When 3D Partial Points Meets SAM: Tooth Point Cloud Segmentation with Sparse Labels

Yifan Liu[1], Wuyang Li[1], Cheng Wang[1], Hui Chen[2], Yixuan Yuan[1(✉)]

[1] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
[2] Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China
yxyuan@ee.cuhk.edu.hk

Abstract. Tooth point cloud segmentation is a fundamental task in many orthodontic applications. Current research mainly focuses on fully supervised learning which demands expensive and tedious manual point-wise annotation. Although recent weakly-supervised alternatives are proposed to use weak labels for 3D segmentation and achieve promising results, they tend to fail when the labels are extremely sparse. Inspired by the powerful promptable segmentation capability of the Segment Anything Model (SAM), we propose a framework named SAMTooth that leverages such capacity to complement the extremely sparse supervision. To automatically generate appropriate point prompts for SAM, we propose a novel Confidence-aware Prompt Generation strategy, where coarse category predictions are aggregated with confidence-aware filtering. Furthermore, to fully exploit the structural and shape clues in SAM's outputs for assisting the 3D feature learning, we advance a Mask-guided Representation Learning that re-projects the generated tooth masks of SAM into 3D space and constrains these points of different teeth to possess distinguished representations. To demonstrate the effectiveness of the framework, we conduct experiments on the public dataset and surprisingly find with only 0.1% annotations (one point per tooth), our method can surpass recent weakly supervised methods by a large margin, and the performance is even comparable to the recent fully-supervised methods, showcasing the significant potential of applying SAM to 3D perception tasks with sparse labels. Code is available at https://github.com/CUHK-AIM-Group/SAMTooth.

Keywords: Weakly-supervised Training · Segment Anything Model · Tooth Point Cloud Segmentation.

1 Introduction

Accurately segmenting teeth in 3D tooth point clouds extracted from Intra-Oral Scanners (IOS) mesh data plays a pivotal role in many orthodontic applications, including detailed analysis of tooth morphology, treatment planning, personalized appliance design, etc [6,17,13,12]. However, existing tooth point cloud segmentation models [29,23,11,3,4,22] rely heavily on large annotated datasets

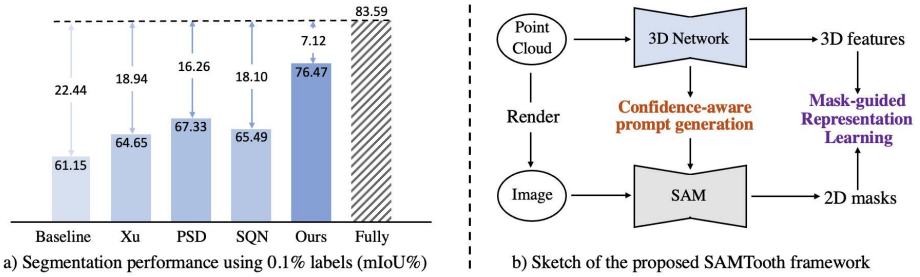


Fig. 1. Illustration of a) segmentation performance of various weakly-supervised methods using sparse labels (0.1%) and b) the proposed SAMTooth framework sketch.

for training, which poses challenges due to the labor-intensive nature of tooth point cloud labeling. For example, it takes around 15 to 30 minutes for an experienced dentist to annotate a half jaw manually [17]. This time-consuming process presents a significant obstacle to establishing large-scale datasets and hinders the generalizability of the diagnosis system.

To address this issue, there has been a growing interest in investigating weakly-supervised alternatives. Among different types of weak labels (scribbles, boxes, partial points, etc.), partial points stand out as a prospective direction due to the annotating efficiency–it only involves labeling a single or several points for each tooth. Existing partial-points-based methods excavate various training constraints from limited labels, such as perturbation consistency [20,30], supervision propagation [24,5], self-supervised pre-training [10,27], pseudo-labeling [16,19], etc, which has achieved great progress in reducing the annotation labor. However, as shown in Fig. 1, after increasing the label sparsity to 0.1% (one point per tooth), we observe that the best entry of existing works [5] only gives marginal 6.18% performance gains over baseline and yields a 22.44% mIoU gap compared with fully supervised oracle, indicating that existing works can not perform well *when the labels are extremely sparse.*

As the first attempt to tackle this issue, we aim to leverage the recent advance of the Segment Anything Model (SAM) [7]. Trained on a large-scale dataset, similar to other vision foundation models [9,2,25,26], SAM can generate fine-grained masks given manually defined visual prompts. As shown in Fig. 1 a), if we render images from the input 3D model, and feed these images to SAM with adequate prompts, we can get 2D object masks of each tooth. As these masks contain explicit shape information, we can use them to complement the extremely sparse supervision. Nevertheless, it is non-trivial to employ the 2D SAM to assist the 3D task directly, which is attributed to the two issues. Firstly, it is tough to prompt 2D SAM automatically to generate the desired masks. The quality of SAM masks heavily relies on appropriate prompts provided by humans, while incorporating human input during model training is not feasible. Secondly, given the significant disparity between 2D images and 3D point clouds,

it is challenging to effectively utilize the 2D masks generated by SAM to enhance model learning in the 3D domain.

To tackle these two challenges, we propose a novel framework SAMTooth, for tooth point cloud segmentation with extremely sparse labels. As shown in Fig. 1 b), the framework consists of two paradigms, including *Confidence-aware Prompt Generation* (CPG) and *Mask-guided Representation Learning* (MRL). To automatically generate appropriate prompts for SAM to use, we propose CPG to aggregate the points of each predicted tooth and project the results to the image plane. As the point predictions may be noisy, the point-wise confidence is further estimated to filter unreliable aggregating candidates. To fully leverage the outputs of SAM for 3D feature learning, we advance MRL to re-project the pixels of SAM's outputs into the 3D space and leverage the contrastive learning to provide training constraints. Considering the background points should also be constrained, we also compute a background mask from SAM's object masks and impose explicit supervision. Extensive experiments demonstrate that SAMTooth can outperform other weakly supervised methods by a large margin and is even comparable to recent fully-supervised methods using only 0.1% annotations.

2 Method

Our framework is designed for weakly-supervised tooth point cloud segmentation, by leveraging the zero-shot capacity of visual foundation model SAM. As shown in Fig. 2, it begins with Image Rendering and Mapping (Sec. 2.1) to render images from the input scan and build the mapping between 3D points and 2D pixels. Then, the input point cloud P is passed to the 3D segmentation network to get coarse predictions Y and point-wise confidence C, which are further passed to Confidence-aware Prompt Generation (Sec. 2.2) to generate adequate point prompts for SAM. After that, SAM processes the generated prompts and rendered images to get object masks M, which are used to constrain the 3D features by Mask-guided Representation Learning (Sec. 2.3). The whole framework is optimized by the segmentation constraints and complementary constraints from SAM's outputs (Sec. 2.4).

2.1 Image Rendering and Mapping

To leverage SAM's outputs for 3D representation learning, we first render images from the 3D IOS mesh as SAM's input. We choose to render from the mesh rather than the point cloud as the mesh contains more textural details and is always available in orthodontic applications. Based on the imaging principle of the pinhole camera, the projected coordinates of each point can be obtained by:

$$[u, v, 1]^T = 1/z \cdot K \cdot T \cdot [x, y, z, 1]^T, \tag{1}$$

where $[u, v]^T$ and $[x, y, z]^T$ are the 2D and 3D coordinates. K and T are the manually defined camera's intrinsic and extrinsic matrices. By using Eq. 1, 2D images can be rendered from the 3D IOS mesh and 2D pixels can also be projected from the 3D space.

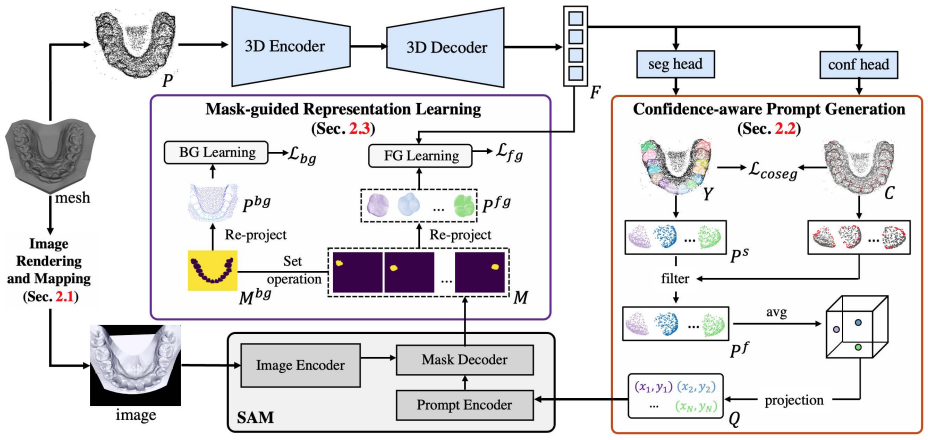


Fig. 2. Illustration of the proposed SAMTooth framework.

2.2 Confidence-aware Prompt Generation

SAM relies on adequate prompts to generate high-quality object masks, which would further influence the subsequent 3D representation learning. Therefore, a suitable prompt-generation strategy should be carefully designed. In this regard, we propose an automatic prompt generation strategy that gets prompts by aggregating 3D points of each coarsely predicted tooth, accompanied by a confidence-aware filtering step to discard those unconfident tooth predictions that would lead to ambiguous prompts.

Point-wise Confidence Estimation. In addition to the original segmentation head, we add a confidence head consisting of several MLP and BN layers. Its role is to estimate point-wise confidence values $C = \{c_1, ..., c_N\} \in \mathbb{R}^N$. To train the two heads simultaneously, we constrain their outputs using the confidence-aware segmentation loss $\mathcal{L}_{coseg}$ on the labeled point set P_{label}:

$$\mathcal{L}_{coseg} = \frac{1}{N} \sum_{p_i \in P_{label}} c_i \cdot \mathcal{L}_{CE}(y_i, y_i^{gt}) + (1 - c_i)^2, \tag{2}$$

where $\mathcal{L}_{CE}$ represents the cross-entropy loss between predictions $y_i \in Y$ and ground truth $y_i^{gt} \in Y^{gt}$. For a certain point p_i that the network is confident about the prediction, the estimated c_i can be large to reduce the second term in Eq. 2 since the first term is already small enough, and vice versa. Therefore, L_{coseg} can encourage the model to generate large c_i for confident predictions and small c_i for unconfident ones.

Confidence-aware Prompt Generation. To generate suitable point prompts for each tooth, we first divide the point cloud P into G subgroups $\{P^{s1}, ..., P^{sG}\}$, where each subgroup shares the same category predictions. Considering noisy points exist in the coarse predictions, we thus use the estimated point-wise confidence as a metric and filter those noisy predictions in a subgroup using a

threshold τ, obtaining filtered subgroups $\{P^{f1}, ..., P^{fG}\}$, which $P^{fi} = \{p_j | c_j > \tau, p_j \in P^{si}\}$. It is easily observed that each subgroup coarsely represents the point cloud of a certain tooth, thus we can get prompts $Q = \{q_1, ..., q_G\}$ by averaging the points belonging to the same subgroup and projecting them to the image coordinates:

$$q_i = Proj(\frac{1}{|P^{fi}|} \sum_{p_k \in P^{fi}} p_k), \tag{3}$$

where $Proj(\cdot)$ projects the 3D coordinates into the 2D image plane as described in Eq. 1. With the guidance of the estimated point-wise confidence, the generated prompts are observed to be closer to the tooth center compared to the simple aggregation, which is more adequate for SAM to produce accurate object masks.

2.3 Mask-guided Representation Learning

With appropriate point prompts, SAM can generate precise object masks, from which we expect to excavate more constraints to complement the sparse supervision of the 3D model. To this end, we propose to re-project 2D object masks into 3D space and utilize contrastive learning for foreground feature discrimination. Considering the background points, i.e., gingiva should also be constrained, we further compute a background mask from the foreground ones and regularize the corresponding features.

Foreground Learning. With previously projected 2D prompts $Q = \{q_1, ..., q_G\}$, where G denotes the number of prompts, we can get SAM's output object masks $M = \{m_1, ..., m_G\}$, where $m_i \in \{0, 1\}^{H \times W}$ is the binary mask of a certain tooth. Then, we extract coordinates of object pixels in m_i as $coord_i = \{(h_j, w_j) | m_i(h_j, w_j) = 1\}$. After that, pixels in $coord_i$ are re-projected to the 3D space, generating the re-projected 3D subgroup P^{fgi} for each m_i:

$$P^{fgi} = \{p_i^{3d} | p_i^{3d} = ReProj(p_i^{2d}), p_i^{2d} \in coord_i\}, \tag{4}$$

Doing so for each mask m_i, we can get the re-projected subgroup set $P^{fg} = \{P^{fg1}, ..., P^{fgG}\}$. As subgroups should contain points of different categories, i.e., different teeth, we leverage contrastive learning to encourage the 3D features among different subgroups distinguishable. Specifically, we extract 3D features of each point in each group, passing them to two consecutive MLP layers with BN, composing $F^{fg} = \{F^{fg1}, ..., F^{fgG}\}$. Then, contrastive loss $\mathcal{L}_{fg}$ is imposed on these features:

$$\mathcal{L}_{fg} = -C \sum_i \sum_j log \frac{exp(f_i^T f_j / t)}{\sum_k exp(f_i^T f_k / t)}, \tag{5}$$

where f_i, f_j are in the same subgroup, f_i, f_k are in the different one, C is the normalization constant, and t is the temperature. The role of Eq. 5 can be treated as the complementary supervision to the extremely sparse labels, which provides massive constraints on the unlabeled points.

Background Learning. To further constrain the background, i.e., gingiva features, we compute the background mask M^{bg} by eliminating the combination of the pixels in foreground masks $M = \{m_1, ..., m_G\}$ generated by SAM:

$$M^{bg} = \overline{m_1} \odot \overline{m_2} \odot ... \odot \overline{m_G}. \tag{6}$$

Then similarly to foreground masks, we also re-project coordinates of pixels in M^{bg} into 3D space to get a background group P^{bg}. As we already know these points should belong to the background class, we can directly constrain the predictions of these background features F^{bg} as 0 (the background label):

$$\mathcal{L}_{bg} = CrossEntropy(SegHead(F^{bg}), 0). \tag{7}$$

2.4 Model Optimization

During training, we first warm up the network using the confidence-ware segmentation loss $\mathcal{L}_{coseg}$ for T epochs, enabling the network to generate coarse segmentation results and point-wise confidence. Then, guided by the output masks of SAM, $\mathcal{L}_{fg}$ and $\mathcal{L}_{bg}$ are used to constrain the foreground and the background 3D features, respectively. The overall optimization objective is:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{coseg} + (\lambda_2 \mathcal{L}_{fg} + \lambda_3 \mathcal{L}_{bg}) \cdot [t > T], \tag{8}$$

where t is the current epoch and $[\cdot]$ is an indicator function that equals 1 if the statement is true else 0.

3 Experiments

3.1 Experiment settings

Datasets and evaluation To evaluate the effectiveness of our proposed method, we conduct experiments on the public 3DTeethSeg [1] dataset. The tooth identification follows the FDI World Dental Federation notation. The 3DTeethSeg [1] is a publicly available tooth segmentation dataset, which contains 1,800 available 3D IOS scans obtained from 900 patients, following a real-world patient age distribution. To make a fair comparison, we use the same split in all experiments where 1,080 scans are randomly selected for training, 360 ones for validation, and the remaining ones for testing. Following previous tooth segmentation methods [14,15,3], we use the Jaccard Index (also known as mIoU), the Dice Similarity Coefficient (DSC), and the point-wise classification accuracy (Acc).

Implementation details We adopt standard ViT-B/16 in [28] as the segmentation backbone. Our framework is trained with AdamW optimizer with a 5e-4 learning rate, 8 batch size, and weight decay of 0.05. We empirically set the confidence thresh-hold τ as 0.6, temperature t as 0.1, warmup epoch T as 10, and the loss weight $\lambda_{1/2/3}$ in 8 as 1/0.1/0.01, respectively. Following previous works [15,14], we sample 16,000 points from the IOS scan to compose the input point cloud and use the three-neighbor-interpolation strategy to upsample the predictions to the original size during the evaluation [3].

Table 1. Quantitative results of different methods on the 3DTeethSeg dataset. The best and second best results are **bold** and underlined.

Ratio	Methods	Incisor	Canine	Premolar	Molar	Gingiva	mIoU%	mAcc%
100%	PointNet++ [18]	71.81	72.31	80.07	80.67	81.34	77.15	89.06
	DGCNN [21]	78.18	78.26	80.07	78.25	75.91	78.88	87.45
	Transformer [28]	83.83	84.35	83.95	81.93	89.31	83.59	91.85
0.1%	Transformer [28]	58.85	60.91	62.35	62.84	61.6	61.15	74.91
	II-Model [8]	64.64	65.24	66.52	64.39	63.34	64.70	77.30
	MT [20]	64.29	64.42	67.14	65.13	61.97	65.12	77.00
	Xu and Lee [24]	63.51	64.53	67.09	64.41	61.52	64.65	76.63
	PSD [30]	67.16	67.31	69.67	66.17	65.17	67.33	78.79
	SQN [5]	63.18	64.68	67.72	67.52	64.38	65.49	78.63
	Ours	**75.94**	**77.33**	**78.02**	**73.54**	**78.52**	**76.47**	**86.64**

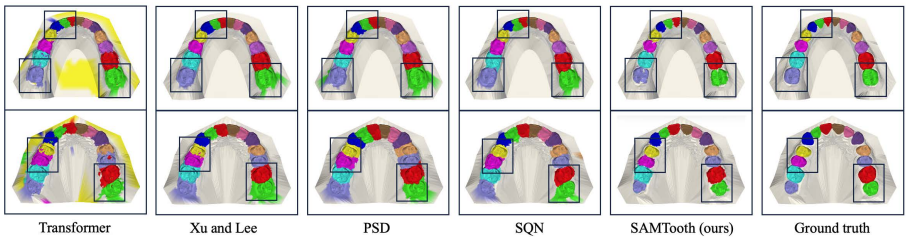


Fig. 3. Results comparison on 3DTeethSeg among previous methods and ours.

3.2 Main results

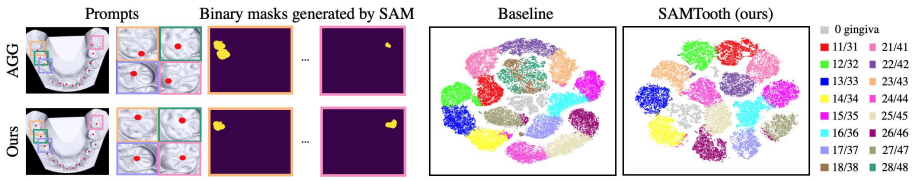
To make a fair comparison with recent state-of-the-art works [8,20,24,30,5], we use the same backbone and we re-produce their methods based on the official repositories. We present the comparison results in 1. SAMTooth achieves 76.47% mIoU and 86.64% mAcc, outperforming previous methods by a large margin. In particular, SAMTooth surpasses II-Model [8], MT [20], Xu and Lee [24], PSD [30], and SQN [5] by 15.32%, 12.47%, 6.47%, 11.35%, 11.82%, 9.14%, and 10.98% in mIoU, respectively. It is also worth noting that with only 0.1% annotations, SAMTooth can achieve comparable performance (76.47% vs 83.59% mIoU) with the fully supervised baseline, which reveals the effectiveness of the proposed framework, and also shows the great potential of SAM for providing training signals for tooth point cloud segmentation with limited labels. We also provide qualitative comparisons in Fig. 3. It can be observed that our method can deliver better segmentation results around boundary regions (black boxes), compared to methods training using other weakly-supervised methods.

3.3 More analysis

Confidence-aware Prompt Generation. To evaluate the effectiveness of CPG, we experiment with another prompt generation strategy *AGG*, which obtains point prompts by simple aggregation among each subgroup. As shown in Table. 2, such a simple aggregation strategy would cause a performance drop of 4.63% in mIoU, revealing the necessity of confidence guidance. We also report

Table 2. Ablation studies of CPG and MRL on 3DTeethSeg.

Methods	w/	Incisor	Canine	Premolar	Molar	Gingiva	mIoU
Baseline	-	57.62	61.11	63.94	63.20	72.58	61.64
Prompt generation	AGG	67.83	71.09	76.48	75.51	87.94	73.11
	CPG	74.03	77.97	80.07	78.87	90.65	77.74
Mask constraints	FL	68.44	72.16	75.46	75.99	88.41	73.55
	BL	61.38	65.63	67.65	67.51	70.96	65.32
	MRL	74.03	77.97	80.07	78.87	90.65	77.74



(a) Prompt comparison between *AGG* and our method. (b) Feature comparison between baseline and our method.

Fig. 4. Illustration of (a) prompt comparison and (b) feature comparison.

qualitative results in Fig. 4 (a), from which we observe prompts generated by *AGG* tend to bias from the center of the tooth, and such prompts would result in mistaken object masks. In contrast, prompts generated by *CPG* are often located around tooth centers and the resulting masks can seamlessly cover each tooth, which can benefit the subsequent representation learning.

Mask-guided Representation Learning. Apart from MRL, we experiment with other constraining strategies, including *FL* and *BL* that solely use foreground and background learning. As shown in Table. 2, using *FL* can already outperform the baseline with 12.40% mIoU gains, due to the complementary constraints for the foreground feature learning. Meanwhile, using *BL* can also bring 3.68% mIoU advancement. Furthermore, combining *FL* and *BL*, i.e., MRL, can improve the performance with the largest 15.32% mIoU improvements over the baseline, revealing the effectiveness of MRL. In addition, We present the T-SNE feature visualizations in Fig. 4 (b). In general, the features of the baseline are scattered with category mixing, e.g., 11/31, 12/32, and 13/33. In contrast, the features of SAMTooth are more intra-class compact with clear boundaries.

4 Conclusion

In this paper, we propose a novel framework for weakly-supervised tooth point cloud segmentation, coined SAMTooth. It leverages the recent advanced promptable foundation model, i.e., SAM, to complement the extremely sparse supervision (one point per tooth). It adopts a Confidence-aware Prompt Generation (CPG) to automatically generate precise prompts for SAM to use, guided by the estimated point-level confidence. Then, it leverages Mask-guided Representation Learning (MRL) to achieve maximal utilization of the fine-grained masks generated by SAM. Extensive experiments on two benchmarks show that the proposed

method shows significant superiority over existing approaches, showcasing the potential of applying SAM for 3D perception tasks.

Acknowledgements. This work was supported by Hong Kong Research Grants Council (RGC) General Research Fund 14204321.

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., 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)
2. Chen, W., Liu, Y., Hu, J., Yuan, Y.: Dynamic depth-aware network for endoscopy super-resolution. IEEE Journal of Biomedical and Health Informatics **26**(10), 5189–5200 (2022)
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. Hao, J., Liao, W., Zhang, Y., Peng, J., Zhao, Z., Chen, Z., Zhou, B., Feng, Y., Fang, B., Liu, Z., et al.: Toward clinically applicable 3-dimensional tooth segmentation via deep learning. Journal of dental research **101**(3), 304–311 (2022)
5. Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., Markham, A.: Sqn: Weakly-supervised semantic segmentation of large-scale 3d point clouds. In: European Conference on Computer Vision. pp. 600–619. Springer (2022)
6. Im, J., Kim, J.Y., Yu, H.S., Lee, K.J., Choi, S.H., Kim, J.H., Ahn, H.K., Cha, J.Y.: Accuracy and efficiency of automatic tooth segmentation in digital dental models using deep learning. Scientific reports **12**(1), 9429 (2022)
7. 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. arXiv preprint arXiv:2304.02643 (2023)
8. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
9. Li, C., Liu, H., Liu, Y., Feng, B.Y., Li, W., Liu, X., Chen, Z., Shao, J., Yuan, Y.: Endora: Video generation models as endoscopy simulators. arXiv preprint arXiv:2403.11050 (2024)
10. Li, M., Xie, Y., Shen, Y., Ke, B., Qiao, R., Ren, B., Lin, S., Ma, L.: Hybridcr: Weakly-supervised 3d point cloud semantic segmentation via hybrid contrastive regularization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14930–14939 (2022)
11. 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)
12. Liu, H., Liu, Y., Li, C., Li, W., Yuan, Y.: Lgs: A light-weight 4d gaussian splatting for efficient surgical scene reconstruction. arXiv preprint arXiv:2406.16073 (2024)
13. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv preprint arXiv:2401.12561 (2024)

14. Liu, Y., Li, W., Liu, J., Chen, H., Yuan, Y.: Grab-net: Graph-based boundary-aware network for medical point cloud segmentation. IEEE Transactions on Medical Imaging (2023)
15. Liu, Y., Liu, J., Yuan, Y.: Edge-oriented point-cloud transformer for 3d intracranial aneurysm segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 97–106. Springer (2022)
16. Liu, Z., Qi, X., Fu, C.W.: One thing one click: A self-training approach for weakly supervised 3d semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1726–1736 (2021)
17. Liu, Z., He, X., Wang, H., Xiong, H., Zhang, Y., Wang, G., Hao, J., Feng, Y., Zhu, F., Hu, H.: Hierarchical self-supervised learning for 3d tooth segmentation in intra-oral mesh scans. IEEE Trans. Med. Imaging (2022)
18. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
19. Tang, L., Chen, Z., Zhao, S., Wang, C., Tao, D.: All points matter: Entropy-regularized distribution alignment for weakly-supervised 3d segmentation. arXiv preprint arXiv:2305.15832 (2023)
20. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
21. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics (tog) **38**(5), 1–12 (2019)
22. Xiong, H., Li, K., Tan, K., Feng, Y., Zhou, J.T., Hao, J., Ying, H., Wu, J., Liu, Z.: Tsegformer: 3d tooth segmentation in intraoral scans with geometry guided transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 421–432. Springer (2023)
23. Xu, X., Liu, C., Zheng, Y.: 3d tooth segmentation and labeling using deep convolutional neural networks. IEEE transactions on visualization and computer graphics **25**(7), 2336–2348 (2018)
24. Xu, X., Lee, G.H.: Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13706–13715 (2020)
25. Yang, C., Liu, Y., Yuan, Y.: Transferability-guided multi-source model adaptation for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 703–712. Springer (2023)
26. Yang, C., Zhu, M., Liu, Y., Yuan, Y.: Fedpd: Federated open set recognition with parameter disentanglement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4882–4891 (2023)
27. Yang, C.K., Wu, J.J., Chen, K.S., Chuang, Y.Y., Lin, Y.Y.: An mil-derived transformer for weakly supervised point cloud segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11830–11839 (2022)
28. Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J., Lu, J.: Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19313–19322 (2022)
29. Zanjani, F.G., Moin, D.A., Claessen, F., Cherici, T., Parinussa, S., Pourtaherian, A., Zinger, S., de With, P.H.: Mask-mcnet: Instance segmentation in 3d point

cloud of intra-oral scans. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. pp. 128–136. Springer (2019)
30. Zhang, Y., Qu, Y., Xie, Y., Li, Z., Zheng, S., Li, C.: Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 15520–15528 (2021)