IOSSAM: Label Efficient Multi-View Prompt-Driven Tooth Segmentation

Xinrui Huang[1,*], Dongming He[2,*], Zhenming Li[2], Xiaofan Zhang[1,3(✉)], and Xudong Wang[2,4,5,6,7,8,9(✉)]

[1] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
{huangxr,xiaofan.zhang}@sjtu.edu.cn
[2] Department of Oral Craniomaxillofacial, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
xudongwang70@hotmail.com
[3] Shanghai Artificial Intelligence Laboratory, Shanghai, China
[4] College of Stomatology, Shanghai Jiao Tong University, Shanghai, China
[5] National Center for Stomatology, Shanghai, China
[6] National Clinical Medical Research Center for Oral Diseases, Shanghai, China
[7] Shanghai Key Laboratory of Stomatology, Shanghai, China
[8] Shanghai Research Institute of Stomatology, Shanghai, China
[9] Research Unit of Oral and Maxillofacial Regenerative Medicine, Chinese Academy of Medical Science, Shanghai, China

Abstract. Segmenting and labeling teeth from 3D Intraoral Scans (IOS) plays a significant role in digital dentistry. Dedicated learning-based methods have shown impressive results, while they suffer from expensive point-wise annotations. We aim at IOS segmentation with only low-cost 2D bounding-boxes annotations in the occlusal view. To accomplish this objective, we propose a SAM-based multi-view prompt-driven IOS segmentation method (IOSSAM) which learns prompts to utilize the pre-trained shape knowledge embedded in the visual foundation model SAM. Specifically, our method introduces an occlusal prompter trained on a dataset with weak annotations to generate category-related prompts for the occlusal view segmentation. We further develop a dental crown prompter to produce reasonable prompts for the dental crown view segmentation by considering the crown length prior and the generated occlusal view segmentation. We carefully design a novel view-aware label diffusion strategy to lift 2D segmentation to 3D field. We validate our method on a real IOS dataset, and the results show that our method outperforms recent weakly-supervised methods and is even comparable with fully-supervised methods.

Keywords: IOS segmentation · Weak annotations · Prompt learning · Foundation model.

* X. Huang and D. He contributed equally to this study.

1 Introduction

In digital dentistry, computer-aided design (CAD) system plays an essential role in various applications, e.g. orthodontic treatment, tooth design or tooth restoration. 3D Intraoral scans (IOS) are commonly used as basic data in dental CAD systems due to their precise dental surface representations. Segmenting teeth and identifying their FDI labels[10] based on IOS data is a fundamental task termed as IOS segmentation. Accurate segmentation and labeling is nontrivial because of the variation of teeth shape and the presence of abnormalities [2].

With the development of deep learning in 3D field, various learning-based automatic IOS segmentation algorithms have been proposed. Based on the difference of input data formats, existing methods can be roughly classified into two categories: processing 2D rendered images [3,12] or processing 3D data (point clouds or meshes) directly [16,29,14,9,10,15,24,28,7]. Although they have shown impressive performances, they are hungry for expensive point-wise annotations. The high cost hinders further enrichment of data which limits the generalization of learning-based methods. Unfortunately, there are few works focusing on IOS segmentation with weak annotations. The DArch proposed by Qiu et al.[21] is the first attempt to segment tooth instance based on weakly-annotated data (specifying all tooth centroids and only a few tooth masks for each IOS). However, labeling centroids of 3D teeth is complex, and their method fails to achieve tooth labeling directly due to the ignorance of the FDI notation.

In recent years, foundation models [26,4,13] have demonstrated extraordinary generalization capacity in few-shot or even zero-shot scenarios. As the recent advance of 2D foundation model, the Segment Anything Model (SAM) [11] has been proposed for promptable, category-agnostic segmentation and achieved amazing performance. Training on over one billion masks endows SAM with an understanding of 'object' which sheds light on a potential way to tooth segmentation with weak annotations. Our main idea is to learn prompts from low-cost 2D annotated data (i.e. bounding-boxes in the occlusal view as Fig.2a) to utilize the pre-trained knowledge about the shape of 'tooth' embedded in SAM for IOS segmentation. While this idea poses some challenges: (1) SAM is sensitive to the accuracy of generated prompts; (2) the segmentation from SAM is category-agnostic but our task requires FDI labels of segmented teeth as well; (3) SAM focuses on the 2D image segmentation, while we call for 3D masks.

To overcome these challenges, we propose a SAM-based multi-view prompt-driven IOS segmentation method (IOSSAM) which first learns category-related prompts for SAM to generate multi-view segmentation, then lifts it to the 3D field. Specifically, we introduce an occlusal prompter which is trained on rendered occlusal images with low-cost bounding-boxes annotations to produce accurate prompts for the occlusal view segmentation. It also predicts category labels for these prompts to enable SAM to produce semantically discernible segmentation. We also develop a dental crown prompter which produces appropriate prompts for the dental crown view segmentation by considering the dental crown length

[10] FDI World Dental Federation numbering system (also 'FDI notation')

prior and the obtained occlusal view segmentation. To lift 2D pixel-level segmentation to 3D point-level segmentation, we design a novel graph-based label diffusion strategy which considers the confidence of segmented teeth in different views during the diffusion process.

To summarize, main contributions of our work include: (1) We propose **a new SAM-based IOS segmentation method (IOSSAM)** which can only rely on weakly-annotated data. To the best of our knowledge, this is the first attempt to utilize SAM to facilitate IOS segmentation. (2) We introduce **the occlusal prompter and the dental crown prompter** to awaken the pre-trained shape knowledge embedded in the foundation model. We further design **a novel view-aware label diffusion strategy** to adapt SAM to the 3D field. (3) We compare our method with several recent methods on a real IOS dataset. Extensive evaluations show that our method can outperform weakly-supervised methods and is comparable with fully-supervised methods.

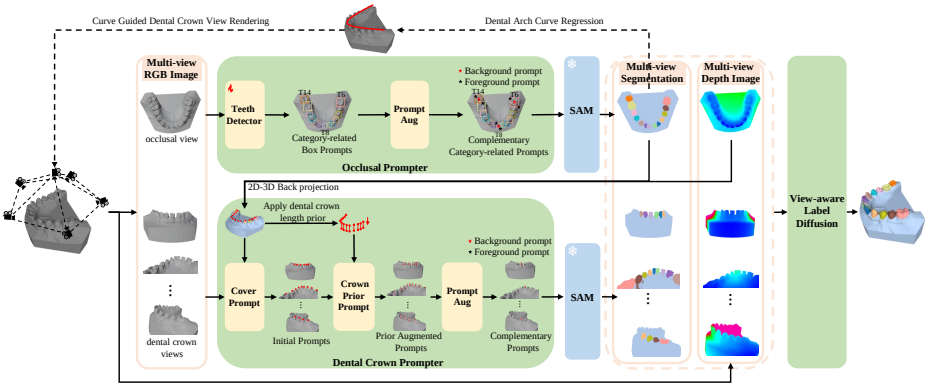


Fig. 1: The pipeline of our proposed IOSSAM for IOS segmentation.

2 Method

2.1 Overview

Our method mainly consists of an occlusal prompter, a dental crown prompter and a view-aware label diffusion module, as shown in Fig.1. For an input IOS, we first obtain its RGB and depth images from the view rendering which is perpendicular to the occlusal plane. The occlusal prompter consumes the RGB image and produces category-related complementary prompts for SAM to segment the rendered image. A dental arch curve is regressed base on tooth masks of the obtained occlusal view segmentation. Then, the IOS model are rendered into RGB and depth images from multiple dental crown views along the dental arch curve. The dental crown prompter generates prompts for the dental crown

segmentation by considering the dental crown length and the segmentation result of the occlusal view. Finally, multi-view segmentation results are lifted to 3D field and merged though our proposed view-aware label diffusion module.

2.2 Occlusal Prompter

In IOS segmentation, we call for not only segmentation masks of teeth but also their FDI notations, while SAM is category-agnostic. Considering that the occlusal view can provide complete category information, we propose an occlusal prompter to obtain category-related prompters from the occlusal view and enable SAM to produce semantic segmentation.

The occlusal prompter consists of two parts: a teeth detector module and a prompt augmentation module. The teeth detector detects each tooth and generate category-related box prompts. We implement it as a query-based detector [6]. The detector is the only trainable module in our method, so we only require low-cost 2D bounding-box annotations on the occlusal view (i.e. two annotated points per tooth). SAM is sensitive to prompts, and the precision and richness of input prompts can significantly influence the quality of the resulting segmentation [8]. We further propose a prompt augmentation module to generate additional foreground and background point-type prompts. We select the center of predicted bounding-boxes as foreground prompts and two corners as background prompts. According to these complementary category-related prompts, SAM generates segmentation for the occlusal view.

2.3 Dental Crown Prompter

The occlusal view covers a portion of the IOS and neglects a large area where teeth meet gum. Integrating segmentation results from more views can help to address this limitation. We render the IOS into RGB and depth images from multiple dental crown views selected along the dental arch curve (more details can be found in supplementary material sec.1), and we propose a dental crown prompter to generate prompts. The dental arch curve has been demonstrated to be approximated using a cubic Bézier curve [19]. We regress the dental arch curve based on tooth cusps found in occlusal segmentation masks. For each dental crown view, our prompter produces complementary prompts for SAM.

Firstly, we back project the occlusal view segmentation into 3D space and calculate the overlap on the IOS between the current dental crown view and the occlusal view. The overlap is projected to the rendered current view, and semantic labels are inherited to form initial masks. The mask centroids of visible teeth in the current dental crown view are used as initial prompts.

Then, we apply statistical priors of the crown length to move each tooth cusp (the point with the minimum depth in the predicted occlusal tooth mask) downward to approximate dental crown centers, and used them as additional foreground prompts. Some researches show that there are approximate proportional relationships between the crown width (CW_t) and the crown length (CL_t) of tooth t (i.e. $\frac{CW_t}{CL_t} = r_t$) [23,17,5]. The CW_t can be approximated by using the

width of predicted bounding box. We assume the z-axis coordinate of a tooth cusp is z_t, so the z-axis coordinate of the corresponding crown center can be approximated as $z_t^{'} = z_t - \frac{CW_t}{2r_t}$. Motivated by [5], we set r_t for central, lateral and canine teeth as 0.75, for premolar teeth as 0.8, for molar teeth as 1.7. Finally, we enhance prompts for each tooth by setting prompts of neighbouring teeth as background prompts. The dental crown view segmentation is generated based on these augmented prompts.

2.4 View-aware Label Diffusion

To assign point-wise labels for the whole IOS, all multi-view 2D segmentation results from SAM should be propagated to the 3D field and merged. Here, we first introduce some notations for illustration. The input IOS with N_{points} points can be denoted as $\mathcal{X} = \{x_i \in \mathbf{R}^3\}_{i=1}^{N_{points}}$. We assume the view set, $\mathcal{V} = \{v_k\}_{k=1}^{N_{views}}$, consists of N_{views} rendered images (v_1 is the occlusal view). Each view v_k contains a pixel coordinate set $\{p_j^{v_k} \in \mathbf{N}^2\}_{j=1}^{N_{pixels}^{v_k}}$ and predicted one-hot labels $\{y_j^{v_k} \in \{0,1\}^C\}_{j=1}^{N_{pixels}^{v_k}}$. We aggregate pixel labels to form a label matrix $L^{v_k} \in \mathbf{R}^{N_{pixels}^{v_k} \times C}$. We additionally initialize a zero matrix $L^{\mathcal{X}} \in \mathbf{R}^{N_{points} \times C}$ as the IOS label matrix. Our goal is to update $L^{\mathcal{X}}$ based on pixel label matrices $\{L^{v_1}, \cdots, L^{v_{N_{views}}}\}$. Inspired by [25,18],we formulate this task as a semi-supervised label diffusion problem on a graph.

We construct a graph by using all 2D pixels and 3D points as nodes and camera projection relationships as edges. The graph consists of multiple subgraph from different views. The construction of a subgraph G^{v_k} is illustrated in Fig.2b. The adjacency matrix $A^{v_k} \in \mathbf{R}^{N_{points} \times N_{pixels}^{v_k}}$ of G^{v_k} is defined as follows:

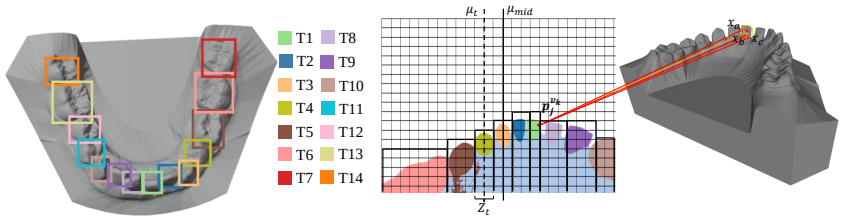
$$A_{i,j}^{v_k} = \begin{cases} \lambda(p_j^{v_k}) & if x_i \in KNN(\delta^{v_k}(p_j^{v_k})) \\ 0 & otherwise \end{cases}, \tag{1}$$

where $\lambda(p_j^{v_k})$ is the diffusion intensity, and $\delta^{v_k}(p_j^{v_k})$ is the 3D coordinate of pixel $p_j^{v_k}$ obtained by back projection. Specifically, the back projection has the form as,

$$\delta^{v_k}(p_j^{v_k})^T = R^{v_k -1} K^{v_k -1} [p_j^{v_k}, 1]^T - R^{v_k -1} T^{v_k}, \tag{2}$$

where $[p_j^{v_k}, 1]$ is homogeneous pixel coordinates. R^{v_k} and T^{v_k} are from the extrinsic camera matrix. K^{v_k} is the intrinsic camera matrix. Different from [25,18], our diffusion intensity is view-aware rather than fixed. This is based on the fact that as a tooth approaches the view center, its segmentation reliability increases due to smaller morphological distortion.

For the dental crown view, we first define 'comfortable zones' (Z_t) based on boundaries of segmentation masks as shown in Fig.2b. Then, we assign decreasing diffusion intensity based on the horizontal distance from the zone center to the



(a) (b)

Fig. 2: (a) Low-cost 2D annotations. (b) The illustration of subgraph construction and diffusion intensity calculation. Each pixel $p_j^{v_k}$ is back projected to 3D space, and its knn neighbors are found to construct pixel-point subgraph. Z_t is formed based on segmentation.

view center. The $\lambda(p_j^{v_k})$ can be formulated as follows:

$$\lambda(p_j^{v_k}) = \begin{cases} 1 - \alpha(\frac{|\mu_{mid} - \mu_t|}{w_{v_k}}) & if p_j^{v_k} \in Z_t \\ \beta & otherwise \end{cases}, \tag{3}$$

where μ_{mid} and μ_t are x-axis coordinates of view and zone center, w_{v_k} is the wide of view image, and α and β are coefficients. For the occlusal view v_1, we simply set $\lambda(p_j^{v_1}) = 1.0$ due to the visibility of all teeth. After getting all subgraphs $\{G^{v_1}, \cdots, G^{v_{N_{views}}}\}$, the IOS label matrix can be calculated as follows:

$$L^{\mathcal{X}} = \sum_{k=1}^{N_{view}} A^{v_k} L^{v_k}. \tag{4}$$

Finally, point-wise labels of the whole IOS are obtained according to:

$$y_i = \underset{c \in \{1, \cdots, C\}}{\arg\max} \; L_{i,c}^{\mathcal{X}} \tag{5}$$

3 Experiments and Results

3.1 Dataset and Evaluation Metrics

We evaluate our method on Teeth3ds dataset [1]. The dataset contains 1800 IOS collected for 900 patients covering upper and lower jaws. Following previous method [10], we use lower jaws (upper jaws share similar dental arches) in our experiments and exclude those IOS containing wisdom teeth (the number of categories, C, is 15 including 14 teeth and one gingiva). We use the train/test split provided by the dataset. To simulate weakly-annotated data, IOS in training set are rendered into RGB images about the occlusal view. Their annotations are generated by projecting ground truth masks to 2D bounding boxes. We use mean Dice Score (mDSC), Overall Accuracy (OA), mean Sensitivity (mSEN), and mean Positive Predictive Value (mPPV) to evaluate the performance.

3.2 Implementation Details

The tooth detector is trained using the AdamW optimizer for 300 epochs with a learning rate $1e^{-4}$ and batch size 24. Other settings keep the same as [6]. We select 14 dental crown views for rendering. In view-aware label diffusion module, coefficients $[\alpha, \beta]$ are empirically set as $[2, 0.1]$. Considering that some baselines are built on meshes, we interpolate the pointwise labels to mesh labels using k nearest neighbor, and $k = 3$ (all evaluations are based on meshes). Our implementation is available [11].

Table 1: The main segmentation results from different methods in terms of the labelwise Dice Score. Segmentation of symmetrical teeth is merged for simplicity, i.e. Gi (gingiva), I_1 (central incisors), I_2 (lateral incisors), Ca (canines), P_1 (1st premolars), P_2 (2nd premolars), M_1 (1st molars), M_2 (2nd molars). All baselines can be sequentially categorized into 3 types: fully supervised general segmentation, fully supervised tooth segmentation, and weakly annotated tooth segmentation.

Method	Gi	I_1	I_2	Ca	P_1	P_2	M_1	M_2	mDSC
PointNet++ [20]	0.908	0.879	0.841	0.846	0.853	0.832	0.857	0.734	0.846
DGCNN [22]	0.956	0.951	0.936	0.924	0.922	0.897	0.905	0.768	0.910
SGTNet [9]	*0.959*	*0.954*	*0.937*	0.928	0.923	0.902	*0.908*	*0.811*	0.917
ToothMeshSeg [10]	**0.968**	**0.967**	**0.954**	**0.941**	**0.938**	*0.907*	0.906	0.805	**0.925**
10×Fewer [27]	0.910	0.913	0.896	0.885	0.892	0.854	0.857	0.648	0.857
Ours	0.915	0.926	0.920	*0.930*	*0.925*	**0.930**	**0.934**	**0.865**	*0.920*

3.3 Quantitative and Qualitative Evaluation

We validate our method by comparing with different methods quantitatively and qualitatively. Baselines can be divided into three categories: (1) universal 3D segmentation [20,22]; (2) domain-specific IOS segmentation [10,9]; (3) weakly-supervised method [27]. For fair comparison, the weakly-supervised method is trained on a similar amount of annotations as our method (i.e. two annotated points per tooth).

The quantitative results are shown in Table 2a. We further show the labelwise results for each type of tooth in Table 1. We can find that our method outperforms previous weakly-supervised method in both global mean metrics and labelwise metrics, and our method trained on low-cost weak annotations can achieve comparable performance to fully-supervised methods. These results demonstrate that utilizing the pretrained knowledge of shape in SAM can facilitate IOS segmentation with weak annotations. We can also observe that our

[11] https://github.com/ar-inspire/IOSSAM

Table 2: The overall results (a) and ablation study (b).

(a) The overall results.

Method	OA	mDSC	mSEN	mPPV
PointNet++	0.876	0.846	0.863	0.848
DGCNN	0.933	0.910	0.922	0.919
SGTNet	*0.941*	0.917	*0.927*	0.923
ToothMeshSeg	**0.944**	**0.925**	**0.943**	*0.927*
10×Fewer	0.886	0.857	0.891	0.856
Ours	0.926	*0.920*	0.912	**0.934**

(b) The ablation study.

Row	Settings			mDSC
	PA	DCP	VDI	
1	-	✓	✓	0.904
2	✓	-	✓	0.916
3	✓	✓	-	0.902
4 (Ours)	✓	✓	✓	**0.920**

method performs better on the posterior dentition than on the anterior dentition (performances of some teeth are even better than fully-supervised methods). We think this is because posterior dentition is larger and more angular in shape which makes it easier to identify on the occlusal view.

The qualitative results are shown in Fig.3. We select some challenging cases for visualization, including missing two 2nd molars (case 1), asymmetrical dental arch (case 2), missing a 2nd premolar (case 3) and missing a canine (case 4). We can observe that our method has advantages in the segmentation and labeling of posterior dentition as well as in the edge segmentation. For case 1, our method distinguishes between 1st and 2nd molars correctly, while some baselines confuse them. We think this is because posterior dentition is easier to identify through the occlusal view detection compared to processing 3D data directly. For case 2, 3 and 4, our method generates uniform internal segmentation with clearer boundaries. We believe that the pre-trained shape knowledge embedded in SAM can provide an overall understanding about 'teeth' which facilitates the consistency of internal segmentation and boundary delineation.

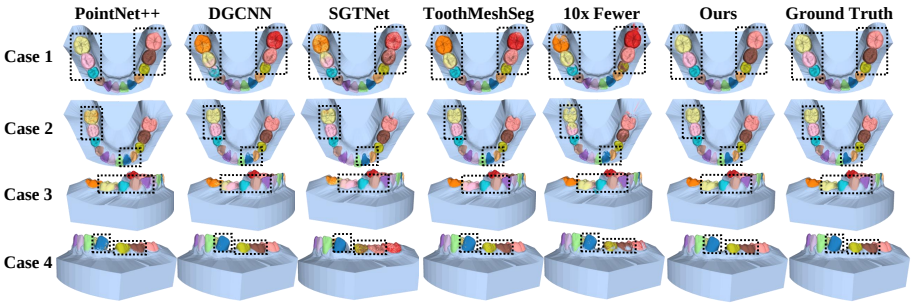


Fig. 3: The qualitative comparison of segmentation results produced by different methods along with the ground truth.

3.4 Ablation Study

In order to verify the effectiveness of different key modules, we perform ablation studies in this subsection. The main results are shown in Table 2b (more detailed label-wise results can be found in supplementary material). Here, PA is prompt augmentation, DCP means the dental crown prior, and VDI represents the view-aware diffusion intensity. '-' and '✓' mean remove or remain the corresponding module. From the table, we can find that these key modules all contribute to the final performance of our proposed method. Using the view-aware diffusion intensity setting is more important. The posterior teeth are more sensitive to the variation of settings compared to anterior teeth.

4 Conclusion

We propose IOSSAM, a novel label efficient IOS segmentation method based on SAM. We introduces two well-designed prompters, the occlusal prompter and the dental crown prompter, to make full use of pre-trained shape knowledge embedded in SAM by learning multi-view category-related prompts from weakly-annotated data. We design a new view-aware label diffusion strategy to lift multi-view 2D segmentation to 3D field and merged. Evaluations on real IOS data show the effectiveness of our proposed method.

Acknowledgments. This work was supported by Cultivation of Interdisciplinary Projects (YG2022ZD014), Shanghai Professional Service Platform of Oral Cranio Maxillofacial Digital Technology Research and Application (21DZ2294600), The Ninth People's Hospital of Shanghai Jiao Tong University School of Medicine "Cross Research Fund Project" (JYJC202129), Shanghai's Top Priority Research Cneter (2022ZZ01017), CAMS Innovation Fund for Medical Sciences(CIFMS) (2019-12M-5-037) and National Natural Science Foundation of China (62301311).

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., et al.: Teeth3ds: a benchmark for teeth segmentation and labeling from intra-oral 3d scans. arXiv preprint arXiv:2210.06094 (2022)
2. 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)
3. Boubolo, L., Dumont, M., Brosset, S., Bianchi, J., Ruellas, A., Gurgel, M., Massaro, C., Del Castillo, A.A., Ioshida, M., Yatabe, M., et al.: Flyby cnn: a 3d surface segmentation framework. In: Medical imaging 2021: Image processing. vol. 11596, pp. 627–632. SPIE (2021)

4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
5. Cao, R., Qiu, P., Ni, J., Xu, H., Pan, H., Cao, Y., et al.: A comprehensive analysis of clinical crowns in young of han nationality with normal occlusion using intraoral scanning. International Journal of Clinical Practice **2023** (2023)
6. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
7. 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)
8. Dai, H., Ma, C., Liu, Z., Li, Y., Shu, P., Wei, X., Zhao, L., Wu, Z., Zhu, D., Liu, W., et al.: Samaug: Point prompt augmentation for segment anything model. arXiv preprint arXiv:2307.01187 (2023)
9. Duan, F., Chen, L.: 3d dental mesh segmentation using semantics-based feature learning with graph-transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 456–465. Springer (2023)
10. Jana, A., Subhash, H.M., Metaxas, D.: 3d tooth mesh segmentation with simplified mesh cell representation. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
11. 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)
12. Leclercq, M., Ruellas, A., Gurgel, M., Yatabe, M., Bianchi, J., Cevidanes, L., Styner, M., Paniagua, B., Prieto, J.C.: Dentalmodelseg: Fully automated segmentation of upper and lower 3d intra-oral surfaces. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
13. Li, J., Li, D., Savarese, S., Hoi, S.: Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. arXiv preprint arXiv:2301.12597 (2023)
14. Li, Z., Liu, T., Wang, J., Zhang, C., Jia, X.: Multi-scale bidirectional enhancement network for 3d dental model segmentation. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
15. Lian, C., Wang, L., Wu, T.H., Liu, M., Durán, F., Ko, C.C., Shen, D.: Meshsnet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3d dental surfaces. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 837–845. Springer (2019)
16. 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)
17. Magne, P., Gallucci, G.O., Belser, U.C.: Anatomic crown width/length ratios of unworn and worn maxillary teeth in white subjects. The Journal of prosthetic dentistry **89**(5), 453–461 (2003)
18. Mascaro, R., Teixeira, L., Chli, M.: Diffuser: Multi-view 2d-to-3d label diffusion for semantic scene segmentation. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 13589–13595. IEEE (2021)
19. Noroozi, H., Hosseinzadeh Nik, T., Saeeda, R.: The dental arch form revisited. The Angle Orthodontist **71**(5), 386–389 (2001)

20. 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)
21. Qiu, L., Ye, C., Chen, P., Liu, Y., Han, X., Cui, S.: Darch: Dental arch prior-assisted 3d tooth instance segmentation with weak annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20752–20761 (2022)
22. Simonovsky, M., Komodakis, N.: Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3693–3702 (2017)
23. Sterrett, J.D., Oliver, T., Robinson, F., Fortson, W., Knaak, B., Russell, C.M.: Width/length ratios of normal clinical crowns of the maxillary anterior dentition in man. Journal of clinical periodontology **26**(3), 153–157 (1999)
24. Sun, D., Pei, Y., Li, P., Song, G., Guo, Y., Zha, H., Xu, T.: Automatic tooth segmentation and dense correspondence of 3d dental model. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 703–712. Springer (2020)
25. Wang, B.H., Chao, W.L., Wang, Y., Hariharan, B., Weinberger, K.Q., Campbell, M.: Ldls: 3-d object segmentation through label diffusion from 2-d images. IEEE Robotics and Automation Letters **4**(3), 2902–2909 (2019)
26. Wang, X., Wang, W., Cao, Y., Shen, C., Huang, T.: Images speak in images: A generalist painter for in-context visual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6830–6839 (2023)
27. 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)
28. Zanjani, F.G., Moin, D.A., Verheij, B., Claessen, F., Cherici, T., Tan, T., et al.: Deep learning approach to semantic segmentation in 3d point cloud intra-oral scans of teeth. In: International Conference on Medical Imaging with Deep Learning. pp. 557–571. PMLR (2019)
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)