

Diff-OSGN: Diffusion-based Occlusal Surface Generation Network with Geometric Constraints

Chen Wang

Shandong University

1500 ShunHua Road, Jinan, China

chen.wang@mail.sdu.edu.cn

Zhiming Cui

ShanghaiTech University

393 Middle Huaxia Road, Shanghai, China

cuizhm@shanghaitech.edu.cn

Guangshun Wei

Shandong University

1500 ShunHua Road, Jinan, China

guangshunwei@gmail.com

Shuyi Lu

Shandong University

1500 ShunHua Road, Jinan, China

lushuyi1sy@gmail.com

Yuanfeng Zhou*

Shandong University

1500 ShunHua Road, Jinan, China

yfzhou@sdu.edu.cn

James Kit Hon Tsoi

the University of Hong Kong

34 Hospital Road, Sai Ying Pun, Hong Kong

jkhtsoi@hku.hk

Zhenpeng Liu

Shandong University

1500 ShunHua Road, Jinan, China

zhenpengliu@mail.sdu.edu.cn

Abstract

Designing a functional occlusal surface for denture crowns is a complex and vital task in prosthodontics. Manual design is time-consuming and heavily relies on the dentist’s experience, as it requires careful consideration of occlusal function. Due to the limitations of manual design, the field has turned to data-driven methods for occlusal surface design. However, many of these methods neglect critical geometric details, such as normals and curvature, impacting the quality of the occlusal surface. In this paper, we introduce Diff-OSGN, a novel denture crown occlusal surface generation network based on a denoising diffusion model, which focuses on generating the detailed geometric structure of denture crowns. We model the occlusal surface as a geometry map based on the occlusal plane, incorporating height and normal maps rasterized from intra-oral scanning crown. Both maps represent occlusal surface geometry, and their combination further enhances these details. Considering the crucial occlusal information, we extract features from the geometry maps of adjacent and occlusal teeth, using them as conditions in the reverse diffusion process to train our network on optimal occlusal function. Additionally, we define three geometric operators and corresponding loss functions as the constraints to better extract geometric features of the target occlusal surface, such as ridges and grooves, for adequate supervision. Our results demonstrate that Diff-OSGN provides quantitatively and qualitatively su-

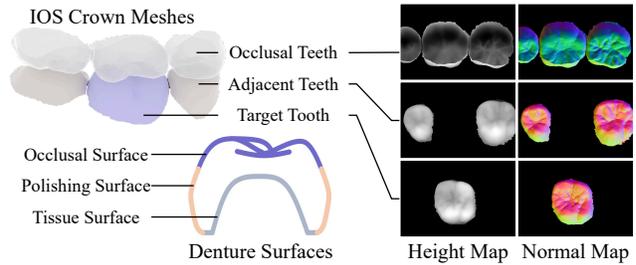


Figure 1. Visualization results of the teeth intra-oral scanning meshes, occlusal surface, and geometry maps. Three rows on the right are the geometry maps rasterized from the occlusal teeth, adjacent teeth and target tooth meshes, showing the height map and normal map separately.

perior performance compared to competing baselines and state-of-the-art methods.

Keywords: *Functional Occlusal Surface Designing, Denoising Diffusion Model, Prosthodontics, Medical imaging.*

1. Introduction

Odonto defect, dentition defect, tooth deformation and edentulous are common and frequently occurring diseases in humans. The main causes of these diseases are caries, periodontal disease, trauma, tumors, and congenital malformations [19]. As the human living environment changes, the incidence rate of oral diseases is increasing, resulting in an increasing demand for oral restorations.

Prosthodontics is a clinical medical science that utilizes artificial devices to restore missing or damaged teeth and maintain their corresponding physiological functions. With the advancement of digital technology, many computer-aided geometric design (CAD) methods [15, 23] and softwares have emerged in the field of dental restoration. This advancement enables dentists to digitally design denture crowns for dental prostheses that meet occlusal function requirements through the analysis of intra-oral scanning (IOS) data. In digital prosthodontics, designing denture crowns is a crucial step, often requiring a large amount of time from dentists or technicians. Technicians must balance occlusal function with aesthetic features, making the process heavily reliant on the dentist’s expertise. As shown in Fig. 1, denture design typically involves three surfaces: the occlusal surface, the polishing surface, and the tissue surface. We focus on the occlusal surface, which is the contact area between the upper and lower teeth and presents significant complexity due to its intricate occlusal functionality requirements.

With the development of artificial intelligence methods, many data-driven occlusal surface design techniques have emerged in the field of prosthodontics. These methods can standardize the design process, improve efficiency, and reduce human errors. Additionally, they can leverage large datasets for learning and optimization, resulting in denture crowns that better meet patient needs. For example, the method proposed by Hwang et al. [12] was the first to utilize a Generative Adversarial Network (GAN) architecture to generate depth maps of occlusal surfaces. Subsequent studies [25, 26] have adopted similar GAN-based approaches, improving the network’s functionality. However, the single-channel depth maps generated by these methods often lack significant geometric details. Accurate representation of ridges and grooves on the occlusal surface is essential for dental restoration, and single-channel maps fail to capture this intricate geometry. Therefore, it is imperative to incorporate more comprehensive geometric information. We propose the concept of teeth geometry map based on the occlusal surface generation method. As shown in Fig. 1, the geometry map contains the height map, which provides precise spatial information and details of the grooves, and the normal map, which includes curvature information. Both of them encompass extensive geometric features to effectively represent the tooth model’s geometry.

Denosing Diffusion Probabilistic Models (DDPM) [11] are a novel class of generative models that can generate structured, high-quality outputs through a sophisticated reverse diffusion process. Compared to GANs, diffusion models typically generate higher-quality images, especially when handling high-resolution data. This is because diffusion models use a multi-step denosing process that more accurately recovers image details, while also ensuring a

more stable training process. This capability makes them ideal for denture crown generation, where the detail is paramount. These models’ ability to handle complex geometries and textures aligns perfectly with the intricate designs of denture crowns, reducing the reliance on individual dentist expertise and ensuring consistent quality across different cases. Additionally, diffusion models can adapt to varied inputs, allowing for the customization of crowns to fit unique dental structures, making them a cutting-edge solution in modern dentistry.

Therefore, we present a diffusion-based geometry map generation network for occlusal surface design. By utilizing the geometry map and extracting the occlusal information of adjacent and occlusal teeth, the diffusion model provides a more detailed and accurate geometric representation of the denture crown, capturing essential features that single-channel depth maps may miss. We propose three kinds of geometric operators and corresponding loss functions to supervise the generation of denture crown geometry maps. The experimental results demonstrate significant improvements in denture crown generation compared to previous methods. To summarize, our contributions are as follows:

- We define and utilize the geometry map of tooth crown data, providing a detailed geometric representation of the occlusal surface. This approach enables more comprehensive modeling of denture structures and improves the understanding of tooth geometry.
- We propose three geometric operators and their corresponding geometric constraints, designed to capture intricate details within the occlusal surface, such as ridges and grooves. This approach effectively represents the essential features for dental restoration, going beyond the capabilities of previous methods.
- We present a novel diffusion-based occlusal surface generation network structure. By extracting features of adjacent and occlusal geometry maps, we capture essential occlusal information and conditioning the diffusion model. By incorporating the geometry map of teeth as constraints, we enhance the precision and realism of the generated occlusal surfaces in prosthesis restoration and outperform previous methods.

2. Related Work

2.1. Digital Dental and Denture Crown Design

With the popularization of computers and the development of CAD methods and software [37], more and more methods have emerged in the interdisciplinary field of medicine and software engineering, with dentistry also receiving widespread attention. With the continuous progress of deep learning, many deep learning methods have also emerged in dentistry [6], which can use accumulated data

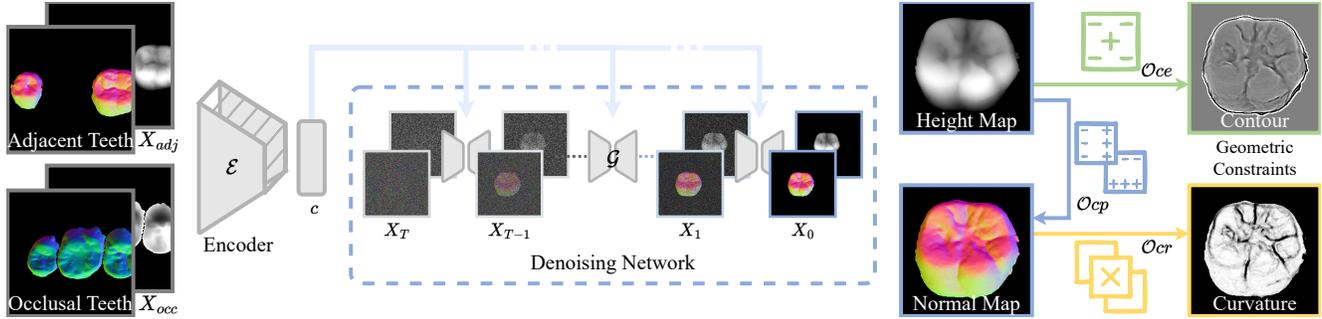


Figure 2. Overview of our pipeline. We first create the local coordinate system by using the IOS crown mesh data, and then rasterize the geometry maps of adjacent teeth and occlusal teeth. We use an encoder module \mathcal{E} to extract the features of two geometry maps as the condition c . Through the reverse diffusion process and denoising module \mathcal{G} , we iteratively fix the noise from Gaussian noise map X_T and predict the geometry map X_0 for the target teeth. We present three geometric operators and their corresponding geometric constraints, derived from the operators and representing detailed features such as contours and curvatures, to effectively leverage geometry information during the training process.

for network training and significantly improve the efficiency of dentists. Among them, there are many methods for dental data analysis, such as cone beam computed tomography (CBCT) segmentation [4], IOS model segmentation [40], tooth anatomical feature detection [7, 31], and teeth model parameterization [1], etc., which can further improve the digitalization process of dentistry. In orthodontics, there are also methods for predicting tooth alignment targets [8, 28, 30] and reconstructing dental crown models from photographs. The detection and classification of dental diseases are also involved [9].

In prosthodontics, there are also related CAD methods [15] and discussions on deep learning methods [2]. Among them, [12] is the first article to use conditional GAN for dental crown design, which only uses simple regression and discriminant loss. Many GAN networks based on depth maps obtained from orthogonal projection [24–26, 35] and 3D-DCGAN based on crown voxelization [5] also emerged.

2.2. Diffusion Model in Medical Image

DDPM [11], a class of generative models, have made significant strides in various deep learning applications [16]. Unlike other generative networks, such as GAN [3, 13] and Variational AutoEncoder (VAE) [18, 27], a diffusion probabilistic model defines a forward diffusion stage where the input data is gradually perturbed over several steps by adding Gaussian noise and then learns to reverse the diffusion process to retrieve the desired noise-free data from noisy data samples [11]. Despite the computational burden, diffusion models are widely applied for their robust pattern coverage and quality of generated samples in medical image processing, such as image-to-image translation [20], image classification [33] and segmentation [33]. Medical anomaly detection is also an important task in computer vision. In dentistry, [9] proposes a novel method for detecting abnormal teeth in panoramic X-rays using a hierarchical multi-

label approach.

Image generation is one of the primary objectives of diffusion models. [17] recently proposed DDM, which takes source and target images and generates intermediate temporal frames along the continuous trajectory. [14] propose the first diffusion-based multi-modality MRI synthesis model, CoLa-Diff, to address the high memory demands and low practicality in multi-modal synthesis.

3. Method

An overview of our method is illustrated in Fig. 2. Our goal is to synthesize the target occlusal surface given the adjacent and occlusal teeth of the target tooth. We represent the adjacent and occlusal teeth as geometry maps and train an encoder to extract their features, which serve as the conditions c for the denoising diffusion model. We then train a denoising network to generate the geometry map of the target occlusal surface through a reverse diffusion process. The generated geometry maps are supervised by reconstruction losses, and we apply three kinds of operators to the generated maps and corresponding geometric constraints to ensure a more accurate and satisfactory occlusal surface.

3.1. Definition and Rasterization of Geometry Map

Towards our goal, we model the occlusal surfaces as the geometry maps $\mathbf{X} = \{\mathbf{n}, \mathbf{h}\}$, where $\mathbf{n} \in (\mathbb{S}^2)^{H \times W}$ represents the normal map and $\mathbf{h} \in \mathbb{R}^{H \times W}$ for height map. Both the input and output geometry maps adhere to the real data distribution. For visualization purposes, all visual effects presented in the paper are converted into the color space after being cropped by the mask, as illustrated in the right part of Fig. 1.

To infer the occlusal plane and rasterize geometry maps, we establish a local coordinate system based on the relative positions of the target tooth, two adjacent teeth, three

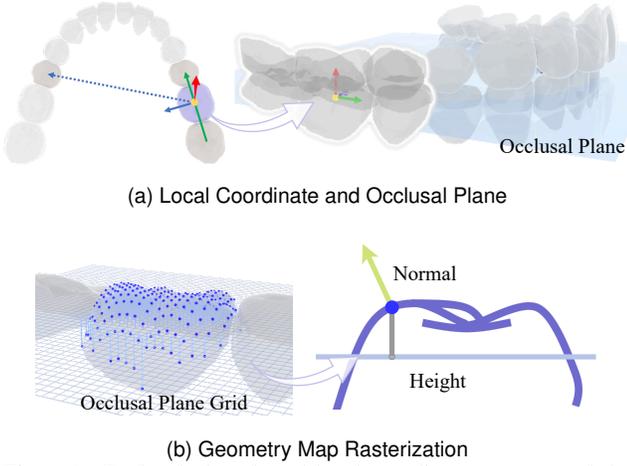


Figure 3. Occlusal plane based local coordinate system, and the geometry map rasterization from IOS meshes. (a) According to the segmentation of meshes, we find the adjacent teeth and occlusal teeth of our target denture crown, then build the coordinate system. The green arrow represent the \vec{x} -axis, the red arrow represent the \vec{z} -axis. The dashed blue arrow represents the temporary \vec{y} -axis, while the solid one represents the \vec{y} -axis. And the blue plane is the inferred occlusal plane. (b) We rasterize the occlusal surface, which represented by the dark blue lines, from occlusal plane grid and get the geometry map, which containing the height value and the normal vector.

occlusal teeth, and one tooth in symmetric position of dentition, using their 3D IOS meshes. Specifically, as shown in Fig. 3, we take the midpoint of two adjacent teeth as the local coordinate origin, and the mesial to distal direction of adjacent teeth as the \vec{x} -axis. We define the direction from the coordinate origin to the symmetrical tooth as the temporary \vec{y} -axis. Then we cross product the \vec{x} -axis and temporary \vec{y} -axis to get the \vec{z} -axis, and then cross product the \vec{x} -axis and the \vec{z} -axis to get the true \vec{y} -axis, thus establishing the local coordinate system. The resulting XOY plane is defined as occlusal plane. We rasterize the crown meshes through the occlusal plane grid, where the height value sampled from each pixel is the z-value perpendicular to the occlusal plane. Geometry maps X_{adj} and X_{occ} are obtained by rasterizing adjacent teeth and occlusal teeth, respectively. The target geometry map X_0 is obtained by sampling the target denture crown.

Due to variations in occlusal planes across cases, depth images generated with the same hyperparameters can be different between datasets. These discrepancies between cases hinder the network’s ability to learn effectively. In contrast to the depth images used in [25, 26], which require numerous hyperparameters, our method directly rasterizes the height map of occlusal surface, capturing the true height values of points relative to the occlusal plane. This approach better reflects the occlusal relationships between teeth, as it allows for both positive and negative values (higher or lower

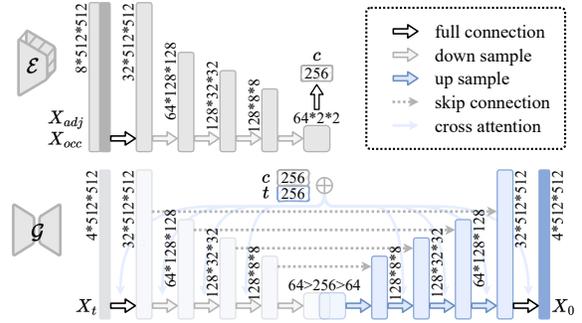


Figure 4. Encoder module \mathcal{E} and denoising module \mathcal{G} . The upper part shows the structure of condition encoder module, which encode the adjacent teeth geometry map X_{adj} and occlusal teeth geometry map X_{occ} as the condition c . The lower part shows the denoising module, which utilizes the condition and time embedding for the reverse diffusion process from X_t to X_0 .

than the occlusal plane), unlike depth maps which are constrained to a 0-1 range. By utilizing these true height values, our method enhances the network’s robustness, making the results less sensitive to the hyperparameters typically associated with depth rasterization and transformation.

3.2. Feature Extraction

The input of our network is the geometry maps from adjacent teeth X_{adj} and occlusal teeth X_{occ} . In order to generate the geometry map of the target tooth X_0 , we define a condition c for denoise diffusion network.

$$c = \mathcal{E}(X_{adj}, X_{occ}). \quad (1)$$

To extract the geometric features of adjacent and occlusal teeth, we propose the encoder module \mathcal{E} , as illustrated in the upper part of Fig. 4. The first fully connected layer consists of a doubleConv structure (comprising a 2D convolution layer, batch normalization, ReLU activation, another 2D convolution layer, and batch normalization). The down-sampling operation is implemented using a 2D max pooling layer followed by two doubleConv layers. The second fully connected layer is applied after flattening the features, followed by a linear layer. The encoder module plays a crucial role in capturing contextual features from the surrounding structures of the target tooth, i.e., the adjacent and occlusal teeth, thereby contributing to precise and functional occlusal surfaces.

3.3. Diffusion Framework

The forward diffusion process, is modeled as a Markov noising process, $\{X_t\}_{t=0}^T$, where X_0 is drawn from the data distribution and it represents the target tooth geometry map in our method. The forward process starts with the clean and initial data sample X_0 and gradually adds noise until it

reaches a standard Gaussian noise distribution \mathbf{X}_T :

$$q(\mathbf{X}_{1:T}|\mathbf{X}_0) := \prod_{t=1}^T q(\mathbf{X}_t|\mathbf{X}_{t-1}), \quad (2)$$

$$q(\mathbf{X}_t|\mathbf{X}_{t-1}) := \mathcal{N}(\mathbf{X}_t; \sqrt{1 - \beta_t}\mathbf{X}_{t-1}, \beta_t\mathbf{I}), \quad (3)$$

where t represents the diffusion step, $\beta_t \in (0, 1)$ is a hyperparameter, and \mathbf{X}_t denotes the distribution at noising step t . Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, the diffusion process can be computed for any step t in a closed form:

$$q(\mathbf{X}_t|\mathbf{X}_0) := \mathcal{N}(\mathbf{X}_t; \sqrt{\bar{\alpha}_t}\mathbf{X}_0, (1 - \bar{\alpha}_t)\mathbf{I}). \quad (4)$$

The training process involves learning a reverse diffusion process that denoises data step by step to recover the original data given the condition c . In our context, conditioned geometry maps synthesis models the distribution $p_\theta(\mathbf{X}_0|c)$ as the reversed diffusion process of gradually denoising the standard Gaussian \mathbf{X}_T , we formulate the reverse diffusion process as a conditional Markov chain as follows:

$$p_\theta(\mathbf{X}_{0:T-1}|\mathbf{X}_T, c) := \prod_{t=1}^T p_\theta(\mathbf{X}_{t-1}|\mathbf{X}_t, c), \quad (5)$$

$$p_\theta(\mathbf{X}_{t-1}|\mathbf{X}_t, c) := \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{X}_t, t, c), \sigma_t^2\mathbf{I}), \quad (6)$$

where c represents the condition, θ represents the neural network parameters, and $\boldsymbol{\mu}_\theta$ is the estimated mean of noise. Instead of predicting ϵ_t as formulated by [11], we follow [21] and predict the unnoised original signal \mathbf{X}_0 itself, by reparameterizing $\boldsymbol{\mu}_\theta$:

$$\boldsymbol{\mu}_\theta = \frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})\mathbf{X}_t + \sqrt{\bar{\alpha}_t}(1 - \alpha_t)\mathcal{G}_\theta(\mathbf{X}_t, t, c)}{1 - \bar{\alpha}_n}. \quad (7)$$

where \mathcal{G}_θ is the denoising module to predict \mathbf{X}_0 , and its structure is illustrated in the lower part of Fig. 4. The first fully connected layer and down-sampling operation are similar to those in the encoder module. Notably, after processing each layer, cross-attention is applied with the condition c and temporal condition t . After passing through a linear layer, the outputs are computed as queries and keys, and cross-attention is performed by multiplying them with the original values. These results are then fed into subsequent modules. In the bottleneck, we utilize doubleConv to process the features. The up-sampling operation consists of bilinear upsampling followed by two doubleConv layers. The final fully connected layer is a single 2D convolution layer, which generates a 4-channel output.

We use cross attention to ensure that the occlusal condition c from the encoder module and the time embedding t are fully integrated and influence the denoising module \mathcal{G} , enabling it to generate realistic occlusal surfaces. Since we directly predict \mathbf{X}_0 in each diffusion process, we can use

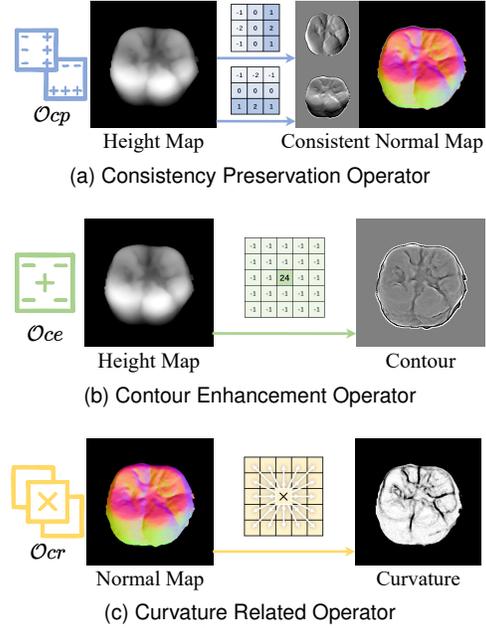


Figure 5. Visualization of geometric operators and corresponding results. Using the \mathcal{O}_{cp} we derive the consistent normal map from height map. Using the \mathcal{O}_{ce} and \mathcal{O}_{cr} for height map and normal map processing separately, we get the contour image and curvature image of target occlusal surface.

the reconstruction loss to supervise the resulting geometry map. Specifically, the reconstruction loss is calculated separately for the height map and the normal map:

$$\mathcal{L}(\theta) := E_{\mathbf{X}_0, t \sim [1, T]} [\mathcal{L}_{recon}(\mathbf{X}_0, \mathcal{G}_\theta(\mathbf{X}_t, t, c))], \quad (8)$$

$$\mathcal{L}_{recon} = \lambda_h \|\bar{\mathbf{h}} - \mathbf{h}^*\|_2^2 + \lambda_n \|\bar{\mathbf{n}} - \mathbf{n}^*\|_2^2, \quad (9)$$

where $\bar{\mathbf{h}}$ and $\bar{\mathbf{n}}$ represent the generated height map and normal map, and \mathbf{h}^* and \mathbf{n}^* represent the ground truth. To constrain our generated geometry maps to match the real distribution, we calculate the norm of the normal map as a mask and comparing it with the actual mask:

$$\mathcal{L}_{mask} = \|\|\bar{\mathbf{n}}\|_2 - \mathbf{M}^*\|_2^2, \quad (10)$$

where \mathbf{M}^* represents the ground truth mask. This approach ensures that the geometry map aligns with the mask and the normal map vectors have the correct magnitude, supporting accurate geometric representation.

3.4. Geometric Operators

We propose three geometric operators to impose geometric constraints on the generated occlusal surface's geometry map. The operators and their corresponding visualization results are illustrated in Fig. 5. It is important to note that our operators work directly on the raw data distribution. For visualization purposes, the results of these operators have been mapped into a color range to enhance intuitiveness.

Consistency Preservation Operator. The Consistency Preservation Operator (\mathcal{O}_{cp} , Fig. 5a) is designed to maintain geometric consistency by aligning predicted normal maps with height map gradients. We use the Sobel filters to the height map to estimate surface normals, thereby enforcing geometric constraints through loss calculation:

$$\mathcal{O}_{cp}(\mathbf{h}(x, y)) = \frac{(\mathcal{S}_x(\mathbf{h}(x, y)), \mathcal{S}_y(\mathbf{h}(x, y)), 1)}{\|(\mathcal{S}_x(\mathbf{h}(x, y)), \mathcal{S}_y(\mathbf{h}(x, y)), 1)\|_2}, \quad (11)$$

$$\mathcal{L}_{cpo} = \|\bar{\mathbf{n}} - \mathcal{O}_{cp}(\bar{\mathbf{h}})\|_2^2, \quad (12)$$

where (x, y) represent the coordinates of the central pixel, \mathcal{O} denotes the corresponding operator, as indicated in the equation above. Additionally, \mathcal{S}_x and \mathcal{S}_y represent the Sobel filters in the two directions, while $\bar{\mathbf{n}}$ and $\bar{\mathbf{h}}$ represent the predicted normal map and height map, respectively.

We calculate the error between predicted normal map and consistent normal map derived from predicted height map as consistency preservation loss \mathcal{L}_{cpo} . This self-supervised approach improves the geometric consistency and accuracy of the generated geometry maps for occlusal surfaces.

Contour Enhancement Operator. During the generation of the occlusal surface, the ridges and grooves are crucial, akin to the high-frequency features in images. We design the Contour Enhancement Operator (\mathcal{O}_{ce} , Fig. 5b), utilizing a Laplacian operator to extract the contours of the height map:

$$\mathcal{O}_{ce}(\mathbf{h}(x, y)) = \mathbf{h}(x, y) - \frac{1}{N} \sum_i^N (\omega_i \cdot \mathbf{h}(x_i, y_i)), \quad (13)$$

$$\mathcal{L}_{ceo} = \|\mathcal{O}_{ce}(\bar{\mathbf{h}}) - \mathcal{O}_{ce}(\mathbf{h}^*)\|_2^2, \quad (14)$$

where \mathbf{h}^* represents the ground truth height map, N represents the number of neighbors of the central pixel, which is set to 24 in this context. We sum the height values of the 24 surrounding pixels at (x_i, y_i) by setting $\omega_i = 1$.

We calculate the error between contours extracted from our height map and the ground truth as contour enhancement loss \mathcal{L}_{ceo} to encourage the network to focus on the high-frequency regions in the height maps. As mentioned above, we map the computed values into a color space to visualize the results, as shown in Fig. 5b. The gray regions indicate areas where the computed value is zero, suggesting no significant changes in height. The lighter regions represent convex areas, while the darker regions indicate concave areas.

Curvature Related Operator. Additionally, the normal map of the tooth contains the required curvature information of the occlusal surface. We design the Curvature Related Operator (\mathcal{O}_{cr} , Fig. 5c), which uses a N-neighbor cross-

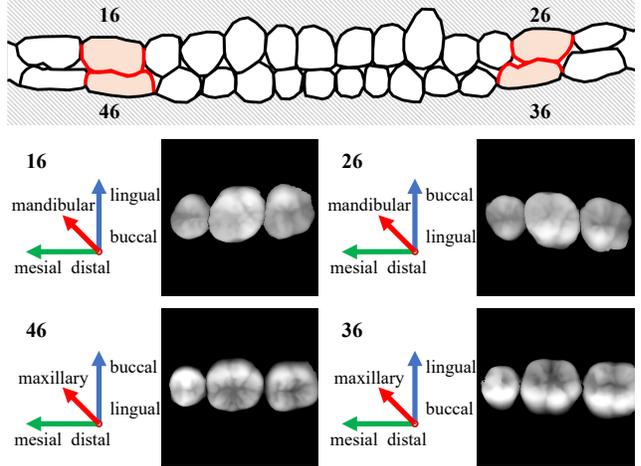


Figure 6. Visualization of four kinds of 1st molar frame and geometry map at different positions. There are differences in buccolingual direction of four positions teeth.

product method to extract the curvature related features:

$$\mathcal{O}_{cr}(\mathbf{n}(x, y)) = \frac{1}{N} \sum_i^N (\mathbf{n}(x, y) \times \mathbf{n}(x_i, y_i)), \quad (15)$$

$$\mathcal{L}_{cro} = \|\mathcal{O}_{cr}(\bar{\mathbf{n}}) - \mathcal{O}_{cr}(\mathbf{n}^*)\|_2^2, \quad (16)$$

where \mathbf{n}^* represents the ground truth normal map, N is set to 24. We calculate the curvature error as the curvature related loss \mathcal{L}_{cro} . This operator and loss function supervise the computed curvature information of occlusal surfaces through the normal map to enforce geometric constraints on the teeth, thereby assessing local geometric variations.

By applying these operators, the geometry maps gain significant detail and accuracy, ensuring that the occlusal surfaces are both visually and functionally realistic. This methodology underscores the robustness of the proposed approach, as the operators help maintain the integrity of the dental features within the raw data distribution.

3.5. Loss Function

During the training process, we use a composite loss function to enable the network to learn reconstruction and enhance geometric details. The total loss function is as follows:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_m \mathcal{L}_{mask} + \lambda_{cpo} \mathcal{L}_{cpo} + \lambda_{geo} \mathcal{L}_{ceo} + \lambda_{geo} \mathcal{L}_{cro}, \quad (17)$$

where λ are the weights of the corresponding loss function. Specifically, the reconstruction loss \mathcal{L}_{recon} is calculated separately for the height map and the normal map. Based on the need to balance the data distribution and empirical experience, ensuring each loss term contributes equally during optimization, we set $\lambda_h = 1$ and $\lambda_n = 10$ as their respective weights. The weights of other losses are $\lambda_m = 0.1$ for

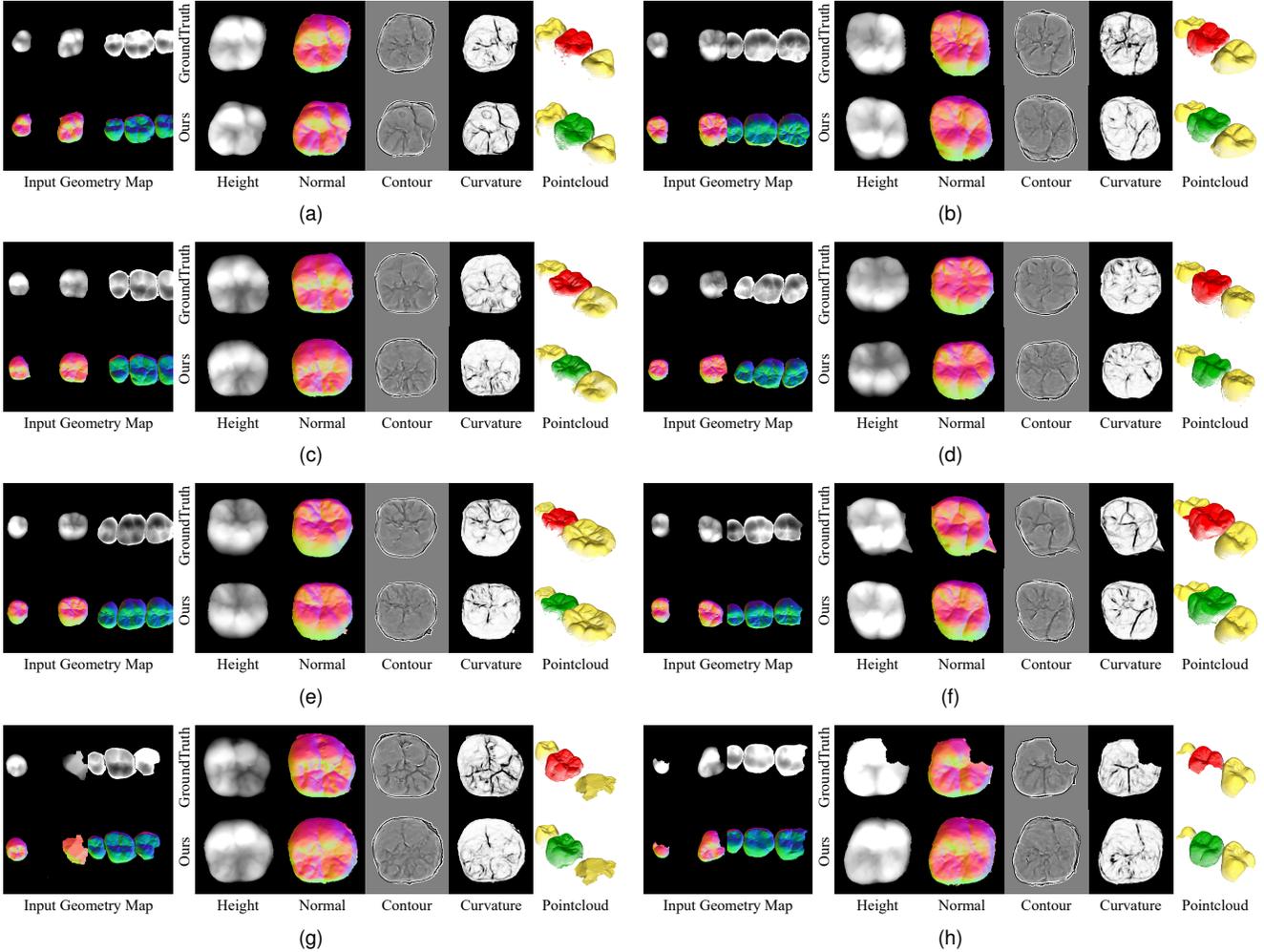


Figure 7. Visualization of the results of our network. (a)-(h) Eight cases from our testing dataset, containing teeth in different positions. In each case, the first columns represent the input geometry maps X_{adj} and X_{occ} . The following five columns display the target tooth’s height map, normal map, contour image, curvature image, and reconstructed point cloud with adjacent teeth. The two rows are the ground truth target tooth and the results from our network. The results are zoomed in to clarify the details.

mask loss \mathcal{L}_{mask} , $\lambda_{cpo} = 0.1$ for consistency preservation loss \mathcal{L}_{cpo} , and $\lambda_{geo} = 0.01$ for contour enhancement loss \mathcal{L}_{ceo} and curvature related loss \mathcal{L}_{cro} . We analyzed the λ_{geo} in the ablation study Section 4.4, which is the most impactful.

4. Experiment

4.1. Dataset and Settings

Dataset. Our dataset was sourced from the intra-oral scanning data of 319 normal occlusion patients. We employed the method described in [40] for semantic segmentation and labeling of our dataset. Specifically, we focused on the first molars (e.g., sixth tooth from the midline), selecting cases where both adjacent and occlusal teeth were present. This resulted in a dataset comprising 197 #16 teeth, 189 #26

teeth, 188 #36 teeth, and 197 #46 teeth. For network training, we randomly divided the dataset into three groups: 70% for training, 15% for testing, and 15% for validation.

After obtaining the semantic information of the teeth, we selected the appropriate segmented tooth meshes. Following the method described in Section 3.1, we performed mesh rasterization to calculate the geometry map. Data pairs were then constructed using adjacent and occlusal teeth as inputs and the target tooth as the output. As shown in Fig. 6, it is worth noting that the local coordinate systems created for different teeth exhibit some variations in the geometry maps.

Implementation Details. We use the Adam optimizer with an initial learning rate of 0.001, which linearly decays to 0 from 15,000 to 30,000 epochs. The network is implemented using PyTorch and trained on a single NVIDIA RTX 3090ti

Table 1. Ablation Study Results. Green shades indicate the optimal results, while light red shades signify a decline in performance.

Metric	RMSE↓		PSNR↑		SSIM↑		Angle↓	Cont↓	Curv↓	FID↓	LPIPS↓
	height	normal	height	normal	height	normal					
w.o. normal	0.340	-	21.17	-	0.903	-	-	0.162	-	177.8	0.105
w.o. \mathcal{L}_{mask}	0.322	0.121	21.77	20.73	0.908	0.883	3.618	0.164	0.262	82.81	0.079
base	0.314	0.125	21.31	20.38	0.906	0.881	3.733	0.164	0.268	80.08	0.084
w.o. $cp\mathcal{O}$	0.297	0.125	22.34	20.76	0.912	0.882	3.415	0.162	0.259	79.67	0.084
w.o. $ce\mathcal{O}$	0.321	0.118	21.65	21.02	0.906	0.885	3.085	0.164	0.245	71.34	0.072
w.o. $cr\mathcal{O}$	0.302	0.120	22.08	20.36	0.908	0.883	3.279	0.163	0.269	75.73	0.079
All	0.293	0.110	23.57	21.21	0.915	0.885	2.987	0.160	0.220	72.17	0.070

(24 GB memory) GPU. The encoder module outputs condition c and time embedding features are of 256 channels. We set the diffusion step t to 200 and the batch size to 8. The geometry map was rasterized at a resolution of 512×512 , corresponding to a spatial resolution of 16 pixels per millimeter.

4.2. Evaluation Metrics

To evaluate the generated crown geometry map, we employed various image-based metrics by projecting the geometric map onto images. It is noteworthy that these metrics are computed on height maps that have undergone a masking process and been converted to grayscale images scaled from 0 to 255. The mask is derived from pixels where the second norm of the normal vector exceeds 0.5. The same process is applied to the normal maps.

We evaluated the height map and normal map separately using Root Mean Square Error (RMSE) [25], Peak Signal Noise Ratio (PSNR) [26], Structural Similarity Index Measure (SSIM) [29]. Additionally, we calculate the mean angle (Angle) between generated normal map $\bar{\mathbf{n}}$ and ground truth \mathbf{n}^* :

$$Angle(\bar{\mathbf{n}}, \mathbf{n}^*) = mean(\arccos \frac{\bar{\mathbf{n}}_i \cdot \mathbf{n}_i^*}{\|\bar{\mathbf{n}}_i\| \cdot \|\mathbf{n}_i^*\|}). \quad (18)$$

For methods lacking a normal map output [5, 25], \mathcal{O}_{cp} will be employed to derive the normal map from generated height map. Using \mathcal{O}_{ce} and \mathcal{O}_{cr} , we can derive contour and curvature maps from the generated target geometry map, which enables us to compute RMSE as evaluation metrics for these maps (Cont and Curv). By reconstructing the point cloud from the height map, we calculated the Chamfer Distance (CD) to evaluate the errors of the reconstruction:

$$CD(\bar{P}, P^*) = \sum_{\bar{p} \in \bar{P}} \min_{p^* \in P^*} \|\bar{p} - p^*\|_2^2 + \sum_{p^* \in P^*} \min_{\bar{p} \in \bar{P}} \|p^* - \bar{p}\|_2^2, \quad (19)$$

where \bar{P} represents reconstructed point cloud from height map, and P^* represents ground truth point cloud. To further evaluate the rationality of the generated model results

in terms of data distribution, we utilized Fréchet Inception Distance (FID) [10] and Learned Perceptual Image Patch Similarity (LPIPS) [36]. Additionally, for 3D methods comparison, we included Hausdorff Distance (HD), Density-aware CD (DCD) [32], and F-score as evaluation metrics. We aim to demonstrate the effectiveness and reliability of our method through multiple evaluation metrics.

4.3. Qualitative Evaluation

As illustrated in the Fig. 7, the visualized results of the geometry map for the occlusal surfaces generated by our method demonstrate its effectiveness. Compared with the ground truth, it is evident that our method produces favorable results across different tooth positions. Additionally, we utilized \mathcal{O}_{ce} and \mathcal{O}_{cr} to compute the contour and curvature images. It can be observed that our results align well with the ground truth in terms of dental grooves. Furthermore, we reconstructed the dental crown point cloud based on the geometry map, demonstrating satisfactory results. As illustrated in Fig. 7, we demonstrate two cases (g, h) involving incomplete teeth as inputs. The experimental results indicate that our network is capable of generating accurate occlusal surfaces even when provided with low-quality input data. Notably, in certain instances, the generated results exhibit superior visual quality compared to the ground truth.

4.4. Ablation Study

In order to verify the effectiveness of the network and loss functions design in this paper, we conducted ablation experiments, as shown in Table. 1. The first row indicates that the input and output include only the height map, excluding the normal map. In the second row, both the height map and normal map are included to build the complete geometry map, along with the reconstruction loss \mathcal{L}_{recon} , but without the mask loss \mathcal{L}_{mask} . The base method in the third row represents the baseline model, where the mask loss function is incorporated alongside the geometry map. Unlike the complete method in the 7th row, which includes all three geometric operators, the fourth to sixth rows demonstrate the results when the complete model has one geomet-

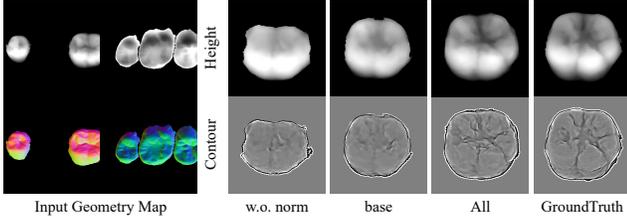


Figure 8. Visualization results of ablation study of our networks. Four columns in right represent the network without normal map, the base network, the network with all geometric constraints, and the ground truth results separately. We visualize the height maps and contour images.

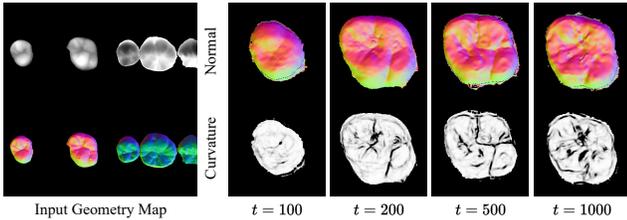


Figure 9. Visualization results of ablation of diffusion step t setting. Four columns in right represent the results of different step t models separately. We visualize the normal maps and Curvature images.

ric operator removed.

As shown in the Table. 1, incorporating the normal map and mask loss significantly improves the generation of the geometry of the occlusal surface. The generated height maps and corresponding contour images of our networks in different structure is shown in Fig. 8. It can be observed that the network without the normal map and associated constraints may result in poor performance, deviating significantly from the true height map in terms of shape, position, and contribution.

This demonstrates that the normal map in our geometry map effectively preserves the geometric information of the occlusal surface. Compared to the network with full geometric constraints, the results from the base network show less pronounced depth variations at the ridges and grooves. Additionally, our geometric operators also play a crucial role in enhancing geometric detail. When the Consistency Preservation Operator \mathcal{O}_{cp} is removed, the consistency between the generated height and normal maps decreases, leading to a drop in normal metrics and data distribution metrics such as FID and LPIPS. Similarly, when the Contour Enhancement Operator \mathcal{O}_{ce} and Curvature Related Operator \mathcal{O}_{cr} are removed, the accuracy of the height map and normal map is adversely affected.

As shown in Fig. 9, we conducted experiments with different diffusion steps t . The results demonstrate that, for our architecture, an appropriate number of diffusion steps, e.g. $t = 200$, not only improves the quality of the gener-

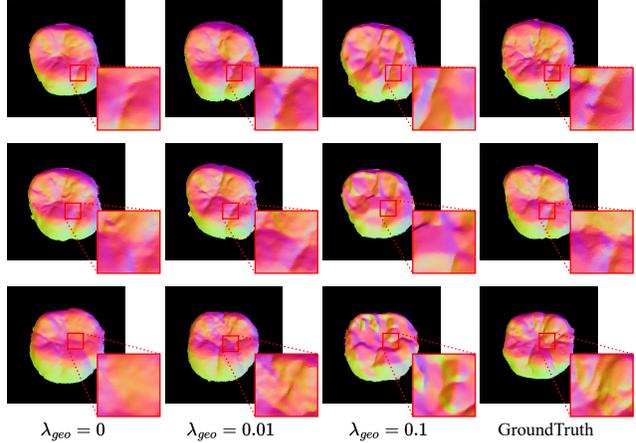


Figure 10. Visualization results of ablation study in different λ_{geo} . Each row presents one cases corresponding to different λ_{geo} setting. Zoom in to focus on the grooves of the generated occlusal surfaces.

ation but also significantly enhances both training and inference efficiency. As the value of t increases, the details gradually become distorted, the generated crowns exhibit noticeable artifacts in regions such as grooves and edges. This phenomenon is attributed to the ability of our network to more effectively capture the fine-grained features of the data, thereby reducing the reliance on a larger number of diffusion steps.

Besides, we conducted ablation experiments to evaluate the impact of our hyperparameter λ_{geo} on the generated occlusal surfaces. The results are summarized in Fig. 10, which shows the changes in the clarity of the grooves on the generated teeth as λ_{geo} varies from 0 to 0.1. The results show that as λ_{geo} increases from 0, the grooves on the generated teeth become clearer, reaching optimal clarity at $\lambda_{geo} = 0.01$. As λ_{geo} continues to increase to 0.1, the grooves start to appear abnormal and overly pronounced. In conclusion, the hyperparameter λ_{geo} plays a crucial role in controlling the clarity and realism of the grooves on the generated occlusal surfaces.

4.5. Comparison

To verify the effectiveness of the method presented in this paper, we modified several existing 2D image generation models and several point cloud completion models to work with our dataset for comparison.

Specifically, we adapted networks based on GANs such as Pix2Pix [13] and CycleGAN [38], networks based on VAEs [18] like VQVAE [27], and DDPM [11]. To ensure a fair comparison, we modified the structure of their output layers: first, we increased the number of output channels to 4, which include the \mathbf{n} normal vectors and the \mathbf{h} height. The modifications involved altering the data input and output of these networks to use our geometry maps. This con-

Table 2. Comparison Results of Different Methods.

Units are *dB* for PSNR, *degrees* for Angle, and *mm* for CD. Darker green shades indicate superior performance.

Metric	RMSE _n ↓	PSNR _h ↑	SSIM _h ↑	Angle ↓	Cont ↓	Curv ↓	CD ↓	FID ↓	LPIPS ↓
Pix2Pix [13]	0.159	16.44	0.884	4.393	0.211	0.382	1.215	242.9	0.161
CycleGAN [38]	0.254	10.52	0.773	7.410	0.265	0.713	2.102	370.7	0.386
VAE [18]	0.115	22.33	0.909	3.340	0.169	0.262	1.184	295.0	0.081
VQVAE [27]	0.108	23.01	0.903	3.472	0.164	0.236	1.046	120.9	0.080
DDPM [11]	0.149	17.88	0.889	4.399	0.162	0.325	1.027	164.1	0.122
3DCGAN [5]	0.136	22.44	0.910	4.248	0.209	0.401	1.301	585.0	0.076
DCPRGAN [25]	0.103	24.27	0.914	3.100	0.205	0.218	1.113	132.2	0.075
Our	0.110	23.57	0.915	2.987	0.160	0.220	0.959	72.17	0.070

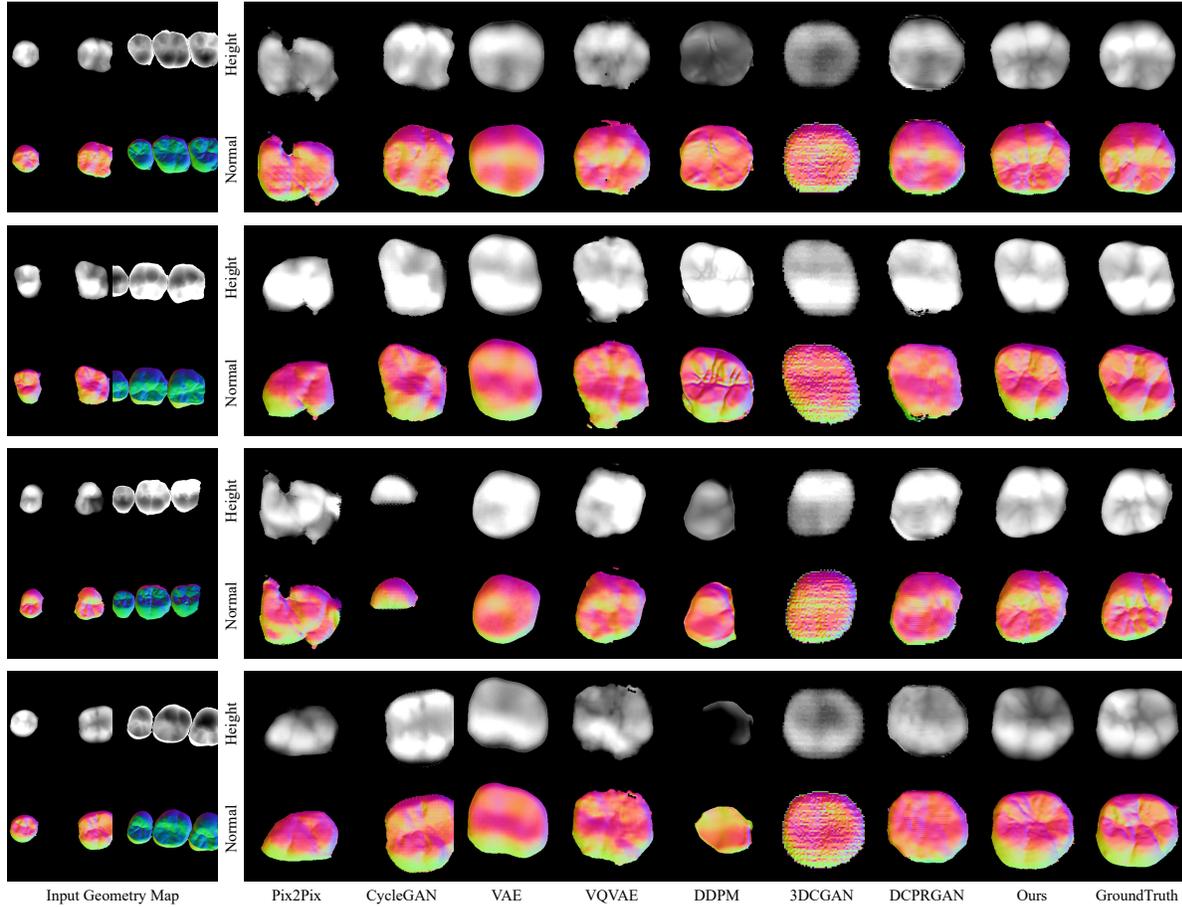


Figure 11. Visualization results of comparison experiments with the other methods. There are four cases from our test dataset, where the left two columns represent the input geometry maps. Each of the right columns represent the geometry map results of the target tooth generated by eight different methods, alongside the ground truth.

version ensured that the output reflected the actual height and normal value distributions, rather than normalizing the results to the 0-1 range used in image generation networks. Then we trained these models on the same dataset to obtain comparable results. Besides, we have modified point cloud completion methods, such as CRA-PCN [22], GeoFormer [34], and SVDFormer [39], to conduct comparative

experiments with 3D-based methods. Specifically, we used the mesh vertices of adjacent teeth and occlusal teeth from the dataset as the initial partial point cloud input. By applying Farthest Point Sampling (FPS), we ensured that the input point cloud consists of 2,048 points. For the output, we sampled 16,384 points from the complete tooth to construct a training dataset for the point cloud completion task.

Table 3. Comparison Results of Point Cloud Completion Methods. Units are *mm* for CD, HD, DCD. The threshold of F-score is *0.3mm*.

Metric	CD↓	HD↓	DCD↓	F-score↑
CRA-PCN [22]	1.018	8.289	0.347	0.324
GeoFormer [34]	1.100	9.928	0.415	0.319
SVDFormer [39]	1.213	10.22	0.411	0.321
OUR	0.959	2.606	0.221	0.401

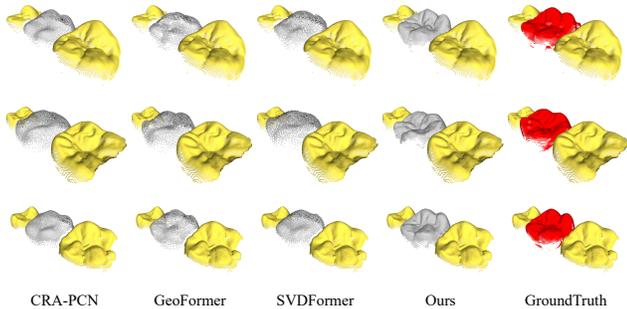


Figure 12. Visualization of three cases compared with the other point cloud based methods. The yellow point cloud are the input adjacent teeth, and the gray point clouds are the output of different methods.

Additionally, we compared two methods for generating occlusal surfaces of dental crowns. The first method, proposed by Ding et al., uses 3D voxelized data and a method based on 3DCGAN [5]. We voxelized our data for training, and it is worth mentioning that we directly used height for voxelization without utilizing normals. The second method, DCPRGAN [25], proposed by Tian et al., predicts occlusal surfaces based on depth maps. We projected our height map into image space as a depth map for training. For both of the above methods, during testing, we projected their generated voxel data or depth images back into our height map space first for evaluation. Since their networks outputs do not include normals, the normal map was derived from the height map.

As shown in the Table. 2, our method performs slightly lower than some methods in pixel-by-pixel metrics such as MSE and PSNR. However, it significantly outperforms other methods in metrics such as FID and LPIPS. Observing the visualized images in Fig. 11, we notice that the pixel-by-pixel indices show a rough similarity in shape to the ground truth, resulting in a lower average error. However, the lack of geometric details makes these images appear more blurry, and traditional models struggle to capture rich geometric details. This is because diffusion models, like other generative models, are weaker in pixel-by-pixel metrics but excel in capturing data diversity and high-frequency detail features. The superior performance in feature-based evaluation metrics, such as FID and LPIPS, suggests that the results obtained by our method are more aligned with

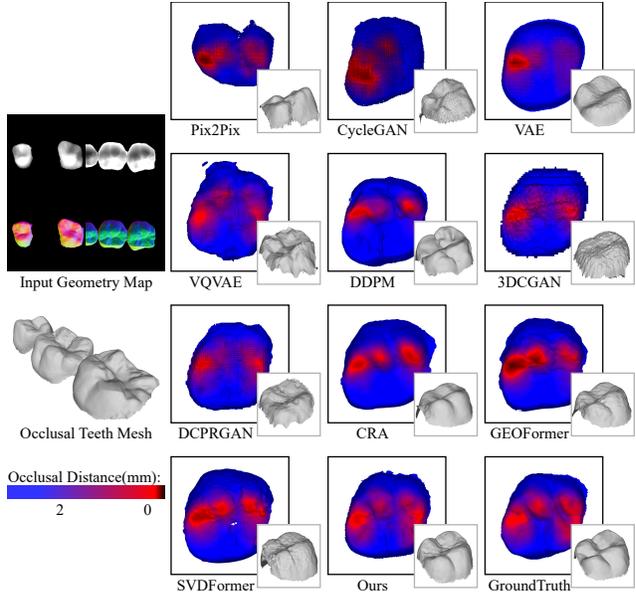


Figure 13. Visualization of occlusal distance compared with the other methods. The left side displays the input geometry map and the mesh of the occlusal teeth. We visualized the occlusal distance, where red areas indicate tight occlusion. A larger red area signifies a better occlusal fit for the denture.

the real data distribution.

As shown in Fig. 12, we visualized the geometry map generated by our method by converting it into a point cloud and compared it with the results of point cloud completion methods. It can be observed that the crowns generated by our method exhibit better performance in capturing details such as grooves. As illustrated in Table. 3, compared to 3D-based methods, although the dimensions differ, our image-based approach achieves higher resolution and places greater emphasis on the occlusal surface, which may correspond to features along a single dimension in 3D representations.

The most crucial aspect of occlusal surface design is the occlusion. We reconstruct the geometry maps generated by various methods into point clouds with normals, then further reconstruct the occlusal surface meshes. To analyze the occlusal function, we calculated the distance between the vertices of the reconstructed denture mesh and the occlusal teeth mesh. The results are shown in the Fig. 13, where red areas represent regions of tight occlusal contact. It is evident that the mesh reconstructed by our method meets the best occlusion, indicating that the geometry produced by our method is more stable and that the occlusal areas match the ground truth more closely, resulting in the optimal occlusal relationship.

Table 4. User Study Results.
Darker green shades indicate greater satisfaction.

Metric	Average Rank↓	Satisfaction Rate↑
Pix2Pix	7.92	14.8%
CycleGAN	8.04	8.9%
VAE	6.10	31.5%
VQVAE	4.58	70.9%
DDPM	3.15	80.1%
3DCGAN	7.15	14.6%
DCPRGAN	3.90	65.3%
Our	2.04	93.4%
Ground Truth	2.12	-

4.6. User Study

We invited 6 professional prosthodontists, 3 technicians, and 12 resident physicians to evaluate the generated occlusal surfaces of eight different methods for denture crowns, including our method and the ground truth. Participants were asked to rank the results of these methods based on their quality and assess whether each result was satisfactory for prosthodontic treatment. The survey results are summarized in Table 4, which shows the average ranking and the percentage of participants who rated each method as the best. Our method received the best average rank and satisfaction rate, indicating that it outperformed both the ground truth and other methods. Specifically, our method achieved a 93.4% approval rating, demonstrating its superior performance and high quality in generating occlusal surfaces.

In conclusion, the survey results, supported by the expertise of professional prosthodontists, demonstrate that our method for generating denture crowns is the most effective, providing results that even surpass the quality of clinical measurement data in some cases.

5. Conclusion

In this study, we introduced a novel approach for designing occlusal surfaces of dental crowns in prosthodontics to address the challenges associated with occlusal functionality. We developed a data-driven method using a diffusion structure for generating occlusal surfaces. By leveraging geometry maps, our method effectively captures and processes the intricate geometric relationships above adjacent teeth, occlusal teeth, and the target denture crown. We define three geometric operators and geometric constraints, which play a crucial role in enhancing the geometric details of the generated occlusal surfaces, and enable the generation of more precise and functional dental crowns to address the intricate demands of occlusal design. Our method not only streamlines the design process but also enhances the precision and consistency of the generated crowns, reduc-

ing the reliance on manual design and the potential for human error. Compared with previous methods, our method is more effective, robust, and explainable in occlusal surfaces design.

In denture design, the occlusal function of the first molar (tooth #6) is particularly important, and it also has a higher incidence of dental issues. Therefore, we focus on the design of the occlusal surface for the first molar. For other molars, our general baseline method can be applied to generate the occlusal surfaces. However, our method cannot be directly used for the design of dentures for incisors and canines. Firstly, the occlusal surfaces of incisors and canines are not parallel to the occlusal plane of the dental arch, which causes our rasterization of the geometry map to become ineffective. Secondly, designing the incisors and canines denture, the occlusal surface is less critical; more attention is needed for the polishing surface, with a greater emphasis on aesthetic considerations. Our future work includes creating a larger dental dataset that encompasses a wider variety of teeth, allowing for different sampling planes in geometry map creation tailored to various types of teeth. And we plan to explore simplified diffusion model computation strategies to achieve faster inference. This will enable efficient deployment in resource-constrained clinical environments. Additionally, we plan to incorporate more anatomy-based priors, such as manually annotated dental grooves, to enhance the precision and applicability of our model.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant 62172257; the Natural Science Foundation of Shandong Province (Major Basic Research) project under Grant ZR2024ZD12. This work is Supported by HPC Platform of ShanghaiTech University.

References

- [1] Y. Chen, S. Gao, P. Tu, and X. Chen. Automatic 3d teeth reconstruction from five intra-oral photos using parametric teeth model. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 3
- [2] Y. Chen, J. K. Y. Lee, G. Kwong, E. H. N. Pow, and J. K. H. Tsoi. Morphology and fracture behavior of lithium disilicate dental crowns designed by human and knowledge-based ai. *Journal of the Mechanical Behavior of Biomedical Materials*, 131:105256, 2022. 3
- [3] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath. Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65, 2018. 3
- [4] Z. Cui, C. Li, N. Chen, G. Wei, R. Chen, Y. Zhou, D. Shen, and W. Wang. Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model. *Medical Image Analysis*, 69:101949, 2021. 3

- [5] H. Ding, Z. Cui, E. Maghami, Y. Chen, J. P. Matinlinna, E. H. N. Pow, A. S. L. Fok, M. F. Burrow, W. Wang, and J. K. H. Tsoi. Morphology and mechanical performance of dental crown designed by 3d-dcgan. *Dental Materials*, 39(3):320–332, 2023. [3](#), [8](#), [10](#), [11](#)
- [6] H. Ding, J. Wu, W. Zhao, J. P. Matinlinna, M. F. Burrow, and J. K. Tsoi. Artificial intelligence in dentistry—a review. *Frontiers in Dental Medicine*, 4:1085251, 2023. [2](#)
- [7] Y. Fan, Q. Ma, G. Wei, Z. Cui, Y. Zhou, and W. Wang. Td-net: tooth axis detection network based on rotation transformation encoding. *Graphical Models*, 121:101138, 2022. [3](#)
- [8] Y. Fan, G. Wei, C. Wang, S. Zhuang, W. Wang, and Y. Zhou. Collaborative tooth motion diffusion model in digital orthodontics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 1679–1687, 2024. [3](#)
- [9] I. E. Hamamci, S. Er, E. Simsar, A. Sekuboyina, M. Gundogar, B. Stadlinger, A. Mehl, and B. Menze. Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays. *arXiv preprint arXiv:2303.06500*, 2023. [3](#)
- [10] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. [8](#)
- [11] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. [2](#), [3](#), [5](#), [9](#), [10](#)
- [12] J.-J. Hwang, S. Azernikov, A. A. Efros, and S. X. Yu. Learning beyond human expertise with generative models for dental restorations. *arXiv preprint arXiv:1804.00064*, 2018. [2](#), [3](#)
- [13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. [3](#), [9](#), [10](#)
- [14] L. Jiang, Y. Mao, X. Wang, X. Chen, and C. Li. Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 398–408, 2023. [3](#)
- [15] X. Jiang, N. Dai, X. Cheng, J. Wang, Q. Peng, H. Liu, and C. Cheng. Robust tooth surface reconstruction by iterative deformation. *Computers in Biology and Medicine*, 68:90–100, 2016. [2](#), [3](#)
- [16] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*, page 102846, 2023. [3](#)
- [17] B. Kim and J. C. Ye. Diffusion deformable model for 4d temporal medical image generation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 539–548, 2022. [3](#)
- [18] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. [3](#), [9](#), [10](#)
- [19] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng. A smart dental health-iot platform based on intelligent hardware, deep learning, and mobile terminal. *IEEE journal of biomedical and health informatics*, 24(3):898–906, 2019. [1](#)
- [20] Q. Lyu and G. Wang. Conversion between ct and mri images using diffusion and score-matching models. *arXiv preprint arXiv:2209.12104*, 2022. [3](#)
- [21] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022. [5](#)
- [22] Y. Rong, H. Zhou, L. Yuan, C. Mei, J. Wang, and T. Lu. Crapcn: Point cloud completion with intra-and inter-level cross-resolution transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4676–4685, 2024. [10](#), [11](#)
- [23] Y. Tegawa and Y. Kinouchi. Dental magnetic attachment: toward third generation devices. *IEEE Transactions on biomedical engineering*, 55(3):1185–1190, 2008. [2](#)
- [24] S. Tian, R. Huang, Z. Li, L. Fiorenza, N. Dai, Y. Sun, H. Ma, et al. A dual discriminator adversarial learning approach for dental occlusal surface reconstruction. *Journal of Healthcare Engineering*, 2022, 2022. [3](#)
- [25] S. Tian, M. Wang, N. Dai, H. Ma, L. Li, L. Fiorenza, Y. Sun, and Y. Li. Dcpr-gan: dental crown prosthesis restoration using two-stage generative adversarial networks. *IEEE Journal of Biomedical and Health Informatics*, 26(1):151–160, 2021. [2](#), [3](#), [4](#), [8](#), [10](#), [11](#)
- [26] S. Tian, M. Wang, F. Yuan, N. Dai, Y. Sun, W. Xie, and J. Qin. Efficient computer-aided design of dental inlay restoration: a deep adversarial framework. *IEEE Transactions on Medical Imaging*, 40(9):2415–2427, 2021. [2](#), [3](#), [4](#), [8](#)
- [27] A. Van Den Oord, O. Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. [3](#), [9](#), [10](#)
- [28] C. Wang, G. Wei, G. Wei, W. Wang, and Y. Zhou. Tooth alignment network based on landmark constraints and hierarchical graph structure. *IEEE Transactions on Visualization and Computer Graphics*, 2022. [3](#)
- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. [8](#)
- [30] G. Wei, Z. Cui, Y. Liu, N. Chen, R. Chen, G. Li, and W. Wang. Tanet: towards fully automatic tooth arrangement. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 481–497, 2020. [3](#)
- [31] G. Wei, Z. Cui, J. Zhu, L. Yang, Y. Zhou, P. Singh, M. Gu, and W. Wang. Dense representative tooth landmark/axis detection network on 3d model. *Computer Aided Geometric Design*, 94:102077, 2022. [3](#)
- [32] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, and D. Lin. Balanced chamfer distance as a comprehensive metric for point cloud completion. *Advances in Neural Information Processing Systems*, 34:29088–29100, 2021. [8](#)
- [33] Y. Yang, H. Fu, A. I. Aviles-Rivero, C.-B. Schönlieb, and L. Zhu. Diffmic: Dual-guidance diffusion network for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 95–105, 2023. [3](#)

- [34] J. Yu, B. Huang, Y. Zhang, H. Li, X. Tang, and S. Gao. Geoforner: Learning point cloud completion with tri-plane integrated transformer. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8952–8961, 2024. [10](#), [11](#)
- [35] F. Yuan, N. Dai, S. Tian, B. Zhang, Y. Sun, Q. Yu, and H. Liu. Personalized design technique for the dental occlusal surface based on conditional generative adversarial networks. *International Journal for Numerical Methods in Biomedical Engineering*, 36(5):e3321, 2020. [3](#)
- [36] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. [8](#)
- [37] Y. Zhao, X. Wang, T. Che, G. Bao, and S. Li. Multi-task deep learning for medical image computing and analysis: A review. *Computers in Biology and Medicine*, 153:106496, 2023. [2](#)
- [38] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. [9](#), [10](#)
- [39] Z. Zhu, H. Chen, X. He, W. Wang, J. Qin, and M. Wei. Svdformer: Complementing point cloud via self-view augmentation and self-structure dual-generator. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14508–14518, 2023. [10](#), [11](#)
- [40] S. Zhuang, G. Wei, Z. Cui, and Y. Zhou. Robust hybrid learning for automatic teeth segmentation and labeling on 3d dental models. *IEEE Transactions on Multimedia*, 2023. [3](#), [7](#)