DentalPSAM: Lifting SAM to Dental Plaque Segmentation with Hybrid 2D-3D Knowledge Fusion

Haoning Jiang[1], Rachel Melissa Fok[1], Xiangcheng Zhang[1], Ruitong Sun[1], Jonathan Hok Kan YEUNG[1], Ke Deng[1⋆], and Liangqiong Qu[1⋆]

The University of Hong Kong, Hong Kong SAR
cdeng@hku.hk, liangqqu@hku.hk

Abstract. Accurate segmentation of dental plaque from intraoral scan (IOS) is crucial for assessing oral hygiene practices. However, the scarcity of 3D dental plaque datasets and the limitations of relying solely on either 3D geometric or 2D appearance information present significant obstacles. To address this, we introduce DentalPSAM, a hybrid 2D-3D model for 3D dental plaque segmentation. DentalPSAM features an interleaved dual-branch architecture with an appearance-enhanced 3D mesh branch and a geometry-enhanced 2D SAM branch, built upon the zero-shot learning capabilities of the Segment Anything Model (SAM). These branches process 3D IOS and its 2D rendered images in parallel, integrating both geometric and appearance information. To achieve better alignment between 2D and 3D features, we employ model stitching to connect SAM encoder with both branches through lightweight adaptation. Moreover, we present the first open-source 3D intraoral scan dataset for dental plaque segmentation. The code is available at https://github.com/HKU-HealthAI/DentalPSAM.

Keywords: Dental Plaque Segmentation · 2D-3D Feature Fusion · Intraoral Scan · Medical Image Segmentation

1 Introduction

Oral diseases, affecting nearly 3.5 billion people worldwide [21], represent a major global health burden. Dental plaque is the primary precursor to caries, gingivitis, and periodontitis [20]. Consequently, accurate and timely detection of dental plaque is paramount for effective clinical intervention and improved patient outcomes [28]. Recent advancements have enabled effective automatic 2D image-based plaque detection, utilizing both conventional photography [16] and fluorescence imaging [6]. While effective, these methods are limited to 2D images and cannot fully capture the three-dimensional nature of dental plaque.

In parallel, the adoption of 3D intraoral scanners (IOS) is rapidly expanding in modern dental practice, offering comprehensive and manipulable 3D representations of oral conditions for applications like impression-taking, orthodontic

⋆ Ke Deng and Liangqiong Qu are co-corresponding authors.

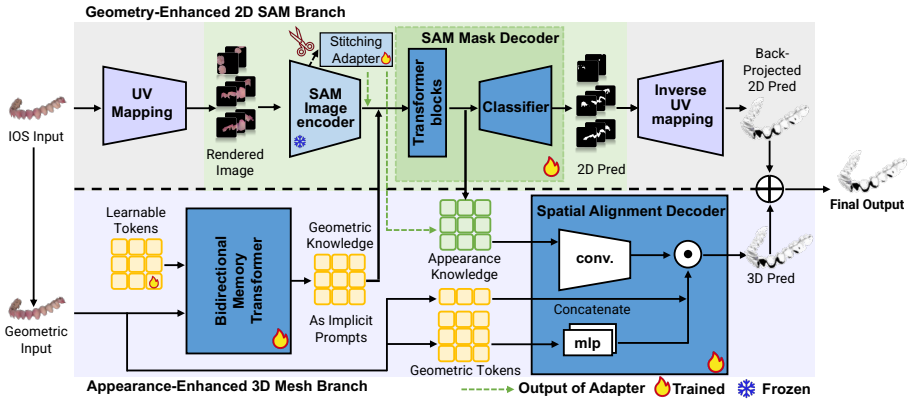

Fig. 1. Overall Architecture of DentalPSAM. Geometry-Enhanced 2D SAM Branch: UV-rendered images are encoded by a frozen SAM encoder; BMT converts 3D geometric coordinates into implicit prompts guiding the SAM mask decoder, yielding back-projected 2D predictions. **Appearance-Enhanced 3D Mesh Branch:** A lightweight spatial alignment decoder fuses SAM appearance features with 3D geometric tokens for direct mesh predictions. Dashed lines show the stitching adapter paths: SAM intermediate features are aligned with the 3D branch, adapter-refined dense prompts are injected into the 2D branch, and adapter-refined appearance features are sent to the 3D branch. The final output combines back-projected 2D and direct 3D predictions.

planning, and prosthodontics [2,11]. Existing 3D IOS segmentation techniques can be broadly categorized into two approaches: (1) **geometric-based methods** [29,17,31] that operate directly on 3D coordinates, normal vectors, and curvature. Despite their geometric precision, these methods rely purely on geometric information and demand substantial annotated 3D data. (2) **2D rendered image-based methods** [3,18,14] that convert the 3D mesh into multi-view 2D images to leverage powerful 2D segmentation models. While effective for appearance modeling, these methods suffer from loss of 3D spatial information during conversion.

Despite the progress in 3D IOS segmentation, directly applying these techniques to dental plaque segmentation faces inherent limitations: 1) **Lack of Annotated Data.** There is a lack of 3D IOS data specifically annotated for dental plaque [15,26,31,23]. Data acquisition requires clinical staining procedures, and expert annotation of a single patient takes 30–60 minutes, along with additional omission checks, making it far more demanding than standard IOS labeling. 2) **Insufficient Representation.** Unlike tooth structure segmentation, effective plaque segmentation requires both appearance (color, texture) and geometric information simultaneously. Geometric methods excel at 3D structure but fall short in appearance modeling, causing false detections at interdental regions where geometry is ambiguous. 2D methods handle appearance well but lose geo-

metric context, leading to missed detections at gum margins. Neither paradigm alone is sufficient for accurate plaque segmentation.

To address these challenges, we first introduce **PlaqueIOS**, the first open-source 3D IOS dental plaque dataset comprising 400 scans from 200 patients with expert clinical annotation, providing a foundation for this and future research. Building on this dataset, we propose **DentalPSAM**, a hybrid 2D-3D knowledge fusion model with an interleaved dual-branch architecture (Figure 1). The **geometry-enhanced 2D SAM branch** leverages the zero-shot learning capabilities of SAM [14], augmented by a bidirectional memory transformer (BMT) that encodes 3D geometric coordinates into implicit prompts, guiding SAM to capture appearance differences with geometric context. Concurrently, the **appearance-enhanced 3D mesh branch** employs a lightweight spatial alignment decoder to fuse appearance features from the 2D SAM branch with 3D geometric coordinates for direct mesh predictions. To further align the heterogeneous 2D and 3D feature spaces, we fine-tune a lightweight bidirectional stitching adapter after joint pre-training, enabling the two branches to mutually enhance each other with minimal additional parameters.

2 Related Work

In this section, we review works closely related to our study, including lifting the Segment Anything Model to 3D scenes and 3D intraoral scan segmentation.

Lifting SAM to 3D Scenes. SAM [14], trained on 11M images with 1B masks, demonstrates strong zero-shot segmentation. Medical adapters [10] and MedSAM2 [19] extend it to volumetric data. Yang et al. [25] merge multi-view SAM masks into 3D but ignore geometric information. Chen et al. [5] adapt SAM with 3D adapters, yet forcing heterogeneous 2D appearance and 3D geometric features to share a single encoder branch causes them to compete for representation capacity, limiting the specialization of each modality.

3D Intraoral Scan Segmentation. Traditional IOS segmentation employs curvature-based [27] and harmonic-field-based [30] methods, requiring extensive domain knowledge. Deep learning follows two paradigms: (1) 3D geometric methods [29,17,31,22] exploiting 3D coordinates and normals, and (2) 2D rendered image methods [3,18] converting meshes to multi-view images. CrossTooth [24] combines a Point Transformer with multi-view rendered color features, where appearance serves only as auxiliary cues to guide geometric boundary detection rather than as a primary modality for plaque-discriminative appearance modeling. None of these methods jointly leverages plaque-discriminative appearance representations and 3D geometry; our DentalPSAM bridges this gap via a dual-branch architecture.

3 Method

DentalPSAM employs an interleaved dual-branch architecture: a **geometry-enhanced 2D SAM branch** that leverages SAM's zero-shot appearance un-

derstanding guided by geometric prompts, and an **appearance-enhanced 3D mesh branch** that fuses 2D appearance features with 3D coordinates for direct per-face prediction. After joint pre-training, a lightweight bidirectional stitching adapter further aligns the two feature spaces.

3.1 Image Rendering via UV Mapping

To bridge 3D IOS data and the 2D SAM encoder, we convert each 3D IOS mesh into 2D rendered images via UV mapping. Each mesh is divided into three anatomically motivated parts: the Upper (occlusal biting surface), Inner (lingual/palatal surface), and Outer (buccal/facial surface), to maximally preserve geometric and color information during projection. The upper (occlusal) part uses planar mapping; the curved inner and outer parts use cylindrical mapping:

$$[u, v]^T = [x, z]^T \quad \text{(planar, upper part)} \tag{1}$$

$$[u, v]^T = [\arctan(z/x),\ y]^T \quad \text{(cylindrical, inner/outer)} \tag{2}$$

where $[u, v]^T$ and $[x, y, z]^T$ are 2D and 3D coordinates, and y aligns vertically with the inner/outer tooth faces. Each IOS yields 3 rendered images from distinct viewpoints. The inverse transformations recover 3D coordinates from 2D pixel positions, enabling bidirectional projection. The resulting rendered image I^R and its inverse-mapped 3D coordinate array I^G serve as paired inputs to the 2D SAM branch described next.

3.2 Geometry-Enhanced 2D SAM Branch

The 2D SAM branch is particularly effective in capturing subtle appearance cues in geometrically non-prominent regions where texture and color, rather than surface curvature, are the primary discriminative signals for plaque. This branch processes a rendered 2D image I^R through a *frozen* SAM image encoder to extract appearance embeddings E^R. Concurrently, the corresponding 3D coordinate array $I^G \in \mathbb{R}^{H\times W\times 3}$, where each valid pixel stores its recovered 3D mesh position via inverse UV mapping, is treated as geometry tokens and fed into the bidirectional memory transformer (BMT, Fig. 2) to generate geometric features F^G. A lightweight projection adapts F^G to the SAM embedding space, after which it is added element-wise to E^R to form geometry-enhanced image embeddings. These enhanced embeddings are passed to the *learnable* SAM mask decoder for 2D mask prediction, which is then back-projected to 3D via inverse UV mapping.

A key challenge is that UV-rendered images contain large background regions with no corresponding 3D geometry. Directly feeding these zero-coordinate background points into BMT would dilute the geometric signal and harm plaque localization. We therefore zero-pad background coordinates and reorder them to the front of I^G, so the LSTM's gating mechanism can selectively suppress them.
Bidirectional Memory Transformer (BMT).

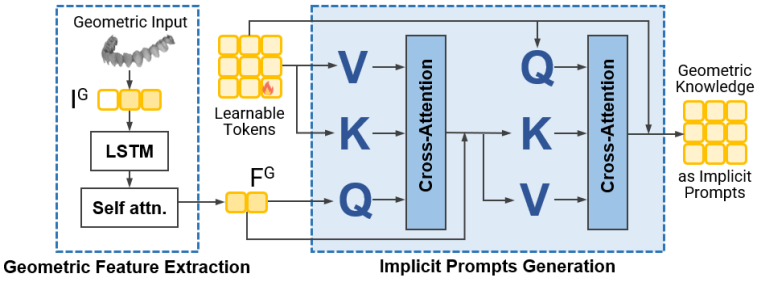

Fig. 2. BMT architecture. **Left (Geometric Knowledge Extraction):** LSTM suppresses zero-padded background entries; self-attention captures global geometric relationships to produce F^G. **Right (Implicit Prompts Generation):** Two bidirectional cross-attention passes project F^G into implicit geometric prompts via learnable tokens.

Step 1: Geometric Knowledge Extraction. We use LSTM [12] to compress I^G into a fixed-size geometric representation. The LSTM input gate $i_t = \sigma(W_i \cdot [h_{t-1}, I_t^G] + b_i)$ naturally suppresses zero background entries (when I_t^G=0, $i_t \approx \sigma(b_i) \to 0$), preserving geometric content without discarding boundary-region points. The compressed sequence is then refined by self-attention to capture global geometric relationships while maintaining permutation invariance. We denote the output as $F^G \in \mathbb{R}^{L\times d}$, a sequence of L=256 compressed geometric tokens each with d-dimensional features.
Step 2: Implicit Prompts Generation. To project F^G into 2D prompt space, we introduce learnable tokens T and apply two cross-attention passes in opposite directions. In the first pass, F^G serves as query and T as keys/values:

$$Q_1{=}W_Q^{(1)}F^G,\ K_1{=}W_K^{(1)}T,\ V_1{=}W_V^{(1)}T;\quad F^G \mathrel{+}= \text{softmax}\left(\frac{Q_1K_1^T}{\sqrt{d_k}}\right)V_1 \tag{3}$$

In the second pass, T serves as query and the updated F^G as keys/values:

$$Q_2{=}W_Q^{(2)}T,\ K_2{=}W_K^{(2)}F^G,\ V_2{=}W_V^{(2)}F^G;\quad T \mathrel{+}= \text{softmax}\left(\frac{Q_2K_2^T}{\sqrt{d_k}}\right)V_2 \tag{4}$$

The updated T evolves into implicit geometric prompts that capture the learned 2D projection of 3D geometry.

3.3 Appearance-Enhanced 3D Mesh Branch

While the 2D SAM branch excels at appearance-based segmentation, its back-projected 3D predictions can suffer from projection artifacts and lack direct 3D structural consistency. The 3D mesh branch addresses this by producing direct per-face predictions from 3D coordinates, enriched with appearance knowledge from the 2D SAM branch via a lightweight spatial alignment decoder.
Features Transformation. Re-projecting 2D appearance features into 3D space requires a spatial transformation. A fixed inverse UV matrix cannot adapt to

tooth morphology variations across patients, so the decoder learns this transformation adaptively: an MLP constructs transformation blocks from 3D geometric coordinates, while convolutional blocks progressively refine the 2D appearance features to align with them. This adaptive alignment enables the decoder to accurately capture complex spatial relationships across different tooth surfaces.
Confidence Filter. To prevent the 3D branch from making overconfident plaque predictions in ambiguous surface regions, a per-face confidence value C=0.5 is applied via element-wise multiplication to the final 3D predictions. This biases uncertain predictions toward negative, improving precision.

3.4 Bidirectional Model Stitching

After joint pre-training, the two branches remain in heterogeneous feature spaces, making direct full fine-tuning expensive and prone to overfitting. We adopt *model stitching* [9]: we measure linear transferability between SAM encoder layers and the mesh feature distribution to identify block 5 as the optimal stitching point, then insert a lightweight LoRA-based adapter [13] at this point. The adapter extracts SAM's intermediate visual features and simultaneously produces two outputs: one injected into the mesh branch to enrich 3D predictions with visual appearance cues, and one injected as dense prompts into the SAM mask decoder to refine 2D predictions. This bidirectional injection simultaneously enhances both branches from a single adapter, distinguishing our stitching from standard fine-tuning. Only the adapter parameters (∼0.87M, 0.76% of total) are updated; both branch parameters remain frozen.

3.5 Model Optimization

During training, both branches are simultaneously optimized with a composite loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{2D} + \lambda_2 \mathcal{L}_{3D}, \quad \lambda_1 = 1.0,\ \lambda_2 = 2.0 \tag{5}$$

The higher weight on $\mathcal{L}_{3D}$ reflects its role as the primary training signal for per-face prediction; $\mathcal{L}_{2D}$ serves as an auxiliary constraint that enforces 2D appearance consistency and stabilizes joint training. $\mathcal{L}_{2D}$ uses **DiceCELoss**:

$$\mathcal{L}_{2D} = 1 - \frac{2|\hat{Y} \cap Y|}{|\hat{Y}| + |Y|} + \alpha \cdot \text{CrossEntropy}(\hat{Y}, Y) \tag{6}$$

where $\hat{Y}$ and Y are the 2D predictions and ground truth before inverse UV mapping. $\mathcal{L}_{3D}$ uses **BCEWithLogitsLoss**:

$$\mathcal{L}_{3D} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\sigma(\hat{y}_i)) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))] \tag{7}$$

where $\hat{y}_i/y_i$ are per-face 3D predictions/ground truth, σ is sigmoid, and N is the number of mesh faces. During inference, we average the back-projected 2D probability map and the direct 3D probability map, $P_{\text{fuse}} = (P_{2D} + P_{3D})/2$, and threshold P_{fuse} at 0.5 for final 3D segmentation.

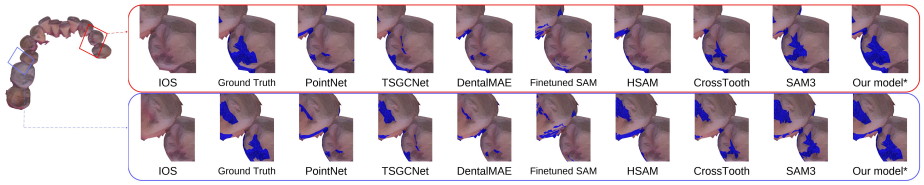

Fig. 3. Qualitative comparison of plaque segmentation results from different methods. Blue regions indicate predicted or annotated plaque areas. Our model shows more accurate plaque localization among the compared methods.

4 Experiments

4.1 Dataset and Implementation

Dataset. Under IRB approval with informed consent, we collected 400 IOSs from 200 patients using 3Shape TRIOS scanner(Copenhagen, Denmark) with Ci Double Plaque Checker disclosure. Labels were annotated in MeshLab [8] by two clinicians (5+ years), finalized at overlap $\geq 85\%$. After gum removal and downsampling to 16K cells, we split 220/60/120 IOSs for training/validation/testing.
Implementation. Adam optimizer (lr=$1e^{-4}$, batch size=4), SAM ViT-B backbone. Baselines: PointNet [22], TSGCNet [29], DentalMAE [1], fine-tuned SAM [14], H-SAM [7], CrossTooth [24], and fine-tuned SAM3 [4].
Evaluation Metrics. Following [28,6], we use **Intersection over Union (IoU)**, **Dice Similarity Coefficient (DSC)**, and **Overall Accuracy (OA)** for evaluation. IoU and DSC are reported per class (plaque, non-plaque).

4.2 Comparison with Baselines

Metrics	Region	[22]	[29]	[1]	[14]	[7]	[24]	[4]	DentalPSAM
IoU↑	Plaque	0.436	0.476	0.447	0.463	0.472	0.480	<u>0.513</u>	**0.556**
	Non-Plaque	0.709	<u>0.758</u>	0.751	0.745	0.700	0.746	0.693	**0.777**
	Overall	0.572	<u>0.617</u>	0.599	0.604	0.586	0.613	0.603	**0.667**
DSC↑	Plaque	0.601	0.626	0.611	0.623	0.635	0.645	<u>0.673</u>	**0.712**
	Non-Plaque	0.824	<u>0.858</u>	0.855	0.851	0.819	0.851	0.814	**0.872**
	Overall	0.712	0.742	0.733	0.737	0.727	<u>0.748</u>	0.744	**0.792**
OA↑	Overall	0.769	<u>0.807</u>	0.799	0.800	0.772	0.801	0.777	**0.831**

Table 1. Performance comparison on PlaqueIOS. IoU and DSC are reported for plaque, non-plaque, and overall regions; OA is reported overall. Higher metrics (↑) represent better performance. The highest performance is highlighted in **bold**, and the second-best is <u>underlined</u>.

As in Table 1, DentalPSAM achieves the best performance across all metrics. CrossTooth [24], which combines a Point Transformer with multi-view rendering

Variant	IoU	DSC	OA	Variant	IoU	DSC	OA
(a) w/o BMT	0.463	0.623	0.800	(b) w/ BMT	0.494	0.656	0.802
(c) 3D only	0.533	0.692	0.826	(d) w/o appear.	0.473	0.634	0.812
(e) w/o Stitch	0.547	0.700	0.830	(f) **Full**	**0.556**	**0.712**	**0.831**

Table 2. Comparison between variants of DentalPSAM. We evaluate six variants of our model: (a) 2D geometry-enhanced SAM branch without BMT. (b) Geometry-enhanced 2D SAM branch. (c) Appearance-enhanced 3D mesh branch. (d) Appearance-enhanced 3D mesh branch without appearance knowledge. (e) DentalPSAM without model stitching. (f) full DentalPSAM. IoU and DSC on plaque regions.

from six orthogonal views, represents a strong SOTA 3D method yet DentalPSAM surpasses it by 7.6% in plaque IoU. Meanwhile, DentalPSAM further outperforms SAM3, the most advanced 2D foundation model, by 4.3% in plaque IoU. Pure geometric methods (PointNet, TSGCNet) struggle with non-prominent plaque regions and pure 2D methods (fine-tuned SAM, H-SAM, SAM3) lose geometric context. DentalPSAM effectively addresses both limitations. Figure 3 further shows that our model achieves highly accurate segmentation among all compared methods in representative cases.

4.3 Ablation Study

Effectiveness of 2D SAM and 3D Mesh Branches. The final output of DentalPSAM is obtained by fusing back-projected 2D predictions from the 2D SAM branch and direct 3D predictions from the 3D mesh branch. To validate each branch, we report individual performances in Table 2 (b) and (c). The full model (f) outperforms each standalone branch by 6.2% and 2.3% in plaque IoU, confirming that the two branches provide complementary geometric and appearance information.

Effectiveness of BMT. We validate BMT by removing it from the 2D SAM branch. As shown in Table 2 (a) and (b), removing BMT drops plaque IoU by 3.1% and DSC by 3.3%, underscoring BMT's critical role in capturing and utilizing geometric knowledge for improved 2D predictions.

Effectiveness of Spatial-Alignment Lightweight Decoder. Our lightweight decoder's key feature is its ability to enrich 3D predictions by incorporating appearance knowledge from the 2D SAM branch. To assess this, we remove the appearance transfer from the 3D decoder (d) and compare against the 3D branch that receives it (c). As shown in Table 2, omitting appearance knowledge causes a -6.0% plaque IoU and -5.8% DSC drop, indicating that appearance features provide essential contextual and texture information that purely geometric data cannot supply.

Effectiveness of Bidirectional Stitching. After pre-training the dual-branch model, fine-tuning only the stitching adapter (e→f) yields $+0.9\%$ plaque IoU and $+1.2\%$ DSC, with only $\sim$0.87M additional trainable parameters (0.76% of total), validating model stitching as an efficient lightweight fine-tuning stage.

5 Conclusion

We propose DentalPSAM, a dual-branch model integrating geometric and appearance information for 3D IOS dental plaque segmentation. Key innovations include a bidirectional memory transformer generating geometric implicit prompts, a lightweight spatial alignment decoder fusing 2D appearance with 3D coordinates, and a model stitching stage fine-tuning bidirectional 2D–3D feature alignment between pre-trained branches. Extensive experiments on real-world 3D dental plaque segmentation datasets validate the effectiveness of our model. Additionally, we introduce the first open-source 3D intraoral scan dataset for dental plaque segmentation, fostering further research in this field.

Acknowledgments. This study was supported by the National Natural Science Foundation of China (62306253), the Early Career Fund (27207025), the Guangdong Natural Science Fund-General Program (2024A1515010233), and the Research Grants Council (17104124) .

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Almalki, A., Latecki, L.J.: Self-supervised learning with masked autoencoders for teeth segmentation from intra-oral 3d scans. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7820–7830 (2024)
2. Angelone, F., Ponsiglione, A.M., Ricciardi, C., Cesarelli, G., Sansone, M., Amato, F.: Diagnostic applications of intraoral scanners: A systematic review. Journal of Imaging **9**(7), 134 (2023)
3. Ben-Hamadou, A., Smaoui, O., Rekik, A., Pujades, S., Boyer, E., Lim, H., Kim, M., Lee, M., Chung, M., Shin, Y.G., et al.: 3dteethseg'22: 3d teeth scan segmentation and labeling challenge. arXiv preprint arXiv:2305.18277 (2023)
4. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P., Feichtenhofer, C.: Sam 3: Segment anything with concepts (2025), https://arxiv.org/abs/2511.16719
5. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., et al.: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. Medical Image Analysis **98**, 103310 (2024)
6. Chen, X., Shen, Y., Jeong, J.S., Perinpanayagam, H., Kum, K.Y., Gu, Y.: Deepplaq: Dental plaque indexing based on deep neural networks. Clinical Oral Investigations **28**(10), 534 (2024)
7. Cheng, Z., Wei, Q., Zhu, H., Wang, Y., Qu, L., Shao, W., Zhou, Y.: Unleashing the potential of sam for medical adaptation via hierarchical decoding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3511–3522 (2024)

8. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., et al.: Meshlab: an open-source mesh processing tool. In: Eurographics Italian chapter conference. vol. 2008, pp. 129–136 (2008)
9. Go, H., Narnhofer, D., Bhat, G., Truong, P., Tombari, F., Schindler, K.: Vist3a: Text-to-3d by stitching a multi-view reconstruction network to a video generator (2025), https://arxiv.org/abs/2510.13454
10. Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.A., Dou, Q.: 3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable medical image segmentation. arXiv preprint arXiv:2306.13465 (2023)
11. Guo, P., Wang, R., Zeng, S., Zhu, J., Jiang, H., Wang, Y., Zhou, Y., Wang, F., Xiong, H., Qu, L.: Exploring the vulnerabilities of federated learning: A deep dive into gradient inversion attacks. IEEE Transactions on Pattern Analysis and Machine Intelligence **48**(1), 438–455 (2026). https://doi.org/10.1109/TPAMI.2025.3646639
12. Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)
13. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
15. Li, J., Cheng, B., Niu, N., Gao, G., Ying, S., Shi, J., Zeng, T.: A fine-grained orthodontics segmentation model for 3d intraoral scan data. Computers in Biology and Medicine **168**, 107821 (2024)
16. Li, S., Guo, Y., Pang, Z., Song, W., Hao, A., Xia, B., Qin, H.: Automatic dental plaque segmentation based on local-to-global features fused self-attention network. IEEE Journal of Biomedical and Health Informatics **26**(5), 2240–2251 (2022)
17. Lin, Z., He, Z., Wang, X., Zhang, B., Liu, C., Su, W., Tan, J., Xie, S.: Dbganet: dual-branch geometric attention network for accurate 3d tooth segmentation. IEEE Transactions on Circuits and Systems for Video Technology (2023)
18. Liu, Y., Li, W., Wang, C., Chen, H., Yuan, Y.: When 3d partial points meets sam: Tooth point cloud segmentation with sparse labels. arXiv preprint arXiv:2409.01691 (2024)
19. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos (2025), https://arxiv.org/abs/2504.03600
20. Marsh, P., Martin, M., Marsh, P., Martin, M.: Dental plaque. Oral microbiology pp. 98–132 (1992)
21. Organization, W.H.: Oral health (2024), https://www.who.int/news-room/fact-sheets/detail/oral-health, accessed: 2024-11-05
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. Shi, Y., Zhang, X., Ji, J., Jiang, H., Zheng, C., Wang, Y., Qu, L.: Hsenet: Hybrid spatial encoding network for 3d medical vision-language understanding (2025), https://arxiv.org/abs/2506.09634
24. Xi, S., Liu, Z., Chang, J., Wu, H., Wang, X., Hao, A.: 3d dental model segmentation with geometrical boundary preserving. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 10476–10485 (2025)

25. Yang, Y., Wu, X., He, T., Zhao, H., Liu, X.: Sam3d: Segment anything in 3d scenes. arXiv preprint arXiv:2306.03908 (2023)
26. Ye, J.Z., Ørkild, T., Søndergaard, P.L., Hauberg, S.: 3Shape FDI 16 Meshes from Intraoral Scans (7 2023). https://doi.org/10.11583/DTU.23626650.v2
27. Yuan, T., Liao, W., Dai, N., Cheng, X., Yu, Q.: Single-tooth modeling for 3d dental model. International journal of biomedical imaging **2010**(1), 535329 (2010)
28. Yüksel, B., Özveren, N., Yeşil, Ç.: Evaluation of dental plaque area with artificial intelligence model. Nigerian Journal of Clinical Practice **27**(6), 759–765 (2024)
29. Zhang, L., Zhao, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D.: Tsgcnet: Discriminative geometric feature learning with two-stream graph convolutional network for 3d dental model segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6699–6708 (2021)
30. Zou, B.j., Liu, S.j., Liao, S.h., Ding, X., Liang, Y.: Interactive tooth partition of dental mesh base on tooth-target harmonic field. Computers in biology and medicine **56**, 132–144 (2015)
31. Zou, B., Wang, S., Liu, H., Sun, G., Wang, Y., Zuo, F., Quan, C., Zhao, Y.: Teethseg: An efficient instance segmentation framework for orthodontic treatment based on multi-scale aggregation and anthropic prior knowledge. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11601–11610 (2024)