TSegAgent: Zero-Shot Tooth Segmentation via Geometry-Aware Vision-Language Agents

Shaojie Zhuang[1], Lu Yin[2⋆], Guangshun Wei[1], Yunpeng Li[3], Xilu Wang[2], and Yuanfeng Zhou[1*]

[1] Shandong University, Jinan 250010, China
[2] University of Surrey, Guildford, Surrey, United Kingdom
[3] King's College London, London, United Kingdom

Abstract. Automatic tooth segmentation and identification from intra-oral scanned 3D models are fundamental problems in digital dentistry, yet most existing approaches rely on task-specific 3D neural networks trained with densely annotated datasets, resulting in high annotation cost and limited generalization to scans from unseen sources. Thus, we propose TSegAgent, which addresses these challenges by reformulating dental analysis as a zero-shot geometric reasoning problem rather than a purely data-driven recognition task. The key idea is to combine the representational capacity of general-purpose foundation models with explicit geometric inductive biases derived from dental anatomy. Instead of learning dental-specific features, the proposed framework leverages multi-view visual abstraction and geometry-grounded reasoning to infer tooth instances and identities without task-specific training. By explicitly encoding structural constraints such as dental arch organization and volumetric relationships, the method reduces uncertainty in ambiguous cases and mitigates overfitting to particular shape distributions. Experimental results demonstrate that this reasoning-oriented formulation enables accurate and reliable tooth segmentation and identification with low computational and annotation cost, while exhibiting strong generalization across diverse and previously unseen dental scans. [4]

Keywords: Zero-shot tooth segmentation · Digital dentistry · FDI identification.

1 Introduction

Automatic tooth segmentation and identification from intra-oral scanned 3D models are essential components of digital dentistry, supporting a wide range of clinical applications such as orthodontic diagnosis, treatment planning, and prosthodontic design. Most existing methods [9,22,3,23,10,21] formulate these tasks as supervised learning problems and rely on task-specific 3D neural networks trained with densely annotated dental datasets like Teeth3DS [1]. While

⋆ Corresponding authors: Yuanfeng Zhou, Lu Yin.
[4] Code is available at: https://github.com/znshje/TSegAgent.

such approaches have shown promising performance under controlled settings, they require substantial annotation effort and often exhibit limited generalization when applied to scans acquired from unseen sources, scanners, or patient populations.

Vision backbones are advancing by improving the reception mechanism from local to global [15,13,14,18]. Transformer [16,4] and mamba [5,6] backbones change the way images are processed. Since large models like SAM [8] became stronger in vision understanding, increasing studies adapted SAM to different areas and tasks. Examples include SAM-Med3D [17] and MedSAM [12], that adapt general-purpose segmentation models to medical data, as well as conversational systems such as ChatIOS [20] that explore vision-language interaction for intra-oral scan analysis. Recent IOSSAM [7] and 3DTeethSAM [11] demonstrate great potential in SAM-based segmentation, but they require training a teeth detector or refiner, and lack the ability of segmentation error detection. With vision-language models advance, SAM3 [2] features a text-prompt segment anything model, enabling the training-free tooth segmentation on images. These methods demonstrate the potential of foundation models to reduce annotation requirements and improve flexibility. However, most existing approaches still treat tooth segmentation and identification as direct prediction problems, focusing on adapting large models to dental data rather than rethinking the underlying problem formulation.

In this work, we propose TSegAgent, a geometry-aware vision-language agent that reformulates tooth segmentation and identification as a zero-shot geometric reasoning problem. Instead of learning dental-specific representations, TSegAgent combines general-purpose foundation models with explicit geometric inductive biases derived from dental anatomy. The core idea is to leverage multi-view visual abstraction for instance extraction while enforcing anatomical consistency through geometry-grounded reasoning at the instance level. By explicitly encoding dental arch structure, volumetric relationships, and symmetry priors, TSegAgent enables anatomically consistent tooth segmentation and identification without task-specific training.

This reasoning-oriented formulation offers two key advantages. First, it significantly reduces annotation and training requirements while avoiding overfitting to particular model shapes or datasets, leading to strong generalization across dental scans from diverse and previously unseen sources. Second, the integration of interpretable geometric priors reduces uncertainty in visually ambiguous cases, which is critical for clinical reliability. Extensive experiments demonstrate that TSegAgent achieves accurate and consistent tooth segmentation and identification with low computational and annotation cost, highlighting the potential of geometry-grounded reasoning for practical dental AI systems. To sum up, our contributions include: (1) We propose a zero-shot framework for tooth instance segmentation and identification from intra-oral scanned 3D models, eliminating the need for task-specific training and dense annotations; (2) We design a geometry-aware and interpretable vision-language agent that integrates dental arch priors, multi-view visual evidence, and volumetric information to achieve

accurate tooth classification; (3) We introduce explicit geometric reasoning and verification mechanisms that reduce ambiguity in tooth identification and enable robust generalization across dental models from arbitrary sources.

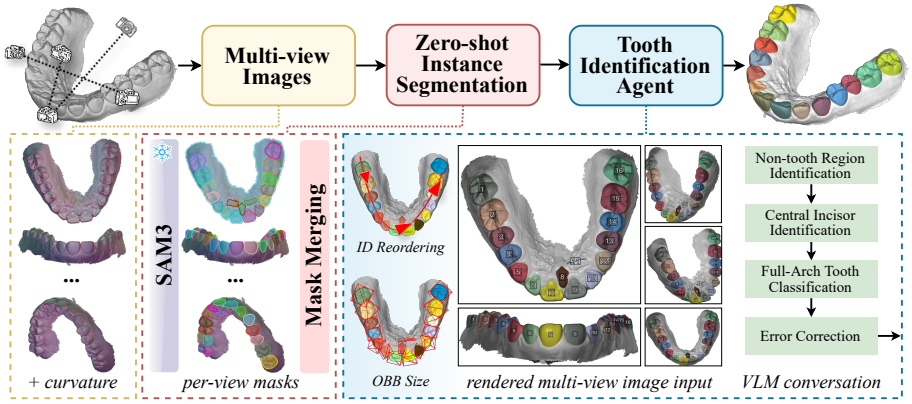


Fig. 1: The pipeline of TSegAgent. Given an intra-oral scanned 3D model, we first perform multi-view rendering (with curvature) and apply SAM3 for zero-shot tooth instance segmentation. The resulting masks are merged into face-level instance labels, which are then reordered based on dental arch geometry. Finally, a vision-language agent identifies tooth instances by multi-round conversation and reasoning over visual cues and geometric constraints.

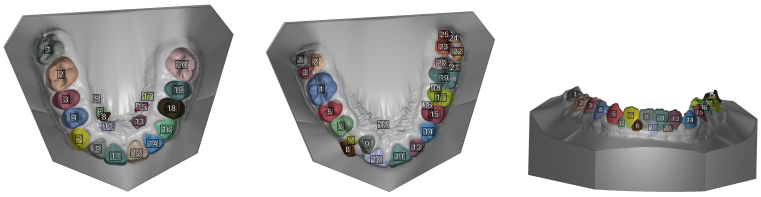
2 Method

2.1 Overview

Given an intra-oral scanned (IOS) model, our goal is to automatically segment individual tooth instances and assign FDI labels in a zero-shot manner. As illustrated in Fig. 1, TSegAgent consists of two components: multi-view zero-shot tooth instance segmentation, and geometry-aware vision-language agent for tooth identification. The entire pipeline avoids task-specific training and global post optimization, relying instead on explicit geometric inductive biases and reasoning to achieve interpretable results.

2.2 Tooth Instance Segmentation by Multi-View Images

The first component of TSegAgent focuses on extracting individual tooth instances from an intra-oral scanned 3D model in a zero-shot manner. Since most general-purpose segmentation models [8,12] are designed for 2D images, we convert the input 3D mesh into a set of multi-view renderings that capture complementary perspectives of the dental arch. As shown in Fig. 1, the rendering



(a) (b) (c)

Fig. 2: Typical challenging cases for tooth classification, including (a) non-tooth regions, (b) tooth over-segmentation into multiple instances due to occlusion or complex morphology, and (c) small gingival papilla between adjacent teeth that may be misclassified as tooth instances.

viewpoints are selected from five directions: perpendicular to the occlusal plane, the patient's anterior-posterior direction, and the left and right lateral sides. For each view, we additionally apply 10° to 30° perturbations on the axes as new views to ensure the model surfaces are covered as comprehensively as possible, while avoiding excessive inter-tooth occlusion that would severely degrade the segmentation results. Meanwhile, we add the curvature heatmap to the image to enhance fuzzy areas. Each rendered image is then predicted independently using SAM3 [2] with a text prompt "tooth", producing candidate instance masks without training.

These per-view predictions are projected back onto the 3D mesh. The masks provide coarse instance proposals, and they are inherently unordered or incomplete due to occlusion or view-dependent ambiguity. To obtain a unified instance segmentation, we design a mask merging strategy that fuses multi-view predictions into face-level instance labels, based on the overlap IoU metric between masks. Specifically, for any two masks M_a and M_b from different views, if their IoU exceeds the threshold, they are merged as one instance; If either $|M_a \cap M_b|/|M_a|$ or $|M_a \cap M_b|/|M_b|$ exceeds the threshold, we select the containment vs. non-containment relationship that occurs most frequently across all views to determine whether M_a contains M_b (or vice versa). We then use this consensus relationship to differentiate instances.

2.3 Tooth Identification Agent

Despite the mask merging step, challenging cases still remain, including non-tooth regions (e.g., gingiva), over-segmentation of teeth, and small gingival papillae between adjacent teeth (Fig. 2). By carefully designing the text prompts and the image inputs, we leverage the capabilities of the vision-language models (VLM) to eliminate these error cases.

Instance ID Reordering. The tooth instances obtained from SAM3 are inherently unordered, which poses a challenge for anatomically consistent tooth

identification. To provide a structural reference for reasoning, we perform instance ID reordering based on explicit dental arch priors. For each segmented tooth instance, we compute its geometric centroid from the associated mesh faces. All centroids are projected onto the XY-plane, where a second-order polynomial curve is fitted to approximate the dental arch. Each instance is assigned a scalar parameter corresponding to its closest point on the fitted curve. By sorting instances according to this parameter, we reorder instance IDs into a contiguous sequence that reflects their spatial arrangement along the dental arch.

This reordering step encodes relative tooth positions in a geometry-driven manner and enables subsequent reasoning processes to exploit sequential and symmetry constraints without relying on learned positional embeddings.

Multi-round Conversation-based Tooth Identification Agent. After instance reordering, tooth identification is performed by a geometry-aware vision-language agent through a multi-round conversation. As illustrated in Fig. 1, the dental mesh is rendered from selected views as image inputs, with temp tooth ID printed on tooth instances. Rather than predicting all tooth labels in a single pass, the agent decomposes the identification task into a sequence of interpretable sub-tasks, each guided by explicit geometric and anatomical constraints. Meanwhile, we add a 3-dimensional size of each tooth oriented bounding-box (OBB) into text prompts, for VLM to easily evaluate tooth shapes. This design allows the agent to progressively reduce ambiguity and improve robustness in zero-shot settings. Detailed text prompt design can be found in our source codes.

Non-tooth Region Identification. In the first reasoning round, the agent identifies non-tooth instances, such as gingiva or interdental papilla. The prompt is designed to check the size, position, and shape of each tooth to identify non-tooth instances. Most challenging cases shown in Fig. 2 can be effectively filtered out at this stage, with the help of provided geometric clues. Non-tooth instances are excluded from subsequent identification steps, simplifying the reasoning space and reducing potential error propagation.

Central Incisor Identification. In the second reasoning round, the agent identifies central incisors, which serve as a critical anatomical reference due to their bilateral symmetry and central position along the dental arch. In the prompt, we guide the VLM to determine the IDs of the central incisors based on dental arch symmetry and size correspondence, rather than relying on left–right pixel balance in the image.

Full-Arch Tooth Classification. Conditioned on the reordered instance sequence and the identified central incisors, VLM infers tooth identities by reasoning over relative position, morphological characteristics, and volumetric relationships. We add detailed tooth features, common errors, and output formats in text prompt. This sequential formulation enables the agent to exploit dental arch continuity and avoid implausible label assignments.

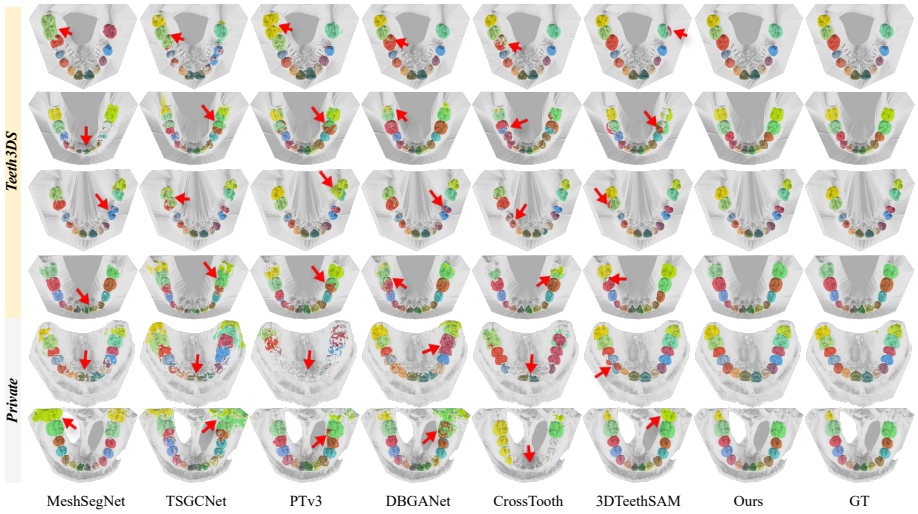


Fig. 3: Qualitative results of candidate methods, from Teeth3DS and private dataset.

Error Detection and Correction. During and after full-arch classification, the agent performs explicit error detection based on bilateral symmetry, sequential consistency along the dental arch, and relative volumetric similarity between corresponding teeth. When violations of these constraints are detected, the agent is prompted to correct its previous predictions through additional conversational rounds. This self-correction enables recovery from early errors without external supervision or post-hoc optimization, with segmentation gains mainly from non-tooth filtering and over-segmentation removal and error correction chiefly improving FDI accuracy.

After the multi-round conversation, the agent outputs the tooth ID to FDI label mapping for each segmented instance, completing the zero-shot tooth identification process.

3 Experiments

3.1 Dataset and Settings

We evaluate candidate methods on the Teeth3DS dataset [1], which contains 1200 3D tooth models. Each model is annotated with an FDI label for each vertex. Meanwhile, to evalutate the generalizability of the methods, we also test them on a private dataset collected from our collaborating dental clinic, which consists of 340 3D tooth models in the testing set with the same annotation scheme as Teeth3DS. The private dataset is more challenging than Teeth3DS, as most are scanned plaster models with more severe defects, interference, and artifacts. For each candidate methods, we train them on the training set of Teeth3DS from

Table 1: Performance comparison on Teeth3DS and private dataset. The best results are highlighted in bold, while the second best results are underlined.

Method	mIoU	TLA	TSA	TIR	$TIR_{=1}$
Teeth3DS Dataset					
MeshSegNet [9]	76.46± 13.2	90.98± 11.8	89.70± 4.5	95.09± 14.3	83.33
PTv3 [19]	83.61± 11.6	<u>94.45± 8.3</u>	95.22± 1.5	95.29± 12.1	79.67
TSGCNet [22]	53.45± 14.4	82.80± 10.8	85.23± 3.7	85.99± 20.3	51.17
DBGANet [10]	83.51± 13.1	95.37± 7.0	96.09± 1.4	95.80± 13.2	<u>86.50</u>
CrossTooth [21]	47.51± 29.0	63.57± 47.2	72.29± 34.69	70.21± 40.0	49.17
3DTeethSAM [11]	<u>92.03± 9.5</u>	93.64± 13.9	**97.12± 2.5**	96.03± 12.0	85.00
TSegAgent (Ours)	**93.37± 2.8**	**96.40± 7.4**	<u>96.76± 1.5</u>	**97.46± 7.6**	**87.17**
Private Dataset					
MeshSegNet [9]	38.69± 22.7	80.83± 36.8	85.69± 9.1	76.33± 24.4	<u>39.41</u>
PTv3 [19]	36.42± 19.3	81.40± 34.8	86.67± 4.6	60.84± 32.7	21.76
TSGCNet [22]	28.75± 10.8	68.21± 12.8	80.70± 4.9	53.63± 22.7	3.83
DBGANet [10]	41.59± 17.4	<u>96.60± 10.7</u>	91.10± 4.2	61.83± 26.9	22.06
CrossTooth [21]	35.17± 22.1	34.48± 46.8	87.68± 6.6	46.99± 29.1	6.76
3DTeethSAM [11]	<u>81.42± 14.1</u>	70.98± 26.3	<u>94.63± 8.5</u>	<u>82.42± 19.6</u>	37.06
TSegAgent (Ours)	**82.10± 11.1**	**96.68± 1.8**	**95.99± 3.1**	**85.41± 17.9**	**51.96**

scratch until convergence. Then we evaluate their performance on the testing set of Teeth3DS and the private dataset. TSegAgent is compatible with various VLM services, and use "ChatGPT 5.2" in our experiments.

3.2 Metrics

We use the same metrics as in Teeth3DS benchmarks [1]: TLA, TSA and TIR. Additionally, we report $TIR_{=1}$, the proportion of samples where all instances are correctly identified. For each GT tooth P_i with centroid $\mathbf{c}_i$ and size $\mathbf{s}_i$, we match it to the closest predicted tooth $\hat{P}_{\pi(i)}$. The metrics are:

$$\begin{aligned} &\mathrm{TLA} = \exp\Big(-\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{\mathbf{c}_i - \hat{\mathbf{c}}_{\pi(i)}}{\mathbf{s}_i}\Big\|_2\Big), \mathrm{TSA} = \mathrm{F1}\big([\mathbf{y}^{gt} \neq 0],\, [\mathbf{y}^{pred} \neq 0]\big)\,, \\ &\mathrm{TIR} = \frac{1}{N}\sum_{i=1}^{N}\Big(\ell_i = \hat{\ell}_{\pi(i)}\Big)\,, \mathrm{TIR}_{=1} = \frac{1}{M}\sum_{m=1}^{M}\Big(\mathrm{TIR}^{(m)} = 1\Big)\,. \end{aligned} \tag{1}$$

Additionally, we report the mean Intersection over Union (mIoU) for tooth-wise segmentation performance. For each GT tooth, the IoU is computed with the predicted tooth with the same label.

3.3 Comparison

Results are shown in Fig. 3 and Tab. 1. On the Teeth3DS dataset, TSegAgent achieves competitive performance and outstanding stability across the metrics.

Table 2: Ablation study of segmentation and reasoning components in TSegAgent, with performance evaluated on Teeth3DS dataset.

Settings				Metrics				
VLM	Conv.	Arch	OBB	mIoU	TLA	TSA	TIR	$TIR_{=1}$
-	-	-	-	-	75.21± 27.1	<u>96.45± 1.7</u>	-	-
✓	✓	-	-	85.49± 14.2	91.67± 15.2	95.59± 4.4	89.64± 18.3	66.72
✓	✓	✓	-	<u>86.20± 16.4</u>	92.26± 13.6	96.02± 2.1	90.43± 20.6	74.83
✓	-	✓	✓	83.14± 21.3	<u>95.73± 8.9</u>	96.27± 1.7	86.36± 26.0	71.17
✓	✓	✓	✓	**93.37± 2.8**	**96.40± 7.4**	**96.76± 1.5**	**97.46± 7.6**	**87.17**

Compared to training-based methods, TSegAgent demonstrates better ability of instance differentiation. Especially, compared to 3DTeethSAM which is also based on SAM, TSegAgent preserves better instance integrity while achieving higher classification accuracy.

More importantly, on the private dataset with a substantial domain shift, TSegAgent demonstrates significantly stronger generalization capability. Compared to supervised methods such as 3DTeethSAM and DBGANet, which exhibit notable performance degradation under cross-domain conditions, TSegAgent maintains robust overall accuracy, highlighting the reliability of the proposed reasoning framework in challenging real-world scenarios.

3.4 Ablation Studies

We conduct ablation experiments on the Teeth3DS dataset to analyze the contribution of each component in TSegAgent in Tab. 2. The segmentation-only baseline (*w/o* VLM) evaluates initial segmentation performance. Although it achieves a high TSA score, TSA does not account for instance-level distinguish; therefore, incorrect segmentations may still lead to an improvement in TSA but low TLA. Introducing VLM significantly improves segmentation quality, as VLM will wipe out any mis- or over-segmentation. Enabling ID reordering (Arch) further improves TIR by providing a consistent positional reference along the dental arch. Adding volumetric cues (OBB) enhances structural awareness and improves robustness in distinguishing anatomically similar or bad teeth. Finally, decomposing tooth identification into structured reasoning steps effectively reduces ambiguity and enforces anatomical consistency, increasing TIR to 97.46% and $TIR_{=1}$ to 87.17%. Overall, the full TSegAgent configuration achieves the best performance, validating the complementary benefits of geometric inductive biases and multi-round reasoning.

4 Conclusion

In this work, we presented TSegAgent, a geometry-aware vision-language framework for zero-shot tooth segmentation and identification, which highlights the potential of integrating foundation models with structured geometric reasoning

for practical dental AI applications. Future work may explore extending the proposed reasoning framework to other anatomical structures and incorporating additional domain-specific priors to further enhance robustness and clinical applicability.

Acknowledgments. This study was funded by the National Natural Science Foundation of China under grant 62572284 and 62502273 (Youth Program); the Natural Science Foundation of Shandong Province under grant ZR2024ZD12 (Major Basic Research project) and ZR2025QC1558 (Youth Program); High-Level Innovation Teams of Universities and Research Institutes (Self-Cultivated under the "20 Measures for Universities" of Jinan) under grant 202534004; China Scholarship Council under grant 202506220152.

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

References

1. Ben-Hamadou, A., Smaoui, O., Chaabouni-Chouayakh, H., Rekik, A., Pujades, S., Boyer, E., Strippoli, J., Thollot, A., Setbon, H., Trosset, C., Ladroit, E.: Teeth3ds: a benchmark for teeth segmentation and labeling from intra-oral 3d scans (2022)
2. 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
3. Cui, Z., Li, C., Chen, N., Wei, G., Chen, R., Zhou, Y., Shen, D., Wang, W.: Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model. Medical Image Analysis **69**, 101949 (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
5. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces (2024), https://arxiv.org/abs/2312.00752
6. Hatamizadeh, A., Kautz, J.: Mambavision: A hybrid mamba-transformer vision backbone. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 25261–25270 (2025)
7. Huang, X., He, D., Li, Z., Zhang, X., Wang, X.: IOSSAM: Label Efficient Multi-View Prompt-Driven Tooth Segmentation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15001. Springer Nature Switzerland (October 2024)
8. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv:2304.02643 (2023)

9. Lian, C., Wang, L., Wu, T.H., Wang, F., Yap, P.T., Ko, C.C., Shen, D.: Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3d intraoral scanners. IEEE transactions on medical imaging **39**(7), 2440–2450 (2020)
10. Lin, Z., He, Z., Wang, X., Zhang, B., Liu, C., Su, W., Tan, J., Xie, S.: Db-ganet: Dual-branch geometric attention network for accurate 3d tooth segmentation. IEEE TCSVT **34**(6), 4285–4298 (2024). https://doi.org/10.1109/TCSVT.2023.3331589
11. Lu, Z., Lou, J., Ma, M., Jin, H., Zheng, Y., Zhou, K.: 3dteethsam: Taming sam2 for 3d teeth segmentation (2025), https://arxiv.org/abs/2512.11557
12. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
13. 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)
14. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 5105–5114 (2017)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 6000–6010 (2017)
17. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., Fu, B., Zhang, S., He, J., Qiao, Y.: Sam-med3d: Towards general-purpose segmentation models for volumetric medical images (2024), https://arxiv.org/abs/2310.15161
18. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Trans. Graph. **38**(5) (oct 2019). https://doi.org/10.1145/3326362, https://doi.org/10.1145/3326362
19. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer v3: Simpler, faster, stronger. In: CVPR. pp. 4840–4851 (2024). https://doi.org/10.1109/CVPR52733.2024.00463
20. Wu, Y., Zhang, Y., Wu, Y., Zheng, Q., Li, X., Chen, X.: Chatios: Improving automatic 3-dimensional tooth segmentation via gpt-4v and multimodal pre-training. Journal of Dentistry **157**, 105755 (2025). https://doi.org/https://doi.org/10.1016/j.jdent.2025.105755, https://www.sciencedirect.com/science/article/pii/S030057122500199X
21. Xi, S., Liu, Z., Chang, J., Wu, H., Wang, X., Hao, A.: 3d dental model segmentation with geometrical boundary preserving. In: CVPR. pp. 10476–10485 (2025)
22. 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)
23. Zhuang, S., Wei, G., Cui, Z., Zhou, Y.: Robust hybrid learning for automatic teeth segmentation and labeling on 3d dental models. IEEE Transactions on Multimedia **27**, 792–803 (2025). https://doi.org/10.1109/TMM.2023.3289760