a neural network. In: Medical Image Computing and Computer Assisted Intervention, pp. 712–719.
Wu, T.-H., Lian, C., Lee, S., Pastewait, M., Piers, C., Liu, J., Wang, F., Wang, L., Chiu, C.-Y., Wang, W., et al., 2022. Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans. IEEE Trans. Med. Imaging, 3158–3166.
Xie, R., Yang, Y., Chen, Z., 2023. Wits: weakly-supervised individual tooth segmentation model trained on box-level labels. Pattern Recognit., 108974.
Xu, X., Liu, C., Zheng, Y., 2018. 3d tooth segmentation and labeling using deep convolutional neural networks. IEEE Trans. Vis. Comput. Graph., 2336–2348.

Zhang, L., Zhao, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D., 2021. Tsgcnet: discriminative geometric feature learning with two-stream graph convolutional network for 3d dental model segmentation. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 6699–6708.
Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., Meng, D., 2020. Tsasnet: tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network. Knowl.-Based Syst., 106338.
Zheng, Y., Chen, B., Shen, Y., Shen, K., 2023. Teethgnn: semantic 3d teeth segmentation with graph neural networks. IEEE Trans. Vis. Comput. Graph. 29, 3158–3168.