CACR-Net: Constraint-Aware Occlusion-Guided Dental Crown Reconstruction with Latent SDF Diffusion

Wei Guo
School of Computer Science
Shenyang Aerospace University
Shenyang, China

Shilin Chen
School of Computer Science
Shenyang Aerospace University
Shenyang, China

Chen Yu
School of Computer Science
Shenyang Aerospace University
Shenyang, China

Ni An
Department of Prosthodontics
School of Stomatology, Capital Medical University
Beijing, China

Dan Meng
Department of Prosthodontics
School of Stomatology, Capital Medical University
Beijing, China

***Abstract*—Computer-aided design (CAD) and computer-aided manufacturing (CAM) of dental crowns often rely on library templates and manual adjustments, while existing deep-learning methods still struggle to reconstruct anatomically accurate occlusal surfaces from intraoral scans. We propose a constraint-aware dental crown reconstruction network (CACR-Net), an occlusion-guided two-stage framework that integrates geometric constraints with diffusion modeling for anatomically accurate and functionally plausible reconstruction. In Stage-1, the Curvature-Guided Mamba Dental Crown Network (CMDen-Net) extracts multi-resolution geometric features via a Serialized-Mamba Network (SMN) to predict an initial crown point cloud, supervised by a curvature-guided occlusal constraint (CGOC) and an SDF-based non-penetration loss. In Stage-2, SDFDiff-Net performs latent-space diffusion conditioned on the stage-1 output and antagonist teeth. It then decodes the denoised representation into a continuous SDF, from which Marching Cubes (MC) extracts the zero-level set to form a closed crown mesh. Evaluation on the public Teeth3DS+ dataset shows average reconstruction errors of 0.74 mm (CD-L1) and 0.59 mm (CD-L2), with higher occlusal accuracy and geometric consistency than existing deep-learning methods. CACR-Net demonstrates potential for integration into CAD workflows and for reducing the need for manual post-editing.**

***Index Terms*—Dental crown reconstruction, occlusal constraint, Serialized-Mamba, latent-space diffusion, prosthodontics.**

I. Introduction

Partial edentulism is a prevalent challenge in oral healthcare. The World Health Organization projects tooth loss in approximately 7% of adults aged $\geq$ 20 years by 2025, increasing to 23% among those over 60 [1]. Clinical surveys indicate that over 35% of adults have tooth-structure defects due to caries, trauma, or functional wear, which reduce masticatory efficiency and may cause tooth migration, alveolar bone resorption, pulp exposure, and root fracture. Fixed dental prostheses are the standard clinical approach. With the advancement of digital dental technologies, CAD/CAM-based crown reconstruction has become mainstream, but template-driven workflows remain time-consuming and require substantial manual adjustment. Therefore, an AI-based solution is needed to improve reconstruction efficiency and quality.

Deep learning has shown considerable promise in 3D dental crown reconstruction. Early works [2]–[5] rely on 2D depth maps of local tooth context: conditional generative adversarial networks (CGANs) use them to generate dental crown depth maps or 3D dental crown geometry, but the intermediate 2D representation provides only partial 3D geometric information and limits the fidelity of complex dental crown anatomy. To address these limitations, recent methods [6]–[8] reconstruct dental crowns directly from 3D intraoral scans by extracting geometric features from intraoral point clouds. These methods improve anatomical realism and reduce dependence on template-based design, but most reported systems still target specific crown types and do not introduce explicit guidance on occlusal surface morphology or non-penetration.

Dental crown reconstruction still faces three challenges: limited solutions that cover all types of dental crowns; difficulty in accurate 3D crown modeling from intraoral scans; and a lack of explicit constraints on occlusal contacts and spatial relationships. To address these challenges, an occlusion-guided two-stage framework (CACR-Net) is introduced. The main contributions are as follows.

- To enable dental crown reconstruction across all tooth types, CACR-Net is designed as a two-stage architecture to produce dental crown meshes that are directly usable in CAD/CAM.
- CACR-Net utilizes a Serialized-Mamba Network (SMN) for efficient point cloud encoding and performs conditional diffusion in a signed distance field (SDF) latent space for accurate, CAD-ready mesh generation.
- To impose explicit constraints on occlusal surface formation and to suppress penetration into neighboring denti-

tion, a curvature-guided occlusal constraint (CGOC) loss and an SDF-based non-penetration loss are proposed in CACR-Net.

II. Related Work

Existing learning-based methods for dental crown reconstruction can be broadly grouped into 2D depth-map based and 3D point-cloud based approaches. For 2D depth maps, CGAN-based methods generate crowns from depth images: Hwang et al. [2] generate 3D dental crowns directly from 2D depth images, Yuan et al. [3] reconstruct 3D crowns from anatomically consistent 2D depth maps, Tian et al. [4] propose DCPR-GAN with groove parsing and occlusal contact constraints, and Tian et al. [5] develop the dual discriminator GAN with dilated convolutions to improve morphological detail. Although these methods improve reconstruction quality, their reliance on 2D depth maps limits the fidelity of complex 3D geometries and fine-detail continuity.

Recent studies reconstruct dental crowns directly from 3D point clouds. Feng et al. [6] combine pose estimation and shape estimation networks for maxillary anterior crowns, but do not explicitly model occlusal-surface structures and generalize poorly across tooth types. Chau et al. [7] use a 3D-GAN for maxillary molars and achieve favorable geometric alignment, but remain constrained to single-tooth reconstruction. Hosseinimanesh et al. [8] integrate Transformer modules with differentiable Poisson surface reconstruction to produce high-resolution crown meshes, yet fail to accurately capture intricate local geometric features and lack explicit mechanisms for generating occlusal surfaces.

III. Methods

CACR-Net is an occlusion-guided two-stage framework for dental crown reconstruction from intraoral scans. In Stage-1, the Curvature-Guided Mamba Dental Crown Network (CMDen-Net) takes a single-arch intraoral scan as input and predicts an initial crown point cloud for the missing tooth site. In Stage-2, the Latent Conditional SDF Diffusion Dental Crown Network (SDFDiff-Net) performs conditional denoising in a signed distance field (SDF) latent space conditioned on the stage-1 predicted crown point cloud and antagonist teeth to reconstruct a closed triangular crown mesh that is directly usable in CAD/CAM workflows. The overall architecture is illustrated in Fig. 1.

A. Curvature-Guided Mamba Dental Crown Network (CMDen-Net)

1) Tooth-wise Iterative Farthest Point Sampling (TIFPS): Dental crown generation requires both global context and occlusal detail. CMDen-Net constructs three tooth-wise point sets (high, medium, and low resolution) from the single-arch intraoral scan. Each point set is obtained by applying tooth-wise iterative farthest point sampling (TIFPS), which performs iterative farthest point sampling independently within each tooth region.

2) Voxelization and Serialized-Mamba Network: Unlike Transformers that scale quadratically, Mamba-based state-space models offer linear-time complexity, making them highly efficient for capturing long-range geometric dependencies in dense, multi-tooth intraoral scans [9], [10]. Therefore, CMDen-Net adopts a Serialized-Mamba Network (SMN) backbone (Fig. 2).

To enhance local geometric learning, each point is represented by normalized coordinates and normal vectors (x, y, z, n_x, n_y, n_z). For each center point i, a dual-branch geometric enhancement module processes its k-nearest neighbors $\mathcal{N}_i$. The coordinate branch applies a pointwise MLP to neighbor coordinates, while the normal branch processes the normal-variation score $s_j = |1 - \langle n_i, n_j \rangle|$. Outputs from both branches are aggregated via max pooling and fused at the point level.

To enable sequence modeling, points are mapped to a unified V^3 voxel grid via $(x^v, y^v, z^v) = \mathrm{int}((x^s, y^s, z^s) \cdot V)$. This discrete grid supports four deterministic traversal schemes (Z-order, Hilbert, zigzag, raster) that flatten the 3D space into 1D sequences. Since each SMN block randomly selects a scheme, an explicit Order Index Prompt (OIP) is introduced. Each sequence is padded with an OIP embedding mapped from its serialization index, allowing the SMN to disambiguate traversal views and effectively exploit spatial cues. In each Mamba block, bidirectional SSMs are applied to capture features along the serialized sequence. A 1×1 Conv1D is applied for channel projection. The serialized sequence and its reversed copy are processed by SSMs in the forward and backward directions, and the two directional representations are fused into a single feature. Residual connections link adjacent blocks. When the input and output channels match, identity addition is applied. Otherwise, a 1×1 Conv1D is placed on the residual path to align dimensions before summation.

3) Hierarchical Decoding: CMDen-Net processes the three resolutions with separate SMNs. The resulting multi-resolution features are concatenated to form a global feature F. The global feature F is further transformed by several MLP layers to generate three features, which are denoted as F_1, F_2, F_3. These features are used in a hierarchical decoding strategy inspired by Feature Pyramid Networks [11] and Point Pyramid Decoder [12]. The high-level feature F_3 generates a low-resolution skeletal point cloud P_L, which represents the coarse structure. The medium-level feature F_2 generates a mid-resolution point cloud and adds it to P_L to form P_M, which is used to improve performance in local details. The low-level feature F_1 generates a high-resolution point cloud and adds it to P_M to form P_H, which contains the fine details.

4) Losses of CMDen-Net: CMDen-Net is optimized with a composite objective that aligns with three task constraints: overall geometric fidelity across resolutions, occlusal-surface accuracy, and non-penetration with neighboring dentition and antagonist teeth. The objective comprises a multi-resolution Chamfer Distance (CD) loss, an SDF-based non-penetration loss, and a curvature-guided occlusal constraint (CGOC) loss.

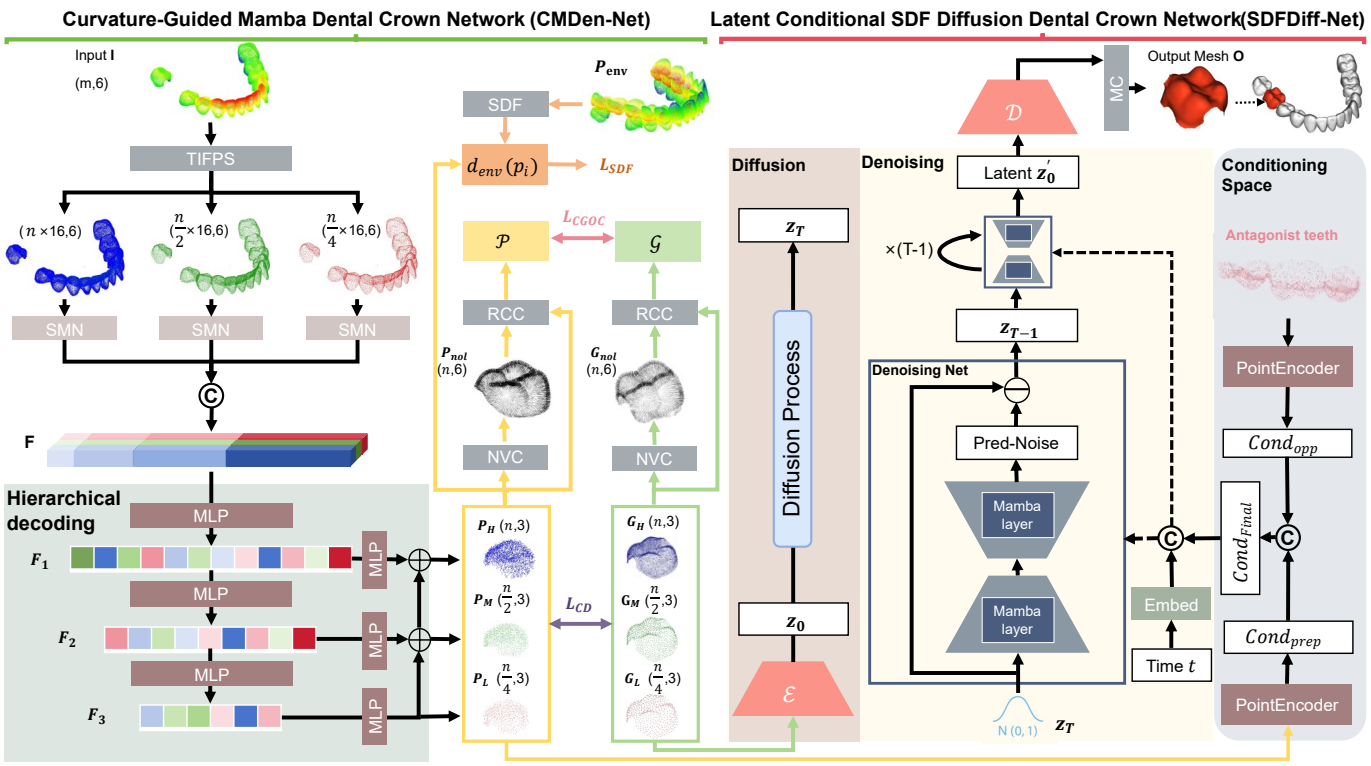

Fig. 1. Architecture of CACR-Net. CMDen-Net produces an initial crown point cloud; SDFDiff-Net performs latent-space denoising conditioned on the stage-1 output and antagonist teeth. Diffusion denotes forward noise injection during training, while denoising denotes iterative refinement during generation. **Abbreviations**: TIFPS: Tooth-wise Iterative Farthest Point Sampling; SMN: Serialized-Mamba Network; RCC: Relative Coordinate Concatenation; NVC: Normal Vector Calculation; MC: Marching Cubes.

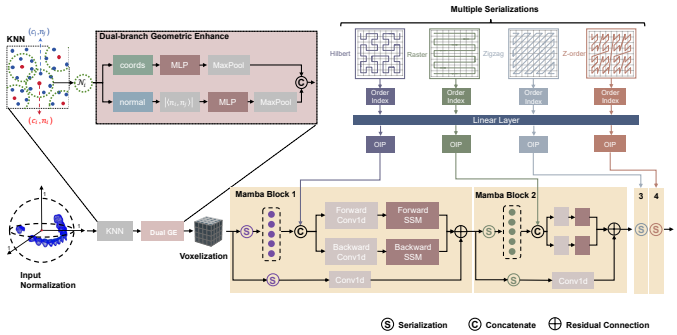

Fig. 2. The architecture of Serialized-Mamba Network. Different colors represent different serialization schemes.

To quantify the geometric discrepancy between the predicted and ground-truth crowns, a multi-resolution CD loss is adopted. For a predicted point set P and a ground-truth point set G, the squared CD loss is defined as:

$$d_{\mathrm{CD}}(P,G) = \frac{1}{|P|}\sum_{x\in P}\min_{y\in G}\|x-y\|_2^2 + \frac{1}{|G|}\sum_{y\in G}\min_{x\in P}\|y-x\|_2^2. \tag{1}$$

CMDen-Net generates predicted point clouds at three resolutions, which are denoted as P_H, P_M, and P_L. The corresponding ground-truth point clouds are obtained by using IFPS on the full-resolution ground-truth, and are denoted as G_H, G_M, and G_L. The losses at the high, medium, and low resolutions are denoted as $d_{\mathrm{CD}}^H, d_{\mathrm{CD}}^M$, and d_{CD}^L. The multi-resolution CD loss is defined as:

$$L_{\mathrm{CD}} = d_{\mathrm{CD}}^H(P_H,G_H) + \alpha_L\, d_{\mathrm{CD}}^L(P_L,G_L) + \alpha_M\, d_{\mathrm{CD}}^M(P_M,G_M), \tag{2}$$

where α_L and α_M are weights for the low-resolution and medium-resolution terms.

To prevent penetration between the predicted crown and surrounding dentition, an SDF-based non-penetration loss is introduced. The SDF is a scalar function defined in 3D space that measures the shortest distance from any query point to a surface, with positive values outside the mesh, zero on the surface, and negative inside. The environment M_{env} is formed by the surfaces of teeth in the target arch and the antagonist tooth surfaces. The environment mesh is used exclusively for loss supervision in the training phase to provide geometric supervision against collisions. It is not provided as an additional input to CMDen-Net. Let $d_{\mathrm{env}}(x) = \mathrm{SDF}_{M_{\mathrm{env}}}(x)$ denote the signed distance field of environment mesh M_{env}. The field is computed with AABB-accelerated nearest-triangle queries, and the sign is determined by the generalized winding

number [13]. The predicted dental crown point cloud is denoted as $P = \{p_i\}_{i=1}^{N}$. The loss L_{SDF} penalizes penetration cases defined by $d_{\mathrm{env}}(p_i) < 0$ as:

$$L_{\mathrm{SDF}} = \frac{1}{N} \sum_{i=1}^{N} \max\left(0, -d_{\mathrm{env}}\left(p_i\right)\right). \tag{3}$$

Occlusal morphology is characterized by high-curvature structures on the occlusal surface (cusps, grooves, and fossae). Pure CD supervision underweights these regions. A CGOC loss is introduced to provide explicit guidance on occlusal accuracy. Surface normal vectors are estimated and normalized to unit length for each point using a fixed-radius neighborhood. For a center point i, the k nearest neighbor set $\mathcal{N}_i$ is constructed. Unit normal vectors at center point i and each neighbor point $j \in \mathcal{N}_i$ are denoted n_i and n_j. Curvature for both predicted and ground-truth is estimated from surface normal vector variation. The curvature indicator C_i at point i is defined as:

$$C_i = \frac{1}{k} \sum_{j=1}^{k} \left(1 - \langle n_i, n_j \rangle\right). \tag{4}$$

After computing the curvature for each point, the predicted and ground-truth point clouds are independently ranked according to the curvature. The top-$\mathcal{T}$ points with the highest curvature are selected to form an occlusal-morphology subset for CGOC supervision. The selected points form the sets $P_{\mathrm{top}} = \{(p_i, C_i^P)\}_{i=1}^{\mathcal{T}}$ and $G_{\mathrm{top}} = \{(g_j, C_j^G)\}_{j=1}^{\mathcal{T}}$. For a given predicted point p_i, the Euclidean distance to each g_j is computed as $\{\,\mathrm{dist}_{ij} = \|p_i - g_j\|_2^2\,\}_{j=1}^{\mathcal{T}}$. These distances are converted into soft matching weights w_{ij} by a temperature-controlled SoftMax function:

$$w_{ij} = \frac{\exp(-\mathrm{dist}_{ij}/\tau)}{\sum_{k=1}^{\mathcal{T}} \exp(-\mathrm{dist}_{ik}/\tau)}, \tag{5}$$

where τ is a temperature parameter that controls the sharpness of matching. Each weight assigns greater importance to closer ground-truth points. The soft matching strategy avoids the gradient instability caused by hard assignments and enables stable training in high-curvature areas. The CGOC loss is computed as a weighted combination of curvature and positional differences between the matched high-curvature regions:

$$L_{\mathrm{CGOC}} = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} w_{ij} \left(\left|C_i^P - C_j^G\right| + \|p_i - g_j\|_2^2\right). \tag{6}$$

The total loss is a weighted sum of the multi-resolution CD loss, SDF loss, and CGOC loss:

$$L = \lambda_{\mathrm{cd}}\, L_{\mathrm{CD}} + \lambda_{\mathrm{cgoc}}\, L_{\mathrm{CGOC}} + \lambda_{\mathrm{sdf}}\, L_{\mathrm{SDF}}, \tag{7}$$

where λ_{cd}, λ_{cgoc}, and λ_{sdf} are weighting factors for L_{CD}, L_{CGOC}, and L_{SDF}.

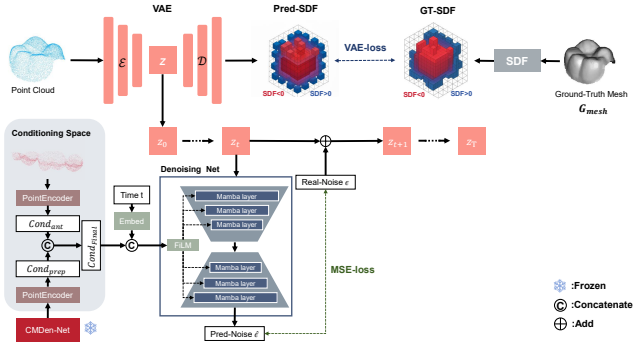

Fig. 3. SDFDiff-Net training workflow. CMDen-Net parameters are frozen.

B. Latent Conditional SDF Diffusion Dental Crown Network (SDFDiff-Net)

Clinical crown design requires a closed and smooth mesh that is directly usable in CAD/CAM workflows. To meet this requirement, SDFDiff-Net adopts conditional diffusion in a decodable latent SDF space [14], [15]. The SDF representation provides a continuous and topology-consistent description of crown geometry [16]. Marching Cubes (MC) extracts the zero-level set ($\mathrm{SDF}(x) = 0$) to form a closed triangular mesh. Conditioning is provided by the stage-1 predicted crown point cloud and antagonist teeth, which provide missing-site anatomical cues and occlusal context, respectively. The training workflow is shown in Fig. 3. During stage-2 training, the parameters of CMDen-Net remain frozen.

A variational autoencoder (VAE) is first trained to establish a decodable latent SDF space on ground-truth crowns. Given a ground-truth crown mesh G_{mesh}, the encoder ε produces a latent representation z. The decoder D maps z to a continuous SDF, denoted $D(z)$. A VAE loss is defined as:

$$\begin{aligned} L_{\mathrm{vae}} = \mathbb{E}_x \left\| D(z) - \mathrm{SDF}(G_{\mathrm{mesh}}) \right\|_1 \\ + \beta\, \mathrm{KL}(q_\phi(z \,|\, G_{\mathrm{mesh}}) \,\|\, p(z)), \end{aligned} \tag{8}$$

where $\mathrm{SDF}(G_{\mathrm{mesh}})$ denotes the ground-truth SDF operator, $q_\phi(z|G_{\mathrm{mesh}})$ denotes the encoder posterior, $p(z)$ is a zero-mean Gaussian prior, and β controls KL regularization.

Conditional diffusion is trained in the VAE latent space. Given a ground-truth crown mesh G_{mesh}, a latent representation $z_0 \sim q_\phi(z|G_{\mathrm{mesh}})$ is encoded. A forward diffusion process adds standard Gaussian noise to the latent representation at discrete timesteps. The denoising network is trained to approximate the reverse diffusion process and predicts the noise at each timestep.

Conditioning uses two point-cloud cues. The stage-1 predicted crown point cloud from CMDen-Net specifies a patient-specific geometric prior at the missing-tooth site. The antagonist teeth point cloud specifies the occlusal context that governs functional surface formation. These inputs reduce geometric ambiguity and encourage occlusal-consistent generation in the latent space. The predicted crown point cloud

and antagonist teeth point cloud are encoded by a PointNet-based encoder [17] into condition embeddings $Cond_{\text{prep}}$ and $Cond_{\text{ant}}$, respectively. Their concatenation forms the final condition $Cond_{\text{final}}$. The timestep embedding is concatenated with $Cond_{\text{final}}$ to form the context vector ctx_{emb}, which conditions the denoising network.

The denoising network comprises Mamba layers modulated by feature-wise linear modulation (FiLM). For the l-th layer, FiLM generates channel-wise scale and shift factors from ctx_{emb} and modulates the Mamba output as:

$$h^{(l+1)} = \text{Mamba}\Big(h^{(l)}\Big)\cdot\sigma(\text{Linear}(ctx_{\text{emb}}))+\text{Linear}(ctx_{\text{emb}})\,, \tag{9}$$

where $\sigma(\text{Linear}(ctx_{\text{emb}}))$ and $\text{Linear}(ctx_{\text{emb}})$ denote the learned scale factor and shift factor.

SDFDiff-Net is optimized by minimizing the MSE loss between the predicted noise $\hat{\epsilon}_\theta(z_t, t, Cond_{\text{final}})$ and the ground-truth noise ϵ:

$$L_{\text{diffusion}}(\theta) = \mathbb{E}_{x_0,\,\epsilon,\,t}\Big[\|\epsilon - \hat{\epsilon}_\theta(z_t,\, t,\, Cond_{\text{final}})\|^2\Big]\,. \tag{10}$$

During denoising, a latent noise $z_T \sim \mathcal{N}(0, I)$ is iteratively denoised from $t = T$ down to $t = 0$. At each step, the network removes a portion of the noise to obtain $\hat{z}_0$. The decoder evaluates the SDF from $\hat{z}_0$, and MC extracts its zero-level set to reconstruct a closed triangular crown mesh.

IV. Experiments

A. Experimental Setup

1) Dataset and Preprocessing: Experiments use the public Teeth3DS+ dataset [18], [19], which contains 1,800 intraoral scans from 900 patients with separate upper- and lower-jaw point clouds. The two arches are provided in a shared coordinate system, so no additional inter-arch rigid registration is applied and the antagonist point cloud is used as provided. Data were acquired and verified by experienced orthodontists and oral surgeons and anonymized to comply with the GDPR. Scans were captured by three intraoral scanners (Primescan, Dentsply; Trios3, 3Shape; iTero Element 2 Plus) with 10–90 μm spatial resolution. We split the dataset at the patient level into training/validation/testing sets (70%/10%/20%) with disjoint patients. Reconstructed and ground-truth crowns are represented by 2,048 uniformly sampled points per tooth, and all point clouds are normalized to $[0, 1]$. Teeth3DS+ provides per-tooth labels in the Fédération Dentaire Internationale (FDI) two-digit numbering system. During training, one labeled tooth is removed to simulate a missing-tooth input, and its point cloud and mesh are used as ground truth.

2) Implementation Details: All experiments are implemented in PyTorch and run on a single-GPU machine equipped with an NVIDIA RTX 4090. Training uses the Adam optimizer with an initial learning rate of 10^{-4}, a batch size of 16, and 300 epochs. For CMDen-Net, the kNN neighborhood size is set to $k = 32$; the voxel grid resolution is set to 64^3; and the OIP embedding dimension is set to 128. In the multi-resolution CD loss in (2), the weighting factors are set to $\alpha_L = 0.2$ and $\alpha_M = 0.4$. For CGOC loss in (6), the top-$\mathcal{T}$ high-curvature points are set to 30% and the SoftMax temperature is set to $\tau = 0.1$ in (5). For the overall training objective in (7), the loss weights are set to $\lambda_{cd} = 0.5$, $\lambda_{cgoc} = 0.3$, and $\lambda_{sdf} = 0.2$. For SDFDiff-Net, the VAE latent dimension is set to 512 and the KL regularization weight is set to $\beta = 10^{-5}$ in (8). The diffusion process follows the DDPM formulation [14] with $T = 800$ steps. During SDFDiff-Net training, the parameters of CMDen-Net remain frozen. Marching Cubes extracts the zero-level set on a 128^3 grid.

B. Evaluation Metrics

Reconstruction quality is assessed by Chamfer Distance (CD)-L1, CD-L2, and Earth Mover's Distance (EMD) [20] on uniformly sampled point sets. CD-L1 uses the $\|\cdot\|_1$ norm and CD-L2 uses the $\|\cdot\|_2$ norm; EMD measures the optimal-transport cost under the optimal one-to-one matching between two point sets. Unless otherwise specified, values in Tables I–IX are computed in normalized $[0, 1]$ space and multiplied by 1000. For clinical interpretability, crowns are denormalized and CD-L1/CD-L2 are recomputed in physical space (mm) without scaling; mesh-level accuracy is evaluated by MSE on 128^3 indicator grids.

V. Results

CACR-Net was evaluated on Teeth3DS+ under the patient-disjoint split and the protocol described in the previous section. Overall reconstruction performance is summarized in Table I.

TABLE I
Performance by tooth category on Teeth3DS+ (FDI).

Category	FDI IDs included	CD-L1	CD-L2	EMD	MSE
Incisors	11–12, 21–22, 31–32, 41–42	12.79	2.17	29.73	0.0016
Canines	13, 23, 33, 43	13.57	2.45	30.87	0.0020
Premolars	14–15, 24–25, 34–35, 44–45	14.12	2.48	31.76	0.0023
Molars	16–18, 26–28, 36–38, 46–48	15.04	2.69	31.92	0.0026
Overall	All test cases	13.92	2.31	30.22	0.0018

As shown in Table I, errors are lower for incisors and higher for molars, consistent with the increased morphological complexity and larger occlusal surfaces in posterior teeth. After denormalization, the average distances are 0.7392 mm (CD-L1) and 0.5923 mm (CD-L2). Fig. 4 shows reconstruction results across tooth types.

A. Comparison with Existing Reconstruction Methods

CACR-Net was compared with PF-Net [12], FSC [21], AdaPoinTr [22], and the dental crown baseline PDCNet [8]. PF-Net uses a multi-resolution encoder-decoder, FSC adds a two-stage WGAN refinement, AdaPoinTr uses a geometry-aware Transformer, and PDCNet integrates completion with Poisson surface reconstruction.

PF-Net, FSC, and AdaPoinTr were trained and evaluated on Teeth3DS+ using the official implementations under the same preprocessing pipeline and data split. Because the code for

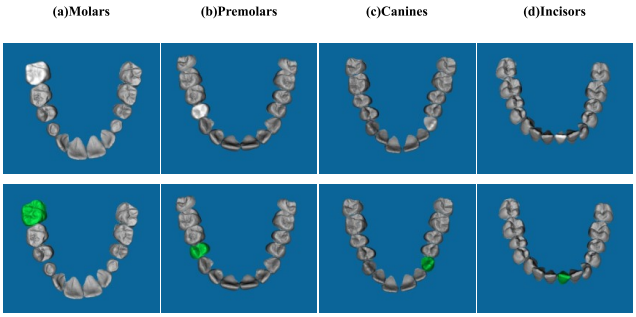

Fig. 4. Reconstruction examples across tooth types: (a) molars, (b) premolars, (c) canines, and (d) incisors. For each case, the top row shows the generated crown and the bottom row shows the ground truth.

PDCNet was not available, its published results are reported as reference values only; strict comparisons are restricted to baselines re-evaluated under the same protocol. The performance comparison is shown in Table II.

TABLE II
COMPARISON WITH EXISTING METHODS. "–" NOT APPLICABLE (NO MESH MSE FOR POINT-ONLY BASELINES).

Method	CD-L1	CD-L2	EMD	MSE
PF-Net	26.64	7.64	41.36	–
FSC	20.49	5.51	38.92	–
AdaPoinTr	19.11	4.80	36.29	–
PDCNet	54.39	8.41	75.31	0.0023
CACR-Net (Ours)	**13.92**	**2.31**	**30.22**	**0.0018**

Among baselines re-evaluated under the same protocol, CACR-Net achieved the lowest CD-L1, CD-L2, and EMD. PDCNet distances are larger because it did not normalize point clouds when computing these metrics, so its values are provided for reference only.

B. Ablation Analysis

Ablation experiments vary one component at a time while keeping all other modules, training settings, and losses identical to the main experiment.

1) Roles of the Two Stages: Three settings are evaluated: CMDen-Net point-cloud output, SDFDiff-Net conditioned only on antagonist teeth, and the full pipeline conditioning SDFDiff-Net on the stage-1 point cloud; meshes are uniformly sampled for point-domain metrics. As shown in Table III, the combined pipeline yielded lower CD-L1, CD-L2, and EMD than either single network. The mesh-level MSE was also lower. The SDFDiff-Net only variant showed higher point-domain errors than the combined pipeline.

2) Roles of SMN Backbone: An ablation experiment evaluated the backbone of CMDen-Net by comparing PointNet++ [23] (hierarchical sampling), CMLP [12] (multi-layer feature aggregation), DGCNN [24] (dynamic kNN EdgeConv), PointNeXt [25] (improved training and scaling), and Point Transformer [26] (self-attention). Results are reported in Table IV, and SMN achieved the lowest errors on all three metrics.

TABLE III
PERFORMANCE OF CMDEN-NET, SDFDIFF-NET, AND THE FULL PIPELINE.

Method	CD-L1	CD-L2	EMD	MSE
CMDen-Net	19.97	5.17	37.42	–
SDFDiff-Net	17.64	3.95	35.19	0.0027
CMDen-Net + SDFDiff-Net	13.92	2.31	30.22	0.0018

TABLE IV
BACKBONE COMPARISON FOR FEATURE EXTRACTION.

Method	CD-L1	CD-L2	EMD
PointNet++	16.24	3.04	34.32
CMLP	15.49	2.84	33.09
DGCNN	15.07	2.79	32.26
PointNeXt	14.83	2.42	31.91
Transformer	14.43	2.49	31.17
SMN	13.92	2.31	30.22

3) Roles of Dual-branch Geometric Enhancement: Three ablation settings were tested to assess the Geometric Enhancement Module. In Single Branch (Coord), coordinates were mapped by a point-wise MLP. In Single Branch (Coord+Normal), coordinates and normal vectors were mapped jointly by a single point-wise MLP. In Dual Branch, coordinates and normal vectors were mapped in separate branches and were fused. Results are summarized in Table V.

TABLE V
ABLATION OF BRANCH DESIGN IN THE GEOMETRIC ENHANCEMENT MODULE.

Method	CD-L1	CD-L2	EMD	MSE
Single Branch (Coord)	17.32	3.69	35.84	0.0029
Single Branch (Coord+Normal)	16.41	3.27	34.57	0.0021
Dual Branch	13.92	2.31	30.22	0.0018

As shown in Table V, Single Branch (Coord+Normal) yielded lower mesh-level MSE than Single Branch (Coord). This indicated that normal vectors provided useful directional cues for reconstruction. The dual-branch design further reduced the errors. A decoupled design was more effective than joint encoding and captured curvature-dominated details more consistently.

4) Roles of OIP: An ablation evaluated prompting for sequence disambiguation. Three variants were tested: No-Prompt, Order Prompt (OP), and OIP. The No-Prompt removed all sequence prompts. The OP baseline used an array prompt as described in [10]. The third ablation used OIP. Results are shown in Table VI. The OIP variant yielded lower CD-L1, CD-L2, and EMD than No-Prompt and OP.

5) Roles of the SDF-based Non-penetration Loss: Two ablation settings were tested: without the SDF loss and with the SDF loss. Penetration rate is defined as the percentage of predicted crown points $p_i \in P$ that satisfy $d_{env}(p_i) < 0$, where P denotes the stage-1 reconstructed dental crown point cloud. Results are reported in Table VII. With the SDF loss,

TABLE VI
PROMPTING ABLATION FOR SEQUENCE DISAMBIGUATION (NO-PROMPT, OP, OIP).

Method	CD-L1	CD-L2	EMD
No-Prompt	17.05	3.50	35.76
OP	16.12	3.16	34.38
OIP	13.92	2.31	30.22

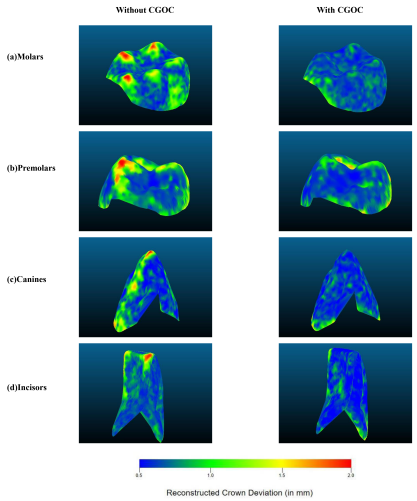

Fig. 5. Point-wise reconstruction error heatmaps with and without CGOC. (A) Molars; (B) Premolars; (C) Canines; (D) Incisors.

penetration rate, CD-L1, CD-L2, and EMD were lower.

TABLE VII
PERFORMANCE OF CACR-NET WITH AND WITHOUT SDF LOSS.

Method	CD-L1	CD-L2	EMD	Penetration (%)
Without SDF loss	16.26	3.19	34.62	10.8
With SDF loss	13.92	2.31	30.22	4.9

6) Roles of CGOC: As shown in Table VIII, CGOC led to lower CD-L1, CD-L2, and EMD. Fig. 5 shows point-wise reconstruction-error heatmaps for both settings. Warmer colors (red) indicate higher errors and cooler colors (blue) indicate lower errors. CGOC reduced errors on the occlusal surface.

TABLE VIII
PERFORMANCE OF CACR-NET WITH AND WITHOUT CGOC.

Method	CD-L1	CD-L2	EMD	MSE
Without CGOC	18.49	3.83	33.09	0.0032
With CGOC	13.92	2.31	30.22	0.0018

7) Roles of Antagonist-teeth Conditioning: Two ablation settings were tested for SDFDiff-Net: with and without antagonist-teeth conditioning. Only the conditional input changed. For Table IX, a z-based occlusal-proxy subset is defined as the top-40% of crown points ranked by their z-coordinates in the normalized-scan coordinate frame. Table IX reports results for the whole crown and the subset. Antagonist-teeth conditioning led to marginal changes in the whole dental

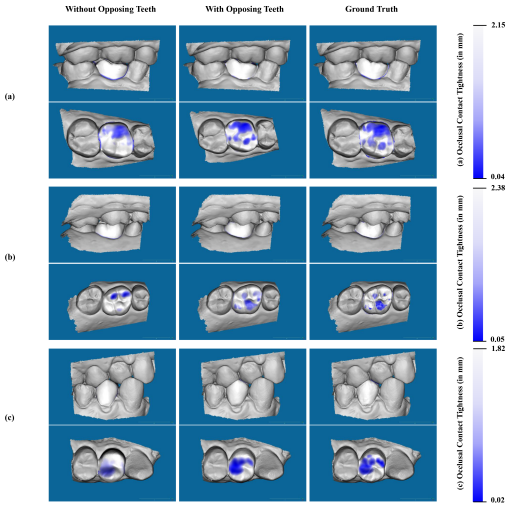

Fig. 6. Impact of antagonist teeth on occlusal surface error distribution. (A–C) Three representative test cases showing heatmap variations.

crown metrics but yielded a substantially larger improvement on the occlusal-proxy subset. That suggests that antagonist geometry primarily benefits occlusal-surface reconstruction. The smaller gains on the whole-crown metrics are consistent with the occlusal surface being a limited subset of the full crown.

Fig. 6 visualizes the minimum point-to-surface distances from each reconstructed crown to the adjacent or antagonist teeth. Compared with SDFDiff-Net without antagonist teeth, the conditioned variant produced heatmaps more similar to the ground-truth heatmaps, indicating improved occlusal-surface reconstruction.

VI. DISCUSSION

The results show that CACR-Net improves dental crown geometry, occlusal fidelity, and mesh-level feasibility under the evaluation protocol. CMDen-Net provides a patient-specific point-cloud prior, and SDFDiff-Net converts it into a CAD/CAM-ready closed mesh via latent SDF diffusion, consistent with the advantage of the full pipeline (Table III). SMN with mixed serialization and OIP improves multi-tooth feature encoding (Table IV, Table VI), while dual-branch geometric enhancement with CGOC emphasizes high-curvature occlusal anatomy and the SDF loss suppresses interpenetration (Table V–VII). Antagonist conditioning mainly benefits occlusal-surface reconstruction (Table IX, Fig. 6). Limitations include validation only on Teeth3DS+ under a single-arch protocol and reliance on labels for TIFPS. Future work will evaluate external datasets, reduce dependence on segmentation, and assess clinical workflow performance.

VII. CONCLUSION

This study presents CACR-Net, an occlusion-guided two-stage framework for dental crown reconstruction. By integrating a Mamba-based point cloud generator (CMDen-Net) with latent SDF diffusion (SDFDiff-Net), our method

TABLE IX
RECONSTRUCTION RESULTS FOR THE WHOLE CROWN AND THE OCCLUSAL PROXY SUBSET WITH AND WITHOUT ANTAGONIST TEETH.

Method	Dental crown			Occlusal-proxy Subset		
	CD-L1	CD-L2	EMD	CD-L1	CD-L2	EMD
Without antagonist teeth	14.22	2.78	31.33	18.62	5.34	38.07
With antagonist teeth	13.92	2.31	30.22	16.24	3.13	35.27

effectively maps intraoral scans to CAD-ready closed meshes. Extensive evaluations on the Teeth3DS+ dataset demonstrate that CACR-Net outperforms representative baselines in CD-L1, CD-L2, and EMD, particularly in high-curvature occlusal regions. Ablation studies validate the efficacy of the SMN backbone, explicit occlusal constraints (CGOC), SDF-based non-penetration loss, and antagonist-teeth conditioning. Ultimately, CACR-Net offers a highly accurate, contact-aware solution that bridges the gap between deep learning and clinical CAD/CAM workflows. The source code and pre-trained models are publicly available at https://github.com/Link10907/CACR-Net.

ACKNOWLEDGEMENTS

The authors used ChatGPT (OpenAI) to assist with English language editing and grammatical improvements in preparing this manuscript. All core concepts, methodology, results, and conclusions are solely the responsibility of the authors.

REFERENCES

[1] World Health Organization, "Oral Health," 2025. Available: https://www.who.int/news-room/fact-sheets/detail/oral-health.

[2] J.-J. Hwang, S. Azernikov, A. A. Efros, and S. X. Yu, "Learning Beyond Human Expertise with Generative Models for Dental Restorations," arXiv:1804.00064, 2018.

[3] F. Yuan et al., "A Full Crown Restoration Approach for Defect Teeth Based on CGAN," Journal of Computer-Aided Design & Computer Graphics, pp. 2113–2120, 2019, doi:10.3724/SP.J.1089.2019.17774.

[4] S. Tian et al., "DCPR-GAN: Dental Crown Prosthesis Restoration Using Two-stage Generative Adversarial Networks," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 1, pp. 151–160, 2022, doi:10.1109/jbhi.2021.3119394.

[5] S. Tian et al., "A Dual Discriminator Adversarial Learning Approach for Dental Occlusal Surface Reconstruction," Journal of Healthcare Engineering, vol. 2022, pp. 1–14, 2022, doi:10.1155/2022/1933617.

[6] Y. Feng et al., "3D reconstruction for maxillary anterior tooth crown based on shape and pose estimation networks," International Journal of Computer Assisted Radiology and Surgery, vol. 18, no. 8, pp. 1405–1416, 2023, doi:10.1007/s11548-023-02841-1.

[7] R. C. W. Chau, R. T.-C. Hsung, C. McGrath, E. H. N. Pow, and W. Y. H. Lam, "Accuracy of artificial intelligence-designed single-molar dental prostheses: A feasibility study," The Journal of Prosthetic Dentistry, 2023, doi:10.1016/j.prosdent.2022.12.004.

[8] G. Hosseinimanesh, A. Alsheghri, J. Keren, F. Cheriet, and F. Guibault, "Personalized dental crown design: A point-to-mesh completion network," Medical Image Analysis, vol. 101, p. 103439, 2024, doi:10.1016/j.media.2024.103439.

[9] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," arXiv:2312.00752, 2024.

[10] T. Zhang et al., "Point Cloud Mamba: Point Cloud Learning via State Space Model," in *Proc. AAAI Conf. Artif. Intell.*, 2025, pp. 10121–10130, doi:10.1609/aaai.v39i10.33098.

[11] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. CVPR*, 2017, doi:10.1109/cvpr.2017.106.

[12] Z. Huang, Y. Yu, J. Xu, N. Feng, and X. Le, "PF-Net: Point Fractal Network for 3D Point Cloud Completion," in *Proc. CVPR*, 2020, doi:10.1109/cvpr42600.2020.00768.

[13] G. Barill, N. G. Dickson, R. Schmidt, D. C. Levin, and A. Jacobson, "Fast winding numbers for soups and clouds," vol. 37, no. 4, pp. 1–12, 2018, doi:10.1145/3197517.3201337.

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," arXiv:2006.11239, 2020, doi:10.48550/arXiv.2006.11239.

[15] G. Chou, Y. Bahat, and F. Heide, "Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions," in *Proc. ICCV*, 2023, doi:10.1109/iccv51070.2023.00215.

[16] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation," in *Proc. CVPR*, 2019, doi:10.1109/cvpr.2019.00025.

[17] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *Proc. CVPR*, 2017, doi:10.1109/cvpr.2017.16.

[18] A. Ben-Hamadou et al., "Teeth3DS: a benchmark for teeth segmentation and labeling from intra-oral 3D scans," arXiv:2210.06094, 2022, doi:10.48550/arxiv.2210.06094.

[19] A. Ben-Hamadou et al., "Teeth3DS+" [Dataset], OSF, 2024. Available: https://osf.io/xctdy/.

[20] H. Fan, H. Su, and L. Guibas, "A Point Set Generation Network for 3D Object Reconstruction from a Single Image," in *Proc. CVPR*, 2017, doi:10.1109/CVPR.2017.16.

[21] X. Wu, X. Wu, T. Luan, Y. Bai, Z. Lai, and J. Yuan, "FSC: Few-Point Shape Completion," in *Proc. CVPR*, pp. 26077–26087, 2024, doi:10.1109/cvpr52733.2024.02464.

[22] X. Yu et al., "PoinTr: Diverse Point Cloud Completion with Geometry-Aware Transformers," in *Proc. ICCV*, pp. 12478–12487, 2021, doi:10.1109/iccv48922.2021.01227.

[23] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," arXiv:1706.02413, 2017, doi:10.48550/arxiv.1706.02413.

[24] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic Graph CNN for Learning on Point Clouds," ACM Transactions on Graphics, vol. 38, no. 5, pp. 1–12, 2019, doi:10.1145/3326362.

[25] G. Qian et al., "PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies," arXiv:2206.04670, 2022.

[26] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," Computational Visual Media, vol. 7, no. 2, pp. 187–199, 2021, doi:10.1007/s41095-021-0229-5.