VBCD: A Voxel-Based Framework for Personalized Dental Crown Design

Linda Wei[1(†)], Chang Liu[2,5(†)], Wenran Zhang[4], Zengji Zhang[6], Shaoting Zhang[5], and Hongsheng Li[1,3(∗)]

[1] Multimedia Laboratory (MMLab), The Chinese University of Hong Kong
[2] College of Biomedical Engineering, Fudan University
[3] Centre for Perceptual and Interactive Intelligence (CPII) under InnoHK
[4] Department of Second Dental Center, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine
[5] Sensetime Research
[6] School of Biomedical Engineering, Shanghai Jiao Tong University
1155230127@link.cuhk.edu.hk, 23110720100@m.fudan.edu.cn,
hsli@ee.cuhk.edu.hk

Abstract. The design of restorative dental crowns from intraoral scans is labor-intensive for dental technicians. To address this challenge, we propose a novel voxel-based framework for automated dental crown design (VBCD). The VBCD framework generates an initial coarse dental crown from voxelized intraoral scans, followed by a fine-grained refiner incorporating distance-aware supervision to improve accuracy and quality. During the training stage, we employ the Curvature and Margin line Penalty Loss (CMPL) to enhance the alignment of the generated crown with the margin line. Additionally, a positional prompt based on the FDI tooth numbering system is introduced to further improve the accuracy of the generated dental crowns. Evaluation on a large-scale dataset of intraoral scans demonstrated that our approach outperforms existing methods, providing a robust solution for personalized dental crown design. The related code is in: https://github.com/lullcant/VBCD

Keywords: Dental Crown Prosthesis · Point to Mesh Generation · Mesh Completion

1 Introduction

Oral diseases are among the most common non-communicable diseases worldwide, affecting an estimated 3.5 billion people [16]. Tooth injuries due to wear, caries, or trauma are the most prevalent forms of dental malfunction. For patients with abraded teeth, prosthodontics, specifically crown restoration, is a typical preferred treatment option [18]. Therefore, designing an artificial crown to restore

---

[†] Equal contribution.

[∗] Corresponding author.

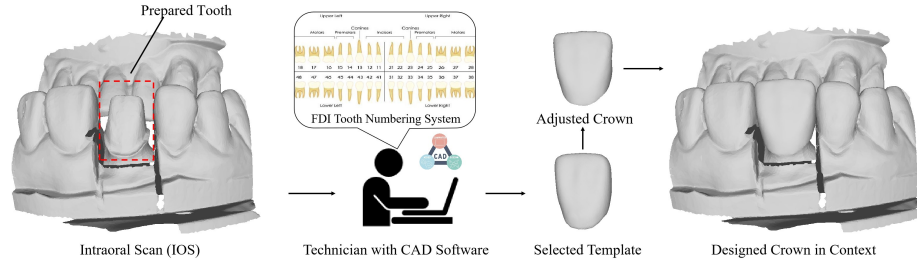[∗∗] This paper has been accepted in MICCAI 2025.

Fig. 1: The CAD workflow for dental crown design initiates with the acquisition of an intraoral scan, upon which the technician adjusts the template crown selected according to Fédération Dentaire Internationale (FDI) tooth numbering system.

dental function is critical to clinical dental practice. Modern dental computer-aided design (CAD) systems have greatly advanced the design of dental crowns [2,10,15]. As illustrated in Fig. 1, dental technicians select an appropriate crown template for the prepared tooth according to the Fédération Dentaire Internationale (FDI) tooth numbering system [9], which is then modified based on the patient's intraoral scan (IOS) to fabricate a restoration crown.

Despite their advantages, the workflow of CAD systems for crown restoration remains labor-intensive due to the lack of customization of crown templates. Dental technicians are required to make extensive and intricate modifications to the templates of target teeth, carefully adjusting them to accommodate the positions and occlusal relationships of adjacent and opposing teeth [21]. Therefore, developing an automated method capable of generating customized crown templates or final crowns is essential to alleviate the workload of dental technicians.

Recently, with the development of artificial intelligence for clinical application [22], the use of deep learning techniques for prosthesis design has emerged as a promising avenue for exploration. For instance, 3D-CNN and GANs have been employed to generate partial dental crowns and occlusal surface reconstructions [3,5,20]. However, these approaches often suffer from low accuracy and a lack of customization, primarily due to their inability to account for antagonist teeth. Recent studies [6,7,25] have framed the task of customized dental crown generation as a mesh completion problem, whose objective is to generate a dental crown mesh that restores the abraded areas with the information from the IOS. Most of the deep neural network models are not well-suited for handling mesh data, so the mesh completion task is usually decomposed into two main steps: point cloud completion and mesh reconstruction.

Although many studies have demonstrated excellent performance in point cloud completion [19,24,26], most of them do not take normal vectors of the completed points into account, which are essential for many mesh reconstruction algorithms. To solve this limitation, Shape as Points (SAP) [17] incorporates a novel module that combines an encoder-decoder model for normal vector prediction with a Differentiable Poisson Surface Reconstruction (DPSR) module. This module facilitates the reconstruction of the mesh surface from the point cloud

by constructing a Poisson Surface Reconstruction (PSR) indicator function grid and applying the Marching Cubes algorithm [14] to it.

Inspired by this study, several works [6,7,25] achieved end-to-end mesh completion for the customized dental crown by combining a point cloud completion network with SAP. Despite their promising performance, there are two primary limitations in previous methods. **First**, the number of points in the generated crown mesh was fixed and limited, while it varies with the type of tooth in clinical practice. **Second**, most studies were conducted on small-scale datasets or only focused on molars. The evaluation of the robustness and generalization was insufficient.

In this paper, we propose **V**oxel-**B**ased network for dental **C**rown **D**esign (VBCD) to address these issues. We first convert the input IOS into volume and utilize a 3D UNet backbone to generate a coarse dental crown in the volume modality. To mitigate the loss of geometric details during the voxelization process, we adopt a coarse-to-fine strategy to refine the coarse dental crown, incorporating a distance-based loss function to further improve the performance. Additionally, we introduce a positional prompt based on the FDI tooth numbering system to improve the customization of the generated crown. Our contributions are summarized as follows:

- We propose VBCD, a coarse-to-fine framework that automatically generates personalized crowns based on an intraoral scan with a prepared tooth. The framework can process an arbitrary number of input and output points, which facilitates the generation of accurate and fine-grained dental crowns.
- We design a tooth position prompt based on the FDI tooth numbering system and utilize the Curvature and Margin Penalty Loss (CMPL) to further enhance the performance of the crown generation framework.
- We construct the most extensive oral scan dataset on dental crown generation tasks to date, encompassing a complete set of tooth types, and perform extensive experiments to demonstrate the robustness and generalization of our framework utilizing the dataset.

2   Methods

The overall architecture of VBCD is shown in Fig. 2. The framework follows a coarse-to-fine paradigm to generate fine-grained dental crowns according to the IOS. The coarse dental crown is initially generated in volume modality. Further optimization is performed on this structure with a point cloud representation.

2.1   Voxelization

The input of our framework is an IOS with the prepared tooth to be restored. To regularize the point cloud input and make our framework adaptable to an arbitrary number of input points, we transform the IOS point cloud data into a volume, which is a structured representation with spatial context for CNN. The
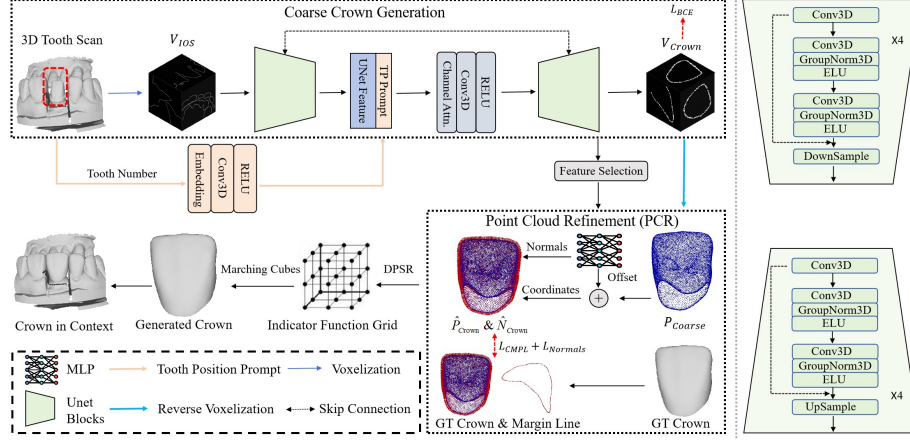
Fig. 2: Overall architecture of our VBCD. Given an IOS with a prepared tooth, our framework first generates a coarse crown in volume modality, then further refines the coarse result in point cloud representation through PCR.

given IOS is voxelized as a volume $V_{\text{IOS}} \in \mathbb{R}^{D \times H \times W}$ with the spacing s. The origin of the IOS bounding box (o_x, o_y, o_z) is also set to be the volume origin. For each point (x, y, z) in the IOS, the value of the corresponding voxel $V_{\text{IOS}}^{(i,j,k)}$ is set to be 1, where the index of voxel is $(i, j, k) = \lfloor \frac{(x,y,z) - (o_x, o_y, o_z)}{s} \rfloor$.

For volume ground truth, the dental crown is voxelized as a volume V_{GT} by applying the same procedure as V_{IOS}.

2.2   Coarse Crown Generation

Volumes voxelized from the IOS data are fed into a 3D UNet backbone for a coarse crown generation. We introduce a tooth position (TP) prompt to incorporate the context of tooth position. The label of the prepared tooth in the FDI tooth numbering system Fig. 1 is explicitly encoded as an embedding of 128 dims and then concatenated to the output of the bottleneck feature of UNet. Subsequently, a channel attention module [23] aggregates the TP prompt and the UNet feature, and a convolution layer is applied to ensure the compatibility of the feature with the shape required by the decoder.

The final output $f \in \mathbb{R}^{C \times D \times H \times W}$ of the backbone is fed into a convolution layer $Conv_{\text{coarse}}$. The output of $Conv_{coarse}$, which is the logits of crown volume, is denoted as $\hat{\mathbf{L}} \in \mathbb{R}^{1 \times D \times H \times W}$. We apply BCE loss to supervise the coarse crown generation, where σ is the sigmoid function.

$$\mathcal{L}_{BCE} = - \left(V_{GT} \cdot \log(\sigma(\hat{\mathbf{L}})) + (1 - V_{GT}) \cdot \log(1 - \sigma(\hat{\mathbf{L}})) \right)$$

The predicted coarse crown volume $V_{\text{Crown}} \in \{0, 1\}^{D \times H \times W}$ is a binary volume derived by thresholding $\hat{\mathbf{L}}$, setting voxels greater than zero to one. We obtain

the points on the coarse crown, denoted as $P_{\text{coarse}} \in \mathbb{R}^{N \times 3}$, through reverse voxelization of V_{Crown}.

2.3   Crown Generation Refinement

Although we can get a coarse generation result with the procedure above, there are two major issues that remain in the coarse crown generation: **1.** Voxelization inevitably results in the loss of fine geometric details from the original mesh point. **2.** BCE loss is a voxel-level loss that lacks the ability to provide distance-aware supervision. To address these issues, we introduced a point cloud refiner (PCR) to further refine the generated coarse crown P_{coarse}. The feature embedding $e \in \mathbb{R}^{N \times C}$ of the coarse crown point is gathered from the final layer feature f of UNet by the mask selection procedure:

$$\mathbf{F} = f \odot \mathbf{M} \in \mathbb{R}^{C \times D \times H \times W}, \quad e = \text{Flatten}(\mathbf{F}[\mathbb{I}_{\{\mathbf{F} \neq 0\}}]) \in \mathbb{R}^{N \times C},$$

in which $\mathbf{M} \in \{0,1\}^{C \times D \times H \times W}$ is the mask obtained from broadcasting V_{Crown} in the channel dimension. N is the number of points on the coarse crown, and $\odot$ is the Hadamard Product. e is employed to predict the offset between P_{Coarse} and the points in the ground truth dental crown P_{GT}. The feature embedding is also utilized to predict the normal vector of each point, which is indispensable for reconstructing the crown mesh.

The fine-grained dental crown point cloud $\hat{P}_{\text{crown}}$ and the normal vectors $\hat{N}_{\text{crown}}$ are computed using the following formula:

$$\hat{P}_{\text{crown}} = P_{\text{coarse}} + \text{MLP}_1(e), \quad \hat{N}_{\text{crown}} = \text{MLP}_2(e)$$

Motivated by [25], we introduce the Curvature and Margin Penalty Loss (CMPL), which better assists the model in generating crown details and ensuring more accurate delineation of the margin line, to supervise the fine-grained prediction. The CMPL is formulated as follows:

$$\mathcal{L}_{\text{CMPL}} = \frac{1}{|\hat{P}_{\text{Crown}}|} \sum_{\mathbf{p} \in \hat{P}_{\text{Crown}}} \left(e^{|\kappa(\mathbf{p})|} + \mathbb{I}_{\{\mathbf{p} \in M(P_{\text{GT}})\}} \right) \min_{\mathbf{q} \in P_{\text{GT}}} \|\mathbf{p} - \mathbf{q}\|_2$$
$$+ \frac{1}{|P_{\text{GT}}|} \sum_{\mathbf{q} \in P_{\text{GT}}} \left(e^{|\kappa(\mathbf{q})|} + \mathbb{I}_{\{\mathbf{q} \in M(P_{\text{GT}})\}} \right) \min_{\mathbf{p} \in \hat{P}_{\text{Crown}}} \|\mathbf{p} - \mathbf{q}\|_2$$

where $\kappa(\mathbf{p})$ is the curvature of point $\mathbf{p}$; $M(P_{GT})$ is the margin line points of P_{GT}, which is illustrated in Fig. 2. The normal vector $\hat{N}_{\text{crown}}$ of each point is supervised by the MSE loss denoted as $\mathcal{L}_{\text{Normals}}$. The ground truth normal vector of each point in the generated crown is defined as the normal vector of its nearest neighbor in the ground truth.

The overall loss of the proposed method is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{CMPL}} + \mathcal{L}_{\text{Normals}}$$

With the generated points and the corresponding normal vectors, it's handy to reconstruct the mesh of the dental crown with some plug-and-play surface reconstruction algorithms [8,11,12]. In this paper, we use the DPSR and Marching Cubes algorithms, which were also utilized by previous research, for fair comparison. Specifically, we first utilize the DPSR module in SAP to estimate an indicator function grid by solving the Poisson equation with the point cloud and its normals. The iso-surface from this indicator function grid is extracted as the reconstructed mesh with the marching cubes algorithm [14].
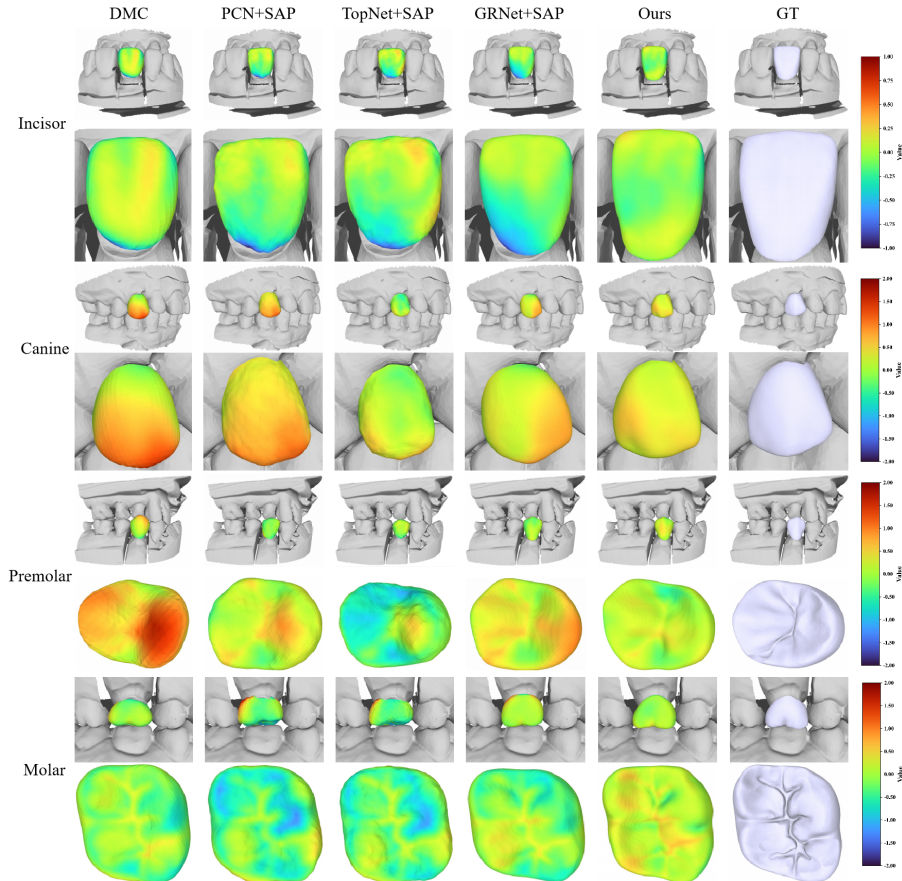
3   Experiments

Fig. 3: Comparison Experiment Results. The color map denotes the distance between a point in the generated crown and its nearest neighbor in the ground truth. No significant compromise of the occlusal relationship was detected.

3.1   Dataset and Implementation Details

Dataset. Our dataset comprises 6,499 oral scans with single-tooth edentulous, covering all the tooth types, including incisors, canines, premolars, and molars. The corresponding restoration crown of each oral scan is also included. For each oral scan, a 2cm × 2cm × 2cm cubic region centered on the crown's center is extracted. The margin line of the crown is defined as all the edges that belong to only one face in the crown mesh. The dataset is split into a training, validation, and test set with a ratio of 7:1:1. We employ stratified sampling to handle different tooth types, ensuring that the distribution of tooth types in the training and test sets remains consistent.

Implementation Details. Our framework was implemented using PyTorch and trained on 2 NVIDIA GTX 4090 GPUs with a batch size of 4 for 720000 iterations. AdamW was used as the optimizer, and the initial learning rate was set to be 10^{-4}. The inference time is about 357ms/case, which is much more efficient compared to the manual design (5-10 min/case)[15]. To minimize the quantization effects of the voxelization and keep geometric details [24], the input intraoral scan was voxelized as a 128^3 volume with a fine-grained spacing 0.15mm. The UNet backbone contained 4 downsampling and upsampling blocks; the base channel of feature C was 64. To enhance the stability of the training process, we employed a two-stage training strategy. During the initial stage, only L_{coarse} was applied as supervision. The $\mathcal{L}_{\text{CMPL}}$ and $\mathcal{L}_{\text{Normals}}$, which were designed for crown refinement, were incorporated in optimization after 400,000 iterations.

3.2   Results

Evaluation Metrics. To better assess the performance, all the distance-related metrics are calculated in the **physical coordinate**, which is measured in **millimeters (mm)**. We use L2 chamfer distance (CD-L2) and F-score to measure the similarity between the generated and the ground truth point cloud, as in the previous works [7,25]. Moreover, to better evaluate the consistency of the generated crown to the ground and reduce the impact of artifacts, we adopted fidelity [26] as an additional evaluation metric.

Comparison Experiment. We compared our method with the open-sourced crown generation method DMC [7] and some widely used point cloud completion networks with SAP. Notably, SAP is required for all methods as it generates the PSR indicator function grid necessary for mesh reconstruction. The qualitative results of the comparative experiments are presented in Table 1a. Our method outperforms previous approaches across all tooth types and evaluation metrics. For tooth types with limited data, such as canines, the performance degradation is minimal, demonstrating the robustness of our approach. The visualization results, shown in Fig. 3, indicate that our model exhibits higher similarity to the ground truth and produces more finely detailed crowns with greater precision than previous methods (Fig. 3 rows 2, 4, 6, 8). Specifically, the margin lines

Table 1: Experimental Results: (a) Comparison Experiment Results on All the Tooth Types, (b) Ablation Study on Tooth Numbering Prompt and CMPL

(a) Comparison Experiment Results on All the Tooth Types.

Metric	Method	Incisor	Canine	Premolar	Molar	Overall
CD-L2 $\downarrow$ (mm^2)	DMC [7]	0.390	0.621	0.363	0.362	0.375
	PCN+SAP [26]	0.367	0.471	0.345	0.347	0.354
	TopNet+SAP [19]	0.505	0.576	0.503	0.532	0.523
	GRNet+SAP [24]	0.300	0.328	0.288	0.285	0.290
	Ours	**0.161**	**0.177**	**0.138**	**0.133**	**0.140**
Fidelity $\downarrow$ (mm)	DMC [7]	0.361	0.458	0.363	0.384	0.377
	PCN+SAP [26]	0.335	0.358	0.318	0.336	0.332
	TopNet+SAP [19]	0.386	0.403	0.396	0.416	0.405
	GRNet+SAP [24]	0.273	0.267	0.258	0.280	0.273
	Ours	**0.217**	**0.225**	**0.216**	**0.210**	**0.213**
F-score $\uparrow$	DMC [7]	0.760	0.631	0.747	0.818	0.785
	PCN+SAP [26]	0.817	0.759	0.816	0.870	0.845
	TopNet+SAP [19]	0.744	0.708	0.700	0.769	0.745
	GRNet+SAP [24]	0.905	0.866	0.901	0.934	0.918
	Ours	**0.928**	**0.928**	**0.949**	**0.970**	**0.957**

(b) Ablation Study (only overall metrics are provided because of limited page count).

| Components | | | | CD-L2 (mm) $\downarrow$ | Fidelity (mm) $\downarrow$ | F-Score $\uparrow$ |
PCR	TP	Prompt	CMPL			
✗	✗		✗	0.230	0.314	0.896
✓	✗		✗	0.198	0.231	0.929
✓	✗		✓	0.154	0.216	0.934
✓	✓		✗	0.156	0.219	0.932
✓	✓		✓	**0.140**	**0.213**	**0.957**

of the crowns generated by our method are significantly more precise, which is critical for clinical applications (Fig. 3 rows 2, 4, 5, 7).

Ablation Study. We performed ablation studies in Table 1b to verify the effectiveness of our design. The baseline uses the UNet backbone to generate a restoration crown (point coordinates and normal vectors) directly. We evaluated the effect of components in our design, including (1) PCR, (2) TP Prompt, and (3) CMPL. For the model without CMPL, we used CPL in [25] as a distance-aware loss. In Table 1b, there is a performance improvement with each component added to the baseline. The VBCD, which is the ensemble of all the components and baseline, outperformed all the other experiment settings.

4   Conclusion

In this paper, we propose a coarse-to-fine framework for generating personalized dental crowns across all tooth types. Our framework leverages voxelization to regularize unordered point clouds and further incorporates CMPL and tooth position prompts to enhance the precision, detail, and fit of the margin lines of the generated crowns. Quantitative experiments and visualization results demonstrate that our method outperforms previous approaches across all metrics and is adaptable to all tooth types. However, our framework is limited by high memory consumption due to high-resolution voxelization [4]. Future work could address this issue by incorporating sparse convolutions [1,13] as the encoder in the UNet, improving computational efficiency.

Disclosure of Interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3075–3084 (2019)
2. Davidowitz, G., Kotick, P.G.: The use of cad/cam in dentistry. Dental Clinics of North America **55**(3), 559–ix (2011)
3. Farook, T.H., Ahmed, S., Jamayet, N.B., Rashid, F., Barman, A., Sidhu, P., Patil, P., Lisan, A.M., Eusufzai, S.Z., Dudley, J., et al.: Computer-aided design and 3-dimensional artificial/convolutional neural network for digital partial dental crown synthesis and validation. Scientific Reports **13**(1), 1561 (2023)
4. Fei, B., Yang, W., Chen, W.M., Li, Z., Li, Y., Ma, T., Hu, X., Ma, L.: Comprehensive review of deep learning-based 3d point cloud completion processing and analysis. IEEE Transactions on Intelligent Transportation Systems **23**(12), 22862–22883 (2022)
5. Feng, Y., Tao, B., Fan, J., Wang, S., Mo, J., Wu, Y., Liang, Q.: 3d reconstruction for maxillary anterior tooth crown based on shape and pose estimation networks. International Journal of Computer Assisted Radiology and Surgery **18**(8), 1405–1416 (2023)
6. Hosseinimanesh, G., Alsheghri, A., Keren, J., Cheriet, F., Guibault, F.: Personalized dental crown design: A point-to-mesh completion network. Medical Image Analysis p. 103439 (2024)
7. Hosseinimanesh, G., Ghadiri, F., Guibault, F., Cheriet, F., Keren, J.: From mesh completion to ai designed crown. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 555–565. Springer (2023)
8. Huang, J., Gojcic, Z., Atzmon, M., Litany, O., Fidler, S., Williams, F.: Neural kernel surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4369–4379 (2023)
9. ISO: Dentistry — designation system for teeth and areas of the oral cavity (2009), https://www.iso.org/standard/41835.html

10. Jain, R., Takkar, R., Jain, G., Takkar, R., Deora, N., Jain, R.: Cad-cam the future of digital dentistry: a review. IP Ann Prosthodont Restor Dent **2**(2), 33–6 (2016)
11. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. vol. 7 (2006)
12. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. ACM Transactions on Graphics (ToG) **32**(3), 1–13 (2013)
13. Liu, B., Wang, M., Foroosh, H., Tappen, M., Pensky, M.: Sparse convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 806–814 (2015)
14. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353 (1998)
15. Miyazaki, T., Hotta, Y., Kunii, J., Kuriyama, S., Tamaki, Y.: A review of dental cad/cam: current status and future perspectives from 20 years of experience. Dental materials journal **28**(1), 44–56 (2009)
16. Organization, W.H.: Global oral health status report: towards universal health coverage for oral health by 2030. World Health Organization (2022)
17. Peng, S., Jiang, C., Liao, Y., Niemeyer, M., Pollefeys, M., Geiger, A.: Shape as points: A differentiable poisson solver. Advances in Neural Information Processing Systems **34**, 13032–13044 (2021)
18. Raigrodski, A.J., Hillstead, M.B., Meng, G.K., Chung, K.H.: Survival and complications of zirconia-based fixed dental prostheses: a systematic review. The Journal of prosthetic dentistry **107**(3), 170–177 (2012)
19. Tchapmi, L.P., Kosaraju, V., Rezatofighi, H., Reid, I., Savarese, S.: Topnet: Structural point cloud decoder. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 383–392 (2019)
20. Tian, S., Huang, R., Li, Z., Fiorenza, L., Dai, N., Sun, Y., Ma, H.: A dual discriminator adversarial learning approach for dental occlusal surface reconstruction. Journal of Healthcare Engineering **2022**(1), 1933617 (2022)
21. Turkyilmaz, I., Wilkins, G.N., Varvara, G.: Tooth preparation, digital design and milling process considerations for cad/cam crowns: Understanding the transition from analog to digital workflow. Journal of Dental Sciences **16**(4), 1312 (2021)
22. Wang, G., Duan, Q., Shen, T., Zhang, S.: Sensecare: a research platform for medical image informatics and interactive 3d visualization. Frontiers in Radiology **4**, 1460889 (2024)
23. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11534–11542 (2020)
24. Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., Sun, W.: Grnet: Gridding residual network for dense point cloud completion. In: European conference on computer vision. pp. 365–381. Springer (2020)
25. Yang, S., Han, J., Lim, S.H., Yoo, J.Y., Kim, S., Song, D., Kim, S., Kim, J.M., Yi, W.J.: Dcrownformer: Morphology-aware point-to-mesh generation transformer for dental crown prosthesis from 3d scan data of antagonist and preparation teeth. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–119. Springer (2024)
26. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: 2018 international conference on 3D vision (3DV). pp. 728–737. IEEE (2018)