DuoDent: Tooth Generation using Dual-Stream Diffusion with Normal Consistency

Doeyoung Kwon[1], Seongjun Kim[1], In-Seok Song[2], and Seung Jun Baek[1(✉)]

[1] Korea University, Seoul, South Korea
{doeyoung, iamsjune, sjbaek}@korea.ac.kr
[2] Korea University Anam Hospital, Seoul, South Korea
densis@korea.ac.kr

Abstract. Generating high-precision 3D dental data is crucial for clinical practice, virtual simulation, and education. However, it is challenging to synthesize smooth and detailed tooth models. In this work, we introduce DuoDent, a dual-stream diffusion-based framework for the synthesis of accurate 3D tooth point clouds followed by a refined mesh generation. Our framework combines Transformer-based diffusion and CNN-based diffusion to capture both global dental structures and fine local features, thereby enhancing surface detail while reducing artifacts such as staircase and rough textures. The generated point clouds are optimized using normal consistency constraints for proper alignment of surface normals, which is key to high-quality mesh reconstruction. In addition, we apply a normal estimation with orientation consistency to the generated point clouds prior to converting them to output meshes, which enables the generation of smoother and anatomically precise tooth models. Extensive experiments validate that our method not only outperforms existing approaches in quantitative metrics but also delivers superior qualitative results, demonstrating its potential to significantly improve tooth modeling in dentistry. Our code is available at https://github.com/kdy-ku/DuoDent.

Keywords: Dual-Stream architecture · 3D Tooth Modeling · Surface Reconstruction · Point cloud Diffusion.

1 Introduction

Rapid advancement of digital dentistry has underscored the critical need for accurate and high-fidelity 3D dental models in clinical diagnostics, treatment planning, and educational simulations [3, 23]. Precise modeling of 3D teeth is essential for improving diagnostic accuracy and patient-specific simulations. In particular, the generation of 3D teeth with high fidelity and anatomical accuracy enables various downstream applications of digital dentistry. For example, 3D teeth generation can be used in VR simulators with haptic and HMDs for training for dental treatment and surgery [16, 19], by synthesizing anatomically accurate teeth for virtual patients [11]. The synthesis can be further conditioned on diseased or complex cases using LLMs. In addition, 3D teeth generation can

assist implant planning by conditioning the generation on the morphology of adjacent and opposing teeth.

However, reconstructing detailed and smooth 3D representations of teeth remains challenging due to the inherent complexity of dental anatomy and the limitations of conventional modeling techniques [1, 6], which often result in artifacts such as staircase effects and surface roughness [7, 23]. Recent progress in deep learning, particularly the emergence of diffusion models, has opened new avenues for generating high-quality 3D representations [17, 18, 20, 24, 28, 30]. By probabilistically modeling data distributions, diffusion-based approaches can synthesize complex geometries from noisy inputs [9, 17, 20, 30]. However, most existing diffusion methods struggle to simultaneously capture both the global structure and the fine local details intrinsic to dental morphology [4, 25]. Consequently, while these models may produce reconstructions that are globally coherent, they often lack the refined anatomical features necessary for reliable clinical applications.

In this paper, we propose a novel dual-stream diffusion-based framework that leverages two complementary diffusion processes to generate a 3D point cloud representation of teeth followed by a refined mesh generation. Specifically, our framework integrates Transformer-based diffusion to capture the overall shape and global context of the tooth, and CNN-based diffusion that focuses on preserving intricate details of local anatomy. This dual-stream approach not only enhances surface detail but also mitigates common reconstruction issues such as staircase artifacts and surface roughness [27]. A key innovation is the incorporation of normal consistency constraints [2] into the training of our model. By using a loss function derived from the constraints, our method improves the alignment of surface normals, leading to high-quality mesh reconstructions. In addition, we propose orientation consistency for normal estimation prior to converting point clouds to meshes in order for smoother mesh surfaces. Experiments show that DuoDent achieves superior performance in the synthesis of tooth meshes compared to baseline methods in both quantitative and qualitative aspects. Our contributions are summarized as follows: (1) introducing a dual-stream diffusion framework that combines a Transformer-based diffusion for capturing global structure and a CNN-based diffusion for preserving local details, (2) applying normal consistency to point cloud generation which can reduce artifacts, (3) enhancing mesh reconstruction through normal estimation with orientation consistency to produce smoother mesh surfaces.

2 Methods

2.1 Overview

A human adult has a total of 32 teeth. In dentistry, each tooth is assigned with a standardized number. Our goal is to generate a high-quality tooth mesh conditioned on the tooth number. DuoDent consists of two stages: Dual-Stream Diffusion and Surface Reconstruction (Fig. 1). In the first stage, a dual-stream

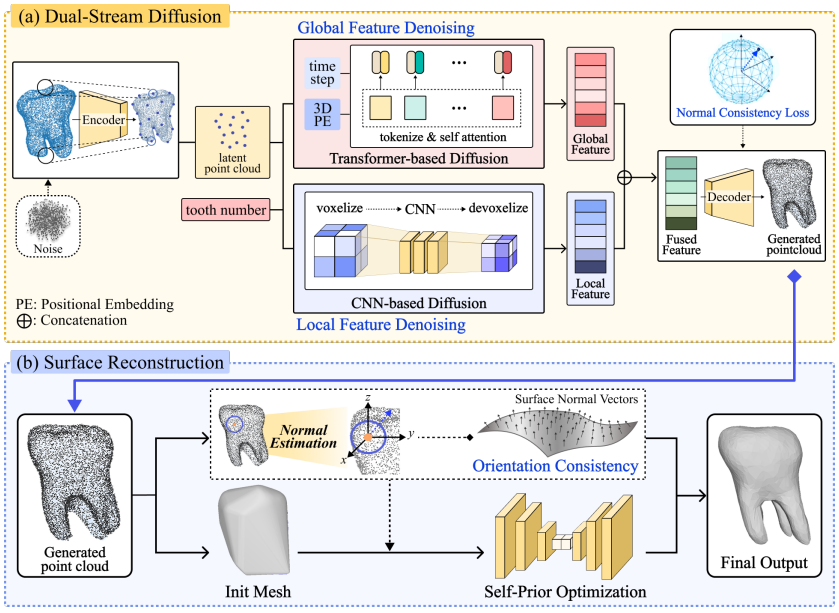

Fig. 1. DuoDent begins with (a) encoding the noisy input point cloud into a latent point cloud and processing it through two parallel branches—a Transformer-based diffusion to capture global structural context and a CNN-based diffusion to extract fine local details—resulting in a high-fidelity tooth point cloud; (b) from the generated point cloud, DuoDent draws an initial mesh and progressively refines it by re-estimating normals and applying vertex adjustments, yielding an accurate 3D tooth surface.

diffusion module integrates a Transformer-based diffusion with CNN-based diffusion to generate tooth structures as point clouds with normal consistency constraints to regulate the normals of generated point clouds. In the second stage, the generated point cloud is processed where its normals are re-estimated and its intrinsic geometric patterns are learned to progressively shrink-wrap an initial mesh, accurately reconstructing the tooth's 3D surface.

2.2 Dual-Stream Diffusion

The input point cloud consists of N points denoted by $X_0 \in \mathbb{R}^{N\times 3}$ where $N = 10^4$. The noisy point cloud $X_t \in \mathbb{R}^{N\times 3}$ is obtained by gradually corrupting X_0 over time steps $t \in [1, T]$ according to the standard DDPM process [9]. We map X_t into latent point cloud representation denoted by $Z_t \in \mathbb{R}^{M\times d}$ using a PointNet++ [22] encoder where the latent points are down-sampled from the original point cloud to $M = 64$ with dimension $d = 256$. The latent points are fed into two streams of denoising modules as follows.

Denoising Global Features. In the Transformer-based diffusion branch, the goal is to model global relationships in noisy 3D point clouds. Transformers use

self-attention mechanisms and thus are better at capturing global and long-range dependencies as compared to, e.g., CNNs [5]. The position-aware self-attention can incorporate positional relationships by reweighting attention scores using a position map derived from the embeddings. The input Z_t is firstly voxelized by discretizing the continuous point cloud into a grid-like structure. The voxels are then tokenized and are input to the Transformer layers where we used the learnable 3D positional encodings. The tooth number is encoded and input as conditions to the denoiser. For the actual implementation of Transformers, we adopted the network from DiT-3D [20] which we empirically found to be well-suited for our task. Finally, the Transformer branch produces a (noisy) global feature $Z_t^{\text{Tr}} \in \mathbb{R}^{M \times d}$ capturing rich global contextual information from the input latent point cloud.

Denoising Local Features. The latent representation Z_t is also input to the CNN-based Diffusion branch. The goal is to learn fine-grained local details from Z_t, exploiting the inductive bias of convolutional layers. For the implementation of CNN-based denoiser, we used Point-Voxel CNN (PVCNN) [14] similar to Point-Voxel Diffusion [30], which was empirically found to be suitable for our task. PVCNN is able to extract local neighborhood information through voxelization and 3D convolutional layers, which are combined with point features. Finally, the CNN branch yields $Z_t^{\text{CNN}} \in \mathbb{R}^{M \times d}$ which encapsulates intricate geometric information by capturing the proximal relationships between points.

Output. The global feature Z_t^{Tr} and the local feature Z_t^{CNN} are concatenated to form a fused feature representation denoted by $Z_t^F = Z_t^{\text{Tr}} \oplus Z_t^{\text{CNN}} \in \mathbb{R}^{M \times 2d}$. The fused feature is passed to a PointNet++ [22] decoder which up-samples the representation to generate the output point cloud conditioned on the given tooth number. The integration of global and local information in Z_t^F allows the decoder to produce point clouds that are both structurally coherent and rich in fine details. Overall, our dual-stream architecture synergistically combines Transformer-based Diffusion for capturing global context with CNN-based Diffusion for extracting local detail, which enhances the model's capability to generate detailed and realistic tooth point clouds.

Optimization. To train DuoDent, we combine two loss functions for the point cloud reconstruction and local surface consistency. The first loss is the MSE loss for denoising, which minimizes the difference between the model output (predicted noise) $\epsilon_\theta(X_t, t)$ and the ground truth noise ϵ:

$$L_{\text{MSE}}(t) = \|\epsilon - \epsilon_\theta(X_t, t)\|_2^2 \tag{1}$$

The second loss is for the local surface smoothness based on Normal Consistency Constraint (NCC) [2]. The NCC measures the variation in the similarity of normal vectors among neighboring points. The NCC loss is applied to the estimated output at all the diffusion timesteps as follows. Consider $\hat{X}_0(t)$ which is an estimated output of input point cloud X_0 at timestep t given by

$$\hat{X}_0(t) = (X_t - \sqrt{1-\bar{\alpha}_t}\ \epsilon_\theta(X_t, t))/\sqrt{\bar{\alpha}_t} \tag{2}$$

where $\bar{\alpha}_t$ is a parameter for the noise schedule [9]. For a point p in $\hat{X}_0(t)$, its normal $n(p)$ is estimated via a local plane fitting process based on Principal Component Analysis (PCA) [26] as follows. Consider local neighborhood $\mathcal{N}(p)$ of p is defined as either all points within a radius of 0.1 units or the 30 nearest neighbors, whichever criterion is satisfied first. The centroid of $\mathcal{N}(p)$, denoted by $\bar{p}$, is computed as the average of the positions of the neighboring points, and the covariance matrix is defined as:

$$C = \frac{1}{|\mathcal{N}(p)|} \sum_{q \in \mathcal{N}(p)} (q - \bar{p})(q - \bar{p})^T \tag{3}$$

The unit-length eigenvector corresponding to the smallest eigenvalue of C is defined as the normal estimate $n(p)$. The NCC loss is defined as:

$$L_{\text{NCC}}(t) = \frac{1}{N} \sum_{p \in \hat{X}_0(t)} \left(\sum_{q \in \mathcal{N}(p)} \left(n(p)^T n(q) - \mu(p) \right)^2 \right)^{\frac{1}{2}} \tag{4}$$

where $n(p)$ and $n(q)$ are the estimated normal vectors at point p and q from $\hat{X}_0(t)$, and $\mu(p)$ is defined by $\mu(p) = 1/|\mathcal{N}(p)| \sum_{q \in \mathcal{N}(p)} n(p)^T n(q)$. In our formulation, the NCC loss is computed at each timestep t to dynamically enforce normal consistency throughout the iterative refinement process. Thus, the NCC loss ensures that the estimated normals remain consistent and the local surface smoothness is maintained throughout the entire diffusion process. The overall loss is a weighted sum of these two components:

$$L = \mathbb{E}_{t \sim [1,T], \epsilon \sim \mathcal{N}(0,I)} \left[\lambda_{\text{MSE}} \cdot L_{\text{MSE}}(t) + \lambda_{\text{NCC}} \cdot L_{\text{NCC}}(t) \right] \tag{5}$$

with hyperparameters λ_{MSE} and λ_{NCC} balance the losses.

2.3 Surface Reconstruction

Normal Estimation with Orientation Consistency. We begin by properly estimating the normals of the point cloud generated from the diffusion stage. The estimated normals will be used later for refining the mesh output. For each point p from the generated point cloud, we compute an initial estimate of its normal $n(p)$ using PCA, or Eq. (3). Next, we propose to apply the *orientation consistency* across the entire point cloud for coherent normal orientations. The orientation consistency refers to the process of adjusting normals so that adjacent normals are uniformly aligned, thereby mitigating local estimation errors [10,12]. The concept can be used to achieve normal alignment robust to noise in the output generated from the diffusion models. We propose a simple method for orientation alignment as follows. For each point p, we refine the orientation of its normal $n(p)$ relative to the normals of its k nearest neighbors denoted by $\mathcal{N}_k(p)$. The adjustment is performed by aligning the normal direction with that of the majority of its neighbors given by

$$n(p) \leftarrow \text{sign} \left(\sum_{q \in \mathcal{N}_k(p)} n(p)^T n(q) \right) n(p) \tag{6}$$

Table 1. Quantitative comparison with baselines using 1-NNA (CD, EMD), Normal Consistency Loss (Normal.C), and F-score. The best values are highlighted in **bold**.

Method	CD (↓)	EMD (↓)	Normal.C (↑)	F-Score (↑)
LION [28]	0.885	0.829	0.816	0.594
SLIDE [18]	0.697	0.621	0.894	0.882
DiT-3D [20]	0.711	0.629	0.914	0.881
PVD [30]	0.637	0.631	0.913	0.823
DuoDent (Ours)	**0.557**	**0.532**	**0.926**	**0.912**

Table 2. Evaluation results on FairyTooth2 [29].

Method	CD (↓)	EMD (↓)	Normal.C (↑)	F-Score (↑)
LION [28]	0.891	0.856	0.798	0.554
SLIDE [18]	0.705	0.711	0.765	0.806
DiT-3D [20]	0.738	0.678	0.889	0.769
PVD [30]	0.686	0.697	0.801	0.795
DuoDent (Ours)	**0.629**	**0.597**	**0.919**	**0.891**

where Eq. (6) ensures that the direction of $n(p)$ is consistent with the predominant orientation in its local neighborhood. This two-stage process, local plane fitting with PCA followed by orientation consistency, produces accurately estimated and uniformly oriented normals suitable for further processing.

Self-Prior Optimization. After estimating the normal vectors, we reconstruct the surface mesh from the point cloud using Point2Mesh [8]. Starting with an initial watertight mesh, the method incrementally refines vertex positions through a CNN-based displacement model. The optimization minimizes a bidirectional Chamfer distance and incorporates a beam-gap loss to align the evolving mesh with the input point cloud [8]. This method employs weight-sharing and a coarse-to-fine refinement strategy to maintain local consistency and capture fine details, enabling mesh reconstruction adaptable to various point cloud conditions without reliance on external priors or extensive pre-training.

3 Experiments

3.1 Experimental Setup

Datasets. We utilize a dataset consisting of 2,255 tooth samples derived from 75 anonymized 3D dental cone-beam computed tomography (CBCT) scans, which are used to represent dental structures in voxel format. The CBCT scans were sourced from Korea University Anam Hospital and approved by the Institutional Review Board at Korea University (IRB No. 2020AN0410). Annotation processes involved rigorous cross-validation and were subsequently reviewed by specialists in oral and maxillofacial surgery. The dataset was divided into training, validation, and test subsets following an 8:1:1 ratio. Each individual tooth was extracted as tensors with dimensions of $96 \times 96 \times 96$, and corresponding

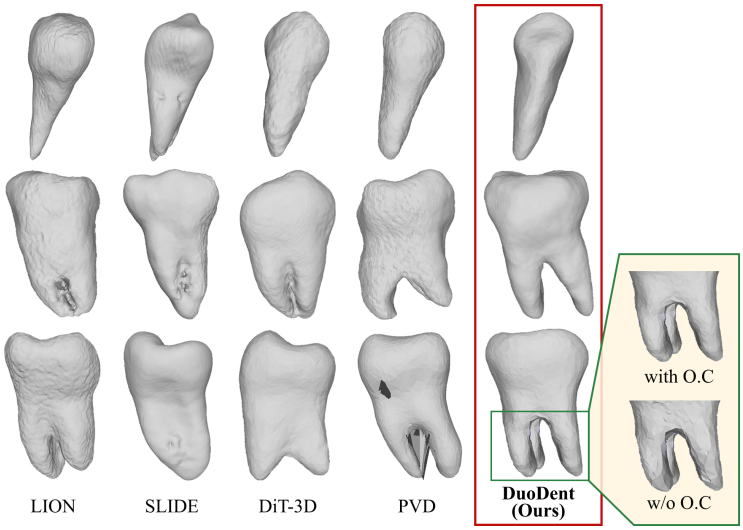

Fig. 2. Surface reconstruction results of the generated tooth. The rightmost part shows a comparison between the results with and without orientation consistency (O.C).

point clouds were densely sampled with 10^4 points. These point clouds were then scaled to fit within a three-dimensional range of $[-1, 1]^3$, enabling precise reconstruction.

Evaluation Metrics. We use four metrics widely used in recent point cloud and mesh generation studies [18, 20, 21]: 1-Nearest Neighbor Accuracy (1-NNA) [15] computed with both Chamfer Distance (CD) and Earth Mover's Distance (EMD), normal consistency, and F-score. The 1-NNA metric, calculated using both CD and EMD, assesses the distributional similarity between generated and ground truth point clouds, assessing both quality and diversity through a leave-one-out accuracy of a 1-NN classifier. Normal consistency evaluates the geometric fidelity of surfaces by measuring the alignment of normal vectors between generated and target point clouds. The F-score offers a balanced assessment of precision and recall, computed over various distance thresholds.

Implementation. The loss function weights in (5) were set to $\lambda_{\text{MSE}} = 1$ and $\lambda_{\text{NCC}} = 0.1$. The optimization was performed using the Adam optimizer [13] with a learning rate of 10^{-4} and a batch size of 16. For surface reconstruction, the number of neighbors (k) was set to 30, and the process was executed over 1,000 iterations. For the baselines, DiT-3D [20] was trained for 3,000 epochs, LION [28] for 35,000 iterations, and PVD [30] for 3,000 epochs. In the SLIDE [18], the latent position DDPM was trained for 56,000 iterations, the latent feature DDPM for 30,000 iterations, and the number of key points was set to 16.

3.2 Experimental Results

Quantitative results. We evaluated our method against state-of-the-art approaches (LION, SLIDE, DiT-3D, and PVD) using Chamfer Distance (CD), Earth Mover's Distance (EMD), Normal Consistency Loss, and F-Score (see Table 1). Our approach achieves a CD of 0.557 and an EMD of 0.532 (computed via 1-NNA), demonstrating superior point-wise accuracy and distribution alignment in reconstructing high-fidelity 3D dental structures. Moreover, our method consistently outperforms the baselines across all metrics. Notably, the highest Normal Consistency Loss (0.926) confirms the effectiveness of our normal optimization strategy in preserving local surface quality, which is crucial for generating meshes that are both geometrically accurate and visually smooth. Additionally, an F-Score of 0.912 underscores the robust structural coherence and precise reconstruction capabilities of our approach. To further validate generalizability of our model, we evaluated the methods on ToothFairy2 [29], which is a public dataset of CBCT images with a resolution similar to our data. As shown in Table 2, DuoDent achieved the best performance, demonstrating its effectiveness on diverse datasets.

Qualitative results. A visual comparison of generated samples (Fig. 2) reveals that DuoDent produces smoother surfaces, reduces staircase artifacts, and enhances anatomical details. Unlike baseline methods which suffer from surface roughness, structural distortions, or local inaccuracies, DuoDent captures both global structural coherence and local curvatures, yielding more realistic dental models. Our dual-stream diffusion framework appears to help the model both learn global structures and preserve fine details. The normal estimation and optimization process helps reduce high-frequency noise while maintaining sharp anatomical boundaries and continuous surfaces. Moreover, the result shows that using orientation consistency leads to smoother surfaces and better preserves structural details, e.g., see the lower right part of Fig. 2. A similar trend for qualitative improvement through orientation consistency was observed in the generated samples. These qualitative results demonstrate the effectiveness of DuoDent in generating accurate, high-resolution 3D dental models suitable for clinical and educational applications.

Ablation study. To validate the contributions of each component of DuoDent, we performed an ablation study (Table 3). The results indicate that the dual-stream architecture and normal consistency constraints collectively enhance the accuracy of reconstruction. Notably, integrating both the Transformer-based diffusion and CNN-based diffusion significantly improves performance over single-stream alternatives, reinforcing the necessity of a hybrid feature extraction.

4 Conclusion

In this study, we introduced a dual-stream diffusion framework for 3D tooth modeling. DuoDent combines a Transformer-based diffusion to capture global shapes and a CNN-based diffusion to preserve local details in generating point

Table 3. Ablation study. The best values are highlighted in **bold**. In this table, "Global" denotes the Transformer-based Diffusion Module, and "Local" denotes the CNN-based Diffusion Module.

Method	CD (↓)	EMD (↓)	Normal.C (↑)	F-Score (↑)
Global+NCC	0.666	0.592	0.903	0.840
Local+NCC	0.676	0.702	0.900	0.763
Global+Local	0.598	0.605	0.921	0.885
DuoDent (Ours)	**0.557**	**0.532**	**0.926**	**0.912**

clouds. Normal consistency constraints are applied to optimize point cloud quality and smooth surface reconstruction. Experiments show that our approach outperforms existing techniques in both quantitative and qualitative evaluations. The method effectively reduced artifacts while preserving essential anatomical features. An ablation study confirmed the complementary benefits of the dual-stream architecture and normal optimization. Overall, our framework offers a reliable solution for accurate and high-quality synthesis of 3D dental data.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2022R1A5A1027646), and the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Creative Consilience Program grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201819), and the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711195279 , RS-2021-KD000009), and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant Number: HI23C0162).

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

References

1. Chochlidakis, K.M., Papaspyridakos, P., Geminiani, A., Chen, C.J., Feng, I.J., Ercoli, C.: Digital versus conventional impressions for fixed prosthodontics: A systematic review and meta-analysis. The Journal of prosthetic dentistry **116**(2), 184–190 (2016)
2. Cui, R., Qiu, S., Anwar, S., Liu, J., Xing, C., Zhang, J., Barnes, N.: P2c: Self-supervised point cloud completion from single partial clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14351–14360 (2023)
3. Dawood, A., Marti, B.M., Sauret-Jackson, V., Darwood, A.: 3d printing in dentistry. British dental journal **219**(11), 521–529 (2015)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)

5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
6. Ender, A., Mehl, A.: Accuracy of complete-arch dental impressions: a new method of measuring trueness and precision. The Journal of prosthetic dentistry **109**(2), 121–128 (2013)
7. Etemad-Shahidi, Y., Qallandar, O.B., Evenden, J., Alifui-Segbaya, F., Ahmed, K.E.: Accuracy of 3-dimensionally printed full-arch dental models: a systematic review. Journal of clinical medicine **9**(10), 3357 (2020)
8. Hanocka, R., Metzer, G., Giryes, R., Cohen-Or, D.: Point2mesh: A self-prior for deformable meshes. arXiv preprint arXiv:2005.11084 (2020)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
10. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Surface reconstruction from unorganized points. In: Proceedings of the 19th annual conference on computer graphics and interactive techniques. pp. 71–78 (1992)
11. Hyun-Jae, C.: Metaverse and changes in oral health. Journal of Korean Academy of Oral Health **45**(4), 175–176 (2021)
12. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. vol. 7 (2006)
13. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Liu, Z., Tang, H., Lin, Y., Han, S.: Point-voxel cnn for efficient 3d deep learning. Advances in neural information processing systems **32** (2019)
15. Lopez-Paz, D., Oquab, M.: Revisiting classifier two-sample tests. arXiv preprint arXiv:1610.06545 (2016)
16. Luciano, C., Banerjee, P., DeFanti, T.: Haptics-based virtual reality periodontal training simulator. Virtual reality **13**, 69–85 (2009)
17. Luo, S., Hu, W.: Diffusion probabilistic models for 3d point cloud generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2837–2845 (2021)
18. Lyu, Z., Wang, J., An, Y., Zhang, Y., Lin, D., Dai, B.: Controllable mesh generation through sparse latent point diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 271–280 (2023)
19. Mauroux, M.: Virtual reality in restorative dentistry preclinical education: is the haptic simulator virteasy dental© a discriminating and valid tool to value student's performance? (2020)
20. Mo, S., Xie, E., Chu, R., Hong, L., Niessner, M., Li, Z.: Dit-3d: Exploring plain diffusion transformers for 3d shape generation. Advances in neural information processing systems **36**, 67960–67971 (2023)
21. 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)
22. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
23. Revilla-León, M., Özcan, M.: Additive manufacturing technologies used for processing polymers: current status and potential application in prosthetic dentistry. Journal of Prosthodontics **28**(2), 146–158 (2019)

24. Romanelis, I., Fotis, V., Kalogeras, A., Alexakos, C., Moustakas, K., Munteanu, A.: Efficient and scalable point cloud generation with sparse point-voxel diffusion models. arXiv preprint arXiv:2408.06145 (2024)
25. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
26. Rusu, R.B., Cousins, S.: 3d is here: Point cloud library (pcl). In: 2011 IEEE international conference on robotics and automation. pp. 1–4. IEEE (2011)
27. Shivakumar, S.S., Nguyen, T., Miller, I.D., Chen, S.W., Kumar, V., Taylor, C.J.: Dfusenet: Deep fusion of rgb and sparse depth information for image guided dense depth completion. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 13–20. IEEE (2019)
28. Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K., et al.: Lion: Latent point diffusion models for 3d shape generation. Advances in Neural Information Processing Systems **35**, 10021–10039 (2022)
29. Wang, Y., Zhang, J., Lee, J.H., et al. (eds.): The MICCAI 2024 Challenges: ToothFairy, 3DTeethLand, STS, and Cephalometric X-Ray, Lecture Notes in Computer Science, vol. 14548. Springer (2024), `https://link.springer.com/book/10.1007/978-3-031-88977-6`, proceedings of the MICCAI 2024 Challenges on Dental Data Analysis
30. Zhou, L., Du, Y., Wu, J.: 3d shape generation and completion through point-voxel diffusion. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5826–5835 (2021)