

Photography-based Dental Plaque Detection and Report Generation among Preschool Children using PlaqueSAM

Jing Hao

The University of Hong Kong

Kaixin Guo

The University of Hong Kong

Zhiyi Shan

The University of Hong Kong

Yanqi Yang

The University of Hong Kong

James Kit-Hon Tsoi

The University of Hong Kong

Phoebe Pui Ying Lam

The University of Hong Kong

Kuo Feng Hung

hungkfg@hku.hk

The University of Hong Kong

Article

Keywords:

Posted Date: March 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9027573/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at npj Digital Medicine on October 11th, 2026. See the published version at <https://doi.org/10.1038/s41746-026-03385-9>.

Photography-based Dental Plaque Detection and Report Generation among Preschool Children using PlaqueSAM

Jing Hao^{1*}, Kaixin Guo^{2*}, Zhiyi Shan², Yanqi Yang², James Kit-Hon Tsoi¹, Phoebe Pui Ying Lam², Kuo Feng Hung^{1✉}

¹ Division of Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China

² Division of Paediatric Dentistry and Orthodontics, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China

* Equal Contribution

✉ Corresponding author (hungkfg@hku.hk)

Abstract

Dental plaque is the primary etiological factor for common oral diseases such as dental caries and gingivitis, particularly affecting susceptible populations including children. Preventing plaque-induced dental diseases in paediatric patients remains a global challenge as certain plaque regions are often missed during daily cleaning. Early detection and removal of plaque are essential but often difficult to achieve consistently. A routine monitoring tool that visualizes plaque accumulation on primary teeth could allow parents to actively track their children's oral health, enabling timely prevention and intervention for dental caries and periodontal diseases. Inspired by these, we present PlaqueSAM, an automatic plaque detection system using intraoral photographs captured by smartphones. By uploading six guided images assisted with alignment overlays, PlaqueSAM can generate a comprehensive plaque detection report with visual references. The PlaqueSAM was developed using 3,402 prospectively collected intraoral photographs from 567 preschool children across 11 kindergartens in Hong Kong using a three-stage training strategy based on a multi-task learning paradigm. The performance of PlaqueSAM was quantitatively evaluated on 81 children, achieving sensitivity, specificity, and accuracy of 76.58%, 82.47%, and 79.60%, respectively. This system demonstrates superior plaque detection capabilities without requiring disclosing agents or plastic retractors typically used by dental professionals, making it highly practical for home use to maintain oral hygiene and reduce the risk of common oral diseases. By enabling dynamic monitoring of daily plaque accumulation and cleaning, PlaqueSAM has significant potential to advance teledentistry, promoting children's oral health remotely, conveniently, and at lower cost.

Introduction

Dental plaque is a structured biofilm predominantly composed of bacterial microorganisms. Prolonged accumulation of dental plaque can lead to the onset and progression of common oral diseases, for instance, caries¹ and gingivitis². From 1990 to 2019, untreated dental caries remained the most prevalent health condition worldwide among children³. These conditions impair oral function and negatively affect oral health-related quality of life⁴. Additionally, the associated treatment costs impose substantial medical and economic burdens on society³. Effective

management of dental plaque is widely recognized as a critical strategy for reducing the prevalence and progression of oral diseases⁵. However, the adherent and colourless nature of dental plaque hampers its detection and removal, particularly in young children. Clear visualization of dental plaque may facilitate the maintenance of oral hygiene and support the development of positive oral care habits.

While disclosing agents are commonly used in clinical settings to visualize dental plaque, they have major limitations, including unpleasant taste or texture, cost, time-consuming administration, potential lack of cooperation from children, and inherent subjectivity, which can lead to variability in results among clinicians^{6,7}. Fluorescence-based imaging has emerged as an innovative tool for detecting dental plaque without the use of chemical disclosing agents⁸. However, its diagnostic performance varies by anatomical site and biofilm maturity, with reduced sensitivity for early-stage plaque formation⁹. Furthermore, the requirement for specialized equipment and trained operators, as well as high demand of child cooperation and attention span, limit its wide application among the paediatric population. These challenges highlight the need for alternative methods that are more cost-effective and child-friendly in promoting oral hygiene.

Teledentistry refers to the use of commercially available telecommunication technologies to facilitate remote oral health care, including patient education, diagnosis, maintenance, and the establishment of a dental home¹⁰. By enabling consultations in familiar environments, teledentistry not only reduces travel time and associated costs but also enhances comfort and reduces stress for young children and their caregivers¹¹. Dental plaque is distinguishable from clean tooth surfaces in intraoral images, typically presenting as a white or pale-yellow film with a rough texture that contrasts with the translucent, glossy appearance of sound enamel. Plaque tends to accumulate at the gingival margin and interproximal surfaces, providing a distinct visual signature for automated oral hygiene assessment in teledentistry applications. However, previous studies have relied on images captured with professional cameras and plastic retractors in clinical settings rather than with widely accessible smartphones in non-clinical environments. Hence, their results may be overly optimistic compared to real-world scenarios¹²⁻¹⁴.

Recent advances in artificial intelligence (AI), particularly deep learning, have made it possible to automatically detect dental plaque from oral photographs¹⁵⁻²². Shi et al.²³ proposed the SDNet model and constructed the SDPSeg dataset for dental plaque segmentation based on images obtained using chemical disclosing agents. Similarly, S. Li et al.¹³ developed a neural network to detect dental plaque from oral endoscope images without the need for disclosing agents. Furthermore, W. Song et al.²⁴ introduced CenterFormer, which enhanced segmentation performance for dental plaque detection in oral endoscope images. However, these approaches have three significant limitations when translated to teledentistry applications. First, these methods either rely on chemical disclosing agents or specialized cameras and instruments to take intraoral photographs. This reliance restricts their generalizability, as these tools are less accessible compared to widely available smartphones. Second, these methods were designed to process images containing only a single individual tooth, necessitating a labour-intensive and time-consuming preprocessing step where users need to extract multiple single-tooth regions from intraoral photographs. This limitation significantly reduces their practicality. Last, these approaches lack the ability to generate intuitive and user-friendly plaque detection reports. Specifically, they fail to predict tooth numbers in accordance with the FDI notation system and

cannot provide a clear visualization of the spatial distribution of plaque across the entire oral cavity. Currently, there are no such end-to-end system available for users to self-detect dental plaque conditions by submitting intraoral photographs captured by widely accessible smartphones that can automatically predict and generate standard and easy-to-understand plaque detection reports. The implementation of plaque detection based on intraoral photographs captured by smartphones, coupled with the generation of comprehensive plaque detection reports, could significantly address the three aforementioned challenges. This idea offers a more convenient and efficient solution that eliminates the need for travel time and associated costs for plaque detection in children, thereby enabling more effective maintenance of oral hygiene. To implement such an automated system, it is essential to adopt a multi-task learning (MTL) ²⁵⁻²⁷ framework, as the system must simultaneously perform three key functions: tooth localization, tooth numbering prediction, and plaque segmentation. MTL is an advanced paradigm designed to leverage shared information across multiple related tasks, thereby improving the generalization performance of all tasks. Furthermore, to enhance the reliability of the system and ensure transparency in its prediction mechanism, it is critical to design the system to closely mimic the diagnostic process undertaken by dentists when analyzing intraoral photographs for plaque detection and report generation. Specifically, when dentists examine an intraoral photograph, they typically localize each tooth, assign a corresponding tooth number, and then identify the plaque regions on the specific tooth surface. This process is repeated for all teeth in the oral cavity to complete a comprehensive plaque examination. Emulating this step-by-step approach would contribute to the system's reliability and interpretability. Additionally, data privacy is a key consideration. To protect users' sensitive information, including intraoral photographs and plaque detection reports, the system should ideally support local deployment. This would minimize the risk of privacy leakage and ensure that users' personal data is securely managed.

Considering the aforementioned factors, we present PlaqueSAM, a plaque detection system designed for intraoral photographs captured by smartphones, as shown in Figure 1. PlaqueSAM introduces innovations on two fronts. Firstly, it is capable of performing three tasks in an end-to-end manner, including tooth localization, tooth numbering prediction, and plaque segmentation, ultimately generating a plaque detection report with visual reference. Secondly, PlaqueSAM offers alignment overlays to assist users in capturing intraoral photographs, thereby ensuring optimal performance for plaque detection. These overlays predefine six photographic angles that comprehensively display the buccal surfaces of all teeth within the oral cavity. To implement PlaqueSAM, we propose a multi-task learning framework built upon the Segment Anything Model²⁸ (SAM) owing to its outstanding segmentation performance and strong generalization ability. For training and evaluation, we constructed an extensive dataset comprising 3,888 intraoral photographs from 648 children across eleven kindergartens in Hong Kong, with annotations provided by two professional dentists. We have designed a three-step training process to develop PlaqueSAM as shown in Figure 4 (b). First, the view decoder was trained to classify image views in order to identify the Teeth of Interest (ToI). Second, the box decoder was trained to predict tooth localization and numbering. Third, the prompt encoder and mask decoder were fine-tuned to segment both plaque and teeth masks. Each decoder dedicated to a specific task was trained independently, while the image backbone remained fixed. This cascade training strategy ensured system stability and reduced the complexity of model optimization. With PlaqueSAM, users can upload intraoral photographs taken with smartphones for automated plaque detection. The system automatically processes the images, identifies tooth localizations and numbers, segments plaque

regions within the ToI, and generates a plaque detection report with visual references (Figure 1). Furthermore, PlaqueSAM supports local deployment and respects user privacy, making it a compelling tool for personal oral hygiene monitoring. To evaluate its performance, we conducted experiments on 486 intraoral photographs from 81 children cases. The results demonstrated a sensitivity of 71.88%, specificity of 87.63%, and accuracy of 79.44%. Although PlaqueSAM does not match the precision of plaque detection with traditional disclosing agents, its convenient and fully automated workflow allows parents to monitor plaque accumulation in children on a regular basis, thereby reducing dental anxiety among preschoolers and eliminating travel time and associated costs. This capability enhances preventive care against dental caries and contributes to improved oral health outcomes. In summary, PlaqueSAM represents a notable advancement in automatic plaque detection and teledentistry, offering a practical and user-friendly approach to oral health monitoring.

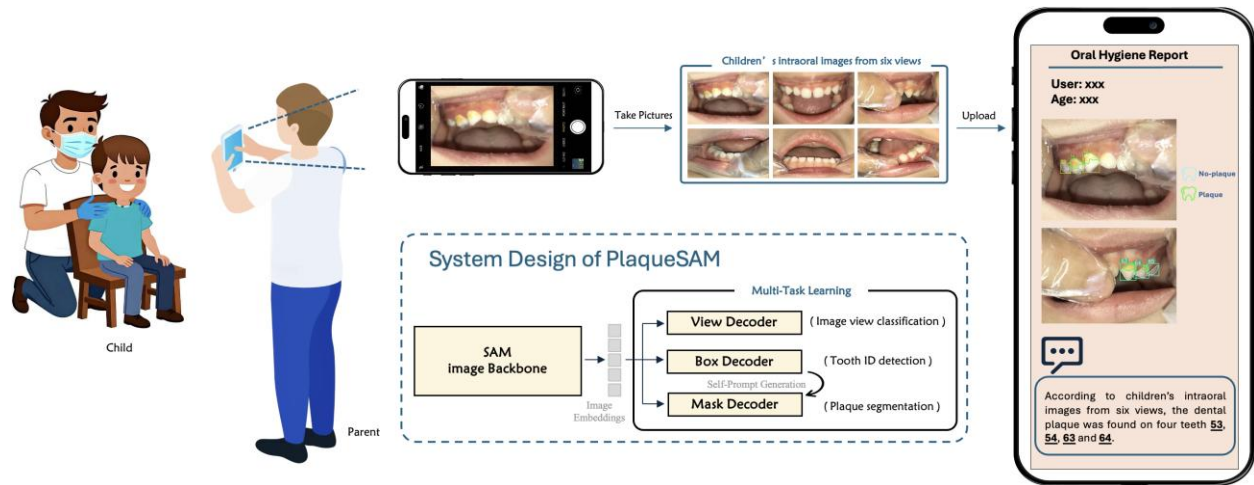


Figure 1 Illustration the process of automatic dental plaque detection. Initially, six intraoral photographs are captured by users utilizing a mobile phone. These images are subsequently uploaded to PlaqueSAM, which automatically generates a comprehensive dental plaque screening report, including both visualizations of the plaque and textual descriptions.

Results

The overall design of PlaqueSAM

PlaqueSAM is an end-to-end system designed to provide a user-friendly tool for screening dental plaque conditions in children, as illustrated in Fig. 1. Six intraoral photographs, capturing the visual appearance of buccal surfaces, are collected by users using a mobile phone. The image capture process is guided by pre-designed alignment overlays corresponding to different views to ensure proper viewing angles. These images are then uploaded to the PlaqueSAM, which could automatically generate a comprehensive dental plaque screening report. The report identifies teeth with dental plaque presence (based on the FDI World Dental Federation notation system) and provides precise visual localization of the dental plaque. Technically, PlaqueSAM processes the input images by first partitioning them into smaller patches and extracting high-dimensional features, which serve as high-level representations for subsequent multi-task predictions. The PlaqueSAM employs a view decoder to predict the orientation of the image view and a box decoder to detect and assign teeth locations and numbers. Additionally, a pre-defined teeth of interest (ToI)

lookup table maps the six distinct intraoral image orientations to their corresponding ToI. Using the ToI lookup table, the predicted view orientation, and teeth detection results, valid bounding boxes of teeth are selected and then used as box prompts. The prompt encoder transforms the self-generated box prompts into feature representations via positional encoding, while the mask decoder integrates the image embeddings and prompt features through cross-attention to predict both the dental plaque mask and the teeth mask for specific teeth identified by their numbering. PlaqueSAM was trained through a three-stage training process using a large-scale, dentist-annotated dataset comprising 3,402 images collected from 567 children. The system could provide a fully automated and user-friendly way to assist users in self-screening dental plaque conditions, promoting oral health management in children.

Two prominent features were prioritized in the design of PlaqueSAM. First, it was essential for the system to be fully automatic and end-to-end, capable of predicting dental plaque distributions across the entire oral cavity. From the perspective of user intent, it is both accurate and reliable in generating the identification of teeth with dental plaque (following the FDI notation system) and the corresponding plaque locations. To achieve this, PlaqueSAM was designed as a unified multi-task prediction framework that integrates image view orientation classification, teeth detection, and plaque mask segmentation tasks into a cohesive system. Moreover, the inference mechanism of PlaqueSAM was designed to mimic the diagnostic process employed by dentists. In real-world dental plaque examination scenarios, dentists first identify the teeth to be examined and assign the corresponding teeth numbers, followed by a focused inspection of the plaque regions within the identifies teeth. Similarly, PlaqueSAM initially predicts the teeth locations and their respective numbers from the input images. It then selects valid ToI bounding boxes based on a predefined ToI lookup table. These bounding box prompts are then passed to the prompt encoder, and the mask decoder predicts the plaque regions within the specified box prompts. Ultimately, the PlaqueSAM generates an identification of the teeth with dental plaque (following the FDI notation system) along with the corresponding plaque locations.

Another critical consideration in the design of PlaqueSAM was ensuring it is a user-friendly system. Unlike professional dentists, users often face challenges in capturing intraoral photographs with ideal viewing angles. Image view orientation is crucial for accurate dental plaque prediction, as even minor changes in orientation can significantly alter the visual characteristics of plaque in the images, thereby compromising plaque prediction performance. To address this challenge, professional dentists designed six tooth alignment overlays (Fig. 4 (e)) to guide users in capturing intraoral photographs at optimal angles. By aligning the teeth of interest (ToI) with the boundary of the corresponding overlays, users can easily capture images with the desired view orientation, reducing the system's operational complexity. In addition, these alignment overlays were encoded into the PlaqueSAM framework to enhance the performance of teeth detection since they could provide prior knowledge of tooth locations. Furthermore, the tools selected for data collection were chosen based on their clinical applicability. Specifically, an iPhone 7 (f/1.8, 28 mm (wide), OIS, flashlight) was utilized instead of a professional camera to ensure that the captured images closely resembled those typically obtained in clinical settings, maintaining comparable pixel quality. Additionally, hands wearing plastic gloves were used instead of a retractor to expose the targeted area for photography, which can better simulate real-world conditions and enhance the practical relevance of the study. These methodological choices were designed to improve the applicability of the PlaqueSAM in daily clinical use.

Quantitative and qualitative analysis

We evaluated PlaqueSAM on the test set consisting of 81 children with a total of 486 intraoral photographs, and the performance is illustrated in Table 1. Overall, the sensitivity, specificity, and accuracy of PlaqueSAM for dental plaque detection were 76.58% (95% C.I. 74.36%–78.68%), 82.47% (95% C.I. 80.44%–84.49%), and 79.60% (95% C.I. 77.43%–80.42%), respectively. PlaqueSAM incorporates three prediction tasks, including image view classification, tooth detection, and dental plaque segmentation, and we performed a detailed evaluation for each task. For image view classification, PlaqueSAM achieved an accuracy of 98.60%, indicating its powerful ability to classify the imaging orientation and thereby provide accurate extraction of the tooth of interest. For tooth detection task, the mean Average Precision (mAP)²⁹ and AP50²⁹ were 75.04% and 93.47%, respectively. At a confidence threshold of 0.5, the precision, recall, and F1-score were 98.87%, 97.22%, and 98.04%, demonstrating excellent performance in tooth identification and localization. For dental plaque segmentation, the mean Intersection over Union (mIoU) was 43.99% (95% C.I. 41.57%–46.42%). We observed that the boundaries between dental plaque regions and normal tooth surfaces were often ambiguous. As a result, the mIoU for plaque segmentation, which serves as a fine-grained indicator of plaque prediction, was lower compared with the other two tasks. However, PlaqueSAM still demonstrated high performance in detecting teeth with dental plaque, achieving a sensitivity of 76.58%.

Table 1 The quantitative metrics of three tasks of PlaqueSAM: image view classification, teeth detection and localization, and dental plaque detection.

Tasks	Metrics (w alignment overlay encoding)				Metrics (w/o alignment overlay encoding)			
1. Image view classification	Accuracy				Accuracy			
	98.60%				98.60%			
2. Teeth detection and localization	AP50	Precision	Recall	F1	AP50	Precision	Recall	F1
	93.47%	98.87%	97.22%	98.04%	89.59%	97.76%	97.04%	97.40%
3. Dental plaque detection	mIoU	Sensitivity	Specificity	Accuracy	mIoU	Sensitivity	Specificity	Accuracy
	43.99%	76.58%	82.47%	79.60%	43.70%	76.10%	82.47%	79.39%

Furthermore, four metrics for dental plaque detection are detailed according to each tooth number, as illustrated in Figure 2. Overall, the sensitivity metric is predominantly above 65%, except for the four types of teeth: 71, 72, 81, and 82. We suspect that light conditions and salivary accumulation may affect plaque recognition. The specificity and accuracy metrics for each tooth are approximately around 80%, demonstrating the superior performance of PlaqueSAM. We also conducted qualitative visualizations, as shown in Figure 3. Additionally, we investigated the effectiveness of alignment overlays for tooth localization and plaque detection tasks, as they can provide prior knowledge of tooth locations. Initially, the alignment overlays are encoded into high-dimensional features using a lightweight CNN feature extractor³⁰, which are then integrated into the box decoder through the cross-attention operation. The experimental results demonstrate that the utilization of alignment overlays improves the AP50 metric for the tooth detection task by 3.88% (from 89.59% to 93.47%). Additionally, more accurate tooth localization further enhances the sensitivity of dental plaque detection by 0.48%. This indicates that incorporating alignment overlays can significantly improve the performance of the PlaqueSAM model in both tooth localization and plaque detection. Detailed metrics can be found in Table 1.

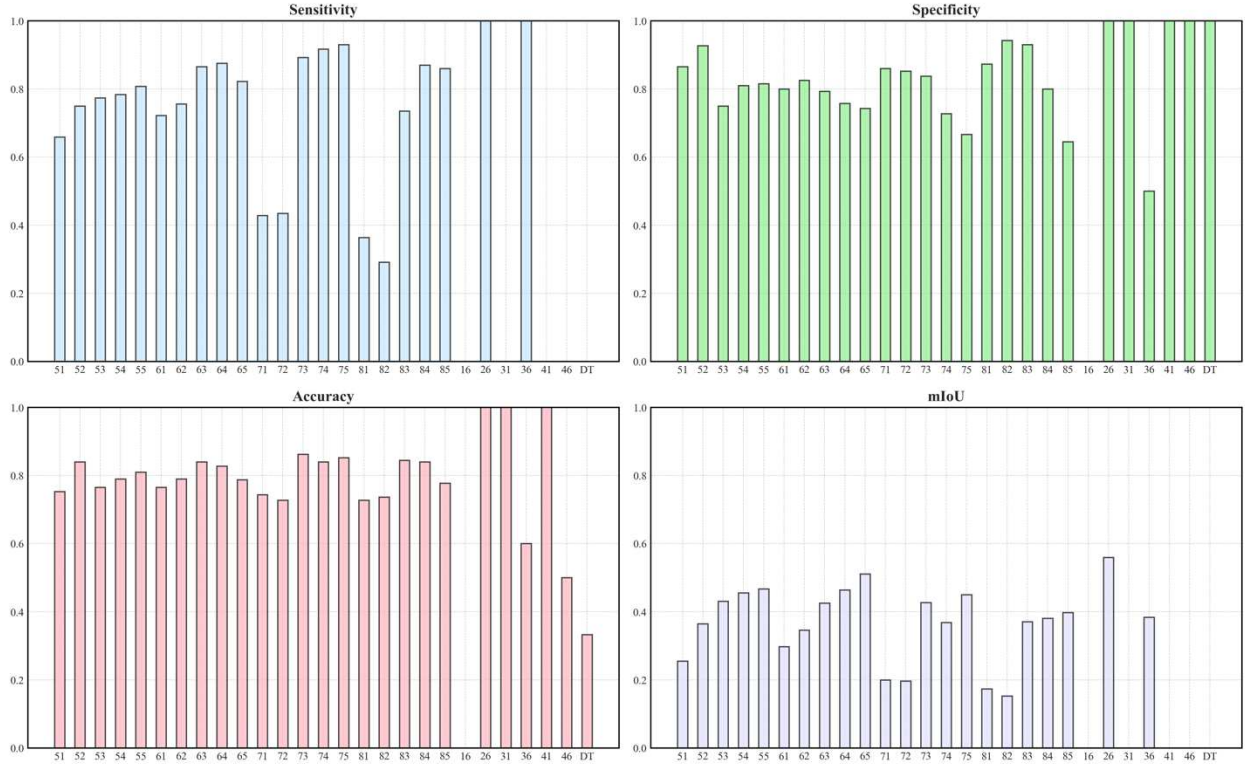


Figure 2 Four metrics for dental plaque detection corresponding to each tooth number.

Ground truth establishment and comparison of clinical examination

To ensure the accuracy and reliability of dental plaque annotations, a calibration session was conducted to assess the agreement between annotators and clinical examination. A random subset comprising 10% of the subject cohort (65 participants) was selected for calibration. An experienced dentist, blinded to the clinical examination results, annotated 390 intraoral photographs for the presence of dental plaque. Agreement between the image-based annotations and clinical examination was quantified using Cohen’s kappa value, yielding $\kappa = 0.802$ ($p < 0.05$), which indicates very good agreement. Based on this result, the dentist’s annotations were adopted as the ground truth for subsequent analyses.

Practical Advantages and Efficiency of PlaqueSAM in Dental Plaque Detection

Compared with clinical examinations, PlaqueSAM-based dental plaque detection offers clear practical advantages (Table 2). Clinical examinations have stricter requirements: they need a space where participants can lie down; an experienced dentist with an assistant, and professional tools such as a disposable mouth mirror and appropriate lighting. In contrast, PlaqueSAM can be easily operated by users without any dental qualifications. It can be conducted in a wide range of settings and requires only a smartphone and disposable gloves. These differences translate into lower personnel and equipment demands, and thus lower overall costs. In our study, the average time to capture a set of intraoral photographs was 94 seconds per child, and PlaqueSAM’s processing time was 0.463 s per child, yielding a total of under 95 seconds. By comparison, a clinical examination

required approximately 300 seconds. Consequently, PlaqueSAM is more time- and cost-efficient and shows strong potential for real-world implementation, particularly in non-clinical settings.

Table 2 The comparison between clinical examination and PlaqueSAM in real world.

Aspect	Clinical examination	PlaqueSAM-based detection
Time per child	About 300s	Photo acquisition: 94 s; processing: 0.463 s; total: less than 95 s
Personnel	An experienced dentist plus an assistant	Two persons without dental qualifications
Equipment	Disposable mouth mirror, appropriate lighting, and other clinical tools	Smartphone and disposable gloves
Space	Clinical area, participant can lie down	No fixed clinical layout required
Cost	High	Low

Advantages of PlaqueSAM over powerful AI Models for Plaque Assessment and Report Generation

To compare the performance of PlaqueSAM and GPT-5, a widely used powerful large language model, in dental plaque assessment and report generation, we randomly selected three children with different levels of oral hygiene according to the clinical examination. For each child, six intraoral images were obtained and submitted to GPT-5 with the following standardized prompt: *“Please generate a dental plaque report based on the six intraoral images provided. These images are of the same child, approximately 6–10 years old. My goal is to perform dental plaque screening for preschool children. Identify and detail the locations of dental plaque, specifying the affected teeth using FDI numbering. Provide a comprehensive description of plaque distribution.”* The reports generated by PlaqueSAM and GPT-5 are provided in the Supplementary Materials. The report generated by PlaqueSAM includes three components, including overall oral hygiene status, visualized plaque distribution, and personalized oral hygiene recommendations. In contrast, the report produced by GPT-5 is largely qualitative and text-based, lacking a systematic indication of which specific teeth require attention.

Discussion

In this paper, we propose PlaqueSAM, a plaque detection system specifically designed for intraoral photographs captured using smartphones. PlaqueSAM employs a multi-task learning framework built upon the SAM and incorporates three tasks: image view classification, tooth detection, and plaque segmentation. The system processes six predefined-angle intraoral photographs from a single child and ultimately generates a plaque detection report accompanied by referenced visualizations. We conducted training and evaluation on a high-quality dataset annotated by professional dentists, comprising 3,888 images collected from 648 children across eleven kindergartens in Hong Kong. The dataset was randomly divided into a training set and a test set, with the training set consisting of 567 child cases and the test set containing 81 cases.

To enhance the practicality and clinical applicability of the PlaqueSAM, data collection was conducted in a manner that closely simulated the typical use of PlaqueSAM in daily dental practice.

Specifically, an iPhone 7 was utilized instead of a professional camera to ensure that the captured images closely resembled those typically obtained in non-clinical settings, while maintaining comparable pixel quality. Additionally, hands wearing plastic gloves were used, rather than a dental retractor, to expose the target area for photography, which can better simulate real-world conditions and enhance the practical relevance of the study. These methodological choices were designed to improve the applicability of the PlaqueSAM in daily clinical use.

To better facilitate intraoral photography for users without professional training in image acquisition angles, six predefined alignment overlays were designed by experienced dentists to assist in identifying the optimal shooting angle. By aligning the teeth of interest (ToI) with the boundaries of the corresponding overlay, users can more easily capture images with the desired orientation, thereby reducing the system’s operational complexity. In addition, these alignment overlays can serve as prior information regarding tooth locations from the algorithm perspective, thereby enhancing the tooth detection performance of PlaqueSAM. An ablation study was conducted to evaluate the impact of incorporating the alignment overlays into the box decoder of PlaqueSAM. As shown in Table 1, encoding the overlays into PlaqueSAM improved the mean Average Precision (mAP) and AP50 for the tooth detection task by 2.0% and 3.88%, respectively, demonstrating that the incorporation of alignment overlays enables more precise localization of tooth. Furthermore, accurate tooth localization allows the prompt-sensitive PlaqueSAM to more precisely delineate plaque regions, resulting in an improvement in plaque detection sensitivity and accuracy by 0.48% and 0.21%, respectively. Thus, the alignment overlays enhance both the applicability and performance of PlaqueSAM, further strengthening its utility.

Table 3 Comparison of various approaches for dental plaque detection. Our PlaqueSAM method does not require disclosing agents and achieves superior performance (accuracy, sensitivity, and specificity) through an advanced multi-task learning framework.

Authors	Image modality	Disclosing agents	Model	# Training set	# Test set	Accuracy	Sensitivity	Specificity	Report Generation
Jeong et al. ³¹	Intraoral image	Yes	U-Net	875	300	63.14%	-	-	No
Pedraza et al. ³²	Photography	Yes	YOLOv11	431	100	-	65.40%	-	No
Chen et al. ³³	Intraoral image	Yes	YOLOv8	1120	280	77.60%	-	-	No
Liu et al. ³⁴	Photography	Yes	Decision tree	-	-	67.10%	68.40%	62.9%	No
Ours	Photography	No	PlaqueSAM	3,402	486	79.60%	76.58%	82.47%	Yes

PlaqueSAM demonstrates superior performance in dental plaque detection compared with multiple prior studies, as summarized in Table 3. Our method shows advantages across five key dimensions: input modality, algorithmic design, training and testing data scale, evaluation performance, and the feature of report generation. First, PlaqueSAM uses smartphone-acquired photographs rather than intraoral images, which require specialized cameras and equipment, and it does not rely on chemical disclosing agents. Second, PlaqueSAM adopts a multi-task learning framework that

enables end-to-end prediction of dental plaque distribution across the entire oral cavity. In terms of dataset scale, the number of samples included in this study substantially exceeds those used in previous experiments, underscoring the reliability of our results. For evaluation performance, PlaqueSAM achieves a sensitivity of 76.58%, specificity of 82.47%, and accuracy of 79.60%, markedly outperforming earlier dental plaque detection methods³¹⁻³⁴. Finally, PlaqueSAM can generate a comprehensive plaque detection report for the entire oral cavity (in Supplementary materials), which is a unique capability not found in existing approaches. This feature enhances the system’s applicability and user interaction in daily clinical practice.



Figure 3 Qualitative visualizations of four types of dental plaque prediction: True Positives (green edges), True Negatives (blue edges), False Positives (yellow edges), and False Negatives (red edges).

PlaqueSAM also shows superior advantages over powerful AI models for plaque assessment and oral hygiene report generation. Specifically, the PlaqueSAM report comprises three components. First, overall oral hygiene status is summarized using the O’Leary Plaque Control Record³⁵. For each child, all buccal surfaces are evaluated for the presence of plaque, and the Plaque Control Record is calculated as: $\text{Plaque Control Record (\%)} = (\text{number of surfaces with plaque} / \text{total number of surfaces examined}) \times 100$. Oral hygiene is then classified as good ($\leq 20\%$), fair (21–40%), or poor ($> 40\%$). The second component provides detailed information on plaque accumulation, combining image-based visualization with textual descriptions. The third component delivers individualized oral hygiene recommendations tailored to the child’s specific plaque distribution. For comparison, the report from GPT-5 had the same three-part structure.

However, its outputs were largely qualitative and text-based, without systematic indication of which specific teeth required attention, which may confuse the users. In contrast, PlaqueSAM is explicitly designed for image-based plaque analysis. It automatically segments plaque on intraoral images and generates reports that integrate visual overlays with precise tooth-level localization using standardized positional references, thereby producing an intuitive “plaque map” for each child. This spatially explicit feedback enables users to identify exactly which areas require improved cleaning and facilitates consistent longitudinal monitoring. These visual and clearly labelled outputs are particularly advantageous for non-dental users, as they translate complex clinical findings into concrete, actionable instructions for daily oral care. Furthermore, PlaqueSAM is specifically optimized for paediatric dentitions and demonstrates higher accuracy in tooth localization than GPT-5, which occasionally misidentified primary teeth as permanent teeth.

This study had some limitations. First, although the intraoral photographs were obtained from 11 kindergartens in both urban and suburban areas of Hong Kong (comprising approximately 90% Southern Chinese preschool children and around 10% ethnic minority groups), the sample represents an Asian population and thus might lack broader racial diversity. The generalizability of our model to more diverse populations should be validated in future prospective studies. Additionally, the precision of plaque detection by PlaqueSAM was comparatively inferior to the conventional visual-tactile examination with probes or chemical disclosing agent methods performed by dental professionals in dental offices. However, as our approach does not require chemical disclosing agents, it is more practical for implementation of dental plaque detection and monitoring in a more cost-effective method and wider populations in non-clinical settings. While there is room for improvement in PlaqueSAM’s accuracy, our model remains more feasible for use at the non-clinical settings compared to previously developed models. The proposed PlaqueSAM demonstrates significant translational potential for the development of mobile software as a tele-dentistry solution for at-home screening, with the aim of identifying individuals who may need to take self-care measures or seek professional dental care for plaque removal. Rather than replacing dental professionals, PlaqueSAM is intended to promote global public oral health, particularly among vulnerable populations who require greater healthcare attention. Future studies are needed to deploy and develop a corresponding mobile application to facilitate real-world testing.

As with any AI system, there is a risk of misinformation or misdiagnosis. This study aimed to develop a practical tool capable of generating comprehensive dental plaque screening reports with detailed visual and textual analyses to advance tele-dentistry and improve children's oral health. Notably, it should be emphasized that PlaqueSAM is intended solely as a screening tool for at-home dental plaque detection, helping to identify individuals who may need to take self-care measures or seek professional dental care for plaque removal, but not as a replacement for dental professionals.

Method

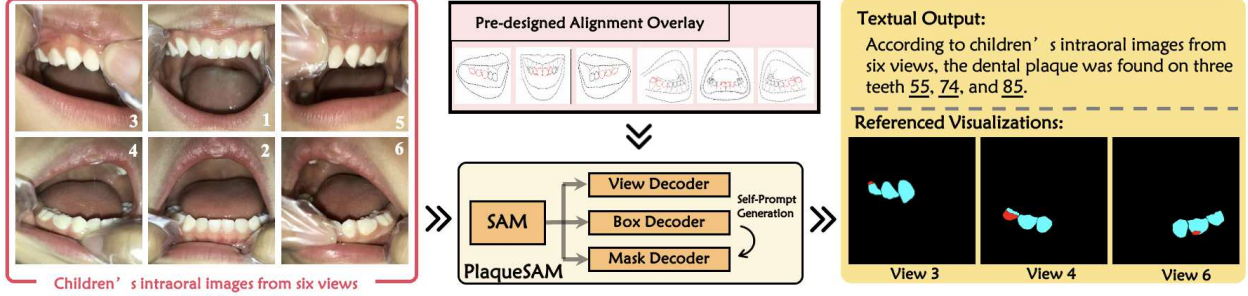
Ethics

The study was conducted in accordance with the Declaration of Helsinki, aiming to explore the possibility of the teledentistry of dental plaque detection. Ethical approval was approved by the Institutional Review Board (protocol number UW24-179). The parents or legal guardians of preschool children had been informed of all relevant details and the voluntary nature of the research study, including its purpose, procedures, anticipated risks and benefits. And written informed consents were obtained from the parents of all participants prior to beginning the study.

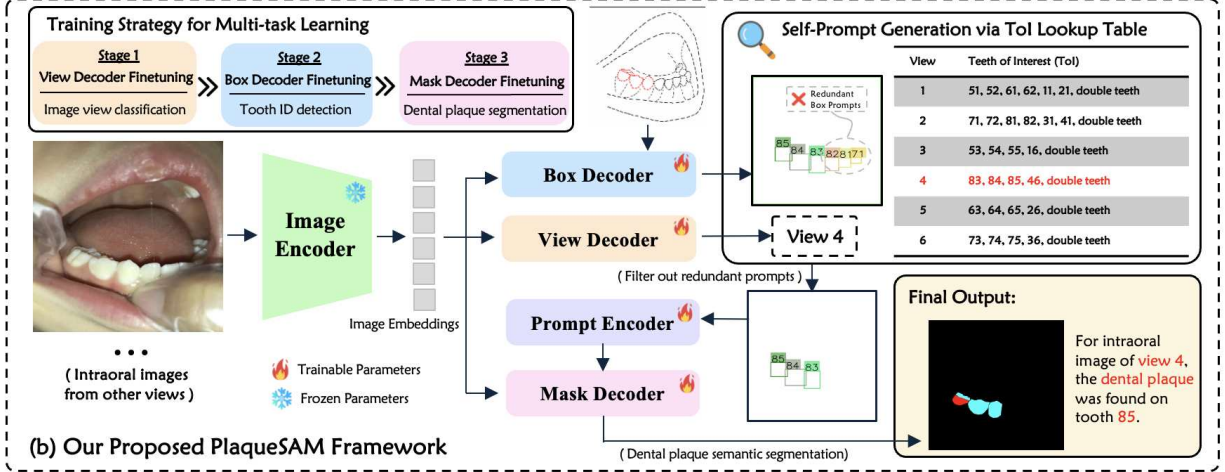
Data acquisition

The dataset collection was conducted across eleven kindergartens in Hong Kong from September 2024 to December 2024. To ensure the validity and reliability of the experiment, inclusion criteria of the children enrolled were established as follows: 1. studied kindergarten grade 1-3 ; 2. had primary or mixed dentition; 3. cooperative behaviour conducive to photographic procedures. Children with fewer than 12 primary teeth, or with serious systemic or behavioural conditions were excluded from the study. Two research assistants were recruited to take intraoral photographs. Both assistants underwent a minimum of one hour of hands-on training prior to the beginning of study. For each participant, a standardized set of six intraoral photographs was obtained, including frontal views of upper and lower teeth apart, quadrant upper right, upper left, lower right, lower left was taken per child. The buccal surfaces of incisors, canines, premolars and molars were photographed using a smartphone (iPhone 7, Apple Inc, Cupertino, CA, f/1.8, 28 mm (wide), OIS), with only room lighting and the built-in flash of the smartphone camera being used. Post-capture, each photograph was reviewed immediately to assess quality. Images that were blurry or out of focus were discarded, and the corresponding photographs were retaken to ensure clarity.

A comprehensive annotation protocol was established to generate ground truth labels for the intraoral image dataset before the formal annotation. Two professional dentists independently annotated each image, providing labels that include the view orientation, the bounding box and category of each tooth, as well as concise masks for both the teeth and dental plaque. The annotations include: (i) the view orientation was determined by anatomical landmarks, including the upper and lower lips and the left and right conner of the mouth. (ii) Tooth identification and categorization adhered to the Fédération Dentaire Internationale (FDI) numbering system. (iii) Segmentation masks for both teeth and dental plaque were based on a standardized set of visual criteria. First, the colour and reflectivity of tooth surfaces were key distinguishing features. Dental plaque appears as a white or pale-yellow film that causes a distinct loss of brightness and specular reflection, contrasting with the translucent, bright surface of a clean tooth which exhibits sharp, white highlights under the flashlight. Second, the surface texture differs significantly, with plaque-covered area demonstrating rough or granular appearance against the smooth and glossy texture of sound enamel. Additionally, the annotations prioritized regions with a high predilection for plaque accumulation, specifically the gingival margin and approximal surfaces. The comprehensive protocol ensures the consistency and reliability of annotations.



(a) The pipeline of dental plaque detection for one child



(b) Our Proposed PlaqueSAM Framework

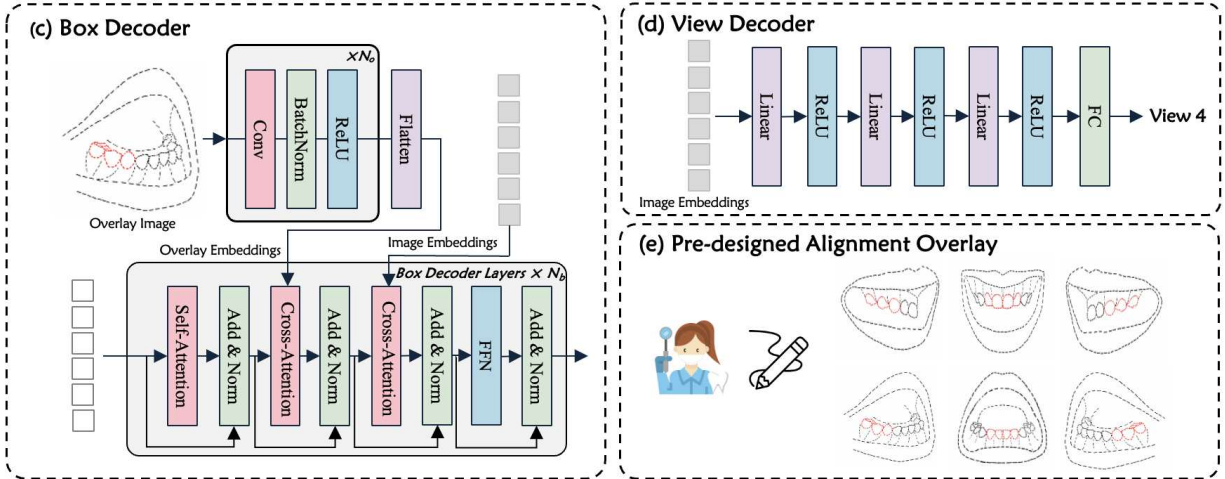


Figure 4 Overview of the proposed PlaqueSAM model. (a) The pipeline of dental plaque detection for an individual child. PlaqueSAM takes six intraoral images together with pre-designed alignment overlays as inputs and produces a dental plaque report with visualized references. (b) The framework of PlaqueSAM, which comprises an image encoder, box decoder, view decoder, prompt encoder, and mask decoder, trained using a three-stage multi-task learning strategy. (c) Architecture of the box decoder. (d) Architecture of the view decoder. (e) The six alignment overlays designed by experienced dentists.

Through this data construction process, a dataset comprising 648 children with a total of 3,888 intraoral photographs was constructed, of which 567 were utilized as the training dataset, and 81 as the test set (Fig. 5 (b)). The dataset comprises a total of 22,482 teeth, of which 10,836 exhibit

dental plaque (Fig. 5 (d)). The gender distribution of the children was approximately balanced (Fig. 5 (a)). The majority of children were within the age ranges of 50 months and 65–80 months (Fig. 5 (c)). Regarding the distribution of teeth with dental plaque, a consistent trend was observed whereby teeth located deeper within the oral cavity—specifically teeth numbered 55, 65, 75, and 85—were more prone to plaque accumulation.

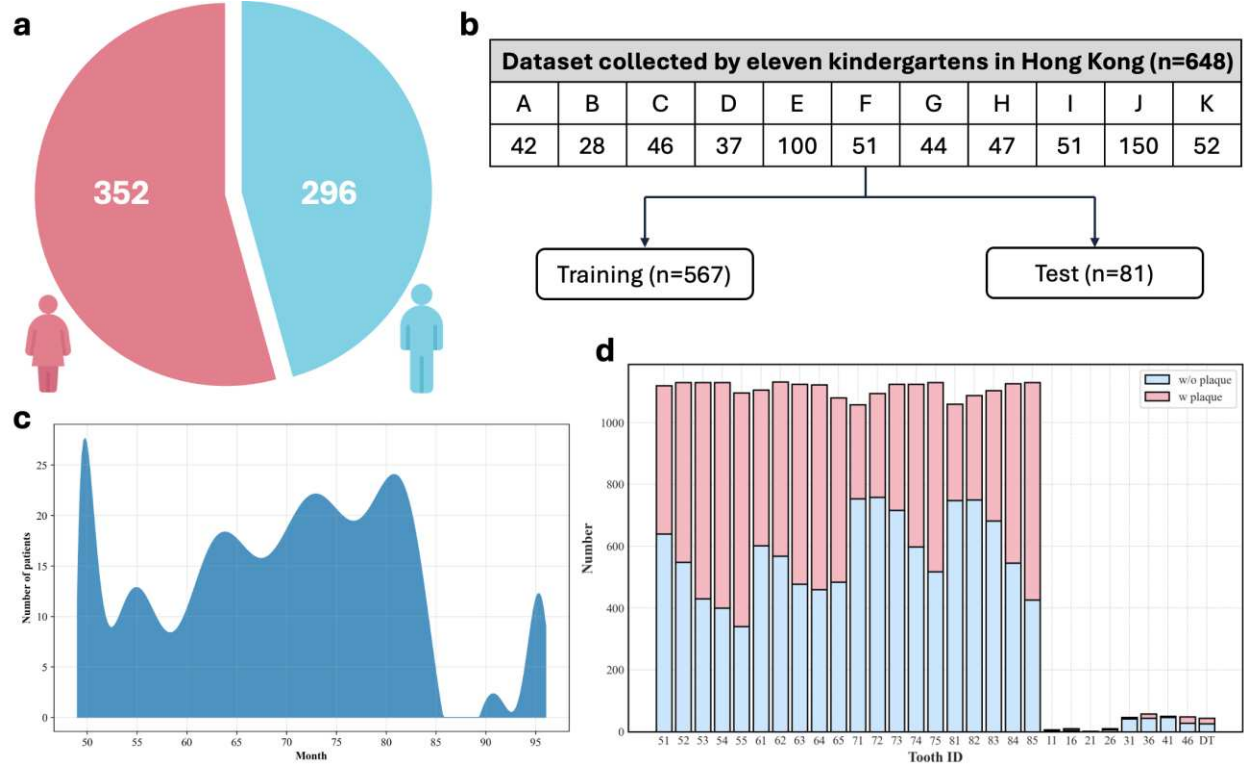


Figure 5 Dataset composition, children characteristics, and dental plaque distributions. a. Gender distribution of children. **b.** Dataset collected from eleven kindergartens in Hong Kong. Of these, 567 and 81 kids were used for training and test sets, respectively. **c.** Age distribution of collected samples. **d.** Distribution of dental plaque corresponding to tooth numbering.

Overview

The proposed model, PlaqueSAM, was designed to predict a child’s dental plaque conditions and provide corresponding segmentation visualizations based on six intraoral photographs captured from different predefined angles. These images were taken using six pre-designed alignment overlays, which ensure that the entire set of teeth within the child’s oral cavity is comprehensively represented. The PlaqueSAM model was built upon the architecture of the Segment Anything Model (SAM), a state-of-the-art interactive segmentation framework that consists of three primary components: an image encoder, a prompt encoder, and a mask decoder. To adapt this framework for dental plaque screening task, we introduced two additional modules: a view decoder for image view orientation classification and a box decoder for tooth bounding box detection and numbering. The workflow of the PlaqueSAM model involves several key steps. First, the image encoder extracts embeddings from the six intraoral photographs, which serve as the foundational feature representation for subsequent prediction tasks. Next, the view decoder predicts the specific view orientation of each image, while the box decoder identifies the bounding boxes for individual teeth

and assigns corresponding tooth IDs. Importantly, for each specific view, only certain predefined teeth are considered relevant for analysis based on the clinical imaging protocol. To ensure that the model focuses solely on these relevant teeth, we employ a predefined Teeth of Interest (ToI) table, which specifies the expected teeth of interest for each image view. Using the outputs of the view decoder and box decoder in conjunction with the ToI table, bounding box prompts that do not correspond to the teeth of interest for a specific view are filtered out. The valid bounding box prompts are then passed to the prompt encoder, and the mask decoder generates both tooth masks and dental plaque masks within the specified bounding box prompts. By iterating through all six images, the PlaqueSAM model predicts the dental plaque mask and assigns a tooth ID for each segment, ensuring that the model captures the full extent of dental plaque distribution across the child’s entire oral cavity. This process culminates in the generation of a comprehensive dental plaque screening report, which includes a detailed visual and numerical analysis of the child’s dental health. Through this approach, the PlaqueSAM model integrates advanced segmentation techniques with domain-specific adaptations, enabling accurate and efficient dental plaque assessment.

By iterating through all six images, the PlaqueSAM model predicts the dental plaque mask and assigns a tooth ID for each segment, ensuring that the model captures the full extent of dental plaque distribution across the child’s entire oral cavity. This process culminates in the generation of a comprehensive dental plaque screening report, which includes a detailed visual and textual analysis of the child’s dental health. Through this approach, the PlaqueSAM model integrates advanced segmentation techniques with domain-specific adaptations, enabling accurate and end-to-end dental plaque assessment.

Image View Orientation Classification

To assess the condition of dental plaque across the child’s entire oral cavity, six intraoral photographs were captured from predefined orientations according to the clinical imaging protocol. These images comprehensively display the visual appearance of all teeth, serving as the foundation for subsequent dental plaque prediction tasks. The clinical imaging protocol specifies a mapping between particular image view orientations and the corresponding teeth of interest (ToI), as illustrated in the ToI lookup table (Fig. 2(b)). Consequently, accurately predicting the image view orientation is essential for determining the teeth of interest associated with the current image. The view decoder was utilized for the image view orientation classification. It decoded the input image embeddings through a multi-layer perceptron³⁶ (MLP) followed by a fully connected layer to output the predicted view orientation. The architecture of the view decoder is depicted in Fig. 2(d).

Overlay-based Teeth Detection

The box decoder was employed to detect the bounding boxes and corresponding numbering of teeth within the image. These bounding boxes serve as prompts for the subsequent mask decoder, thereby the accuracy of teeth detection directly determines the performance of subsequent dental plaque segmentation. To further enhance the performance of teeth detection, we propose an overlay-based box decoder module that incorporates the manually designed overlay, which provides prior information of teeth positions, into the teeth box decoding process.

The detailed architecture of the box decoder is presented in Fig. 2(c). Inspired by the DINO^{37,38}, our box decoder is composed of N_b box decoder layers, with each layer following a structured sequence of self-attention³⁹, cross-attention³⁹, and feed-forward network³⁹ (FFN) operations to iteratively refine the features of object queries $Q \in R^{100 \times 256}$. Additionally, overlay embeddings F_o are introduced via cross-attention after the self-attention operation. This mechanism integrates prior knowledge about teeth locations and the overall structural outline of the image into the object queries, thereby enhancing their feature representation and improving the performance of teeth detection. The overlay embeddings are derived by extracting high-dimensional features from the overlay image $I_{overlay}$ through a lightweight convolutional neural network, followed by a flattening operation. The process of each box decoder layer can be formulated as:

$$\begin{aligned}
F_o &= \text{Flatten}(\text{ReLU}(\text{BatchNorm}(\text{Conv}(I_{overlay})))) \times N_o, \\
Q &= \{q_1, q_2, \dots, q_n\}, \\
H' &= \text{Norm}(\text{Self_Attention}(Q) + Q), \\
H'' &= \text{Norm}(\text{Cross_Attention}(H', F_o) + H'), \\
H''' &= \text{Norm}(\text{Cross_Attention}(H', F_l) + H''), \\
H^o &= \text{Norm}(\text{FFN}(H''') + H'''),
\end{aligned}$$

where $\text{Conv}()$ refers to the convolutional operation, the N_o indicates the number of operation layers, F_l means the image embeddings from the image encoder, and $H^o \in R^{100 \times 256}$ is the output of each box decoder layer.

Self-Prompt Dental Plaque Segmentation

SAM²⁸ is an interactive segmentation model that supports flexible prompts (e.g., points, boxes, and text) to guide the segmentation of target objects. However, SAM has two key limitations when performing the dental plaque segmentation. First, SAM is limited to class-agnostic segmentation, which prevents it from accurately delineating dental plaque regions on tooth surfaces. These regions are characterized by semantic ambiguity, with no clear boundary between the tooth area and the dental plaque area. To address this issue, we fine-tuned the prompt encoder and mask decoder using the curated large-scale dental plaque semantic segmentation dataset to enable SAM to perform class-aware semantic segmentation in a supervised manner. This allows the model to distinguish between the semantic characteristics of teeth and dental plaque and to make corresponding predictions regarding their categories and spatial locations.

Second, SAM relies on user-provided prompts to segment regions of interest, which limits its ability to perform end-to-end automated assessments of dental plaque distribution across a child's entire oral cavity. To achieve fully automated dental plaque assessment, we designed a prompt self-generation strategy. This approach leverages predictions from the view decoder and box decoder, combined with a predefined ToI table, to generate valid box prompts. These prompts are then fed into the prompt encoder, simulating the manual input of box prompts by users. As a result, this strategy enables fully automated predictions of dental plaque conditions.

Training strategy and protocol

The PlaqueSAM model adopted a three-stage training approach. Initially, the view decoder was fine-tuned to classify image view orientations, and the cross-entropy loss was used for supervision. Subsequently, the box decoder was fine-tuned to detect the bounding boxes and the numbering of teeth, where the loss function is formulated as a linear combination of focal loss⁴⁰, L1 loss, and GIoU loss⁴¹. Finally, the prompt encoder and mask decoder were fine-tuned to perform semantic segmentation of teeth and dental plaque in a self-prompt manner, utilizing both focal loss and dice loss for supervision. It was noteworthy that the ViT-Tiny⁴² served as the image encoder and was initialized with the pre-trained weights of SAM's backbone, as SAM had been trained on a large-scale corpus and its backbone had demonstrated exceptional generalization capability in image feature extraction. Throughout the three-stage training process, the image encoder remained frozen.

All intraoral photographs were resized to $512 \times 512 \times 3$ during both training and inference phase, with no data augmentation applied. The network was optimized by AdamW⁴³ optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of $1e-4$ and a weight decay of 0.01 . The global batch size was 6 and the model was trained on 1 NVIDIA A100 (80G) GPU with 100 epochs for each stage. The number of trainable parameters at each stage was 527,878, 15,352,709, and 22,053,584, respectively.

Evaluation metrics

We employed multiple metrics to evaluate performance across different tasks. For the image view classification task, we adopted the accuracy metric, which is defined as:

$$Accuracy = \frac{N_{correct}}{N_{images}}$$

Where $N_{correct}$ denotes the number of correctly classified images and N_{images} denotes the total number of images in the evaluation set.

For the tooth detection task, we used the mean Average Precision (mAP) and AP50 metrics, which are widely used in the field of object detection. For an Intersection over Union (IoU) threshold τ , the Average Precision measures the area under the precision–recall (P–R) curve which reflecting the trade-off between precision and recall:

$$AP(\tau) = \int_0^1 p(r, \tau) dr$$

where $p(r, \tau)$ is the precision as a function of recall r at IoU τ .

The mAP is the mean of $AP(\tau)$ over multiple thresholds $\tau \in \{0.50, 0.55, \dots, 0.95\}$:

$$mAP = \frac{1}{T} \sum_{t=1}^T AP(\tau_t)$$

The AP50 is a special case with $\tau = 0.50$, following the PASCAL VOC criterion²⁹. While AP50 reflects coarse detection performance, the mAP metric offers a stricter and more comprehensive evaluation by assessing detection accuracy across multiple localization tolerances.

For the dental plaque detection task, we adopted four evaluation metrics: mean Intersection over Union (mIoU), sensitivity, specificity, and accuracy. The mIoU metric provides a comprehensive measure of the spatial overlap between the predicted segmentation and the ground truth, which is

particularly important for assessing the quality of segmentation masks. Sensitivity reflects the model's ability to correctly identify teeth with plaque, whereas specificity quantifies the capability to correctly recognize non-plaque teeth. Accuracy offers an overall measure of the proportion of teeth that are correctly classified as either with plaque or without plaque. Together, these metrics enable a thorough and balanced evaluation of the model's performance on plaque detection.

Code availability

The code and pretrained weight used in this research is available in the GitHub repository (<https://github.com/isbrycee/PlaqueSAM>), which is publicly accessible to anyone. Information regarding usage, modification, and distribution of the code is specified in the LICENSE file within the repository. The code is intended for research purposes only and may be restricted to commercial use. Additionally, users must cite the related manuscript and code repository when utilizing this work in their research.

Data availability

The dataset collection was conducted across eleven kindergartens in Hong Kong from September 2024 to December 2024. The data can be obtained through a reasonable request to the corresponding author. Due to privacy and ethical restrictions, this data is limited to research use only.

Competing Interests Statement

The authors declare no competing interests.

Author contributions

Conceptualisation: K.F.H., P.P.Y.L., and J.K.H.T.; Methodology: J.H., K.X.G.; Data curation and investigation: J.H., K.X.G.; Writing—original draft preparation, J.H., K.X.G.; writing—review and editing, Z.Y.S, Y.Q.Y, J.K.H.T, K.F.H., P.P.Y.L. All authors have read and approved the manuscript.

Acknowledgements

This study was supported by the Seed Fund for Basic Research for New Staff at the University of Hong Kong (103036002).

Reference

- 1 Marsh, P. D. Microbiology of dental plaque biofilms and their role in oral health and caries. *Dent Clin North Am* **54**, 441–454 (2010). <https://doi.org/10.1016/j.cden.2010.03.002>
- 2 Murakami, S., Mealey, B. L., Mariotti, A. & Chapple, I. L. C. Dental plaque-induced gingival conditions. *J Periodontol* **89 Suppl 1**, S17–s27 (2018). <https://doi.org/10.1002/jper.17-0095>
- 3 Wen, P. Y. F., Chen, M. X., Zhong, Y. J., Dong, Q. Q. & Wong, H. M. Global Burden and Inequality of Dental Caries, 1990 to 2019. *Journal of Dental Research* **101**, 392–399 (2022). <https://doi.org/10.1177/00220345211056247>
- 4 Martins, L. G. T. *et al.* Impact of dental caries on quality of life of school children. *Pesquisa Brasileira em Odontopediatria e Clínica Integrada* **16** (2016).
- 5 Gurenlian, J. R. The role of dental plaque biofilm in oral health. *American Dental Hygienists' Association* **81**, 116–116 (2007).

- 6 Fasoulas, A. *et al.* Detection of dental plaque with disclosing agents in the context of preventive oral hygiene training programs. *Heliyon* **5**, e02064 (2019). <https://doi.org/10.1016/j.heliyon.2019.e02064>
- 7 Harris, N., Garcia-Godoy, F. & Nathe, C. N. *Primary Preventive Dentistry*. (Pearson Education, 2013).
- 8 Liu, Z., Gomez, J., Khan, S., Peru, D. & Ellwood, R. Red fluorescence imaging for dental plaque detection and quantification: pilot study. *J Biomed Opt* **22**, 1–10 (2017). <https://doi.org/10.1117/1.Jbo.22.9.096008>
- 9 Ganss, C., Glanz, A., Glanz, T., Schlueter, N. & Ruf, S. Red fluorescence of plaque in the dentition—a comparison of Quantitative Light-induced Fluorescence-Digital (QLF-D) images and conventional images of disclosed plaque. *Photodiagnosis Photodyn Ther* **32**, 102063 (2020). <https://doi.org/10.1016/j.pdpdt.2020.102063>
- 10 Alavi, S. *et al.* Assessment of teledentistry in improving access to dental care: a systematic review. *Australian Dental Journal* **70**, 4–41 (2025).
- 11 Surdu, A. *et al.* Telemedicine and Digital Tools in Dentistry: Enhancing Diagnosis and Remote Patient Care. *Medicina* **61**, 826 (2025).
- 12 Andrade, K. M., Silva, B. P. M., de Oliveira, L. R. & Cury, P. R. Automatic dental biofilm detection based on deep learning. *J Clin Periodontol* **50**, 571–581 (2023). <https://doi.org/10.1111/jcpe.13774>
- 13 Li, S. *et al.* Automatic dental plaque segmentation based on local-to-global features fused self-attention network. *IEEE Journal of Biomedical and Health Informatics* **26**, 2240–2251 (2022).
- 14 Tez, B. Ç., Güzel, Y., Kızıltan Eliaçık, B. B. & Aydın, Z. Deep-Learning-Based AI-Model for Predicting Dental Plaque in the Young Permanent Teeth of Children Aged 8–13 Years. *Children* **12**, 475 (2025).
- 15 Hao, J. *et al.* T-Mamba: a unified framework with long-range dependency in dual-domain for 2D & 3D tooth segmentation. *arXiv preprint arXiv:2404.01065* (2024).
- 16 Hao, J., Liu, M., Yang, J. & Hung, K. F. GEM: Boost simple network for glass surface segmentation via vision foundation models. *IEEE Transactions on Multimedia* (2025).
- 17 Hao, J. *et al.* A semi-supervised transformer-based deep learning framework for automated tooth segmentation and identification on panoramic radiographs. *Diagnostics* **14**, 1948 (2024).
- 18 Hao, J. *et al.* Characteristics, licensing, and ethical considerations of openly accessible oral-maxillofacial imaging datasets: a systematic review. *npj Digital Medicine* **8**, 412 (2025).
- 19 Liu, Z. *et al.* The performance of large language models in dentomaxillofacial radiology: a systematic review. *Dentomaxillofacial Radiology*, twaf060 (2025).
- 20 Hao, J. *et al.* Towards Better Dental AI: A Multimodal Benchmark and Instruction Dataset for Panoramic X-ray Analysis. *arXiv preprint arXiv:2509.09254* (2025).
- 21 Hao, J. *et al.* in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 110–121 (Springer).
- 22 Hao, J. *et al.* OralGPT-Omni: A Versatile Dental Multimodal Large Language Model. *arXiv preprint arXiv:2511.22055* (2025).
- 23 Shi, J. *et al.* Semantic decomposition network with contrastive and structural constraints for dental plaque segmentation. *IEEE Transactions on Medical Imaging* **42**, 935–946 (2022).
- 24 Song, W. *et al.* Centerformer: a novel cluster center enhanced transformer for unconstrained dental plaque segmentation. *IEEE Transactions on Multimedia* (2024).
- 25 Zhang, Y. & Yang, Q. A survey on multi-task learning. *IEEE transactions on knowledge and data engineering* **34**, 5586–5609 (2021).
- 26 Yang, J. Multi-task learning for medical foundation models. *Nature Computational Science* **4**, 473–474 (2024).
- 27 Chen, Q. *et al.* Cooperative multi-task learning and interpretable image biomarkers for glioma grading and molecular subtyping. *Medical Image Analysis* **101**, 103435 (2025).
- 28 Kirillov, A. *et al.* in *Proceedings of the IEEE/CVF international conference on computer vision*. 4015–

- 4026.
- 29 Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. The pascal visual object classes (voc) challenge. *International journal of computer vision* **88**, 303–338 (2010).
 - 30 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
 - 31 Jeong, J.-S. *et al.* Deep learning for automated dental plaque index assessment: validation against expert evaluations. *BMC Oral Health* **25**, 1000 (2025).
 - 32 Ramírez-Pedraza, A. *et al.* Deep Learning in Oral Hygiene: Automated Dental Plaque Detection via YOLO Frameworks and Quantification Using the O'Leary Index. *Diagnostics* **15**, 231 (2025).
 - 33 Chen, X. *et al.* DeepPlaq: Dental plaque indexing based on deep neural networks. *Clinical Oral Investigations* **28**, 534 (2024).
 - 34 Liu, Z., Gomez, J., Khan, S., Peru, D. & Ellwood, R. Red fluorescence imaging for dental plaque detection and quantification: pilot study. *Journal of biomedical optics* **22**, 096008–096008 (2017).
 - 35 Doi, K. *et al.* Comparison of plaque control record measurements obtained using intraoral scanner and direct visualization. *Clinical and Experimental Dental Research* **10**, e852 (2024).
 - 36 Taud, H. & Mas, J.-F. in *Geomatic approaches for modeling land change scenarios* 451–455 (Springer, 2017).
 - 37 Zhang, H. *et al.* Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605* (2022).
 - 38 Hao, J. & Chen, S. Language-aware multiple datasets detection pretraining for DETRs. *Neural Networks* **179**, 106506 (2024).
 - 39 Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).
 - 40 Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. in *Proceedings of the IEEE international conference on computer vision*. 2980–2988.
 - 41 Rezatofighi, H. *et al.* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 658–666.
 - 42 Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
 - 43 Zhou, P., Xie, X., Lin, Z. & Yan, S. Towards understanding convergence and generalization of AdamW. *IEEE transactions on pattern analysis and machine intelligence* **46**, 6486–6493 (2024).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialsSubmitted.docx](#)