Journal of Dental Sciences 20 (2025) 1110–1117

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

SegmentAnyTooth: An open-source deep learning framework for tooth enumeration and segmentation in intraoral photos

Khoa Dang Nguyen [a†], Hung Trong Hoang [b†], Thi-Phuong Hong Doan [a], Khai Quang Dao [b], Ding-Han Wang [a*], Ming-Lun Hsu [a**]

[a] College of Dentistry, National Yang Ming Chiao Tung University, Taipei, Taiwan
[b] Faculty of Odonto-Stomatology, University of Medicine and Pharmacy at Ho Chi Minh City, Viet Nam

Received 7 December 2024; Final revision received 3 January 2025
Available online 17 January 2025

KEYWORDS
Artificial intelligence;
Deep learning;
Intraoral photography;
Preventive dentistry;
Tooth segmentation;
Tooth enumeration

Abstract *Background/purpose:* Preventive dentistry is essential for maintaining public oral health, but inequalities in dental care, especially in underserved areas, remain a significant challenge. Image-based dental analysis, using intraoral photographs, offers a practical and scalable approach to bridge this gap. In this context, we developed SegmentAnyTooth, an open-source deep learning framework that solves the critical first step by enabling automated tooth enumeration and segmentation across five standard intraoral views: upper occlusal, lower occlusal, frontal, right lateral, and left lateral. This tool lays the groundwork for advanced applications, reducing reliance on limited professional resources and enhancing access to preventive dental care.

Materials and methods: A dataset of 5000 intraoral photos from 1000 sets (953 subjects) was annotated with tooth surfaces and FDI notations. You Only Look Once 11 (YOLO11) nano models were trained for tooth localization and enumeration, followed by Light Segment Anything in High Quality (Light HQ-SAM) for segmentation using an active learning approach.

Results: SegmentAnyTooth demonstrated high segmentation accuracy, with mean Dice similarity coefficients (DSC) of 0.983 ± 0.036 for upper occlusal, 0.973 ± 0.060 for lower occlusal, and 0.920 ± 0.063 for frontal views. Lateral view models also performed well, with mean DSCs of 0.939 ± 0.070 (right) and 0.945 ± 0.056 (left). Statistically significant improvements over baseline models such as U-Net, nnU-Net, and Mask R-CNN were observed (Wilcoxon signed-rank test, $P < 0.01$).

Conclusion: SegmentAnyTooth provides accurate, multi-view tooth segmentation to enhance dental care, early diagnosis, individualized care, and population-level research. Its open-

* Corresponding author. College of Dentistry, National Yang Ming Chiao Tung University, No. 155, Sec. 2, Linong St., Taipei 112, Taiwan.
** Corresponding author. College of Dentistry, National Yang Ming Chiao Tung University, No. 155, Sec. 2, Linong St., Taipei 112, Taiwan.
E-mail addresses: dhwang@nycu.edu.tw (D.-H. Wang), mlhsu@nycu.edu.tw (M.-L. Hsu).
† These two authors contributed equally to this work.

https://doi.org/10.1016/j.jds.2025.01.003

source design supports integration into clinical and public health workflows, with ongoing improvements focused on generalizability.

Introduction

Access to dental care remains a significant global challenge, particularly in underserved rural areas. The density of dentists per 10,000 population varies drastically: Africa has only 0.2 dentists, South-East Asia 1.5 dentists, while Europe boasts 6.3 dentists.[1] These disparities underscore critical inequalities in dental care access, with urban areas typically having more dentists while rural regions face severe shortages.[2—4] Moreover, significant workforce shortages and heavy workloads often compel dentists to focus on patient treatment rather than public health activities like surveillance, prevention, and early detection.

To address these challenges, scalable solutions are needed. We propose image-based dental analysis, which means using intraoral images taken at local clinics or community centers, which can be automatically analyzed by advanced deep learning models. This reduces the need for dentists to conduct surveys or preventive tasks in person and allows for broader monitoring of the population.

With this vision, intraoral photo sets of five standard views—upper occlusal, lower occlusal, frontal, right lateral, and left lateral—now hold significant potential for image-based dental analysis. Leveraging single or multiple views, artificial intelligence (AI) models can assist in detecting gingivitis, caries, plaque, calculus, and fissure sealants, further advancing the feasibility of this approach.[5—10]

Tooth enumeration (identifying teeth) and tooth segmentation (outlining teeth) are critical first steps in image-based dental analysis but remain challenging tasks, hindered by variations in lighting, positioning, and anatomy. Early methods, such as edge detection and thresholding, were limited to single-view photos and often produced suboptimal results.[11—13] More recently, AI-driven models have shown promise: El Basat et al. evaluated deep learning models for tooth enumeration and segmentation in upper occlusal views,[14] while Park et al. employed U-Net for tooth segmentation to aid in caries detection.[15] Yoon et al. used Mask R-CNN for enumeration in all five views, also aiding in caries detection.[16]

However, most of these models are limited to specific views or primarily focus on enumeration, limiting precise tooth positioning and shape evaluation. Furthermore, the lack of accessible source code significantly restricts broader adoption and replication in subsequent studies.

To address these limitations, we developed SegmentAnyTooth, an open-source deep learning framework for accurate tooth segmentation and enumeration across five standard intraoral views. SegmentAnyTooth automates multi-view intraoral image analysis, providing a scalable and accessible solution for clinical workflows, oral health research, and large-scale public health surveys. By leveraging AI-driven automation, the framework facilitates early detection, preventive care, and population-level oral health monitoring, particularly in underserved regions with limited access to dental care, helping to reduce inequalities in oral health outcomes.

Materials and methods

Data collection

a. Intraoral photos collection

This study was approved by the Board of Ethics in Biomedical Research at University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam (Approval No. 599/UMP-BOARD). A total of 5000 intraoral photographs, covering five standard views (upper occlusal, lower occlusal, frontal, right lateral, and left lateral), were randomly selected from the archives of the Department of Dental Public Health, Faculty of Odonto-Stomatology, University of Medicine and Pharmacy at Ho Chi Minh City. The photos were captured between 2017 and 2022 using various devices, including smartphones and DSLR cameras, with and without flash.

Only images of permanent teeth were included. Exclusion criteria included photos with primary or supernumerary teeth, dentures, or fixed appliances (except for lingual retainers, orthodontic attachments, and bands), or poor quality due to lighting or blurriness.

b. Data annotation

Two calibrated dentists (Dentist K., and H., each with over five years of experience) annotated the images using the Computer Vision Annotation Tool (CVAT, version 2.4.1, CVAT.ai Corp., Palo Alto, CA, USA).[17] Each image was annotated with all visible tooth surfaces and the corresponding Fédération Dentaire International (FDI) notations. Teeth from the second premolar to the third molar were excluded in lateral views due to their typical absence.

To prevent fatigue-relating errors, photos were divided into sets of 50 photos. Each set was annotated by one dentist and reviewed by the other to minimize errors. Disagreements unresolved by consensus led to case exclusion. The two dentists annotated 300 photos manually for initial model training, then refined the model's predictions using an active learning approach at intervals of 300, 500, 800, and 1000 photos, reducing manual effort over time.

c. Data splitting

The dataset was randomly divided into training (80 %) and testing (20 %) sets while ensuring all photos from the same subject were in the same set. Each view used 800 photos for training and 200 for testing.

Deep learning model for tooth enumeration and segmentation

a. Model architecture

SegmentAnyTooth combines two models: You Only Look Once 11 nano (YOLO11 nano, Ultralytics, Frederick, MD, USA), an object detection model, to draw bounding boxes around each tooth and assign tooth numbers, and Light Segment Anything in High Quality (Light HQ-SAM) for precise tooth segmentation, with the bounding boxes serving as prompts.[18,19] These "nano" and "light" versions prioritize efficiency for real-world applications.

As shown in Fig. 1A, four YOLO11 models handle tooth enumeration for upper occlusal, lower occlusal, frontal, and lateral views, while a single Light HQ-SAM performs

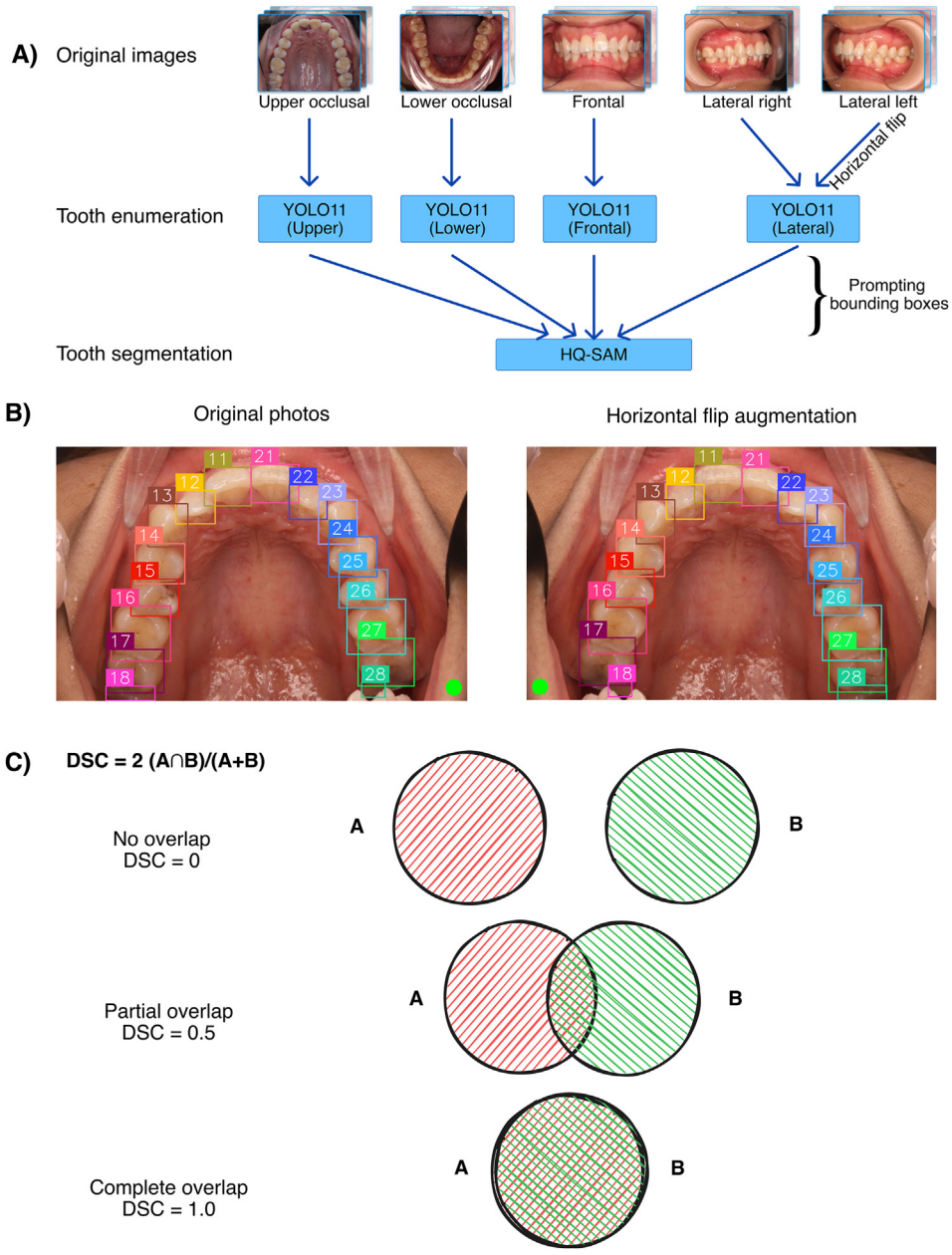

Figure 1 SegmentAnyTooth model architecture and methods. A) SegmentAnyTooth model architectures: original images are processed by YOLO11 for bounding box generation, followed by HQ-SAM for segmentation results. B) Illustration of custom horizontal flip augmentation applied to an upper occlusal photo. The green dot at the bottom indicates the photo's orientation. C) Dice similarity coefficient (DSC) calculation, demonstrating cases of no overlap, partial overlap, and complete overlap between ground truth (A) and predicted (B) segments. Abbreviation: YOLO11 (You Only Look Once version 11), HQ-SAM (Segment Anything in High Quality). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

segmentation. The right lateral view is the primary reference, with left lateral images flipped during training and inference for alignment.

b. Model training

YOLO11 models were fine-tuned using Ultralytics library version 8.3.18, with multiple key adjustments.[19] The input image size was set to 1024x1024 pixels, a batch size of 16, and a maximum of 2000 epochs, applying early stopping after 300 epochs without improvement. Horizontal flip augmentation was disabled; instead, images were manually flipped with adjusted tooth numbers (for example, tooth 11 was swapped with 21 in upper occlusal views). Fig. 1B illustrates this process.

For Light HQ-SAM, only the image encoder and mask decoder were fine-tuned, while the prompt encoder was frozen, following the procedure outlined by Kirillov et al.[20] Training used a 1024x1024 pixels input size, capped at 300 epochs, and early stopping after 30 epochs without improvement in Dice similarity coefficient (DSC). Augmentations included changes in hue, saturation, brightness, contrast, and random transformations.

All training was conducted on a workstation with an Intel Core i9-14900 K processor (Intel Corp., Santa Clara, CA, USA), 128 GB RAM, and dual NVIDIA GeForce RTX 4090 GPUs (NVIDIA Corp., Santa Clara, CA, USA) using PyTorch and Python 3.10.[21,22]

c. Model evaluation

Performance was assessed using DSC to evaluate the overlap between predicted segmentation and ground truth segmentations (Fig. 1C).[23,24] A macro average, representing the unweighted mean across all classes, was calculated. SegmentAnyTooth was compared with U-Net, nnU-Net, and Mask R-CNN for benchmarking.[25–27]

d. Statistical analysis

The Shapiro–Wilk test indicated non-normal data distribution, one-sided Wilcoxon signed-rank tests were used to compare SegmentAnyTooth with the other networks, with the alternative hypothesis that SegmentAnyTooth performs better. Analysis was performed using JASP 0.19.1 (JASP Team, Amsterdam, The Netherlands), with a significance level of 0.05.

Results

Data distribution

The dataset consisted of 5000 intraoral photographs from 953 Vietnamese subjects, organized into 1000 sets. The mean age of subjects was 25.1 ± 5.7 years (range: 12–63 years), with 313 males (32.8 %) and 640 females (67.2 %). Disregarding third molars, 637 sets contained all 28 teeth. Missing teeth distribution included 144 sets missing 1 tooth, 71 missing 2 teeth, 33 missing 3 teeth, 103 missing 4 teeth, and 9 missing 5 teeth. Only 1 image set was missing between 6 and 8 teeth. Fig. 2 details tooth distribution across sets.

Performance evaluation of SegmentAnyTooth

SegmentAnyTooth achieved high segmentation performance across all views, with the highest mean DSC in upper occlusal (0.983 ± 0.036) and lower occlusal (0.973 ± 0.060) views, as shown in Table 1. Performance in the frontal view (0.920 ± 0.063) and lateral views (right: 0.939 ± 0.070, left: 0.945 ± 0.056) was slightly lower but consistent. Fig. 3 shows example segmentations.

In occlusal views, most teeth were segmented with high precision, with mean DSC values ranging from 0.971 ± 0.141 to 0.995 ± 0.003 for teeth 21–23 (Table 2). The lower

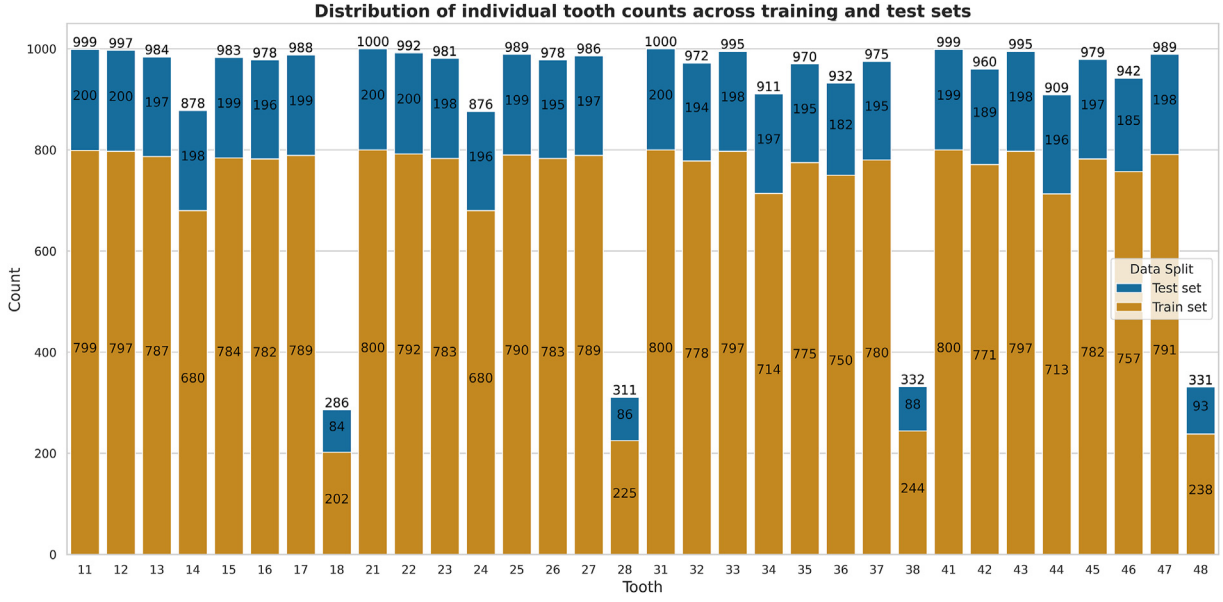

Figure 2 Distribution of individual tooth counts across training and test sets.

Table 1 Dice similarity coefficient of SegmentAnyTooth performance on the test set (n = 200 photos) across intraoral photo views.

View Model		Upper occlusal	Lower occlusal	Frontal	Lateral right	Lateral left
SegmentAnyTooth	Mean ± SD	0.983 ± 0.036	0.973 ± 0.060	0.920 ± 0.063	0.939 ± 0.070	0.945 ± 0.056
	Median (IQR)	0.995 (0.004)	0.993 (0.007)	0.931 (0.073)	0.952 (0.081)	0.953 (0.079)
U-Net	Mean ± SD	0.921 ± 0.031	0.908 ± 0.056	0.779 ± 0.069	0.895 ± 0.065	0.890 ± 0.057
	Median (IQR)	0.933 (0.007)***	0.929 (0.011)***	0.793 (0.088)***	0.906 (0.077)***	0.901 (0.069)***
nnU-Net	Mean ± SD	0.814 ± 0.285	0.738 ± 0.346	0.847 ± 0.204	0.904 ± 0.070	0.911 ± 0.059
	Median (IQR)	0.935 (0.066)***	0.931 (0.131)***	0.909 (0.080)***	0.917 (0.064)***	0.940 (0.070)***
Mask R-CNN	Mean ± SD	0.830 ± 0.109	0.847 ± 0.095	0.751 ± 0.112	0.817 ± 0.103	0.813 ± 0.104
	Median (IQR)	0.853 (0.164)***	0.863 (0.125)***	0.767 (0.158)**	0.835 (0.132)***	0.828 (0.149)***

One-sided Wilcoxon signed-rank test; the alternative hypothesis specifies that SegmentAnyTooth outperforms the other networks.
P values: **: $P < 0.01$, ***: $P < 0.001$.
Abbreviations: SD — standard deviation; IQR — interquartile range.

incisors exhibited a slight decline, with mean DSC values between 0.956 ± 0.182 and 0.974 ± 0.122. Similarly, third molars showed reduced performance, likely due to their lower prevalence in the dataset and poorer lighting conditions in the posterior regions.

In lateral views, SegmentAnyTooth performed well for teeth on the same lateral side. In the lateral right view, upper right quadrant teeth achieved mean DSC ranging from 0.961 ± 0.174 (tooth 16) to 0.990 ± 0.070 (tooth 11), while the lower right quadrant ranged from 0.925 ± 0.240 (tooth 42) to 0.968 ± 0.155 (tooth 45). Opposite-side teeth showed lower accuracy (DSCs: 0.811 ± 0.356 to 0.839 ± 0.333). Interestingly, first premolars from the opposite side outperformed the adjacent canines from the same quadrant, such as tooth 33 (0.839 ± 0.333) compared to tooth 34 (0.952 ± 0.206). The lateral left view exhibited similar trends.

In the frontal view, anterior teeth achieved high DSC values of 0.995 ± 0.004 for tooth 21 and 0.989 ± 0.072 for tooth 11. However, accuracy declined for posterior teeth, with mean DSC values ranging from 0.774 ± 0.370 (tooth 26) to 0.900 ± 0.266 (tooth 36) in the first and second molars regions. Interestingly, third molars outperformed other molars within the same quadrant.

Comparison with existing networks

SegmentAnyTooth outperformed U-Net, nnU-Net, and Mask R-CNN across all views (Table 1). For upper occlusal views, SegmentAnyTooth achieved a median DSC of 0.995 (interquartile range (IQR): 0.004), significantly higher than U-Net (0.933, IQR: 0.007, $P < 0.001$), nnU-Net (0.935, IQR: 0.066, $P < 0.001$), and Mask R-CNN (0.853, IQR: 0.164, $P < 0.001$). In the challenging frontal view, SegmentAnyTooth scored 0.931 (IQR: 0.073), surpassing U-Net (0.793, IQR: 0.088, $P < 0.001$), nnU-Net (0.909, IQR: 0.080, $P < 0.001$), and Mask R-CNN (0.767, IQR: 0.158, $P < 0.01$). Similar trends were observed in the lower occlusal and lateral views, with SegmentAnyTooth scoring 0.993 (lower occlusal), 0.952 (lateral right), and 0.953 (lateral left). The Wilcoxon signed-rank test confirmed these differences were statistically significant.

Discussion

This study introduced SegmentAnyTooth, a deep learning model built on the novel HQ-SAM framework, which was trained on SA-1B, the largest segmentation dataset to date.[18,20] HQ-SAM's transformer-based architecture and prompt-based segmentation enable precise and efficient tooth segmentation across five intraoral views, surpassing CNNs such as U-Net, nnU-Net, and Mask R-CNN. Unlike CNNs, which rely on local feature extraction, HQ-SAM captures global image context and dynamically adapts to prompts, improving segmentation accuracy, generalizability, and robustness across diverse dental structures.[18,28] As shown in Table 1, SegmentAnyTooth consistently outperformed U-Net, nnU-Net, and Mask R-CNN, achieving higher median DSC values across all views, such as 0.995 in the upper occlusal view and 0.931 in the frontal view. These benchmarking models have been widely used in previous studies for tooth enumeration and segmentation.[14–16]

The model demonstrated high accuracy, with the best performance in the upper occlusal view (DSC 0.920 ± 0.063) and slightly lower results in the frontal view (DSC 0.920 ± 0.063), as detailed in Tables 1 and 2. Its strong performance can be attributed to the inherent symmetry of dental structures and the clear positional cues for tooth enumeration in the images. The accuracy and efficiency of this model highlight its potential utility in oral public health and preventive dentistry, where large-scale dental health analysis is critical for population-level initiatives and early interventions.

Occlusal models consistently outperformed smooth surface models, likely due to the distinct features of occlusal surfaces. Additionally, smooth surface models face greater complexity, partly because they must handle a larger number of teeth—24 for lateral views and 32 for frontal views, compared to just 16 for occlusal views.

Analyzing cases of unusually high or low performance relative to other teeth reveals important insights into the model's strengths and limitations. In occlusal views, performance on third molars was noticeably lower, likely due to their low prevalence in the dataset and the difficulty in

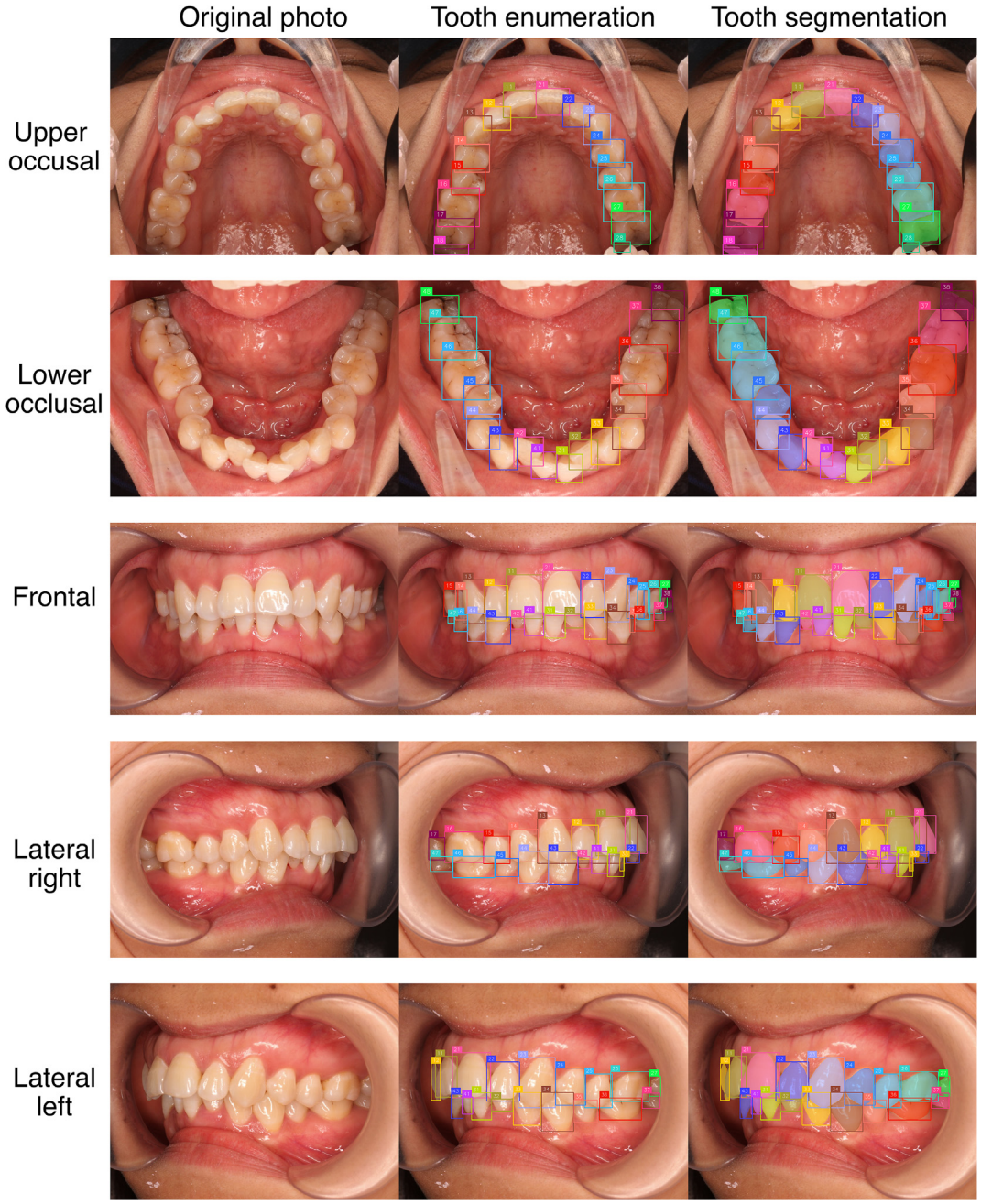

Figure 3 Example segmentation results from SegmentAnyTooth across various intraoral views. The first column shows the original images, the second column displays tooth enumeration results processed by YOLO11 nano, and the third column presents the tooth segmentation results generated by Light HQ-SAM.

capturing them because of saliva, mirror condensation, and patient gagging reflex. Incorporating half-arch photos taken with a small mirror could improve their analysis and serve as a focus for future research. Interestingly, in frontal and lateral views, third molars exhibited higher DSC values compared to second molars, which can be attributed to the DSC metric assigning a value of 1 when both the ground truth and predicted segmentations omit the third molars.[24] Similarly, in the lateral view, this calculation inflates performance for the first premolar on the opposite side. In the frontal view, the upper second premolars (tooth 15, DSC: 0.835 $\pm$ 0.317) performed worse than the upper first premolars (tooth 14, DSC: 0.936 $\pm$ 0.203), likely due to the palatal inclination of the dental arch reducing their visibility (Table 2).

The model's focus on permanent dentition limits its applicability to younger patients, highlighting the need for future research on primary and mixed dentitions. Additionally, the dataset's age distribution while spanning 12–63 years, is concentrated around 25.1 $\pm$ 5.7 years, which may reduce generalizability to other age groups. Future studies should include a more diverse age range to improve the model's robustness across broader demographic groups.

Table 2 Mean and standard deviation of Dice similarity coefficient for SegmentAnyTooth performance on the test set for each tooth across intraoral photo views.

View	Occlusal		Frontal		Lateral right		Lateral left	
Tooth	Mean ± SD	n	Mean ± SD	n	Mean ± SD	n	Mean ± SD	n
11	0.980 ± 0.121	200	0.989 ± 0.072	200	0.990 ± 0.070	200	0.974 ± 0.128	199
12	0.983 ± 0.103	200	0.975 ± 0.135	196	0.969 ± 0.155	198	0.816 ± 0.347	138
13	0.996 ± 0.003	197	0.969 ± 0.155	198	0.981 ± 0.105	197	0.832 ± 0.334	103
14	0.991 ± 0.071	198	0.936 ± 0.203	199	0.982 ± 0.102	198	0.950 ± 0.218	6
15	0.994 ± 0.030	199	0.835 ± 0.317	185	0.979 ± 0.121	196	n/a	0
16	0.982 ± 0.105	196	0.782 ± 0.354	176	0.961 ± 0.174	193	n/a	0
17	0.975 ± 0.124	199	0.847 ± 0.323	109	0.963 ± 0.157	125	n/a	0
18	0.957 ± 0.183	83	0.990 ± 0.099	2	0.975 ± 0.156	5	n/a	0
21	0.995 ± 0.003	200	0.995 ± 0.004	200	0.985 ± 0.072	198	0.988 ± 0.077	199
22	0.995 ± 0.003	200	0.985 ± 0.074	200	0.817 ± 0.341	132	0.988 ± 0.075	199
23	0.995 ± 0.003	198	0.953 ± 0.194	196	0.815 ± 0.354	101	0.986 ± 0.099	198
24	0.991 ± 0.070	196	0.935 ± 0.193	195	0.975 ± 0.156	4	0.973 ± 0.139	197
25	0.991 ± 0.070	199	0.880 ± 0.264	187	n/a	0	0.968 ± 0.155	196
26	0.985 ± 0.100	195	0.774 ± 0.370	183	n/a	0	0.967 ± 0.148	196
27	0.980 ± 0.103	198	0.803 ± 0.367	111	n/a	0	0.953 ± 0.185	127
28	0.948 ± 0.197	86	0.985 ± 0.122	3	n/a	0	0.980 ± 0.140	4
31	0.965 ± 0.155	200	0.967 ± 0.142	199	0.935 ± 0.216	195	0.956 ± 0.183	195
32	0.962 ± 0.159	194	0.978 ± 0.100	194	0.811 ± 0.356	164	0.957 ± 0.183	194
33	0.972 ± 0.139	198	0.963 ± 0.169	197	0.839 ± 0.333	142	0.984 ± 0.102	199
34	0.980 ± 0.121	197	0.934 ± 0.225	196	0.952 ± 0.206	10	0.979 ± 0.123	197
35	0.974 ± 0.140	195	0.920 ± 0.209	188	n/a	0	0.974 ± 0.139	192
36	0.978 ± 0.125	182	0.900 ± 0.266	180	n/a	0	0.960 ± 0.161	181
37	0.971 ± 0.144	195	0.849 ± 0.317	133	n/a	0	0.949 ± 0.183	150
38	0.972 ± 0.142	88	0.963 ± 0.184	16	n/a	0	0.975 ± 0.156	6
41	0.974 ± 0.122	199	0.955 ± 0.174	197	0.933 ± 0.226	197	0.943 ± 0.205	198
42	0.956 ± 0.182	189	0.941 ± 0.217	195	0.925 ± 0.240	190	0.824 ± 0.344	182
43	0.971 ± 0.141	198	0.956 ± 0.184	196	0.965 ± 0.170	198	0.862 ± 0.300	148
44	0.983 ± 0.102	196	0.952 ± 0.184	195	0.963 ± 0.171	193	0.951 ± 0.207	19
45	0.982 ± 0.105	197	0.904 ± 0.245	181	0.968 ± 0.155	193	n/a	0
46	0.989 ± 0.051	185	0.876 ± 0.300	172	0.953 ± 0.184	181	n/a	0
47	0.983 ± 0.102	198	0.821 ± 0.344	150	0.938 ± 0.201	165	n/a	0
48	0.954 ± 0.195	93	0.926 ± 0.255	19	0.954 ± 0.207	9	n/a	0

Abbreviations: SD — standard deviation; n/a - not available.

Future enhancements could include post-processing logic to address misassigned tooth numbers by applying rules-based corrections.[29] This could include removing low-confidence detections, resolving overlaps, and reassigning teeth based on their positional relationships within the dental arch. Additionally, efforts will focus on enabling functions such as querying clinical databases, generating dental charts, and integrating tooth enumeration with lesion detection systems to fully implement our model in clinical settings.[5,30–32]

In conclusion, SegmentAnyTooth is a tool capable of accurate tooth enumeration and segmentation across multiple intraoral views, and our team will release it as open source. Its efficiency, accessibility, and readiness for integration into diagnostic workflows aim to provide useful support to the dental community. By sharing the model freely, we hope to encourage collaboration and innovation, contributing to progress in preventive dentistry. The model can be accessed at: https://github.com/thangngoc89/SegmentAnyTooth.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgments

This work was supported by the Ministry of Health and Welfare; the National Science and Technology Council (NSTC), Taiwan (Grant numbers NSTC 113-2314-B-A49-019, NSTC 113-2314-B-A49-020-MY2, NSTC 113-2314-B-A49-006); the Higher Education Sprout Project of the Yen Tjing Ling Medical Foundation; and National Yang Ming Chiao Tung University and Ministry of Education (MOE), Taiwan.

References

1. World Health Organization. *The global health observatory data: dentists (per 10 000 population)*. Geneva, Switzerland.

Available online: who.int/data/gho/data/indicators/indicator-details/GHO/dentists-(per-10-000-population). [Accessed December 17, 2024].
2. Willie-Stephens J, Kruger E, Tennant M. Public and private dental services in NSW: a geographic information system analysis of access to care for 7 million Australians. *N S Wales Public Health Bull* 2014;24:164—70.
3. Zhang Y, Lu Z, Cheng R, Liu L. Current state of allocation of oral health human resources in northern China and future needs. *Int J Dent Hyg* 2015;13:268—72.
4. World Health Organization. *Retention of the health workforce in rural and remote areas: a systematic review*, 1st ed. World Health Organization, 2020.
5. Zhang JW, Fan J, Zhao FB, Ma B, Shen XQ, Geng YM. Diagnostic accuracy of artificial intelligence-assisted caries detection: a clinical evaluation. *BMC Oral Health* 2024;24:1095.
6. You W, Hao A, Li S, Wang Y, Xia B. Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health* 2020;20:141.
7. Wang C, Zhang R, Wei X, Wang L, Wu P, Yao Q. Deep learning and sub-band fluorescence imaging-based method for caries and calculus diagnosis embeddable on different smartphones. *Biomed Opt Express* 2023;14:866.
8. Moharrami M, Farmer J, Singhal S, et al. Detecting dental caries on oral photographs using artificial intelligence: a systematic review. *Oral Dis* 2024;30:1765—83.
9. Wen C, Bai X, Yang J, Li S, Wang X, Yang D. Deep learning based approach: automated gingival inflammation grading model using gingival removal strategy. *Sci Rep* 2024;14:19780.
10. Schlickenrieder A, Meyer O, Schönewolf J, et al. Automatized detection and categorization of fissure sealants from intraoral digital photographs using artificial intelligence. *Diagnostics* 2021;11:1608.
11. Fernandez K, Chang C. Teeth/palate and interdental segmentation using artificial neural networks. In: Mana N, Schwenker F, Trentin E, eds. *Artificial neural networks in pattern recognition. Lecture notes in computer science*, 9351. Berlin Heidelberg: Springer, 2012:175—85.
12. Na SD, Lee G, Lee JH, Kim MN. Individual tooth region segmentation using modified watershed algorithm with morphological characteristic. *Bio Med Mater Eng* 2014;24:3303—9.
13. Lin CC, Lee HJ. *Detection and Segmentation of Teeth in dental digital images. MSD thesis*. Hsinchu, Taiwan: National Yang-Ming Chiao Tung University; 2004. Available online: http://hdl.handle.net/11536/74179. [Accessed 14 June 2023].
14. El Bsat AR, Shammas E, Asmar D, et al. Semantic segmentation of maxillary teeth and palatal rugae in two-dimensional images. *Diagnostics* 2022;12:2176.
15. Park EY, Cho H, Kang S, Jeong S, Kim EK. Caries detection with tooth surface segmentation on intraoral photographic images using deep learning. *BMC Oral Health* 2022;22:573.
16. Yoon K, Jeong HM, Kim JW, Park JH, Choi J. AI-based dental caries and tooth number detection in intraoral photos: model development and performance evaluation. *J Dent* 2024;141:104821.
17. CVAT.ai Corporation. *Computer vision annotation tool (CVAT) (v2.4.1)*. *Zenodo*. Available from:. 2023. https://doi.org/10.5281/zenodo.7803312.
18. Ke L, Ye M, Danelljan M, et al. Segment anything in high quality. In: *Proceedings of the 37th international conference on neural information processing systems*. Curran Associates Inc., 2023:29914—34.
19. Jocher G, Qiu J, Chaurasia A. YOLO11 by Ultralytics. Available online: github.com/ultralytics/ultralytics. [Accessed October 20, 2024].
20. Kirillov A, Mintun E, Ravi N, et al. *Segment anything*. Paris, France: 2023 IEEE/CVF International Conference on Computer Vision, 2023:3992—4003.
21. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. In: *Proceedings of the 33rd international conference on neural information processing systems*. Curran Associates Inc., 2019:8026—37.
22. Bradley EL. *Medical and Surgical Management*, 2nd ed. 1982:72—95, Philadelphia: Saunders.
23. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imag* 2015;15:29.
24. Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297—302.
25. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical image Computing and computer-assisted intervention — MICCAI 2015. Lecture notes in computer science*, 9351. Springer International Publishing, 2015:234—41.
26. Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203—11.
27. He K, Gkioxari G, Dollár P, Girshick R. *Mask R-CNN*. Venice, Italy: 2017 IEEE international conference on computer vision, 2017:2980—8.
28. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems*. Curran Associates Inc., 2017:6000—10.
29. Chen H, Zhang K, Lyu P, et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep* 2019;9:3840.
30. Zhang X, Liang Y, Li W, et al. Development and evaluation of deep learning for screening dental caries from oral photographs. *Oral Dis* 2022;28:173—81.
31. Thanh MTG, Van Toan N, Ngoc VTN, Tra NT, Giap CN, Nguyen DM. Deep learning application in dental caries detection using intraoral photos taken by smartphones. *Appl Sci* 2022;12:5504.
32. Ding B, Zhang Z, Liang Y, et al. Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm. *Ann Transl Med* 2021;9:1622.