CROWNGEN: CLINICALLY APPLICABLE DENTAL CROWN GENERATION VIA PHYSICAL-REWARD GENERATIVE MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

We present CrownGen, a fully automatic framework for clinically applicable single-crown generation from raw intraoral scans. Dental crown synthesis requires both high-fidelity anatomical morphology and precise functional relations to neighboring and opposing teeth, which are difficult to capture with reconstruction-driven generative objectives alone. CrownGen combines margin-aware latent diffusion, ambiguity-preserving morphology refinement, and reward-guided clinical applicability adaptation. The morphology generation stage produces anatomically coherent crowns with high-frequency dental structures, while the adaptation stage uses non-differentiable physical rewards from proximal contact and occlusal clearance to improve functional validity. Experiments and professional user studies demonstrate superior geometric fidelity, anatomical realism, functional validity, and clinical acceptability over existing learning-based baselines.

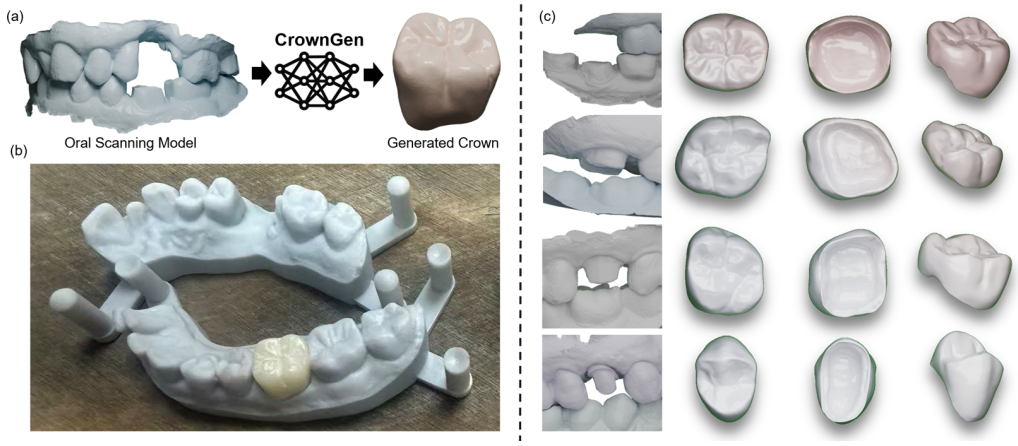


Figure 1: We present CrownGen, the first fully automatic framework for single-crown generation from oral scans that produces functionally plausible crown geometries without manual intervention. (a) Given an oral scan as input, our method achieves fully automatic crown generation. (b) A crown generated by CrownGen and fabricated via 3D printing. The generated crown fits onto the jaw while preserving proper *proximal contact* with neighboring teeth and *occlusal clearance* with opposing teeth. (c) Additional results produced by our method, showing diverse crown morphologies across different cases, with inputs on the left and corresponding generated crowns on the right.

1 INTRODUCTION

Recent advances in diffusion-based 3D generative models have demonstrated strong capability in synthesizing high-fidelity geometry from large-scale data, providing increasingly powerful priors for geometric content creation (Xiang et al., 2025; Tencent-Hunyuan, 2025). However, most existing methods primarily optimize geometric plausibility and reconstruction fidelity, with limited ability to enforce sparse, task-specific functional or physical constraints during generation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Dental crown synthesis is a representative example of such constrained geometric modeling. In standard clinical workflows, crown restoration comprises two essential phases: digital CAD/CAM design and in-mouth chairside delivery. While minor chairside adjustments by dentists remain clinically inevitable due to dynamic oral environments and soft tissue deformations, a functionally valid digital CAD design is crucial to minimize this manual effort.

A suitable digital design must not only resemble a plausible tooth, but also satisfy patient-specific preparation fitting, proximal contact, occlusal clearance, and anatomical morphology. In particular, the margin line determines the preparation boundary, proximal contact affects stability and prevents food impaction, and occlusal clearance avoids interference with the opposing dentition. Even geometrically plausible crowns may therefore remain unsuitable at the digital design stage when these functional relations are violated. Therefore, within our generation framework, *clinical applicability* specifically refers to CAD/CAM design-stage readiness—ensuring the synthesized geometry satisfies physical and functional constraints prior to manufacturing, rather than claiming a zero-adjustment in-mouth deployment.

Existing digital crown design either relies heavily on manual CAD/CAM adjustment or learns geometric mappings from relatively limited dental datasets. Recent large-scale 3D generative models provide substantially stronger geometric priors, but their reconstruction-oriented objectives, such as Chamfer distance, voxel reconstruction, or mesh regression, mainly supervise geometric similarity and do not directly capture sparse and partly non-differentiable functional relations such as proximal contact and occlusal clearance.

To address this limitation, we formulate dental crown synthesis as a constrained geometric generation problem involving two complementary objectives: high-fidelity anatomical morphology generation and CAD-ready clinical applicability adaptation. Based on this formulation, we propose CrownGen, a two-stage framework that first generates anatomically coherent crown geometry using margin-aware latent diffusion with morphology refinement, and then adapts the generated crown using clinically grounded physical rewards. The reward-guided stage directly optimizes proximal-contact and occlusal-clearance objectives that are difficult to express using conventional differentiable reconstruction losses, bridging the gap between pure geometric generation and clinical design standards.

Overall, CrownGen combines large-scale generative priors with patient-specific geometric context and clinically meaningful physical constraints, enabling anatomically realistic and functionally valid crown synthesis under complex dental conditions.

Our contributions are:

- We reformulate dental crown synthesis as a constrained geometric generation problem involving high-fidelity morphology synthesis and clinical applicability adaptation tailored for CAD-stage functional requirements under non-differentiable functional objectives.
- We propose a morphology generation framework combining margin-aware latent diffusion and ambiguity-preserving refinement for anatomically coherent and detail-preserving dental crown synthesis.
- We introduce a reward-guided functional adaptation strategy that enables non-differentiable physical rewards to guide diffusion-based geometric generation toward design-stage functional suitability.

2 Related Work

Dental Crown Design Dental crown design aims to restore tooth function and aesthetics through patient-specific geometry. Early approaches mainly relied on CAD-based template deformation and Statistical Shape Models (SSMs) (Roffmann et al., 2025; Steinbrecher & Gerth, 2008; Mehl et al., 2005; Blanz et al., 2004), which were limited by linear geometric priors and oversmoothed morphology. Recent learning-based methods leverage generative models to learn complex 3D anatomical distributions for crown synthesis, including GANs (Ding et al., 2023), Transformers for mesh completion (Hosseinimanesh et al., 2023), diffusion models for point-cloud generation (Zhang et al., 2024), and voxel-based frameworks (Wei et al., 2026). Two-stage methods such as ToothCR (Zhu et al., 2022) further improve geometric adaptation through post-processing refinement. However,

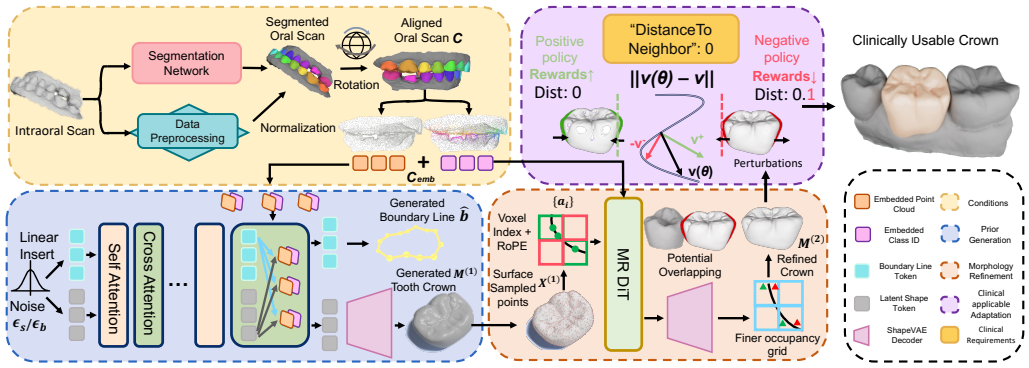


Figure 2: Overview of CrownGen. Patient-specific oral context conditions margin-aware crown generation, followed by morphology refinement and reward-guided adaptation using proximal-contact and occlusal-clearance objectives.

these approaches are typically trained on relatively small dental datasets and often struggle to preserve high-fidelity anatomical details and clinically valid functional relations.

3D Generative Models Recent large-scale 3D generative models have significantly improved scalable geometric synthesis using structured latent representations, voxel sets, point clouds, and other compact 3D representations (Siddiqui et al., 2024; Zhu et al., 2024; Chen et al., 2024; Wei et al., 2025; Ren et al., 2024; Wu et al., 2024; Li et al., 2025). Systems such as TRELLIS (Xiang et al., 2025) and Hunyuan3D (Tencent-Hunyuan, 2025) further demonstrate strong geometric priors learned from large-scale datasets. Their potential for dental reconstruction, however, remains largely unexplored. We leverage such large-scale generative priors for patient-specific crown synthesis while introducing dental-specific conditioning and refinement to preserve anatomical detail and preparation consistency.

Reinforcement Learning for Generative Modeling Reinforcement learning has recently been extended from large language models to diffusion and multimodal generation, enabling optimization toward complex non-differentiable objectives beyond conventional reconstruction losses (Zheng et al., 2025; Ding & Ye, 2025). Recent work has also explored reward-guided optimization for 3D generation, demonstrating its potential to improve geometric fidelity and semantic alignment (Tang et al., 2025). Different from prior RL-based 3D generation that mainly targets perceptual or semantic quality, our method introduces clinically grounded physical rewards for functional dental geometry. Specifically, proximal contact and occlusal clearance provide direct feedback for adapting generated crowns toward clinically valid spatial relations.

3 METHOD

Given patient-specific intraoral scans, our goal is to generate anatomically realistic and functionally suitable crown designs. To achieve both high-fidelity morphology and clinical applicability, we propose a two-stage framework for constrained dental crown generation.

The first stage generates anatomically coherent crown geometry $M^{(1)}$ using a margin-aware latent diffusion model conditioned on patient-specific oral context $\mathcal{C}$, encoded as conditioning tokens C_{emb}. An ambiguity-preserving morphology refinement module further enhances high-frequency anatomical structures, including grooves, fissures, and cusp ridges, yielding the refined crown $M^{(2)}$.

The second stage performs clinical applicability adaptation via physical rewards. Instead of relying solely on differentiable geometric reconstruction objectives, we introduce diffusion adaptation driven by normalized non-differentiable clinical physical rewards $\hat{r}_j$, with proximal and occlusal signed-distance scores $d_{\text{adj}}^{(j)}$ and $d_{\text{occ}}^{(j)}$ providing the basis for the rewards.

Together, these two stages combine high-fidelity anatomical morphology generation with clinical applicability adaptation, enabling clinically valid crown synthesis under complex geometric constraints.

3.1 Dental Morphology Generation

3.1.1 Margin-aware Crown Morphology Learning

Given patient-specific oral context $\mathcal{C}$, we generate an initial crown $M^{(1)}$ using a margin-aware latent diffusion model. The objective of this stage is not to directly optimize functional contact relations, but to learn globally coherent anatomical morphology conditioned on patient-specific oral context. Specifically, we employ a 3D Variational Autoencoder (3D-VAE) and a Diffusion Transformer (DiT) to model crown geometry in latent space.

Given a ground-truth crown mesh M, a VAE encoder V maps the geometry into a compact latent representation $v_s = V(M)$. Then a SDF decoder D_ψ predicts a signed distance field $s(p|v_s) = D_\psi(p, v_s)$, where $p \in \mathbb{R}^3$ is a query point. Finally the crown mesh $\bar{M}$ is reconstructed through Marching Cubes: $\bar{M} = \mathrm{MC}\left(\{p \mid D_\psi(p, v_s) = 0\}\right)$.

From the preprocessed intraoral scan, we construct a semantic point-cloud context $\mathcal{C} = \{(p_i, s_i)\}_{i=1}^{N_c}$, where $p_i \in \mathbb{R}^3$ denotes a sampled point and s_i corresponds to its predicted tooth-position or gingival label. The point coordinates and semantic labels are embedded and fused as

$$c_i = \mathrm{MLP}([\mathrm{PE}(p_i) \parallel \mathrm{Emb}(s_i)]), \qquad C_{\mathrm{emb}} = \{c_i\}_{i=1}^{N_c}, \tag{1}$$

where N_c is the number of context points, $\mathrm{PE}()$ denotes sinusoidal positional encoding, $\mathrm{Emb}()$ is a learnable semantic embedding table, and $\parallel$ denotes concatenation operation. The DiT attends to C_{emb} through cross-attention. At inference, predicted labels and a desired tooth ID identify the single target tooth; no manual labels, target mask, margin, or ground-truth crown is provided.

Because crown morphology depends on the preparation margin, we jointly model crown geometry and margin geometry using two coupled branches. Specifically, the two branches jointly model the crown morphology latent v_s and a margin representation v_b. The preparation margin is represented by a closed curve $b = (b_1, \ldots, b_{256})$, with $b_k \in \mathbb{R}^3$ obtained by interpolation along the annotated margin. The margin branch models this curve jointly with the crown morphology branch, allowing both crown geometry and the margin to be generated at inference time without a manually annotated margin as input.

For a random time step $t \in [0, 1]$, Gaussian perturbations are applied:

$$\tilde{v}_s(t) = (1-t)v_s + t\epsilon_s, \qquad \tilde{v}_b(t) = (1-t)v_b + t\epsilon_b, \tag{2}$$

where ϵ_s and ϵ_b are independent standard Gaussian noise variables with the same shapes as v_s and v_b, respectively. Under this interpolation convention, $t = 0$ corresponds to the data representations and $t = 1$ to Gaussian noise. The target velocities are

$$u_s(t) = \frac{\mathrm{d}\tilde{v}_s(t)}{\mathrm{d}t} = \epsilon_s - v_s, \qquad u_b(t) = \frac{\mathrm{d}\tilde{v}_b(t)}{\mathrm{d}t} = \epsilon_b - v_b. \tag{3}$$

The perturbed latents are concatenated $\tilde{v}_{cat}(t) = [\tilde{v}_s(t) \| \tilde{v}_b(t)]$.

Conditioned on the semantic oral context C_{emb}, the DiT predicts the corresponding flow directions:

$$F(\tilde{v}_{cat}(t), C_{\mathrm{emb}}, t) \to (\hat{u}_s(t), \hat{u}_b(t)). \tag{4}$$

The model is optimized using the joint flow-matching objective:

$$L_{prior} = \mathbb{E}\left[\|\hat{u}_s - u_s\|_2^2 + \lambda_b \|\hat{u}_b - u_b\|_2^2\right]. \tag{5}$$

Here, $\hat{u}_s$ and $\hat{u}_b$ are the predicted velocities at time t, and λ_b controls the relative contribution of the margin-branch loss. The expectation is taken over training examples, sampled times, and Gaussian noise.

During inference, both branches are initialized with independent Gaussian noise at $t = 1$, and the learned joint velocity field is integrated backward to $t = 0$, conditioned on C_{emb}. We denote the

resulting endpoint representations by $v_s^{(0)}$ and $v_b^{(0)}$, where the superscript (0) denotes flow time rather than a mesh-stage index. The shape endpoint $v_s^{(0)}$ is decoded to obtain the initial crown mesh $M^{(1)}$:

$$M^{(1)} = \mathrm{MC}\left(\{p \mid D_\psi(p, v_s^{(0)}) = 0\}\right). \tag{6}$$

3.1.2 High-frequency Morphology Refinement

Although latent diffusion effectively captures global anatomical structure, latent-space generation tends to smooth high-frequency dental morphology due to latent compression and stochastic denoising.

However, realistic dental anatomy contains highly structured local geometric patterns, including cusp ridges, fissure continuity, occlusal grooves, and sharp morphological transitions. Recovering these structures requires preserving local anatomical uncertainty rather than directly memorizing coarse surface geometry.

We therefore introduce an ambiguity-preserving morphology recovery module. Instead of directly conditioning on exact coarse surface coordinates, we convert the generated crown mesh into voxelized spatial anchors that provide approximate geometric localization while intentionally removing precise surface correspondence.

Let N_s denote the number of surface points sampled from the initial crown mesh $M^{(1)}$, and let $x_i \in \mathbb{R}^3$ denote the coordinate of the i-th sampled point:

$$X^{(1)} = \{x_i\}_{i=1}^{N_s}, \tag{7}$$

each point is assigned to a voxel anchor:

$$a_i = \mathrm{clip}\left(\left\lfloor \frac{x_i + 1}{2} R \right\rfloor, 0, R-1\right), \tag{8}$$

Here, R denotes the number of voxels along each axis, and $a_i \in \{0, \ldots, R-1\}^3$ is the integer voxel index assigned to x_i. The mapping assumes coordinates in $[-1, 1]^3$; the floor and clipping operations are applied component-wise.

Unlike direct point-wise conditioning, voxel anchors only indicate local spatial occupancy regions, forcing the refinement generator to reconstruct anatomically plausible high-frequency structures rather than simply copying coarse geometry.

The voxel anchors are encoded as structured spatial tokens and processed by a morphology refinement DiT operating in the same latent geometry space as the first-stage generator.

Importantly, this stage focuses exclusively on recovering anatomical morphology and does not explicitly optimize functional relations such as proximal contact or occlusal clearance. This decoupling prevents sparse functional objectives from suppressing fine-scale dental structures during generation.

We denote the crown mesh obtained after high-frequency morphology refinement of $M^{(1)}$ by $M^{(2)}$. This refined mesh is the candidate geometry evaluated during the subsequent reward-guided adaptation.

3.2 Clinical Applicability Adaptation Via Physical Reward

Although the morphology generation stage produces anatomically realistic crowns, morphology realism alone does not guarantee clinical usability. In practice, generated crowns may still exhibit proximal penetration, insufficient contact, or occlusal interference. A key challenge is that these functional relations are inherently sparse and non-differentiable. Small geometric perturbations may abruptly change contact topology or collision states, making conventional differentiable geometric losses unreliable for direct functional supervision. We therefore introduce a reward-guided physical adaptation stage that incorporates non-differentiable clinical evaluators directly into the geometric generation process. This stage is performed as a global post-training procedure using training-set oral contexts. At inference, the model parameters remain fixed. For each test case, CrownGen generates the crown through forward inference, without test-time parameter updates or patient-specific fine-tuning.

3.2.1 FUNCTIONAL REWARD ESTIMATION

For each oral context, we generate K_c candidate crowns from independent noise initializations $\{M_j^{(2)}\}_{j=1}^{K_c}$, where K_c is the number of independently initialized candidates sampled for each oral context. We use $K_c = 24$ by default and examine candidate-count sensitivity in Section B.1. A functional relation evaluator computes sparse geometric relation scores based on signed distances between crown surfaces and neighboring anatomical structures:

$$E_{\text{func}}(M_j^{(2)}, C) \to (d_{\text{adj}}^{(j)}, d_{\text{occ}}^{(j)}), \tag{9}$$

where C denotes the patient-specific reference geometry used by the evaluator, including neighboring teeth and opposing dentition. The quantities $d_{\text{adj}}^{(j)}$ and $d_{\text{occ}}^{(j)}$ summarize the signed proximal and occlusal distances, respectively, measured in millimeters. Unlike C_{emb}, C denotes geometric evaluator inputs rather than embedded conditioning tokens. Both distances aggregate the closest 5% signed surface samples. For occlusion, C uses all available surfaces in a local crop of the same patient's opposing-jaw scan in the common canonical frame. We define $r_j = -w_{\text{adj}}|d_{\text{adj}}^{(j)} - d_{\text{adj}}^{\star}| - w_{\text{occ}}|d_{\text{occ}}^{(j)} - d_{\text{occ}}^{\star}|$ and normalize it as

$$\hat{r}_j = \text{clip}\left(\frac{r_j - \mu_r}{2(\sigma_r + \epsilon)} + 0.5,\, 0,\, 1\right). \tag{10}$$

where μ_r and σ_r are computed over sampled rewards, while $clip(x, 0, 1) = min\{1, max\{0, x\}\}$.

3.2.2 REWARD-GUIDED DIFFUSION ADAPTATION

Following recent diffusion reinforcement optimization frameworks(Zheng et al., 2025), we optimize the velocity-field prediction objective under implicit positive and negative policies derived from reward-weighted geometric preferences. For each diffusion timestep t, the frozen reference model predicts $v_{old}(x_t, c, t)$, while the trainable model predicts $v_\theta(x_t, c, t)$.

The implicit positive and negative policies are defined as:

$$v_\theta^+(x_t, c, t) = v_\theta(x_t, c, t), \qquad v_\theta^-(x_t, c, t) = 2v_{old}(x_t, c, t) - v_\theta(x_t, c, t). \tag{11}$$

The reward-guided adaptation objective is:

$$L(\theta) = \hat{r}\|v_\theta^+ - v\|_2^2 + (1 - \hat{r})\|v_\theta^- - v\|_2^2. \tag{12}$$

Unlike conventional reconstruction-driven optimization, this formulation allows sparse clinical evaluators to directly guide geometric adaptation without requiring differentiable approximations of contact topology or occlusal relations. As a result, the generated crowns become not only anatomically realistic, but also physically applicable under clinically meaningful functional constraints.

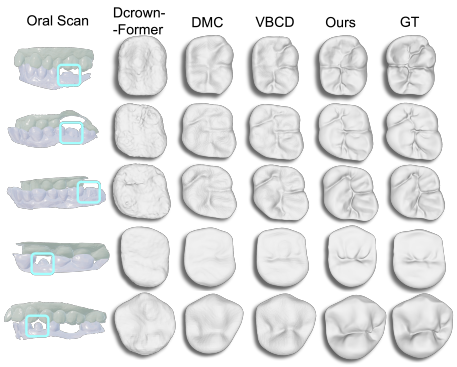
4 EXPERIMENTS

4.1 EVALUATION METRICS

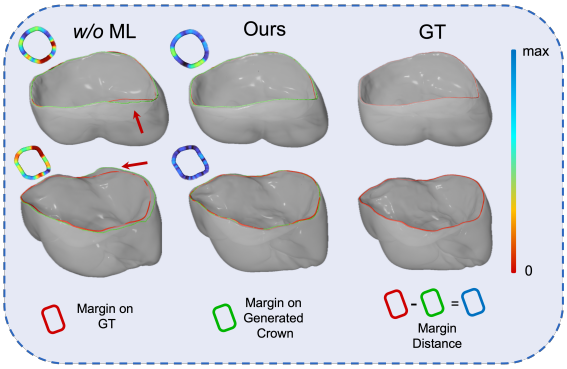
We evaluate the generated crowns from three complementary perspectives: geometric fidelity, functional validity, and expert-perceived clinical usability. Geometric metrics measure the similarity between generated crowns and expert-designed ground truth. Functional metrics evaluate whether the generated crowns satisfy clinically relevant proximal and occlusal relations. User study metrics further assess clinical acceptability from the perspective of dental professionals.

4.1.1 GEOMETRIC FIDELITY METRICS

To evaluate geometric fidelity, we report five metrics: **Chamfer Distance (CD)**, **F-score** with a threshold of 1×10^{-4}, **Surface FPD (S-FPD)**, **Mean Absolute Error (MAE)**, and **Margin-Line Distance (MLD)**. CD and F-score are computed using 50k uniformly sampled surface points. Following (Zhang et al., 2023), S-FPD measures distributional similarity using a pretrained PointNet++ (Qi et al., 2017) feature extractor. MAE is computed on discretized occupancy grids. MLD is computed as the Chamfer distance between the expert-annotated margin line and the generated crown surface, serving as a proxy for marginal fit. All geometries are normalized to the unit cube $[0, 1]^3$ before evaluation, and CD values are reported as $CD \times 10^3$. For fair comparison with baselines, all quantitative results are reported at 128^3 resolution.



(a) Visualization of the comparison. Blue bounding boxes denote the locations of the generated crowns.



(b) Back view of generated crown for ablation study on margin-line awareness. Red: the ground-truth margin; Green: the generated margin. The red arrows highlight regions with poor fitting.

Table 1: Quantitative comparison between our method and state-of-the-art methods on geometric fidelity and functional validity.

Method	CD↓	F-score↑	S-FPD↓	MAE↓	DNT Err.↓	DTA Err.↓
DCrownFormer (Yang et al., 2024)	2.177 ± 0.0300	0.328 ± 0.0047	10.049 ± 0.0866	0.0414 ± 0.0205	0.042 ± 0.033	0.603 ± 1.430
DMC (Hosseinimanesh et al., 2023)	1.101 ± 0.0283	0.596 ± 0.0041	2.838 ± 0.0721	0.0380 ± 0.0156	0.061 ± 0.044	0.606 ± 1.400
VBCD (Wei et al., 2026)	0.981 ± 0.0224	0.640 ± 0.0038	2.661 ± 0.0663	0.0343 ± 0.0113	0.056 ± 0.040	0.689 ± 1.615
Ours	$\mathbf{0.879 \pm 0.0173}$	$\mathbf{0.762 \pm 0.0029}$	$\mathbf{1.010 \pm 0.0640}$	$\mathbf{0.0307 \pm 0.0086}$	$\mathbf{0.008 \pm 0.001}$	$\mathbf{0.350 \pm 0.124}$

4.1.2 Functional Validity Metrics

While geometric metrics evaluate global shape similarity, they do not fully characterize clinical usability. A crown can be close to the ground-truth mesh under CD or F-score but still exhibit inappropriate contact with neighboring or opposing teeth. We therefore report two relation-level functional metrics estimated by our automatic functional evaluator: **Distance to Neighboring Teeth (DNT)** and **Distance to Antagonist (DTA)**.

For each generated crown M_i, the evaluator estimates the distance to neighboring teeth $d_{\text{adj}}^{(i)}$ and the distance to the antagonist $d_{\text{occ}}^{(i)}$ in millimeters. DNT corresponds to proximal contact, while DTA corresponds to occlusal clearance. Given the reference targets $d_{\text{adj}}^{\star} = 0.0\,\text{mm}$ and $d_{\text{occ}}^{\star} = 0.4\,\text{mm}$, we compute

$$\text{DNT} = \frac{1}{N}\sum_{i=1}^{N}\left|d_{\text{adj}}^{(i)} - d_{\text{adj}}^{\star}\right|, \qquad \text{DTA} = \frac{1}{N}\sum_{i=1}^{N}\left|d_{\text{occ}}^{(i)} - d_{\text{occ}}^{\star}\right|.$$

Lower values indicate that the generated crowns better satisfy the desired proximal and occlusal functional relations. These are absolute target-deviation errors, not raw signed distances; positive errors therefore do not exclude penetration. Slightly negative d_{adj} indicates adjustable tight contact, whereas larger negative values indicate excessive tightness or seating risk.

4.2 Comparison with State-of-the-art Methods

We compare our method with three representative learning-based baselines for automated dental crown design: **(a) DCrownFormer**: A point-to-mesh framework that deforms a template mesh conditioned on the surrounding jaw context. **(b) DMC**: A point-cloud completion method based on Shape as Points (Peng et al., 2021), formulating crown design as missing-geometry completion. **(c) VBCD**: A voxel-based framework with margin-aware geometric supervision for efficient crown completion.

All methods use the same split, preprocessing, normalization, and metrics; learning-based baselines are retrained on our data and tuned only on the validation set.

As shown in Table 1, our method achieves the best overall performance in both geometric fidelity and functional validity. Compared with prior approaches, our framework not only improves global

Table 2: Ablation study of CrownGen. Input preprocessing is identical across all variants.

Components				Metrics						
ML	Seg	S2	PR	CD↓	F-score↑	S-FPD↓	MAE↓	MLD↓	DNT Err.↓	DTA Err.↓
✓	✓	✓	✓	$\mathbf{0.879 \pm 0.0173}$	$\mathbf{0.762 \pm 0.0029}$	$\mathbf{1.010 \pm 0.0640}$	$\mathbf{0.0307 \pm 0.0086}$	$\mathbf{0.012 \pm 0.002}$	$\mathbf{0.008 \pm 0.001}$	$\mathbf{0.350 \pm 0.124}$
×	✓	✓	✓	1.082 ± 0.0210	0.743 ± 0.0032	2.449 ± 0.0700	0.0369 ± 0.0108	0.062 ± 0.017	0.008 ± 0.001	0.370 ± 0.140
✓	×	✓	✓	0.902 ± 0.0182	0.762 ± 0.0030	2.653 ± 0.0735	0.0332 ± 0.0101	0.045 ± 0.007	0.009 ± 0.002	0.450 ± 0.165
✓	✓	×	×	1.073 ± 0.0205	0.715 ± 0.0035	1.881 ± 0.0680	0.0362 ± 0.0105	0.014 ± 0.003	0.024 ± 0.017	0.601 ± 0.242
✓	✓	✓	×	0.946 ± 0.0188	0.728 ± 0.0033	1.564 ± 0.0710	0.0339 ± 0.0096	0.012 ± 0.002	0.013 ± 0.011	0.400 ± 0.194

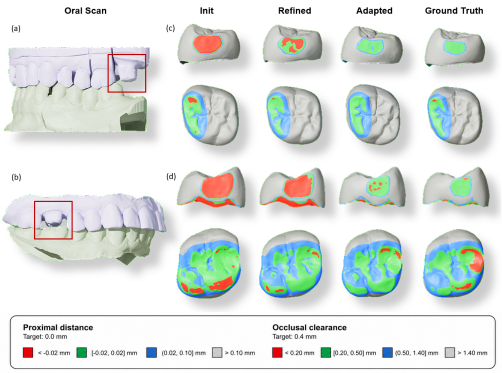


Figure 4: Ablation visualization of high-frequency morphology refinement and clinical applicability adaptation.

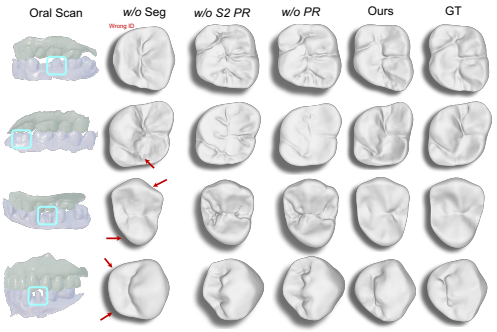


Figure 5: Visualization of the ablation study. Blue bounding boxes denote the locations of the generated crowns, while red arrows highlight the regions where w/o Seg performs suboptimally.

anatomical similarity, but also produces more reliable proximal and occlusal relations, highlighting the benefit of physical reward adaptation for clinically applicable crown synthesis.

Figure 3a presents qualitative comparisons. Although baseline methods can generate plausible crown shapes, they often exhibit over-smoothed occlusal anatomy, distorted cusp structures, or unstable contact relations. In contrast, our method produces crowns with richer anatomical details and more stable proximal and occlusal relations.

4.3 Ablation Studies

We evaluate the contributions of margin-line awareness (**ML**), tooth-ID conditioning (**Seg**), morphology refinement (**S2**), and physical reward-guided adaptation (**PR**). Table 2 reports their effects on geometric fidelity, margin consistency, and functional validity, with input preprocessing held fixed across these variants.

Removing **ML** causes a clear degradation in margin-related accuracy, indicating that explicit supervision on the boundary points v_b is essential for adapting the crown to the shoulder area of the prepared tooth. As shown in Figure 3b, the full model produces a margin line closer to the ground truth (GT). Removing **Seg** leads to moderate performance drops; although the F-score remains relatively stable, Figure 5 shows that the absence of tooth-ID cues can result in mismatched tooth morphologies.

Removing both **S2** and **PR** degrades geometric fidelity and functional validity. Comparing this variant with the model retaining **S2** but omitting **PR** shows that morphology refinement improves both geometric and functional metrics. Adding **PR** further reduces DNT Err. and DTA Err., supporting the benefit of physical reward-guided adaptation.

Contact-region agreement. On 100 randomly sampled test cases, we compare proximal contact regions with expert CAD crowns using voxel IoU, relative contact-area error, and centroid distance (Table 3; mean ± SD). The full model achieves the best overlap and localization, whereas morphology refinement alone does not improve contact localization.

Table 3: Contact-region agreement with expert CAD crowns over 100 cases (mean ± SD).

Setting	IoU↑	Area err.↓	Centroid (mm)↓
Stage 1	0.037 ± 0.042	0.469 ± 0.158	0.871 ± 0.597
+ Morph.	0.290 ± 0.116	0.227 ± 0.187	1.085 ± 0.764
Full + RL	$\mathbf{0.565 \pm 0.110}$	$\mathbf{0.224 \pm 0.201}$	$\mathbf{0.296 \pm 0.263}$

Table 4: Design-stage acceptability rates (%) over 128 expert evaluations per method (64 cases, two experts each).

Method	Proximal (%)	Occlusal (%)	Overall (%)
Ours	**95.3**	**95.3**	**95.3**
Ours w/o RL	84.4	79.7	79.7
Stage 1 only	65.6	70.3	50.0
VBCD	35.9	18.8	4.7
DMC	40.6	34.4	21.9
DCrownFormer	51.6	53.1	34.4

Table 5: Secondary subjective ratings, reported as mean (standard deviation).

Method	Plausibility	Adjustment
Ours	**4.85 (0.49)**	**4.80 (0.52)**
Ours w/o RL	4.45 (0.69)	4.35 (0.81)
Stage 1 only	4.10 (0.72)	4.00 (0.79)
VBCD	2.26 (1.54)	2.19 (0.74)
DMC	2.72 (1.07)	2.13 (0.64)
DCrownFormer	3.30 (1.02)	3.20 (0.70)

5 Results and Analysis

5.1 User Study

We conducted a user study to evaluate the design-stage acceptability of crowns generated by different methods in terms of anatomy, proximal contact, occlusion, and expected adjustment effort.

Participants and Materials. We evaluated 64 shared test cases spanning 16 posterior tooth positions, with four sampled cases per position. Thirty-two dental professionals, including practicing dentists and senior technicians, participated, with two distinct experts evaluating each case and no missing evaluations. For each case, six sets of candidate crowns were evaluated: our full framework, two ablated variants (Ours w/o RL and Stage 1 only), and three learning-based baselines (VBCD, DMC, and DCrownFormer), yielding 128 expert evaluations per method.

Evaluation Protocol. Anonymized crown meshes were randomized and shown with their jaw context to blinded evaluators. Proximal and occlusal acceptability were binary; overall acceptance required both within the same expert's evaluation of a candidate crown. Each expert's judgment was counted separately. Experts also rated anatomical plausibility and adjustment effort on 1–5 scales, where higher scores indicate better anatomy and less required adjustment.

Results and Discussion. Table 4 shows that our method achieves the highest proximal, occlusal, and overall acceptability (95.3% each). Overall acceptability is 34.4% for the strongest baseline, DCrownFormer, versus 79.7% without reward-guided functional adaptation and 50.0% for Stage 1 only.

The secondary subjective ratings show a similar trend. Our method also obtains the highest anatomical plausibility score of 4.85 and manual adjustment score of 4.80, indicating that the generated crowns are perceived as both anatomically realistic and close to clinical readiness. These results indicate that our full pipeline improves both perceived morphology and clinically relevant functional relations.

6 Conclusion

We presented CrownGen, a fully automatic framework for clinically applicable single-crown generation from intraoral scans. Our method combines high-fidelity morphology synthesis with clinical applicability adaptation, improving proximal and occlusal relations through non-differentiable physical rewards. Experiments and design-stage professional assessments demonstrate superior geometric fidelity, anatomical realism, and functional acceptability over existing learning-based baselines. These results support digital CAD/CAM design readiness rather than in-mouth performance; gingival compression and emergence-profile acceptability require future clinical study. Future work will extend the framework to multi-crown and full-arch restoration.

AI USE STATEMENT

In this work, we used GPT-5.5 and GPT-5.6 for the following tasks:

- **Paper writing.** Assisting with drafting and revising text, improving wording and grammar, and writing LaTeX code.
- **Literature retrieval and discovery.** Helping identify, organize, and summarize related work.
- **Code implementation and analysis.** Assisting with implementing and optimizing parts of the code, checking arithmetic consistency in reported statistics, and providing feedback on evaluation methodology and the interpretation of results.

The authors take responsibility for the final manuscript, including all claims, numerical results, references, and AI-assisted content.

REFERENCES

V. Blanz, A. Mehl, T. Vetter, and H.-P. Seidel. A statistical method for robust 3d surface reconstruction from sparse data. In *Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004.*, pp. 293–300, 2004. doi: 10.1109/TDPVT.2004.1335212.

Sijin Chen, Xin Chen, Anqi Pang, Xianfang Zeng, Wei Cheng, Yijun Fu, Fukun Yin, Zhibin Wang, Jingyi Yu, Gang Yu, BIN FU, and Tao Chen. MeshXL: Neural coordinate field for generative 3d foundation models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=Gcks157FI3.

Hao Ding, Zhiming Cui, Ebrahim Maghami, Yanning Chen, Jukka Pekka Matinlinna, Edmond Ho Nang Pow, Alex Siu Lun Fok, Michael Francis Burrow, Wenping Wang, and James Kit Hon Tsoi. Morphology and mechanical performance of dental crown designed by 3d-dcgan. *Dental Materials*, 39(3):320–332, 2023. ISSN 0109-5641. doi: https://doi.org/10.1016/j.dental.2023.02.001. URL https://www.sciencedirect.com/science/article/pii/S0109564123000416.

Zheng Ding and Weirui Ye. Treegrpo: Tree-advantage grpo for online rl post-training of diffusion models. *arXiv preprint arXiv:2512.08153*, 2025.

Golriz Hosseinimanesh, Farnoosh Ghadiri, Francois Guibault, Farida Cheriet, and Julia Keren. From mesh completion to ai designed crown. In Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 555–565, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43996-4.

Zhihao Li, Yufei Wang, Heliang Zheng, Yihao Luo, and Bihan Wen. Sparc3d: Sparse representation and construction for high-resolution 3d shapes modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=yslRXs9gcJ.

Albert Mehl, Volker Blanz, and Reinhard Hickel. Biogeneric tooth: a new mathematical representation for tooth morphology in lower first molars. *European Journal of Oral Sciences*, 113 (4):333–340, 2005. doi: https://doi.org/10.1111/j.1600-0722.2005.00224.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0722.2005.00224.x.

Songyou Peng, Chiyu "Max" Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: a differentiable poisson solver. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

Xuanchi Ren, Jiahui Huang, Xiaohui Zeng, Ken Museth, Sanja Fidler, and Francis Williams. Xcube: Large-scale 3d generative modeling using sparse voxel hierarchies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4209–4219, June 2024.

Oliver Roffmann, Meike Stiesch, Christof Hurschler, and Andreas Greuling. Automatic adjustment of dental crowns using laplacian mesh editing. *Journal of the Mechanical Behavior of Biomedical Materials*, 163:106878, 2025. ISSN 1751-6161. doi: https://doi.org/10.1016/j.jmbbm.2024.106878. URL https://www.sciencedirect.com/science/article/pii/S1751616124005101.

Yawar Siddiqui, Matthias Niessner, Siyu Ma, Angela Dai, et al. Meshgpt: Generating hierarchical meshes with autoregressive transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19263–19274, 2024.

Tillmann Steinbrecher and Maik Gerth. Dental inlay and onlay construction by iterative laplacian surface editing. *Comput. Graph. Forum*, 27:1441–1447, 07 2008. doi: 10.1111/j.1467-8659.2008.01284.x.

David Stutz and Andreas Geiger. Learning 3d shape completion from laser scan data with weak supervision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2018.

Yiwen Tang, Zoey Guo, Kaixin Zhu, Ray Zhang, Qizhi Chen, Dongzhi Jiang, Junli Liu, Bohan Zeng, Haoming Song, Delin Qu, et al. Are we ready for rl in text-to-3d generation? a progressive investigation. *arXiv preprint arXiv:2512.10949*, 2025.

Tencent-Hunyuan. Hunyuan3d-2.5, 2025. URL https://github.com/Tencent-Hunyuan/Hunyuan3D-2.1.

Linda Wei, Chang Liu, Wenran Zhang, Zengji Zhang, Shaoting Zhang, and Hongsheng Li. Vbcd: A voxel-based framework for personalized dental crown design. In James C. Gee, Daniel C. Alexander, Jaesung Hong, Juan Eugenio Iglesias, Carole H. Sudre, Archana Venkataraman, Polina Golland, Jong Hyo Kim, and Jinah Park (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, pp. 627–636, Cham, 2026. Springer Nature Switzerland. ISBN 978-3-032-04984-1.

Si-Tong Wei, Rui-Huan Wang, Chuan-Zhi Zhou, Baoquan Chen, and Peng-Shuai Wang. Octgpt: Octree-based multiscale autoregressive models for 3d shape generation, 2025. URL https://arxiv.org/abs/2504.09975.

Shuang Wu, Youtian Lin, Feihu Zhang, Yifei Zeng, Jingxi Xu, Philip Torr, Xun Cao, and Yao Yao. Direct3d: Scalable image-to-3d generation via 3d latent diffusion transformer, 2024. URL https://arxiv.org/abs/2405.14832.

Jianfeng Xiang, Xiaoxue Chen, Sicheng Xu, Ruicheng Wang, Zelong Lv, Yu Deng, Hongyuan Zhu, Yue Dong, Hao Zhao, Nicholas Jing Yuan, and Jiaolong Yang. Native and compact structured latents for 3d generation. *Tech report*, 2025.

Su Yang, Jiyong Han, Sang-Heon Lim, Ji-Yong Yoo, SuJeong Kim, Dahyun Song, Sunjung Kim, Jun-Min Kim, and Won-Jin Yi. DCrownFormer: Morphology-aware Point-to-Mesh Generation Transformer for Dental Crown Prosthesis from 3D Scan Data of Antagonist and Preparation Teeth . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006. Springer Nature Switzerland, October 2024.

Ao Zhang, Zhen Shen, Jian Yang, Qihang Fang, Gang Xiong, and Xisong Dong. Conditional diffusion guided by part-level latent for dental crown point cloud generation. pp. 284, 06 2024. doi: 10.1117/12.3034176.

Biao Zhang, Jiapeng Tang, Matthias Nießner, and Peter Wonka. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Trans. Graph.*, 42(4):92:1–92:16, August 2023. URL https://doi.org/10.1145/3592442.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Ming-Yu Liu. Diffusionnft: Online diffusion reinforcement with forward process. *arXiv preprint arXiv:2509.16117*, 2025.

Haoyu Zhu, Xiuyi Jia, Changdong Zhang, and Tingting Liu. Toothcr: A two-stage completion and reconstruction approach on 3d dental model. In João Gama, Tianrui Li, Yang Yu, Enhong Chen, Yu Zheng, and Fei Teng (eds.), *Advances in Knowledge Discovery and Data Mining*, pp. 161–172, Cham, 2022. Springer International Publishing. ISBN 978-3-031-05981-0.

Junsheng Zhu, Shuzhe Tang, Meihao Chen, Jun Gao, Saining Xie, and Sanja Fidler. Llama-mesh: Unifying language models with 3d mesh generation. *arXiv preprint arXiv:2411.09595*, 2024.

A Experimental Setup

A.1 Dataset

The dataset consists of 13,325 intraoral scans, split into 9,860 training, 2,465 validation, and 1,000 test samples. Each case includes an intraoral scan, an expert-annotated preparation margin, and an expert-designed digital CAD ground-truth crown rather than an in-mouth adjusted restoration.

For preprocessing, we apply global normalization based on crown centroids and coordinate bounds. During training, the ground-truth crown serves as the reference. At inference, when the ground-truth crown is unavailable, we first obtain an initial crown prediction using VBCD (Wei et al., 2026) and use this prediction only as the reference for normalization. The resulting transform is applied consistently to the entire oral context. Then we convert crown meshes into watertight manifolds using mesh-fusion (Stutz & Geiger, 2018) for VAE fine-tuning.

Following expert-suggested design practice in our clinical setting, we use $d^{\star}_{\mathrm{adj}} = 0.0\,\mathrm{mm}$ and $d^{\star}_{\mathrm{occ}} = 0.4\,\mathrm{mm}$ as soft reference values for functional adaptation.

A.2 Implementation Details

Our framework is implemented in PyTorch and optimized with AdamW. Crown meshes are extracted as zero-level SDF isosurfaces using Marching Cubes on a 256^3 grid over $[-1, 1]^3$.

For **margin-aware crown morphology learning**, we fine-tune the 3D-VAE $\mathcal{V}$ from the Hunyuan3D 2.1 checkpoint using a learning rate of 1×10^{-4}. The encoder takes 20,480 sampled surface points, including 10,240 curvature-aware sharp points. The DiT $\mathcal{F}$ is trained with flow matching on 8 NVIDIA A100 GPUs for 9,000 epochs using a batch size of 35. The learning rate is linearly warmed up to 1×10^{-4} for the first 500 epochs and then cosine-decayed to 1×10^{-5}. Classifier-free guidance with a scale of 2.0 is used during generation. For semantic conditioning, a lightweight Point Transformer segmentation network predicts tooth-position labels in FDI notation and gingival labels for points in the local intraoral scan. We initially sample 10,000 points and apply farthest point sampling to retain $N_c = 4096$ context points.

For **high-frequency morphology refinement**, we voxelize the generated crown with anchor resolution $R = 32$ and anchor budget $K_a = 8192$. The adapted VAE decoder is fine-tuned with off-surface SDF supervision, while the refinement generator is trained in the adapted latent space using the same flow-matching objective.

For **clinical applicability adaptation**, we initialize from the trained refinement generator. At each outer epoch, we generate $K_c = 24$ candidate crowns for each oral context and evaluate their rewards. After all sampling batches have been collected, we group the samples by their conditioning-context identifiers (UIDs) and compute the reward mean and standard deviation within each group. We then normalize the rewards using these group-wise statistics to obtain per-sample advantages, which are stored in a buffer. Each sample's advantage remains fixed throughout the subsequent inner-loop optimization and is reused across all inner epochs, mini-batches, and diffusion timesteps. At the next outer epoch, we generate a new set of candidates with updated generator and recompute the group-wise statistics and advantages. This sampling-and-update procedure is performed exclusively during post-training on training-set contexts; model parameters remain fixed at inference. Candidate crowns are evaluated using a functional relation estimator measuring proximal and occlusal distances in physical millimeter scale. We use $d^{\star}_{\mathrm{adj}} = 0.0\,\mathrm{mm}$ and $d^{\star}_{\mathrm{occ}} = 0.4\,\mathrm{mm}$ as target values.

Inference efficiency. The full pipeline takes 25.04 seconds per crown on an NVIDIA H100 and 13.17 seconds on an NVIDIA B200, compared with approximately 10–15 minutes for manual design.

A.3 Contact-region Evaluation Protocol

We evaluate contact-region agreement on 100 randomly sampled test cases. For generated and expert CAD crowns, contact regions are extracted separately for each adjacent tooth using area-uniform crown-surface samples. Samples are retained when their signed distance to the same-jaw adjacent surface lies in $[-0.10, 0.10]$ mm and their unsigned distance is at most 0.15 mm; negative

signed distances indicate penetration and positive values indicate gaps. Samples within 1.5 mm of the annotated finish-line vertices in Euclidean distance are excluded.

Contact IoU is computed using a shared 0.20-mm voxel grid in the original physical coordinate system. Contact area is estimated as total crown surface area multiplied by the fraction of contact samples, and relative area error is $|A_{\mathrm{gen}} - A_{\mathrm{GT}}|/A_{\mathrm{GT}}$. Contact centroids are the arithmetic means of the corresponding surface samples, with centroid distance measured in millimeters. Results are reported as mean $\pm$ sample standard deviation across cases.

A.4 User Study Protocol

The study in Section 5.1 assesses applicability at the digital design stage, not in-mouth performance. Thirty-two dental professionals evaluated 64 shared cases (four per posterior tooth position), with two distinct experts assessing each case. Each expert evaluated all six methods for the assigned case using anonymized crown meshes shown with jaw context in randomized order. All evaluations were complete, yielding 128 expert judgments per method. Proximal and occlusal acceptance were recorded separately as binary judgments; overall acceptance required both within the same expert's evaluation of the same crown. Each judgment was counted separately when computing acceptance rates.

Anatomical plausibility and anticipated manual adjustment effort were rated separately on 1–5 scales, with higher scores indicating better anatomy and less required adjustment, respectively. The adjustment score describes expected manual effort; it is not converted into a binary acceptance label. Together, these assessments characterize design-stage acceptability and the anticipated need for manual adjustment.

A.5 Proximal and Occlusal Penetration

Negative signed distances indicate geometric penetration. In our digital design-stage interpretation, slight proximal overlap is treated as tight contact rather than an automatic failure, without implying that arbitrary overlap is acceptable. The proximal reward nevertheless penalizes deviations from $d^{\star}_{\mathrm{adj}} = 0$ in either direction. For occlusion, $d^{\star}_{\mathrm{occ}} = 0.4\,\mathrm{mm}$ is a positive-clearance reference, and penetration is not a desired outcome. These are soft reward targets, not hard collision constraints. DNT/DTA report absolute target deviations and therefore do not, by themselves, establish the absence of penetration. The contact-region extraction thresholds above are evaluation definitions, not universal acceptability thresholds.

B Additional Experiments

B.1 Effect of Candidate Count

We examine how the number of candidates sampled for each oral context relates to proximal and occlusal fit during reward-guided crown adaptation. Candidate count controls the set of geometric alternatives available to the model, while dental fit requires compatibility with both neighboring and opposing teeth. Figure 6 presents contact-error curves for $K_c \in \{12, 18, 24, 30, 36\}$.

At 1,800 optimizer updates, $K_c = 24$ has displayed proximal and occlusal values of 0.00937 and 0.35106 mm, respectively. Relative to $K_c = 18$, whose corresponding values are 0.00961 and 0.37391 mm, these are lower by 2.50% and 6.11%. Compared with $K_c = 12$, the reductions are 11.13% and 10.19%; compared with $K_c = 30$, they are 4.23% and 8.46%. The $K_c = 36$ curve ends at a slightly lower proximal value of 0.00932 mm, but its occlusal value is 0.42052 mm, which is higher than that of $K_c = 24$. Thus, the endpoint comparison in the figure favors $K_c = 24$ for occlusal fit while retaining a small proximal trade-off relative to $K_c = 36$.

B.2 The distribution of tooth positions across the dataset

The distribution of tooth positions across the dataset is illustrated in Figure 7.

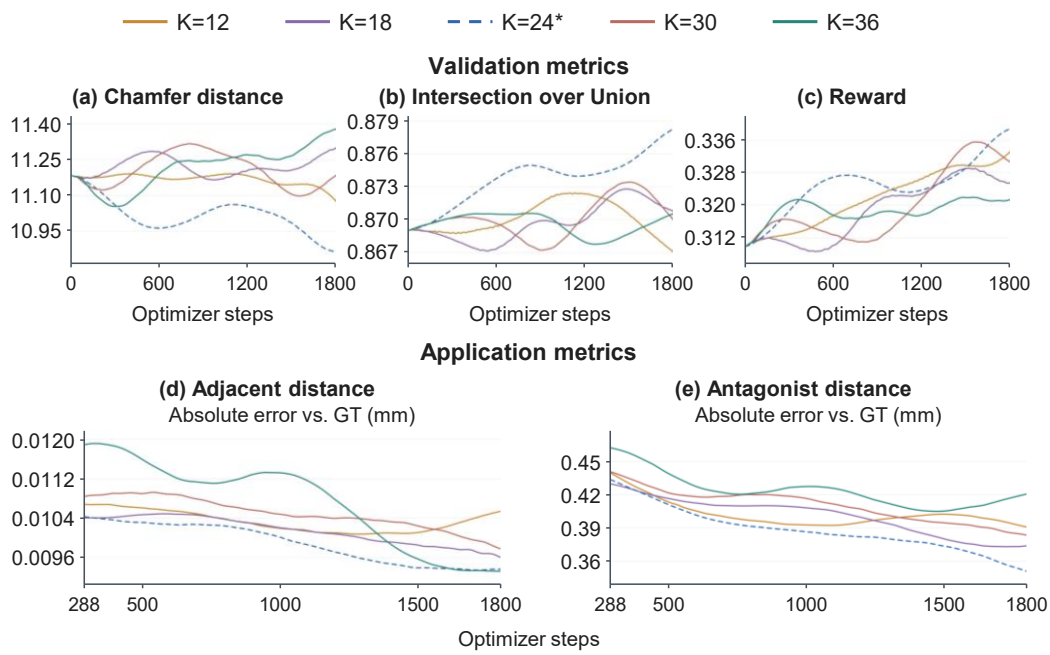


Figure 6: Candidate-count sensitivity of dental contact accuracy for $K_c \in \{12, 18, 24, 30, 36\}$. (d) Proximal and (e) occlusal absolute distance errors relative to the ground-truth crown, $|d_{\text{gen}} - d_{\text{GT}}|$, in millimeters (lower is better), versus optimizer updates. Results average five generation seeds within each of 12 common cases, then average across cases.

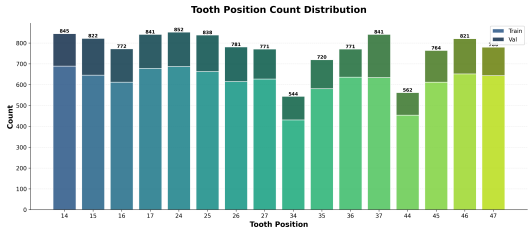


Figure 7: Distribution of training/validation samples across different tooth positions in the dataset

C Dental Background Related to Crowns

A clinically valid crown must satisfy several morphological and functional criteria. Figure 8 shows an example of a *good crown* with proper marginal adaptation, occlusal anatomy, and proximal contacts.

D Gallery

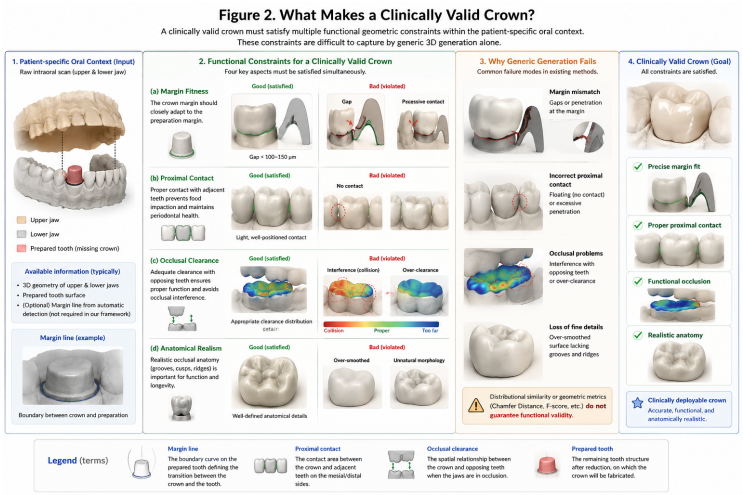


Figure 8: What Makes a Clinically Valid Crown?

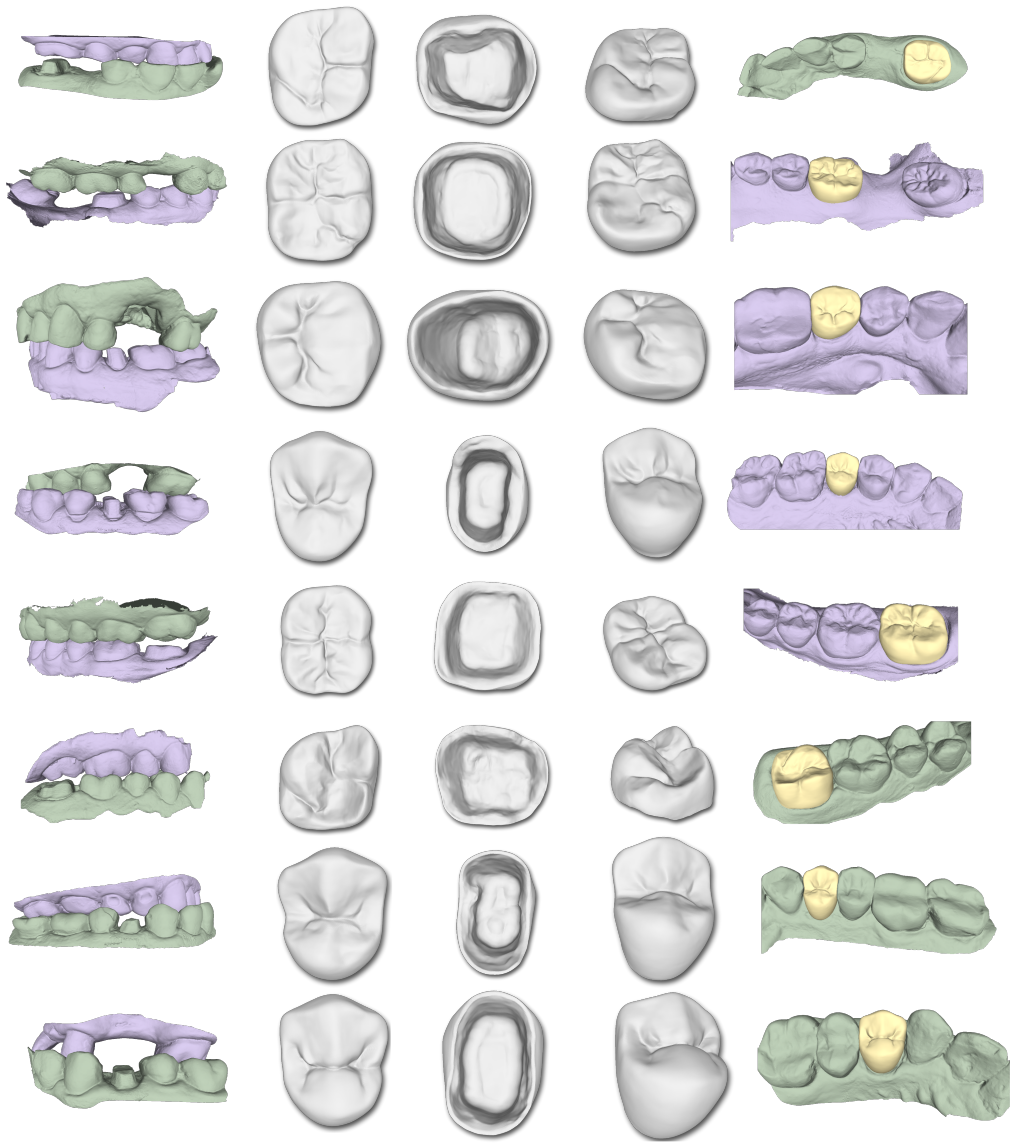


Figure 9: Visual gallery of our CrownGen. The first column illustrates the target anatomy, where green and purple denote the upper and lower jaws, respectively. Columns 2-4 present multi-view renderings of our generated dental crowns. The final column showcases the integrated results, demonstrating the crowns fitted onto their corresponding jaws.