electronics

MDPI

Article

YOLO-DentSeg: A Lightweight Real-Time Model for Accurate Detection and Segmentation of Oral Diseases in Panoramic Radiographs

Yue Hua †, Rui Chen † and Hang Qin *

School of Computer Science, Yangtze University, Jingzhou 434000, China; yuehua@yangtzeu.edu.cn (Y.H.); chanry@yangtzeu.edu.cn (R.C.)

* Correspondence: hangqin@yangtzeu.edu.cn

† These authors contributed equally to this work.

Abstract: Panoramic radiography is vital in dentistry, where accurate detection and segmentation of diseased regions aid clinicians in fast, precise diagnosis. However, the current methods struggle with accuracy, speed, feature extraction, and suitability for low-resource devices. To overcome these challenges, this research introduces a unique YOLO-DentSeg model, a lightweight architecture designed for real-time detection and segmentation of oral dental diseases, which is based on an enhanced version of the YOLOv8n-seg framework. First, the C2f(Channel to Feature Map)-Faster structure is introduced in the backbone network, achieving a lightweight design while improving the model accuracy. Next, the BiFPN(Bidirectional Feature Pyramid Network) structure is employed to enhance its multi-scale feature extraction capabilities. Then, the EMCA(Enhanced Efficient Multi-Channel Attention) attention mechanism is introduced to improve the model's focus on key disease features. Finally, the Powerful-IOU(Intersection over Union) loss function is used to optimize the detection box localization accuracy. Experiments show that YOLO-DentSeg achieves a detection precision (mAP50(Box)) of 87%, segmentation precision (mAP50(Seg)) of 85.5%, and a speed of 90.3 FPS. Compared to YOLOv8n-seg, it achieves superior precise and faster inference times while decreasing the model size, computational load, and parameter count by 44.9%, 17.5%, and 44.5%, respectively. YOLO-DentSeg enables fast, accurate disease detection and segmentation, making it practical for devices with limited computing power and ideal for real-world dental applications.

Keywords: oral pathology; multi-object detection and segmentation; panoramic radiography; digital dentistry; YOLOv8

Academic Editor: Luca Mesin

Received: 29 December 2024
Revised: 9 February 2025
Accepted: 13 February 2025
Published: 19 February 2025

Citation: Hua, Y.; Chen, R.; Qin, H. YOLO-DentSeg: A Lightweight Real-Time Model for Accurate Detection and Segmentation of Oral Diseases in Panoramic Radiographs. *Electronics* **2025**, *14*, 805. https://doi.org/10.3390/electronics14040805

1. Introduction

Oral diseases represent one of the most widespread public health challenges globally, with considerable social and economic consequences. Based on the most recent Global Oral Health Status Report by the World Health Organization (WHO), approximately 3.5 billion people worldwide are affected by various oral diseases. Common conditions such as dental caries and periodontal disease not only affect patients' quality of life but also impose a significant economic burden and strain on healthcare resources worldwide [1]. As such, swift and reliable diagnosis and treatment are fundamental to improving public health and reducing healthcare costs.

Panoramic radiography, also known as panoramic tomography, serves as a non-invasive technique for imaging and is extensively applied in the screening and diagnosis

of various oral diseases. This technique enables clinicians to visualize patients' teeth, jawbones, and surrounding anatomical structures, facilitating the precise evaluation of disease types and lesion severity, thereby providing crucial information for optimal treatment planning. However, traditional manual analysis interpretation of panoramic radiographs relies heavily on the subjective judgment of clinicians. This approach requires significant labor and time and is also prone to human error, particularly due to factors such as fatigue or distraction, which can compromise diagnostic accuracy and overall reliability [2]. Consequently, there is an urgent demand for the development of quick and accurate automated methods for interpreting panoramic radiographs.

In recent years, radiomics has emerged as a cutting-edge technology in the field of medical image analysis and interpretation, offering a robust framework for extracting high-dimensional quantitative features from images, such as texture, shape, and intensity. These features can capture subtle disease-related patterns that may not be discernible through traditional visual assessment, thus enhancing diagnostic accuracy and enabling personalized treatment strategies [3,4]. The integration of radiomics with advanced artificial intelligence (AI) models, especially deep learning frameworks, has shown significant promise in improving detection, segmentation, and classification tasks in medical imaging [5,6]. This synergy provides an enriched data-driven foundation for the interpretation of complex medical images, including panoramic radiographs, ultimately paving the way for more precise and efficient healthcare solutions.

In automated analysis, involving the interpretation of panoramic radiographs, accurately detecting and segmenting areas of interest related to oral pathologies play a crucial role [7], providing clinicians with intuitive visualizations to aid in further diagnosis and treatment planning. However, panoramic radiographs often exhibit noise, low contrast, and uneven exposure [8]. Additionally, similar tissue densities and artifacts (e.g., dental fillings and implants) or temporary obstructions (e.g., orthodontic braces, impacted teeth, crowding, spacing, and missing teeth) can obscure the boundaries of regions of interest, making the detection and segmentation process more complex and challenging [9]. Therefore, establishing an accurate and robust model for detecting and segmenting dental pathologies in panoramic radiographs is essential.

Over recent years, significant research efforts have been dedicated to automatic detection and segmentation methods for panoramic radiographs. The representative approaches can generally be classified into traditional methods and deep-learning-based methods. The traditional methods encompass thresholding-based, region-based, edge detection, and machine learning methods. Thresholding-based segmentation approaches, such as Otsu's method, maximum entropy segmentation, and iterative thresholding, divide pixel regions according to grayscale thresholds [10,11]. Although these methods are simple and practical, they are highly sensitive to noise and threshold selection, often resulting in under- or over-segmentation. Region-based methods (e.g., region growing and split-and-merge) segment regions based on pixel intensity, which works well for images with clear boundaries but is sensitive to noise, computationally expensive, and highly dependent on the choice in seed points [12]. Edge-based methods (e.g., Roberts, Sobel [13], and Canny) detect discontinuities in pixel values to identify boundaries [14], but they are vulnerable to noise, lack edge continuity, and have limited effectiveness. Machine learning methods surpass the aforementioned techniques in preserving image features and spatial information. However, these methods often require extensive human intervention, are computationally intensive, and generally yield unsatisfactory performance. Driven by the significant progress in artificial intelligence and deep learning, the application of deep learning methods has become increasingly common in medical image detection and segmentation owing to their superior feature representation capabilities [15]. Among them,

convolutional neural networks (CNNs) have shown tremendous potential for automatically detecting and segmenting oral pathologies from high-resolution images, becoming a dominant trend in this field [16]. CNN-based models significantly enhance the accuracy of detection and segmentation tasks while also optimizing clinical decision-making by reducing human errors. However, complex large-scale neural networks require substantial computational resources and rely heavily on high-performance graphics processing units (GPUs) to enhance segmentation and detection performance [17]. These approaches are also hindered by challenges including poor feature extraction, oversized models, significant computational burden, and slow speed in detection and segmentation. These issues are particularly prominent in rural hospitals lacking advanced computational resources. The high cost of high-performance GPUs also hinders the widespread application of such technologies in healthcare systems. Additionally, training large-scale models demands considerable computational power and time, further delaying the progress of updating and validating models for oral disease detection and segmentation tasks. Therefore, developing lightweight detection and segmentation models that can operate effectively under limited computational resources is crucial for practical applications.

To tackle the issues highlighted above, this study focuses on common oral dental diseases and proposes an innovative method for detecting and segmenting dental diseases in panoramic radiography, termed YOLO-DentSeg. This method enhances the YOLOv8n-seg framework by incorporating advanced strategies, including the C2f-Faster module, BiFPN Feature Pyramid Network, EMCA attention mechanism, and PowerfulIOU loss function. These enhancements not only improve the system's performance in capturing important features from oral dental disease but also significantly increase detection and segmentation accuracy while accelerating processing speed. Furthermore, the approach effectively minimizes the system's parameters, computational load, and overall size, achieving a better balance between network overhead and performance. This research not only contributes to improving the detection and diagnosis of common oral diseases in dental medicine but also provides technical references for intelligent healthcare in areas such as auxiliary diagnosis, automated medical image processing, and the optimization of medical resource allocation.

The principal contributions of this study are outlined below as follows:

(1) We construct a panoramic radiograph dataset. Owing to the limited availability and inconsistent quality of publicly accessible panoramic radiograph datasets, this study compiled images from the publicly available DENTEX dataset and images collected from Jingzhou First People's Hospital. After expert selection and annotation, a high-quality dataset comprising 2720 images of typical oral pathologies was constructed.

(2) We propose a YOLO-DentSeg model, a lightweight, real-time model for detecting and segmenting dental pathologies, built on an enhanced YOLOv8n-seg framework. The C2f-Faster module is adopted within the backbone network to optimize the structure, cutting down on the parameter count and the computational overhead. The BiFPN structure is used in the neck network to integrate deep and shallow features, strengthening the model's capability to process complex details in panoramic radiographs. The network design incorporates the proposed EMCA attention module, which enhances the model's attention to key features. The conventional loss function is replaced by Powerful-IOU to optimize the bounding box localization accuracy and significantly accelerate loss convergence.

(3) We carried out a wide range of experiments on the dataset we constructed. The experimental outcomes highlight that the our proposed model outperforms the existing models with regard to detection and segmentation accuracy, processing speed, parameter count, computational load, and its ability to run on devices with limited computational resources. These results affirm the success of the YOLO-DentSeg model algorithm in

enabling real-time and accurate detection and segmentation of dental diseases on low-resource devices.

2. Related Work

2.1. Image Processing for Personalized Dental Diagnosis

The traditional image processing algorithms for dental diagnosis in oral medicine include thresholding-based, region-based, edge detection, and machine learning methods that have been widely used in dental medical imaging in recent years.

Threshold-based methods operate by selecting a threshold based on the pixel intensity values in the image. The pixels above the threshold are assigned to one region, and those below are assigned to another. For example, Tikhe et al. [18] used a thresholding method to identify caries in enamel and proximal regions, effectively segmenting the image. Lin et al. [19] applied Otsu thresholding and boundary tracking to segment teeth in periapical radiographs. They later developed a hybrid thresholding method to locate regions of alveolar bone loss [20]. However, thresholding methods often ignore spatial features, making them sensitive to noise and threshold selection, leading to under- or over-segmentation in real-world images with varying grayscale levels.

Region-based methods partition an image into multiple regions by detecting discontinuities in pixel intensity. For instance, Lurie et al. [21] used this approach to segment panoramic dental X-ray images, assisting in osteoporosis detection. Modi et al. [22] applied a region-based method for bitewing dental X-ray segmentation, while Indraswari et al. [23] proposed a 3D region-merging technique for CBCT images. Although effective with clear boundaries, region-based methods are computationally expensive, sensitive to noise, and highly dependent on seed points and similarity metrics. Furthermore, these methods have limited semantic segmentation performance due to their reliance on intensity or texture features.

Edge-based methods identify points and edges in an image by detecting discontinuities in color or pixel intensity. For example, Gan et al. [24] employed a tooth shape propagation strategy to segment tooth contours, while Wang et al. [25] proposed a hybrid active contour model for root segmentation. Despite their accuracy, edge-based methods struggle with edge continuity and closure, often producing false edges and being sensitive to noise. In noisy images or those with unclear boundaries, their performance deteriorates.

Machine-learning-based methods, often regarded as a branch of artificial intelligence (AI), have proven to be powerful tools for computer-aided diagnostic tasks. For example, Fernandez et al. [26] used neural networks for tooth segmentation in palatal views, while Prakash et al. [27] developed an SVM-based defect analysis system. Mortaheb et al. [28] employed SVM and mean shift algorithms for tooth segmentation, addressing metal artifacts. Tuan et al. [29] combined fuzzy clustering techniques for radiograph segmentation, and Geetha et al. [30] used backpropagation neural networks for caries diagnosis. Although these methods have shown success, they require extensive manual intervention and rely heavily on low-level image information. They are also affected by brightness, contrast, and occlusion, thus impacting accuracy and robustness, making them less suitable for practical use.

While these traditional image processing methods have contributed to the detection and segmentation of dental features, their limitations, including manual intervention, noise sensitivity, and slower speeds, hinder their widespread practical application.

2.2. AI-Enabled Detection and Segmentation of Digital Dentistry

With the fast-paced evolution of deep learning technologies and computer vision, an increasing number of artificial intelligence models are being applied to digital dental diag-

nostics in the field of dentistry. For example, Lo Casto et al. [31] evaluated ResNet-152 and VGG-19 for assessing the spatial relationship between the mandibular third molar (MM3) and mandibular canal (MC) in panoramic radiographs. Their findings demonstrated that both CNN models outperformed inexperienced observers in diagnostic accuracy, highlighting the potential of deep learning to enhance clinical decision-making in dental radiology. This study provides valuable insights into the application of AI-assisted diagnostics in dental practice. Mureșanu et al. [32] developed a YOLOv8-based AI model for detecting dental conditions, with a specific focus on identifying teeth at risk prior to radiotherapy. Their model, trained on 1628 annotated panoramic radiographs and externally validated on 180 images, demonstrated strong performance in detecting implants and surgical devices (precision and recall > 0.8). However, its generalizability declined in external validation, particularly for periapical lesions and bone loss, requiring further improvements.

Lakshmi et al. [33] proposed a semi-automatic tooth image segmentation method that uses a graph cut technique to distinguish foreground tissue regions from background tooth regions, followed by segmentation and detection using the AlexNet network. Banar et al. [34] applied the YOLO [35] model to divide images into adjacent non-overlapping blocks, detecting blocks that contain the third molar and identifying their geometric centers, after which a region of interest (ROI) is extracted around each center. A U-Net network is then used to segment the third molar within the ROI, achieving a fully automated analysis of the third molar developmental stages. Wang et al. [36] presented a refined Mask R-CNN network for the segmentation of periapical lesion regions, which achieved more than 0.97 pixel accuracy on average. The network uses ResNet as its backbone to capture rich feature representations from the input images and a region proposal network (RPN) for top-down instance segmentation and mask prediction. However, although these multi-stage methods improve accuracy to some extent, they suffer from slow segmentation speeds.

Patil et al. [37] devised a caries detection technique that merges convolutional neural networks (CNNs) with a multidimensional projection variational method, which improves detection accuracy over traditional CNN methods but still requires extensive training data and a longer processing time. Zhu et al. [38] developed CariesNet, a deep learning model built upon U-Net, to detect caries in panoramic radiographs. The model incorporates a full-scale axial attention module, achieving an accuracy of 93.61%, outperforming traditional diagnostic methods. However, the similarity in texture and color between diseased and healthy dental tissues complicates the accurate segmentation of various regions and their edges. In [39], Ma et al. introduced an improved image cascade network (ICNet) method for segmenting various types of dental pathologies, including calculus and gingivitis. By integrating an attention module into the ICNet structure, the model enhances relevant features while suppressing irrelevant ones, addressing segmentation inaccuracies in lesion areas. However, this approach has high model complexity and computational requirements. Bağ et al. [40] employed the YOLOv5 model to automatically identify key anatomical structures in pediatric panoramic X-rays. Busra Beser et al. [41] evaluated how well YOLOv5 performs in the automatic detection, segmentation, and identification of primary and permanent teeth in children with mixed dentition panoramic radiographs. The study concluded that the YOLOv5-based model demonstrated excellent accuracy and precision in detecting and segmenting both primary and permanent teeth in this specific patient group.

Despite the significant advancements in performance achieved by these models, they generally require substantial computational resources and still face issues such as high miss rates and slow detection speeds in complex environments. Particularly in the process of handling real-time detection and segmentation, the computational demands of the existing models exceed the processing capabilities of terminal devices with limited computing

power. Therefore, designing a model that not only ensures detection accuracy but also operates efficiently on low-power devices has become a focus of current investigations.

3. A General System Model for Automated Oral Disease Detection and Segmentation

In this section, an improved YOLOv8n-seg model is developed for dental pathology detection and segmentation on panoramic radiographs. First, the system workflow is described, followed by an introduction to the overall framework of our proposed model, YOLO-DentSeg, with detailed explanations of each enhancement. The symbols used in this study and their corresponding meanings can be found in Table 1.

Table 1. List of notations.

Symbol	Definition
H	Input feature map's height
W	Input feature map's width
C	Input feature map's channel count
k	Convolution kernel size
$*$	Convolution calculation
X	Input feature maps
x_{ij}	Feature map's value at position (i, j)
σ	Sigmoid activation function
$C1D_k$	1D convolution process with a kernel of size k
$\lvert t \rvert_{odd}$	Take the odd number closest to t
Y	Feature map after recalibration
ω	Attention weight
$\odot$	Element-wise multiplication (Hadamard product)
$IntersectionArea$	Overlapping area between the predicted and ground truth bounding boxes
$UnionArea$	Area of the union between the predicted and ground truth bounding boxes
IoU	Intersection and concurrency ratio of predicted and true bounding boxes
ρ	Euclidean distance between the true disease box and the predicted disease box center points
c	The diagonal distance of the smallest enclosing box for the predicted and true bounding disease boxes.
v	Aspect ratio penalty
α	Weighting factor
P	Penalty factor
λ	Hyperparameters controlling the behavior of the attention function
η	Learning rate
num_epochs	Model training rounds
$batch\_size$	Training batch size
B	Weight decay factor
θ	Model parameter initialization values
W	Learnable convolutional kernel
b	Bias entry
F_i	Feature maps for different layers
α_i	Fusion weights
y_i	True category labeling of the target
$\hat{y}_i$	Probability that the model predicts the category
M_i	True segmentation mask label
$\hat{M}_i$	Probability that a pixel predicted by the model belongs to the target region

3.1. System Description

This section introduces YOLO-DentSeg, an enhanced model based on YOLOv8n-seg, designed for the detection and segmentation of dental pathologies in panoramic radiographs. As shown in Figure 1, the detection and segmentation process of YOLO-DentSeg consists of four key steps. The first step involves constructing a dataset of dental pathologies in panoramic radiographs. A set of diverse panoramic radiograph images are selected, each containing observable dental pathology targets. These images are then resized to ensure compatibility and consistency, providing the necessary input for the YOLO-DentSeg model.

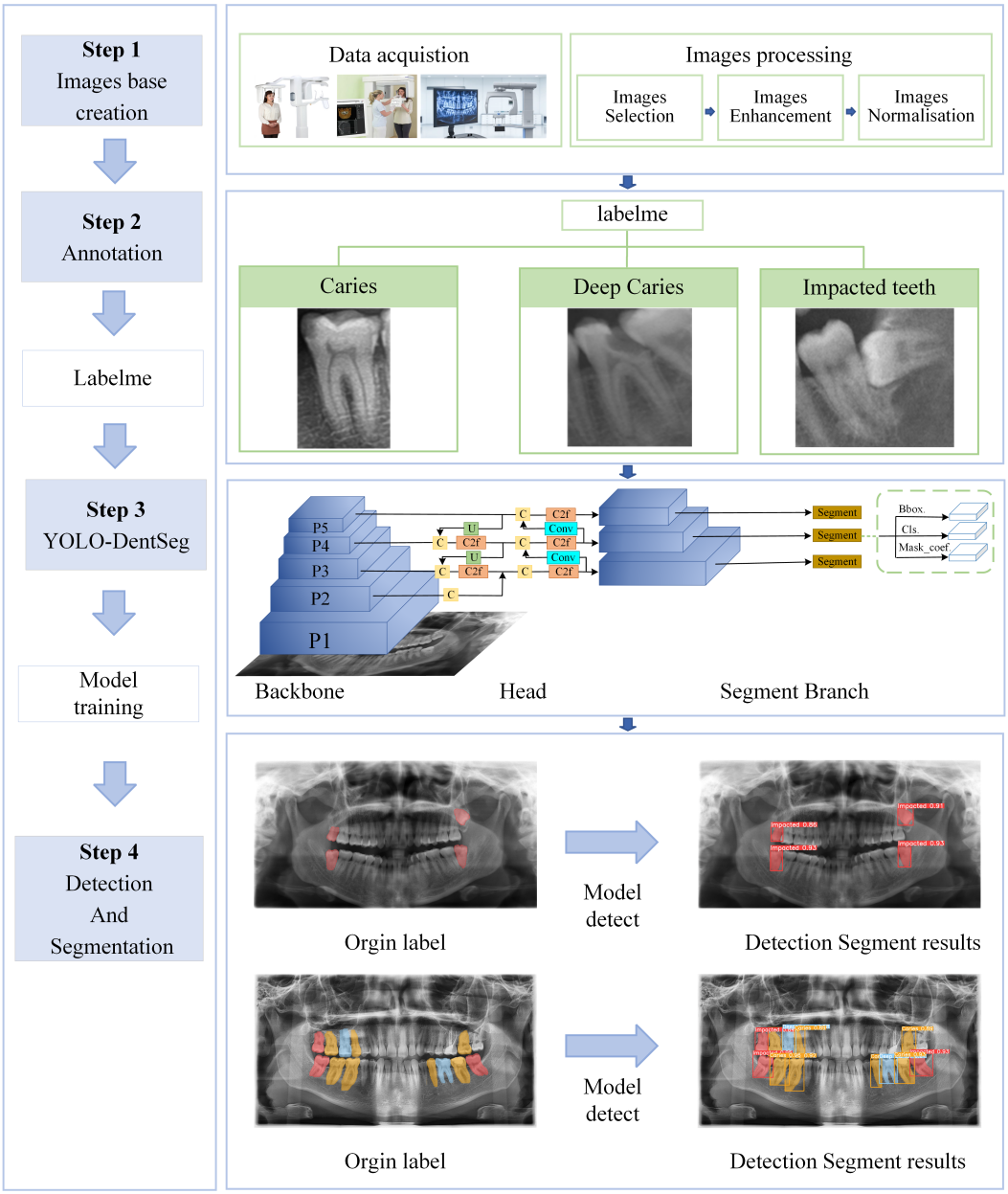

Figure 1. Schematic diagram of oral disease detection and segmentation.

The second step involves annotating the constructed dataset of panoramic radiographs. For this study, three common dental pathologies—caries, deep caries, and impacted teeth—were selected for annotation. Each panoramic radiograph is meticulously annotated by experts to highlight specific pathologies. Once annotated, the dataset is exported in a

YOLOv8-compatible format. Subsequently, the image dataset is proportionally categorized in three sets consisting of training set, verification set, and test set.

The third step involves the actual training process. The training dataset is fed into the YOLO-DentSeg framework, enabling it to learn the characteristics of dental pathologies by analyzing the annotated panoramic radiographs. Leveraging the information obtained from the training set, the algorithm refines its parameters to improve its capability to detect various dental pathologies. Upon completion of the training process, the model undergoes testing with the test set to measure its performance. The YOLO-DentSeg framework generates predictions for dental pathologies in the test set, and these predictions are compared with the actual labels to gauge the effectiveness and accuracy of the model.

The fourth step involves the actual deployment for prediction. The model that demonstrated the best performance in step three was exported and packaged for practical dental pathology detection and segmentation on panoramic radiographs.

3.2. Proposed Method

The range of YOLO (You Only Look Once) algorithmic frameworks stand out among various detection and segmentation methods due to their fast detection and segmentation capability and high accuracy [42]. With the continuous updating and iteration of the model framework, the YOLO series has become a popular real-time target detection and segmentation model, which is widely used in accurate automated medical image analysis, including oral dental disease detection and segmentation.

The Ultralytics team released the latest YOLOv8 object detection algorithm, which evolved from YOLOv5 and features an anchor-free, single-stage detection framework. The complete architecture of the network includes four key components: input, backbone, neck, and head. Input side resizes the image to the appropriate dimensions for training. The backbone, based on the improved CSPDarknet architecture, adopts the ELAN design philosophy and integrates C2f feature fusion with optimized computational performance, enhancing feature extraction, gradient flow, and accuracy in multi-scale object detection. The neck utilizes the Path Aggregation Network (PAN) [43] and Feature Pyramid Network (FPN) [44] to enhance feature fusion across different spatial scales. The head adopts a decoupled design, separating the tasks of classification and detection. YOLOv8-seg is developed from an improved iteration of the YOLO target detection framework, which, in combination with the YOLACT [45] network, is capable of performing accurate pixel-level instance segmentation tasks. This integration enhances the traditional object detection capabilities, transforming it into a model that simultaneously performs both object detection and segmentation. In addition to detecting objects, the model can generate precise mask labels for each object instance in the image, offering finer-grained object recognition and segmentation capabilities, thereby enabling more comprehensive analysis. Therefore, YOLOv8-seg can serve as a foundational model for detecting and segmenting dental conditions in oral healthcare applications.

For improvements in accuracy regarding dental pathology detection and segmentation in panoramic radiographs under complex backgrounds, while maintaining model simplicity and real-time processing, we propose a modified YOLOv8n-seg-based model named YOLO-DentSeg, as illustrated in Figure 2. In the model, C2f-Faster replaces the C2f module in the backbone, cutting down number of parameters and computational overhead. In neck, BiFPN replaces FPN and PAN, combining deep and shallow image features for innovative multi-scale fusion. By integrating the proposed EMCA attention module, the network is empowered to better prioritize key features, thus enhancing the model's performance. Additionally, the loss function is replaced with the PowerIOU loss

function, which optimizes bounding box localization accuracy and significantly accelerates loss convergence.

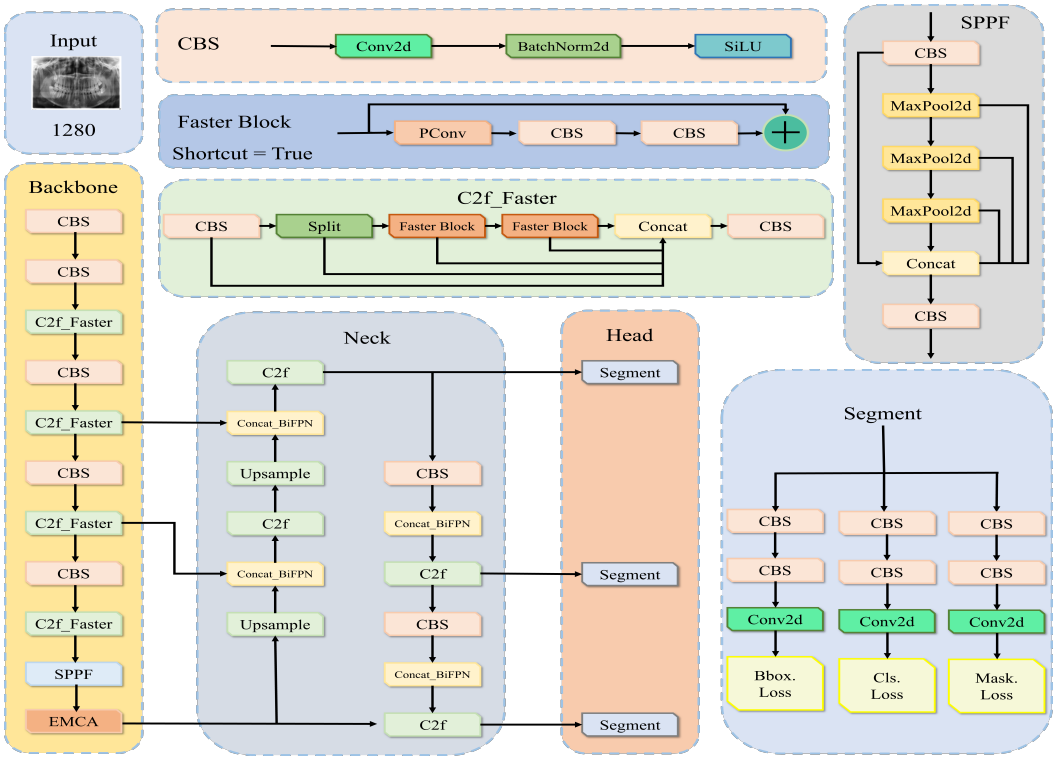

Figure 2. YOLO-DentSeg model structure diagram.

3.2.1. Channel to Feature Map-Faster (C2f-Faster) Module

The C2f module introduced in YOLOv8 effectively leverages information from feature maps at different scales by stacking additional Bottleneck structures, which substantially improves model's feature representation capability. However, the Bottleneck structure includes multiple convolution operations, and stacking more Bottleneck structures while improving multi-scale feature utilization inevitably leads to a rise in computational load and complexity. With greater model depth, the feature maps expand in terms of channels, potentially causing redundancy as different maps may carry similar or even identical information. This redundancy results in a waste of computational resources and a decrease in computational efficiency. To address this issue and optimize resource usage, we utilize the partial convolution (PConv) strategy proposed in the FasterNet lightweight backbone network [46]. Specifically, in PConv, convolution is performed on only a fraction of the input feature map channels (e.g., 1/4), with the other (e.g., 3/4) channels remaining intact. The unprocessed channels are preserved and utilized through 1 × 1 pointwise convolutions, maintaining consistency between input and output feature map dimensions. This approach minimizes unnecessary computations and greatly improves processing efficiency. By retaining the unprocessed channels, the 1 × 1 pointwise convolution can effectively extract useful information from them, thereby reducing the computational load while maintaining model accuracy to the greatest extent possible. The optimization strategy based on PConv effectively reduces computational load, optimizes resource utilization, improves overall computational efficiency, and lowers model complexity. Figure 3 displays a schematic of partial convolution.

Assume the input is a feature map with a height of H, a width of W, and C channels, where the convolution kernel has size k. The computational load for standard convolu-

tion is denoted as F_1 and for partial convolution as F_2. The formulas for calculating the computational load are as follows:

$$F_1 = k \times k \times H \times W \times C \times C, \tag{1}$$

$$F_2 = k \times k \times H \times W \times \frac{C}{4} \times \frac{C}{4}, \tag{2}$$

$$\frac{F_2}{F_1} = \frac{1}{16}. \tag{3}$$

Equation (3) demonstrates that the computational load of using partial convolution is merely 1/16 that of using regular convolution. This shows that partial convolution greatly reduces computational overhead and enhances floating-point operation efficiency.

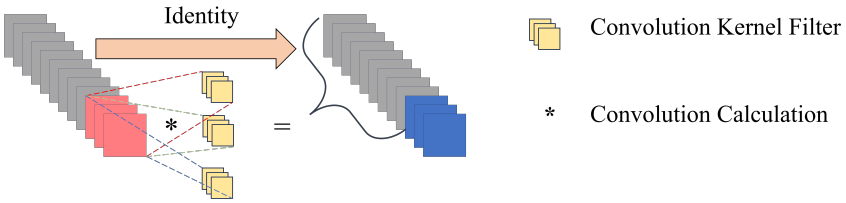

Figure 3. PConv schematic.

This research leverages the partial convolution (PConv) concept from FasterNet to design a Faster-Block module, which replaces Bottleneck structure within the C2f module, leading to a significant reduction in computational requirements. Figure 4a shows Faster-Block module.

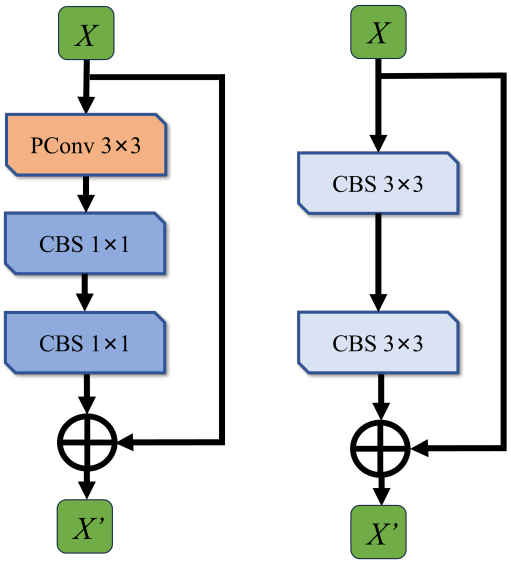

Figure 4. Comparison of Faster-Block and Bottleneck structures.

In Faster-Block module, partial convolution is used. First, a standard convolution operation is performed on one-quarter of the input feature map's channels, with the other three-quarters left unchanged. Then, the convolved channels are concatenated with the unchanged three-quarters of the channels to produce the output. This approach reduces redundant computations while preserving the original channel information to the greatest extent possible. Although three-quarters of the channels are not convolved, they are preserved as the following 1 × 1 convolution layer is capable of extracting valuable information from them. To further reduce potential feature map information loss from

partial convolution, the 1 × 1 pointwise convolution in the CBS module doubles the output channels, maximizing the use of unprocessed channels and preventing information loss. The subsequent 1 × 1 convolution layer brings the channels back to their initial quantity, ensuring that the input (X) from the shortcut branch and the output (X') processed by the main path remain consistent in terms of dimensions and size (note: CBS is a basic convolutional module composed of convolution layers, batch normalization layers, and activation functions).

The Bottleneck module in C2f enhances the representational capacity of feature maps and improves detection accuracy by integrating multi-scale feature information. However, as the number of modules increases, the computational overhead grows significantly. To address this issue, we substituted all Bottleneck modules in C2f with the Faster-Block structure, resulting in the C2f-Faster module. Since YOLOv8 contains a large number of Bottleneck structures, this modification significantly reduces the computational load, thereby greatly improving inference speed, which has been validated in subsequent experiments. The C2f-Faster module is depicted in the schematic shown in Figure 5 (note: Split is a module that splits the number of input channels in half, and Concat is a feature connection module).

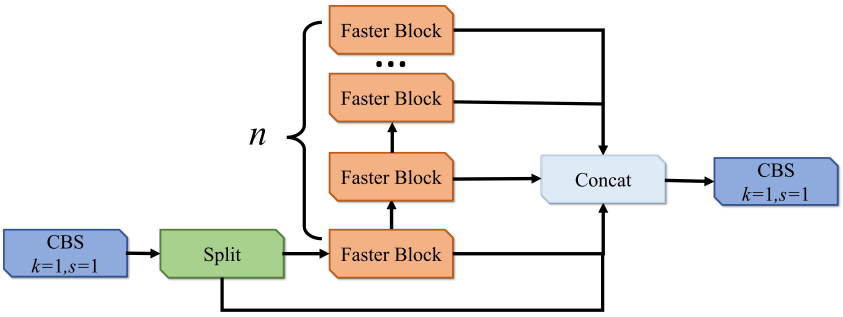

Figure 5. C2f-Faster schematic.

Table 2 compares the feature extraction network structure of YOLOv8n-seg with that of Model 1, which incorporates the C2f-Faster improvements. The original C2f modules in the YOLOv8n-seg feature extraction network were replaced with C2f-Faster modules in Model 1. A comparison of the parameter count between the corresponding C2f and C2f-Faster modules shows that C2f-Faster reduces the parameter count by approximately 51.1% (the details of Model 1 are described in subsequent ablation study).

3.2.2. Bidirectional Cross-Scale Weighted Feature Pyramid (BiFPN) Structure

The YOLOv8n-seg model employs PANet as its neck component. Compared to FPN, PANet introduces a bottom-up path, enabling bidirectional fusion of features. This method allows feature maps to retain high-level semantic details and detailed positional information, overcoming limitations of unidirectional information flow and effectively addressing the issue of significant shallow feature loss in FPN. However, due to the prevalence of elongated and irregularly impacted teeth as well as small-sized caries in oral panoramic radiographic datasets, the corresponding feature regions occupy a small area within the image and carry limited feature information. After undergoing multiple layers of convolution, downsampling, and pooling, some of this feature information is lost. Therefore, for the irregularly shaped lesions in oral panoramic radiographs, PANet still has limitations in capturing detailed image features.

Table 2. Comparing the number of parameters before and after model modification.

Model	Layer	Kernel Size	Output Channel	Parameter
YOLOv8n-seg	Conv	3 × 3	16	
	Conv	3 × 3	32	
	C2f	1 × 1	32	7360
	C2f	1 × 1	64	49,664
	Conv	3 × 3	128	
	C2f	1 × 1	128	197,632
	C2f	1 × 1	256	37,248
	SPPF	1 × 1	256	
	Conv	3 × 3	16	
	Conv	3 × 3	32	
Model 1	C2f-Faster	1 × 1	32	3920
	Conv	3 × 3	64	
	C2f-Faster	1 × 1	64	22,144
	Conv	3 × 3	128	
	C2f-Faster	1 × 1	128	93,184
	Conv	3 × 3	256	
	C2f-Faster	1 × 1	256	23,488
	SPPF	1 × 1	256	

This study proposes replacing the neck structure in YOLOv8n-seg with a BiFPN (Bidirectional Feature Pyramid Network) [47] for the following reasons: (1) in the neck structure, repetitive upsampling and downsampling operations change the image resolution, causing important information to be lost, thereby impairing the feature learning process for oral panoramic radiographs and potentially leading to false or missed detections. The BiFPN structure, through skip connections, enhances information flow between feature maps at different network levels, providing more positional and detailed information, thereby helping the model to better focus on small lesions in oral panoramic radiographs. (2) The BiFPN structure contains fewer network nodes compared to PANet, and the skip connections do not introduce additional fusion nodes. For the proposed panoramic radiograph dental lesion segmentation model, the incorporation of BiFPN offers significant performance improvements with minimal computational load. The network structures of FPN, PANet, and BiFPN are illustrated in Figure 6. The subsequent experimental section will provide a detailed evaluation of their performance regarding dental lesion detection and segmentation.

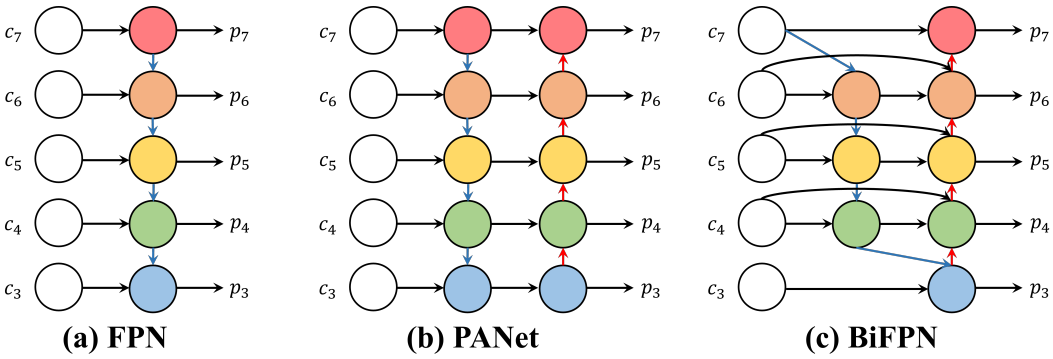

Figure 6. FPN, PANet, and BiFPN structures.

3.2.3. Efficient Multi-Channel Attention Mechanism (EMCA) Attention Module

In oral panoramic radiographs, dental lesions are typically small in size compared to the overall image, and the presence of background noise and redundant information is

substantial. Common challenges include overlapping mechanical artifacts with carious lesions and image blurriness caused by limitations in imaging techniques. These issues greatly complicate the identification of dental lesions, making it difficult to extract relevant features or resulting in extracted features being contaminated with irrelevant information, ultimately leading to reduced detection and segmentation accuracy. To address the difficulties in feature extraction and the low precision in detection and segmentation, this study incorporates the ECA (Efficient Channel Attention) mechanism [48], which focuses attention on specific regions of interest, thereby enhancing feature extraction capability. By utilizing global average pooling (GAP), ECA mechanism extracts comprehensive global features from the feature maps and employs 1D convolution to generate channel attention weights, thus improving the expressiveness of the features.

However, ECA has certain limitations. First, it relies solely on average pooling, which does not fully exploit the local salient features that max pooling can offer, potentially leading to the neglect of important details. Additionally, the feature fusion in ECA is relatively simplistic and lacks the integration of diverse feature information, which may impair segmentation performance when dealing with complex backgrounds. Given that dental lesions are small and the images contain significant background noise and redundant information—such as overlapping artifacts and blurred images due to imaging limitations—these challenges further hinder feature extraction. To resolve these issues, the study presents an improved EMCA (Enhanced Multi-Channel Attention) mechanism. EMCA integrates both global average pooling and max pooling, enabling it to capture both global contextual information and local features simultaneously. This mechanism effectively focuses on the critical information needed for the task at hand, filtering out irrelevant and distracting information from the inputs, thereby significantly enhancing the ability to extract key features related to dental lesions. Figure 7 below illustrates the architecture of EMCA.

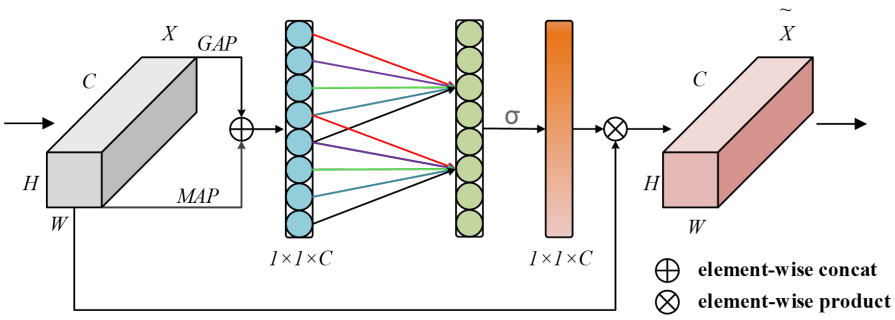

Figure 7. EMCA schematic of attention mechanisms.

The EMCA (Enhanced Multi-Channel Attention) module achieves effective attention computation and feature enhancement through the following steps. The detailed workflow is as follows:

(1) Global Pooling and Feature Fusion: The EMCA module first applies both global average pooling (AdaptiveAvgPool2d) and global max pooling (AdaptiveMaxPool2d) to obtain global feature representations for every channel in the input features:

$$AvgPool(x) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} x_{ij}, \tag{4}$$

where x represents a specific channel in the input feature map, W and H refer to its width and height, and x_{ij} represents the value located at coordinates (i, j) with the feature map.

$$MaxPool(x) = \max_{i,j} x_{ij}, \tag{5}$$

The results from both pooling operations are then summed to enhance global information and capture multi-scale features:

$$Z - Pool = AvgPool(x) + MaxPool(x). \tag{6}$$

(2) Cross-Channel Interaction: Next, by applying a 1D convolution to the fused feature map, the channel attention weights are obtained as follows:

$$\omega = \sigma(C1D_k(Z - Pool)), \tag{7}$$

where σ represents the Sigmoid activation function and $C1D_k$ is the 1D convolution operation with a kernel size of k.

(3) Adaptive Kernel Size Selection: The 1D convolution kernel size k is adjusted according to the number of channels C, as per the following:

$$k = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{\text{odd}}, \tag{8}$$

where C is the number of channels, and γ and b are hyperparameters, typically set to 2 and 1, respectively, with $|t|_{odd}$ denoting rounding to the nearest odd number.

(4) Feature Recalibration: Recalibrate the features using channel attention weight as follows:

$$Y = \omega \odot X. \tag{9}$$

where Y denotes the recalibrated feature map, ω represents channel attention weights, X represents input feature map, and $\odot$ represents element-wise multiplication (Hadamard product).

The EMCA (Enhanced Multi-Channel Attention) module optimizes ECA (Efficient Channel Attention) by introducing a fusion of global max pooling and global average pooling, which enhances the global perspective of features and improves feature extraction across different scales. The primary advantages of EMCA include Enhanced Global Information: by combining global average pooling and global max pooling, EMCA effectively captures both global context information and features across multiple scales, thereby improving the overall feature extraction capability. Efficient and Lightweight: compared to traditional attention mechanisms, EMCA offers significant performance improvements without substantially increasing the number of parameters or computational load, making it a more efficient solution. Long-Range Dependency Capture: through the use of 1D convolutions and adaptive kernel size selection, EMCA can effectively capture long-range dependencies between channels, further enhancing its ability to focus on important features. A detailed evaluation of EMCA's performance in dental lesion detection and segmentation will be provided in the subsequent experimental section of this study.

3.2.4. Powerful IOU

In the process of detecting and segmenting dental disease lesion targets in oral panoramic radiographs, challenges such as occlusion between dental disease lesion targets and significant overlap of different dental disease lesion targets often arise. These issues present difficulties for both the accuracy of detection and segmentation and the speed of bounding box regression. To enhance detection and segmentation precision and accelerate bounding box regression, this study proposes replacing the original bounding box loss function, CIoU (Complete Intersection over Union), in the YOLOv8-seg model with the Powerful-IoU loss function. The structure of Powerful-IoU [49] is shown in Figure 8b.

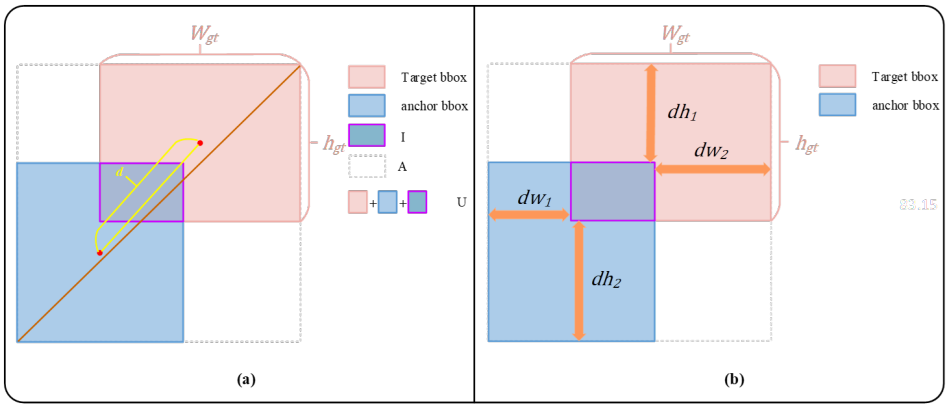

Figure 8. Schematics of CIOU and PowerIOU. (**a**) The structure of the original YOLOv8 boundary box loss function, CIoU (Complete Intersection over Union); (**b**) The structure of the proposed boundary box loss function, Powerful-IoU.

CIoU loss function in bounding box regression only considers the shape loss. In cases where the predicted bounding box for oral dental diseases has the same aspect ratio as the ground truth but differs in size, CIoU becomes ineffective. This hinders the model's ability to converge quickly and impacts the precision of lesion localization. CIoU loss function is provided in Equation (14):

$$IoU = \frac{IntersectionArea}{UnionArea}, \tag{10}$$

$$\rho^2(b, b^g) = (b_x - b_x^g)^2 + (b_y - b_y^g)^2, \tag{11}$$

$$c = \sqrt{(c_x - c_x^g)^2 + (c_y - c_y^g)^2}, \tag{12}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w_g}{h_g} - \arctan\frac{w}{h}\right)^2, \tag{13}$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^g)}{c^2} + \alpha v. \tag{14}$$

In this context, $IntersectionArea$ refers to overlapping area between oral disease's predicted bounding box and oral disease's ground truth bounding box, $UnionArea$ refers to area of union between the two boxes, IoU calculates the ratio of the area where the boxes intersect to their combined area, ρ indicates the Euclidean distance between the two box centers, and c represents the diagonal length of the minimum enclosing bounding boxes for the two boxes. Variable v measures the aspect ratio consistency, and α is a weighting factor.

In contrast, PIoU solves these problems by incorporating a penalty factor that adapts to the size, along with a gradient adjustment function that depends on the quality of the anchor boxes. Loss function of PIoU is formulated as follows:

$$P = \left(\frac{dw_1}{w_{gt}} + \frac{dw_2}{w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}}\right)/4, \tag{15}$$

$$f(x) = 1 - e^{-x^2}, \tag{16}$$

$$PIoU = IoU - f(P), \quad -1 \le PIoU \le 1, \tag{17}$$

$$L_{PIoU} = 1 - PIoU = L_{IoU} + f(P), \quad 0 \le L_{PIoU} \le 2. \tag{18}$$

In the above formulas, dw_1, dw_2, dh_1, and dh_2 represent absolute distances of the edges from the predicted and ground truth bounding boxes for oral diseases. The terms $w_g t$

and $h_g t$ correspond to the ground truth bounding box's width and height.Variable P is the penalty factor, which adjusts for the variation in target box size, thereby preventing the expansion of anchor boxes. Given that the denominator of P is based only on the target box's size and is independent of the enclosing rectangle of both the anchor and target boxes, the loss function can more effectively quantify the gap between predictions and actual values. To optimize the performance of the loss function even more, the PIoU introduces an adaptive gradient adjustment function $f(x)$, which modifies the gradient according to the quality of the bounding region to address the problem of expansion.

By employing the Powerful-IoU loss function, we can accelerate model convergence while maintaining accuracy and reducing training time. Moreover, Powerful-IoU effectively handles overlapping and severely occluded pathology targets, enhances boundary box regression accuracy, and improves the precision of pathology target detection and segmentation.

To further strengthen the focusing mechanism of Powerful-IoU, a PIoUv2 loss function is proposed. In PIoUv2, a dynamic attention mechanism layer is introduced, and its function is determined by a single hyperparameter λ. This mechanism enables PIoUv2 to pay more attention to intermediate-quality anchor boxes; this results in the final Powerful-IoU loss function.

Here, λ is a hyperparameter that controls attention function behavior. Formulas defining PIoUv2 are as follows:

$$q = e^{-p}, \quad q \in (0, 1], \tag{19}$$

$$u(x) = 3x \cdot e^{-x^2}, \tag{20}$$

$$L_{PIoU_{v2}} = u(\lambda q) \cdot L_{PIoU} = 3 \cdot (\lambda q) \cdot e^{-(\lambda q)^2} \cdot L_{PIoU}. \tag{21}$$

In summary, Powerful-IoU loss function addresses the core issues of current IoU-based loss functions by introducing a size-dependent penalty and a function for adjusting the gradient. This results in faster convergence and improved detection accuracy. The non-monotonic focusing mechanism further strengthens the model's ability to manage anchor boxes with different levels of quality. Powerful IoU exhibits significant advantages in improving model performance, particularly in complex scenarios and in accelerating convergence. The subsequent experiments in this paper will evaluate its practical effectiveness in the task of dental pathology segmentation.

4. Real-Time Panoramic Radiograph Dental Pathology Detection and Segmentation Design

In this section, we outline the complete workflow for detecting and segmenting dental pathologies in panoramic radiographs in real time, along with the model training steps; please refer to Algorithms A1 and A2 in Appendix A. Subsequently, the modeling algorithms are discussed after the experiments.

4.1. A Real-Time Training Algorithm for Dental Anomalies and Diseases

As can be seen from Algorithm A1, the overall workflow begins by initializing the model parameters and setting hyperparameters, such as the number of training epochs, batch size, learning rate, and weight decay. Next, data augmentation is applied to the panoramic radiograph image dataset to increase data variety and improve the model's generalization capability. During each training epoch, the model selects a batch of images from the dataset along with their corresponding ground truth labels for the training process. If model performance does not improve over 20 consecutive epochs, an early stopping strategy can be triggered to terminate training prematurely. The training process

persists until either the maximum number of epochs is achieved or the early stopping condition is triggered. After the training is complete, the model outputs the detection and segmentation results.

As is evident from Algorithm A2, in the specific model training process, the first step is forward propagation. The model uses lightweight partial convolution (PConv) to replace standard convolution for feature extraction, reducing computational complexity. Subsequently, the BiFPN structure is applied to fuse features from different layers, ensuring an effective combination of deep semantic features with shallow detail features. The EMCA attention mechanism is also introduced to strengthen pathological regions' feature representation through weighted operations. After passing through these modules, the segmentation head is used to generate the prediction results. Next, the loss function is computed, which includes localization loss, classification loss, and segmentation mask loss. Localization loss uses the Powerful-IOU function to measure the accuracy of bounding box localization. Categorization loss is calculated using binary cross-entropy, which assesses accuracy of class predictions, while segmentation mask loss assesses accuracy of mask predictions. These losses are then weighted and summed to obtain the total loss, with an additional L2 regularization term added to prevent overfitting. Gradients are computed through backpropagation, and model parameters are adjusted using SGD optimizer, progressively improving performance in pathology detection and segmentation tasks of the model.

4.2. Discussion

This research proposes an optimized lightweight model, DentSeg-YOLO, developed specifically for accurate detection and segmentation of dental pathologies in panoramic radiographs. This model provides a technical reference for intelligent healthcare applications in dental pathology detection and segmentation, addressing the issues of low efficiency and accuracy in the traditional methods.

Limitations in accuracy, detection efficiency, and lightweight design have been reported in prior research. In response, this study developed a C2f-Faster lightweight structure to refine backbone and boost the attention paid to pathological areas. In order to improve attention regarding pathological targets, the proposed EMCA attention mechanism was integrated. Additionally, the BiFPN structure and PowerfulIoU loss function were employed to boost the accuracy of detecting and segmenting objects. From the subsequent experimental results, with these improvements, the DentSeg-YOLO model surpasses other detection and segmentation methods, enhancing the accuracy in dental pathology identification and segmentation without compromising its lightweight structure.

5. Personalized Digital Dental Disease Diagnosis and Detection

5.1. Source of Experimental Datasets

This study combines dental panoramic radiograph data from two sources, with strict adherence to medical ethics and data privacy regulations. The first portion of the data comes from the Stomatology Department of Jingzhou First People's Hospital, covering 1645 panoramic radiographs collected between 2022 and 2023. These images underwent rigorous de-identification processing to protect patient privacy while retaining their diagnostic value. The use of these clinical data was approved by the Institutional Ethics Review Board of the School of Computer Science, Yangtze University, and written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

The second portion of the data is sourced from the publicly available DENTEX dataset [50], which includes 1075 images. The integration of this dataset was designed to

enhance diversity, thereby improving dataset quality and model generalization. All data usage complies with the terms and conditions specified by the DENTEX dataset providers.

To ensure the dataset's quality and clinical relevance, senior specialists from the Stomatology Department of Jingzhou First People's Hospital meticulously reviewed and annotated the collected images. Ultimately, 2720 high-quality images were selected for the experiment. These images were divided into three subsets: 80% allocated for training, 10% for validation, and 10% for testing, corresponding to 2176 images in the training set and 272 in each of the validation and test sets. The training set facilitated model training, the validation set was utilized for hyperparameter optimization and initial evaluation, and the test set was used to measure the model's detection and segmentation accuracy, along with its generalization capability.

For disease classification, three categories were determined based on the clinical complexity and treatment approaches for different dental conditions: caries, deep caries, and impacted teeth. Superficial and moderate caries were grouped into a single category due to their relatively simple treatment requirements, whereas deep caries were categorized separately as they often involve pulp pathology and require complex multi-step treatments such as root canal therapy and tooth restoration.

This dataset and classification methodology establish a scientific foundation for subsequent model training and application, ensuring the reliability of the experiment and the medical relevance of the results. All procedures involving human data were conducted in compliance with ethical standards, and patient confidentiality was rigorously maintained throughout the study.

5.2. Image Pre-Processing

In this paper, a comprehensive data augmentation strategy is employed for lesion detection and segmentation in panoramic dental radiographs, with special consideration regarding the unique characteristics and varied imaging conditions of such radiographs. To more accurately simulate the clinical image acquisition environment, two methods were selected: brightness enhancement and linear contrast adjustment. These techniques were intentionally selected, while avoiding augmentation strategies like rotation or vertical flipping that could potentially distort critical positional information of the teeth.

Due to variations in real-world clinical imaging conditions, the brightness enhancement method adjusts image brightness to simulate different lighting conditions. Brightness enhancement is achieved by adjusting the pixel values through an additive factor. For each pixel $I(x, y)$, the brightness-adjusted pixel value $I'(x, y)$ is calculated using the following:

$$I'(x, y) = I(x, y) + f, \tag{22}$$

where f is a randomly chosen additive factor that controls the increase or decrease in brightness. If $f > 0$, the image becomes brighter; if $f < 0$, the image becomes darker. To simulate real-world conditions, a random value of f is selected within the range of -10 to 10 after multiple trials and fine-tuning.

Linear contrast adjustment modifies the linear threshold of the image to adjust the contrast, further enhancing image details. The expression for this process is provided by

$$I'(x, y) = 127 + \alpha \times (I(x, y) - 127), \tag{23}$$

where α is a randomly sampled adjustment factor that determines the degree of contrast adjustment. After multiple trials to simulate real-world conditions, the final sampling range for α is set between [0.4, 1.6]. In this formula, 127 is a constant representing the midpoint of the grayscale range used to maintain the central brightness of the image.

By combining these two enhancement methods, the image data can better simulate the potential variations encountered in real-world imaging scenarios. Brightness enhancement simulates the effects of different lighting conditions, while linear contrast adjustment optimizes contrast, especially in improving the visibility of details in both darker and brighter regions. This hybrid augmentation approach not only increases the dataset's diversity and reduces the risk of model overfitting but also enhances the ability to capture important information of the model in the images. By simulating real-world image conditions, this comprehensive data augmentation strategy helps to train a more robust detection model that is better suited to adapt to various environmental changes, ultimately leading to improved performance and accuracy for lesion detection and segmentation in panoramic dental radiographs. The images shown in Figure 9 are those before and after data augmentation.

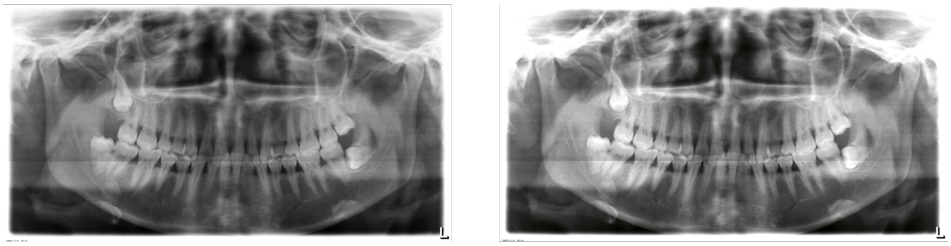

Figure 9. The images before and after data augmentation.

5.3. Environment Configuration

The configuration for the training environment in this research is as follows: Professional Windows 11 as the operating system, version 23H2, with a GeForce RTX 4060 graphics card (8GB VRAM, manufactured by NVIDIA Corporation, Santa Clara, CA, USA), an AMD Ryzen 7 6800H processor (manufactured by Advanced Micro Devices, Inc., Santa Clara, CA, USA), and 16 GB of memory(manufactured by Samsung Electronics Co., Ltd., Suwon, Republic of Korea). PyTorch framework, CUDA 12.1.

To maximize the deep learning model's performance and ensure efficiency during the training process, while also taking into consideration the detailed information in panoramic dental radiographs, such as key features like dental caries, the input image resolution for all experiments was set to 1280 $\times$ 1280. This decision was made based on hardware limitations and the need to preserve critical details in the original images.

In terms of experimental design, the dataset was augmented by randomly adjusting brightness and contrast of training images, effectively doubling size of dataset before training process began. Consequently, data augmentation techniques related to brightness and contrast in the YOLOv8 framework were disabled during training to avoid issues such as over-brightening or over-darkening the pre-processed images, ensuring the quality of the training data. Moreover, to prevent distortion of spatial and structural information in the images, geometric transformation techniques, such as rotation and flipping, were also disabled.

Ensuring the reliability and stability of the results, all hyperparameters for the improved YOLOv8 model were meticulously configured. Included in this configuration are a 0.01 starting learning rate, a batch size of 4, a momentum factor of 0.937, and 100 training epochs. The hyperparameter settings were determined through multiple rounds of tuning and training on the baseline YOLOv8 model. Analysis of the baseline model showed that optimal weights often emerged around the 50th epoch. Based on this observation, an early stopping parameter was set to 20 epochs in this study. This early stopping strategy helps to minimize the consumption of computational resources while effectively preventing overfitting.

5.4. Evaluation Indicators

This study utilizes several metrics to determine how well the model performs in detection and segmentation tasks, thereby enabling a comprehensive evaluation of its advancements as follows:

(1) Precision:

Precision measures the proportion of all samples identified as positive by the model that are actually positive. A higher accuracy rate indicates that the model makes fewer false predictions:

$$P = \frac{TP}{TP + FP}, \tag{24}$$

where TP denotes the number of true positives (samples correctly classified as positive), and FP indicates the number of false positives (samples mistakenly identified as positive).

(2) Recall:

Recall represents the proportion of actual positive samples among those identified as positive by the model:

$$R = \frac{TP}{TP + FN}, \tag{25}$$

where FN is the number of false negatives (samples mistakenly identified as negative).

(3) Mean Average Precision (mAP):

Mean Average Precision (mAP) represents the Mean Average Precision (AP) values computed at various recall thresholds and averaged across all classes. It serves as a crucial metric for assessing the comprehensive effectiveness of a model, particularly in multi-class object detection and segmentation tasks.

For a single category, AP is obtained by calculating the area beneath the precision–recall curve, which reflects model's precision and recall performance at various threshold settings. AP is computed using numerical integration over the precision–recall curve:

$$AP = \int_0^1 P(R)\, dR. \tag{26}$$

In this equation, $P(R)$ represents the precision as a function of recall R, and the integral calculates the area under the curve, summarizing how the model performs across different levels of recall. mAP considers the precision and recall across various thresholds, comprehensively evaluating the model's stability and accuracy under different conditions. A higher mAP indicates that the algorithm performs more accurately and consistently under varying detection scenarios:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i. \tag{27}$$

(4) Parameters:

This indicates the total count of parameters within the model, which directly influences the storage requirements and computational complexity. A larger number of parameters generally indicates a more expressive model, but it also increases the computational load.

(5) Giga Floating Point Operations per Second (GFLOPs):

The computational load $(GFLOPs)$ refers to the quantity of floating-point computations executed per second, measured in billions of operations. A high $GFLOP$ value suggests that more computational capacity is needed for execution, which can be beneficial in tasks that require high precision and sophisticated computations as complex models often lead to better accuracy or performance. However, in simpler scenarios, a lower

GFLOP value is preferable, indicating a lighter model with lower computational resource requirements and faster execution.

(6) The frame rate (FPS):

FPS measures how many image frames the system can handle each second, serving as a metric for the algorithm's processing speed:

$$FPS = \frac{\text{Framenum}}{\text{ElapsedTime}}. \tag{28}$$

(7) Model Size:

This indicates the memory or storage footprint of the model, typically measured as MB or GB. A smaller model size facilitates faster loading and execution, which is especially beneficial in environments with limited resources.

6. Numerical Results

The YOLOv8-seg family includes five pre-trained models—n, s, m, l, and x. Among them, YOLOv8n-seg is the smallest, offering the fastest inference, whereas YOLOv8x-seg, being the largest, achieves the highest accuracy; however, the slower inference speed of the system limits its effectiveness in applications involving real-time detection and segmentation. Considering the deployment constraints on edge devices and the limitations on model parameters, the subsequent experiments and analysis focus on YOLOv8n-seg as it achieves a good compromise between speed and performance, meeting the needs of real-time tasks on resource-limited devices. Table 3 illustrates the comparison of YOLOv8-Seg models at different scales.

Table 3. Comparison between YOLOv8-seg models of varying scales.

Model	Depth	Width	Parameters (M)
YOLOv8n-seg	0.33	0.25	3.26
YOLOv8s-seg	0.33	0.50	11.79
YOLOv8m-seg	0.67	0.75	25.89
YOLOv8l-seg	1.00	1.00	42.90
YOLOv8x-seg	1.00	1.25	67.0

6.1. Evaluation Setup

We compared the YOLO-DentSeg model to the YOLOv8n-seg model across several evaluation indicators, including parameters, computational demands, detection speed, model size, mAP50 (Box), and mAP50 (Seg). As shown in Table 4, the YOLO-DentSeg model achieves a 44.3% reduction in parameters and lowers the computational load by 17.5%, improves the detection speed to 90.3 FPS, and reduces the model size by 41.5%. Additionally, mAP50 (Box) increased by 1.8 percentage points, and mAP50 (Seg) improved by 2 percentage points. Therefore, YOLO-DentSeg outperforms the original YOLOv8n-seg model in all aspects, achieving higher efficiency and accuracy while being more lightweight.

This work also compares the detection and segmentation accuracy of the YOLOv8n-seg model and YOLO-DentSeg model. As depicted in Figure 10, regarding segmentation accuracy, the average segmentation accuracy for deep caries showed the largest improvement, with an increase of 3.4 percentage points; the caries segmentation accuracy improved by 1.7 percentage points; and the impacted teeth segmentation accuracy increased by 0.9 percentage points. Regarding detection accuracy, deep caries achieved the largest improvement, with an increase of 4.3 percentage points, while impacted teeth detection accuracy increased by 0.4 percentage points, and caries detection accuracy improved by

0.8 percentage points. Overall, the improved model demonstrates higher accuracy in identification, with enhanced detection and segmentation precision.

Table 4. Comparison of evaluation indicators before and after model improvement.

Model	Parameters (M)	GFLOPs (G)	FPS	Size (MB)	mAP50 (Box)	mAP50 (Seg)
YOLOv8n-seg	3.26	12.0	90.1	6.5	0.854	0.835
YOLO-DentSeg	1.81	9.9	90.3	3.8	0.872	0.855

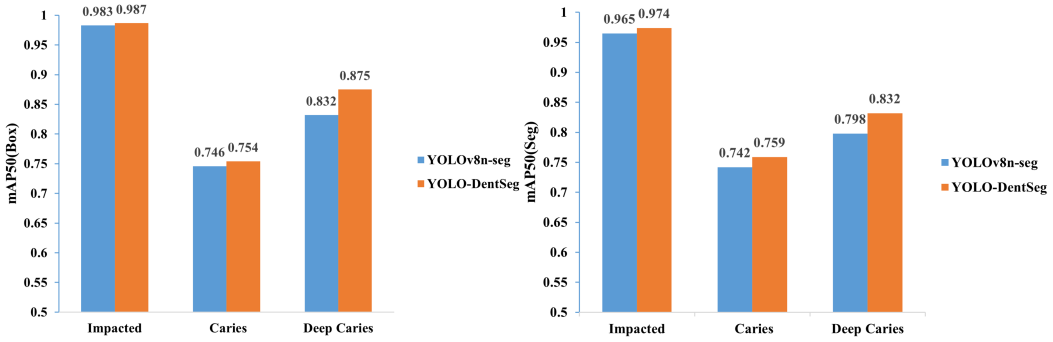

Figure 10. Comparison of detection and segmentation accuracy averages prior to and following model enhancement.

6.2. Ablation Experiments

This section presents an ablation experiment using the panoramic radiograph dataset to analyze the impact of each enhancement strategy—C2f-Faster, BiFPN, EMCA, and PowerfulIOU—on the performance of the YOLOv8-seg model. All the experiments were conducted under consistent experimental conditions as outlined in Section 5.3 to ensure a fair and impartial analysis. The model setups are outlined as follows: Model 1 replaces the original C2f structure in the backbone network with C2f-Faster; Model 2 builds upon Model 1 by using the BiFPN architecture to optimize the neck network; Model 3 adds the proposed EMCA attention module on top of Model 2; and, finally, YOLO-DentSeg (our model) incorporates the PowerfulIOU loss function into Model 3, resulting in the complete YOLO-DentSeg model.

As shown in Table 5, Model 1, based on YOLOv8-seg with the C2f-Faster module in the backbone, shows improvement across multiple metrics. The Mean Average Precision (mAP) for object detection, mAP (Box), increases by 0.7 percentage points, while the instance segmentation mAP, mAP (Seg), improves by 0.9 percentage points, and the detection speed rises by 0.2 FPS. Additionally, the parameter count, computational load, and model size decrease by 21.8%, 15%, and 20%, respectively. Model 2 is based on Model 1, in which the mAP50 (Box) and mAP50 (Seg) continue to increase by 0.7 and 0.1 percentage points after using the BiFPN module in the neck network, although the frame rate decreases slightly. The parameter count, computational load, and model size further decrease by 29.1%, 3.0%, and 26.9%, respectively. Model 3, which incorporates the EMCA attention mechanism into Model 2, achieves further increases in mAP, with mAP50 (Box) improved by 0.2 percentage points, while mAP50 (Seg) demonstrated an improvement of 0.3 percentage points, although there is no significant reduction in parameter count, computational load, or model size. YOLO-DentSeg, the final model based on Model 3 with the PowerfulIOU loss function, shows additional gains, with mAP50 (Box) improved by another 0.2 percentage points, while mAP50 (Seg) demonstrated an improvement of another 0.7 percentage points, and a notable improvement in detection and segmentation speed.

In the end, our proposed YOLO-DentSeg model exhibits superior performance in detection and segmentation of dental pathologies, achieving an average precision for disease

detection (mAP (Box)) of 87.2% and an average precision for disease segmentation (mAP (Seg)) of 85.5%. In addition, the model contains just 1.81 M parameters, the computational load is 9.9 G, the model size is 3.8 MB, and the inference speed reaches 90.3 FPS, which is very suitable for practical application scenarios with low resource consumption. These results indicate that the improvement approach we implemented is effective. Figure 11 presents the experimental curves for ablation experiments (note: here, mAP50 (Box) refers to the Mean Average Precision of disease recognition, while mAP50 (Seg) represents the Mean Average Precision of segmentation masks in disease segmentation. In this context, A denotes the introduction of C2f-Faster, B represents the integration of BiFPN, C indicates the incorporation of EMCA, and D signifies the use of Powerful-IoU).

Table 5. Ablation experiments.

Model	Improvement Strategies				mAP50 (Box)	mAP50 (Seg)	Parameters (M)	GFLOPs (G)	Size (MB)	FPS
	A	B	C	D						
YOLOv8-seg					0.854	0.835	3.26	12.0	6.5	89.4
Model 1	✓				0.861	0.844	2.55	10.2	5.2	89.6
Model 2	✓	✓			0.868	0.845	1.81	9.9	3.8	89.1
Model 3	✓	✓	✓		0.870	0.848	1.81	9.9	3.8	90.1
Ours	✓	✓	✓	✓	0.872	0.855	1.81	9.9	3.8	90.3

To conclude, the YOLO-DentSegNet model, with its lightweight design, demonstrates superior performance compared to the YOLOv8-seg model, excelling in both efficiency and accuracy for detection and segmentation. The model demonstrates advancements in several key metrics: mAP50 (Box), mAP50 (Seg), and FPS have increased by 1.8 percentage points, 2.0 percentage points, and 0.9 FPS, respectively. In addition, the model's computational complexity (measured in GFLOPs), parameters, and model size have decreased by 2.1 G, 1.45 M, and 2.7 MB, respectively, fully demonstrating its efficiency and effectiveness in real-world applications.

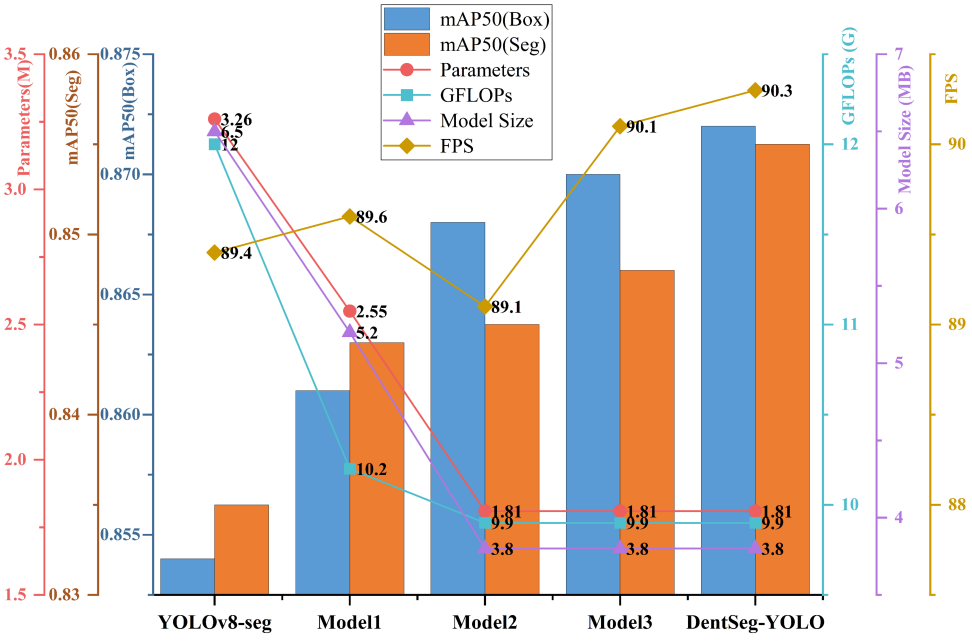

Figure 11. Experimental curves for ablation experiments.

6.3. Comparative Experiments with Different Attention Mechanisms

The introduction of attention mechanisms could greatly increase both the accuracy and efficiency of detection and segmentation. To test the efficacy of the EMCA attention mechanism, we embedded different attention mechanisms at the same location to examine their effects on the model performance. Under the premise of guaranteeing that other parameters remain unchanged, we respectively added the current mainstream attention mechanisms after the SPPF layer of Model 2, including CBAM [51], CA [52], SimAM [53], Triplet [54], ECA, and our proposed EMCA.

According to the results in Table 6, embedding the CBAM, Triplet, or ECA attention mechanisms leads to varying degrees of decline in both mAP50 (Box) and mAP50 (Seg) values. In contrast, embedding the CA, SimAM, or EMCA attention mechanisms results in minor fluctuations in mAP50 (Box) values, while mAP50 (Seg) values improve to varying extents, yielding an overall positive effect on the system's performance. Among these, EMCA achieves the most notable enhancement, mAP50 (Box) improved by 0.2 percentage points, and mAP50 (Seg) demonstrated an improvement of 0.3 percentage points while also boosting FPS by 1.2. This indicates that EMCA is the most suitable attention mechanism for integration into this model architecture. Figure 12 shows the experimental curves for the models with different attention mechanisms added.

Table 6. Evaluation of different attention modules.

Model	mAP50 (Box)	mAP50 (Seg)	Parameters (M)	GFLOPs (G)	Size (MB)	FPS
Model 2	0.868	0.845	1.81	9.9	3.8	89.1
Model 2+CBAM	0.863	0.842	1.88	9.9	3.9	84.3
Model 2+CA	0.868	0.846	1.81	9.9	3.8	89.2
Model 2+SimAM	0.867	0.847	1.81	9.9	3.8	88.1
Model 2+Triplet	0.867	0.842	1.81	9.9	3.8	88.3
Model 2+ECA	0.863	0.840	1.81	9.9	4.0	89.4
Model 2+EMCA	0.870	0.848	1.81	9.9	4.0	90.3

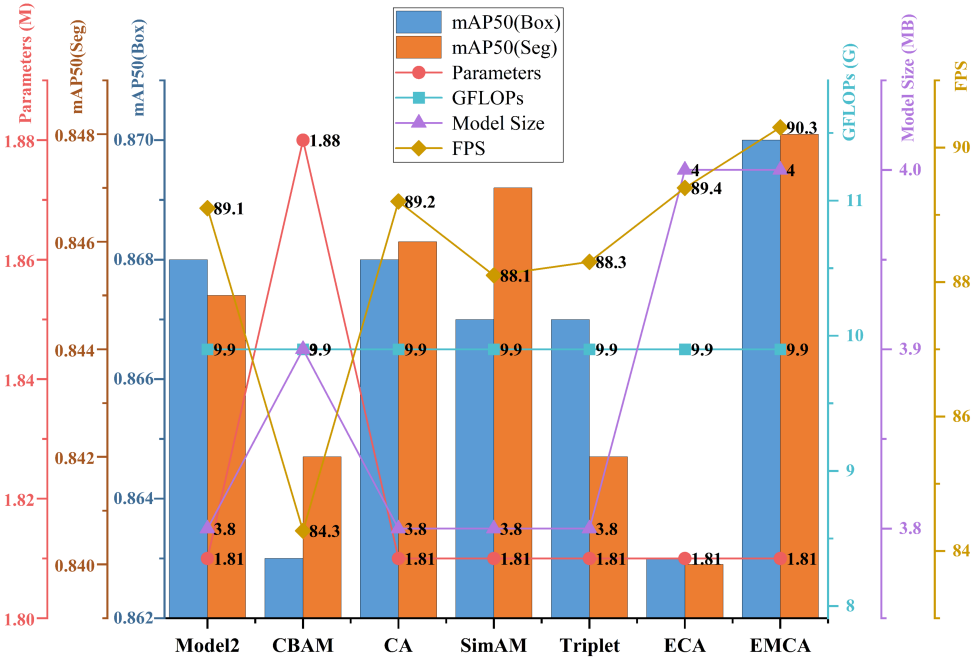

Figure 12. Adding experimental curves for different attention modules.

6.4. Comparative Experiments with Various IoU Loss Functions

To assess the impact of the Powerful-IoU loss function on enhancing detection and segmentation accuracy, a comparative experiment was conducted with CIoU, DIoU, SIoU, EIoU, WIoU, and MPDIoU loss functions. Table 6 displays the results.

According to Table 7, the model achieves the highest levels of average detection precision, average segmentation precision, and processing speed with the PowerfulIoU loss function. In comparison to the conventional CIoU for bounding box detection, PowerfulIoU improves the average detection precision, average segmentation precision, and processing speed by 0.2%, 0.7%, and 0.2 FPS, respectively. These improvements demonstrate that PowerfulIoU significantly enhances model performance, enabling fast and accurate detection and segmentation of complex dental pathology targets. Figure 13 presents the experimental curves obtained with various loss functions.

Table 7. Comparison of results obtained with various loss functions.

Model	mAP50 (Box)	mAP50 (Seg)	FPS
Model 3+CIOU	0.870	0.848	90.1
Model 3+DIOU	0.870	0.852	90.2
Model 3+SIOU	0.865	0.849	88.2
Model 3+EIOU	0.868	0.845	90.2
Model 3+WIOU	0.865	0.843	90.2
Model 3+MPDIOU	0.869	0.845	90.1
Model 3+Powerful-IOU	0.872	0.855	90.3

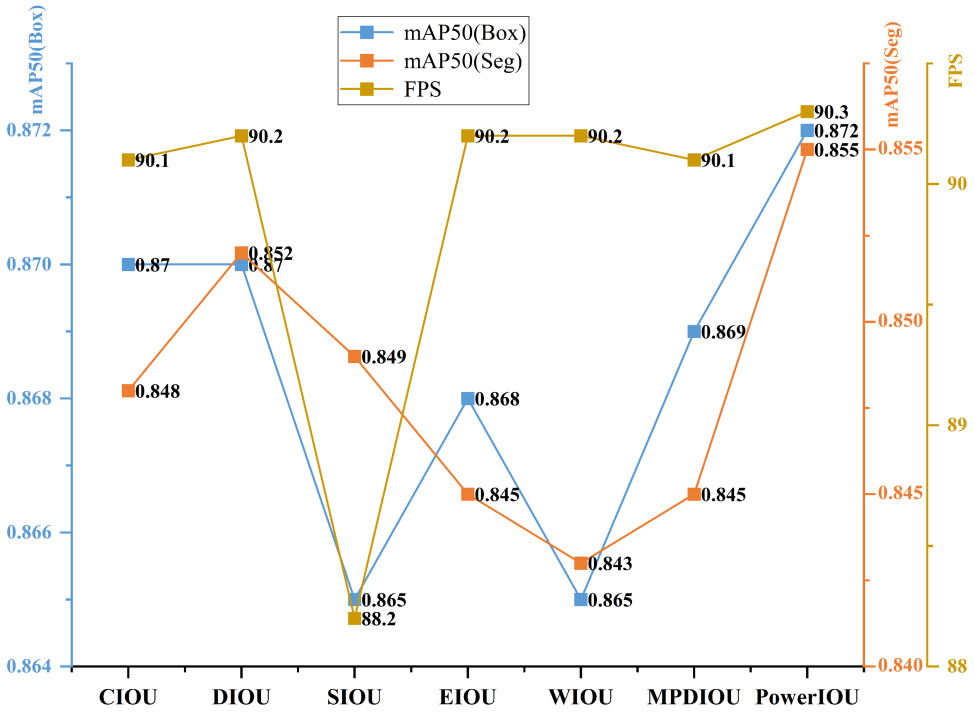

Figure 13. Experimental curves with various employed loss functions.

6.5. Comparative Experiments with Various Models

To comprehensively evaluate the performance of YOLO-DentSeg in detecting and segmenting oral dental diseases, this study conducted comparative experiments with various mainstream detection and segmentation models, including the YOLO series (YOLOv5-seg, YOLOv6-seg, YOLOv7-seg, YOLOv8-seg, YOLOv9-seg, and YOLOv11-seg), RT-DETR, Mask R-CNN, SOLOv2, and YOLOACT. As shown in Table 8, YOLO-DentSeg

demonstrates significant advantages in detection accuracy, computational efficiency, and model lightweighting.

Table 8. Comparative experiments with different models.

Model	mAP50 (Box)	mAP50 (Seg)	Params (M)	GFLOPs (G)	FPS	Size (MB)
Mask R-CNN	0.848	0.845	63.4	235.0	20.6	122.4
SOLOv2	-	0.812	46.6	179.0	24.9	89.4
YOLOACT	0.743	0.730	35.3	68.7	54.2	43.9
YOLOv5-seg	0.853	0.830	2.8	11.0	95.3	5.9
YOLOv6-seg	0.847	0.826	4.4	15.2	98.6	9.1
YOLOv7-seg	0.844	0.824	37.9	141.9	30.9	76.2
YOLOv8-seg	0.861	0.835	3.3	12.0	89.4	6.9
YOLOv9-seg	0.861	0.839	57.5	368.6	30.91	116.6
YOLOv11-seg	0.859	0.84	2.8	13.2	89.9	6.1
RT-DETR	0.864	0.843	27.3	56.8	10.9	56.8
Ours	0.870	0.855	1.8	9.9	90.3	3.8

In terms of detection and segmentation accuracy, YOLO-DentSeg achieves the highest mAP50 (Box) and mAP50 (Seg) scores of 0.870 and 0.855, respectively, among all the models. Compared to Mask R-CNN (0.848 for detection and 0.845 for segmentation), YOLO-DentSeg improves detection accuracy by 2.6% and segmentation accuracy by 1.2%. When compared to SOLOv2 (0.812 for segmentation) and YOLOACT (0.743 for detection and 0.730 for segmentation), the advantages in detection and segmentation accuracy are even more pronounced, with improvements of up to 17.1% and 17.1%, respectively. Even when compared to the latest YOLO series models, YOLOv8-seg (0.861/0.835) and YOLOv11-seg (0.859/0.840), YOLO-DentSeg still enhances detection accuracy by 1.04% and 1.28% and segmentation accuracy by 2.4% and 3.57%, respectively. This indicates its superior capability in locating lesions and delineating edges in complex oral tomographic images.

In terms of model lightweighting, YOLO-DentSeg has only 1.8 M parameters, which is 97.1% fewer than Mask R-CNN (63.4 M), 96.1% fewer than SOLOv2 (46.6 M), and 94.9% fewer than YOLOACT (35.3M). It even has fewer parameters than the lightweight designs of YOLOv5-seg (2.8 M) and YOLOv11-seg (2.8 M). The model size is only 3.8 MB, which is 96.9% smaller than Mask R-CNN (122.4 MB) and 95.7% smaller than SOLOv2 (89.4 MB), making it highly feasible for deployment on medical edge devices.

In terms of computational efficiency, YOLO-DentSeg requires only 9.9 GFLOPs, which is 95.8% less than Mask R-CNN (235.0 GFLOPs), 94.5% less than SOLOv2 (179.0 GFLOPs), and 85.6% less than YOLOACT (68.7 GFLOPs). Its computational load is even lower than that of YOLOv5-seg (11.0 GFLOPs) and YOLOv11-seg (13.2 GFLOPs), validating the effectiveness of its algorithm design in eliminating computational redundancy.

In terms of real-time performance, YOLO-DentSeg achieves an inference speed of 90.3 FPS, which is 338% faster than Mask R-CNN (20.6 FPS), 263% faster than SOLOv2 (24.9 FPS), and 66.6% faster than YOLOACT (54.2 FPS). Although slightly lower than YOLOv5-seg (95.3 FPS) and YOLOv6-seg (98.6 FPS), it still shows significant improvements of 192% and 192% over YOLOv7-seg (30.9 FPS) and YOLOv9-seg (30.91 FPS), respectively, fully meeting the requirements for real-time diagnosis.

In summary, YOLO-DentSeg, through lightweight architecture design and computational optimization, maintains high accuracy while reducing the number of parameters and computational load to less than 5% of traditional models, with inference speeds meeting clinical real-time standards. Compared to two-stage models like Mask R-CNN, it achieves breakthroughs in both accuracy and efficiency; compared to other single-stage models like SOLOv2 and YOLO, it demonstrates a better balance between accuracy and speed. These characteristics make YOLO-DentSeg highly suitable for applications in portable oral

diagnostic devices and low-power embedded systems. Figure 14 presents the experimental results in the form of scatterplots for various models.

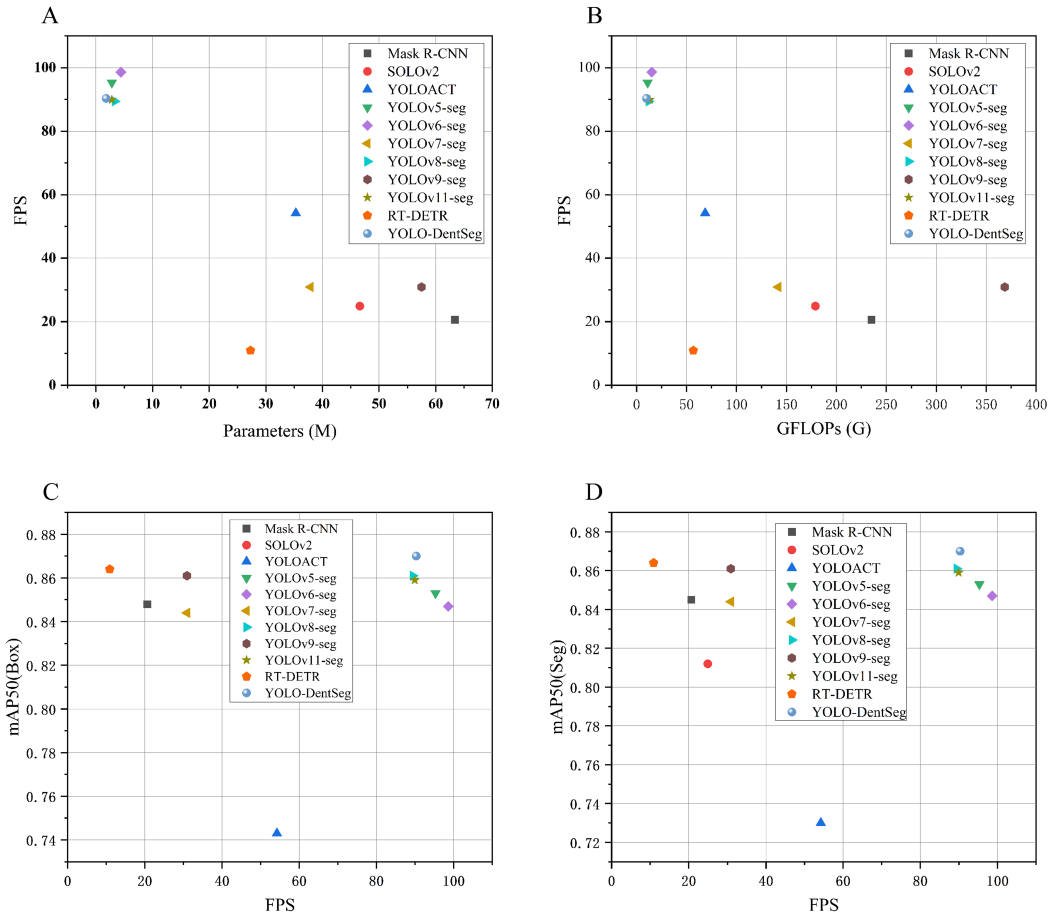

Figure 14. Scatterplots of different model experiments. (**A**) The relationship between the number of parameters and FPS (Frames Per Second) for each model; (**B**) The relationship between computational complexity (FLOPs) and FPS for each model; (**C**) The relationship between FPS and mAP50 (Box) for each model; (**D**) The relationship between FPS and mAP50 (Seg) for each model.

6.6. Visualization Results

Figure 15 displays the inference results from different models, demonstrating that the YOLO-DentSeg model outperforms the other models in detecting and segmenting occluded and multiple disease targets. In the figure, red bounding boxes represent detected impacted teeth, yellow bounding boxes represent detected caries, and blue bounding boxes represent detected deep caries. In contrast, YOLOv5, YOLOv6, and YOLOv8 exhibit significant issues in multi-target disease detection and segmentation, such as missed detections and false positives. Although YOLOv7 and YOLOv9 have lower rates of missed detections, they suffer from repeated detections and segmentations, potentially impacting the overall accuracy and increasing the processing time. The latest models, RT-DETR and YOLOv11, show relatively good performance in detection and segmentation but still experience issues with false positives, making them suboptimal choices. Additionally, the segmentation masks generated by these models generally display jagged edges and irregular shapes, further reducing their practicality.

In summary, the YOLO-DentSeg model is better suited to address common challenges in dental disease detection and segmentation from panoramic tomographic images, such as occlusions between disease targets and severe overlap among different pathological

regions. This model significantly improves the accuracy of identifying, localizing, and segmenting dental diseases, offering greater potential for practical applications.

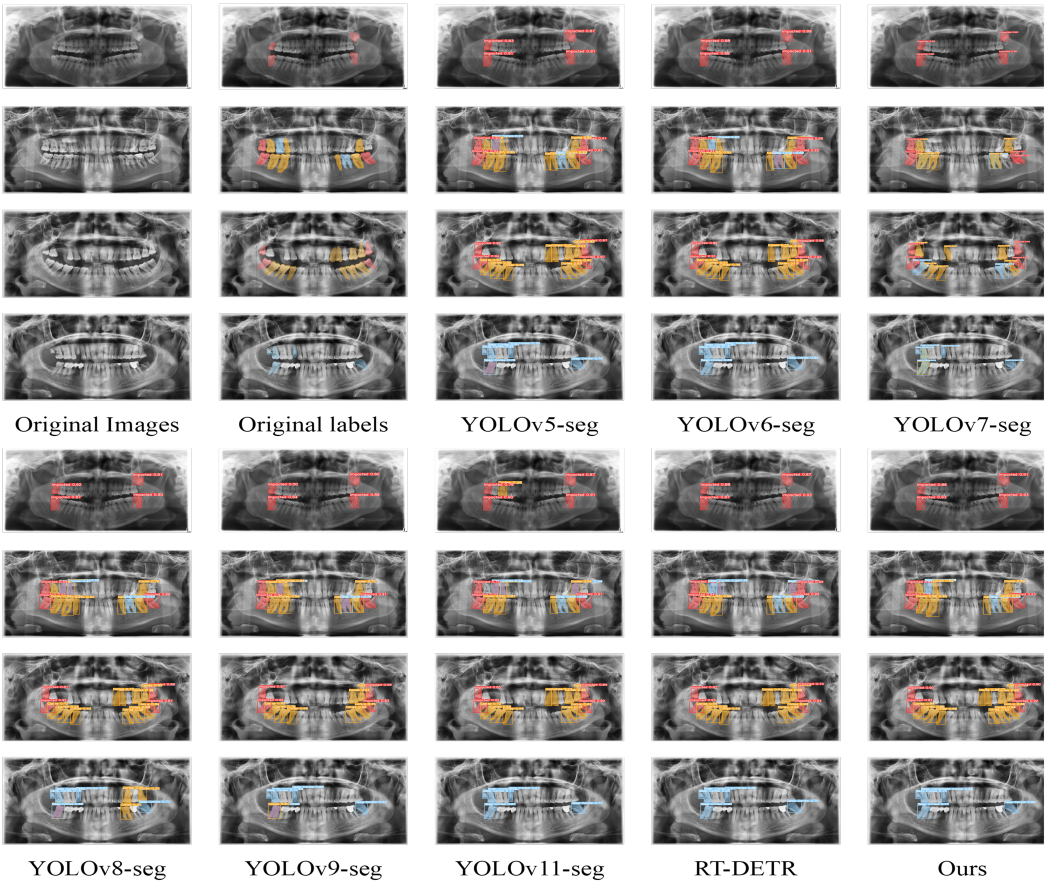

Figure 15. Detection segmentation results for different models.

7. Concluding Remarks

In response to the issues of low detection and segmentation accuracy, slow speed, insufficient feature extraction capabilities, and the inability to run efficiently on devices with limited resources regarding the current methods for oral panoramic tomographic disease detection and segmentation, a lightweight segmentation model for dental diseases, YOLO-DentSeg, was proposed. This model stands out due to its compact structure and high level of accuracy. The model successfully overcame challenges such as poor detection and segmentation accuracy for occluded disease targets and slow speeds, offering valuable insights for the development of visual systems in intelligent medical robotics. The experimental conclusions are as follows: first, the use of the C2f-Faster structure effectively minimized the parameters and reduced the computational load, resulting in a more lightweight model. Then, with the use of the BiFPN architecture and the EMCA attention mechanism, the model strengthened its feature fusion and focus on significant features, leading to more rapid and accurate detection and segmentation. Finally, the Powerful-IOU loss function further enhanced the model's detection and segmentation precision. To assess the performance advantages offered by the YOLO-DentSeg model, it was tested against mainstream models such as Mask R-CNN, SOLOv2,YOLOACT, YOLOv5, YOLOv6, YOLOv7, YOLOv8, YOLOv9, YOLOv11, and RT-DETR. The results reveal that YOLO-DentSeg outperformed all the other models. In conclusion, the YOLO-DentSeg model achieved a lightweight design while improving detection and segmentation accuracy. It effectively met the demands of real-time performance and precision, demonstrating

significant application value and providing a useful reference for disease detection and segmentation in oral panoramic tomographic images.

While the proposed YOLO-DentSeg model has demonstrated excellent performance in the detection and segmentation of oral diseases, we recognize that there is still room for improvement in terms of dataset diversity and model robustness. Future work will focus on several key areas to further enhance the model's generalization ability and clinical applicability.

Firstly, we plan to collaborate with multiple medical institutions to construct a multi-center dataset that encompasses a wider range of imaging devices (e.g., cone–beam CT and multi-detector CT) and imaging conditions (e.g., varying resolutions, contrast levels, and noise levels). This approach, similar to the work by Park et al. [55] in multi-center CT image segmentation, will provide a more comprehensive evaluation of the model's performance across different clinical environments. By incorporating data from diverse sources, we aim to better validate the model's ability to generalize to real-world scenarios.

Secondly, we intend to expand the dataset to include a broader spectrum of oral pathologies beyond caries and impacted teeth, such as periodontal disease, jaw cysts, and odontogenic tumors. This will enable us to assess the model's performance under more complex and varied pathological conditions, thereby increasing its practical utility in clinical settings.

To address the variability in imaging conditions across different medical centers, we will explore cross-center transfer learning techniques. By pre-training the model on data from one center and fine-tuning it on data from other centers, we can improve the model's adaptability to new data domains. This approach will help to mitigate the issue of data distribution mismatch and enhance the model's generalization capabilities.

Additionally, we will investigate more advanced data augmentation techniques, such as synthetic data generation using generative adversarial networks (GANs). By generating synthetic data with diverse imaging characteristics, we can further improve the model's robustness without relying solely on large amounts of real-world data.

Finally, we will continue to optimize the computational efficiency and lightweight design of the model to make it more suitable for deployment on resource-constrained edge devices, such as portable medical equipment. We will also explore more efficient attention mechanisms and feature fusion strategies to further enhance the model's detection and segmentation accuracy while maintaining real-time performance.

Through these efforts, we aim to ensure that the YOLO-DentSeg model can be effectively applied in a wider range of clinical scenarios, providing reliable support for the intelligent diagnosis and treatment planning of oral diseases.

Author Contributions: Conceptualization, Y.H. and R.C.; methodology, Y.H. and H.Q.; software, Y.H.; validation, H.Q. and R.C.; formal analysis, Y.H.; investigation, Y.H. and R.C.; writing—original draft preparation, Y.H.; writing—review and editing, H.Q.; visualization, Y.H.; supervision, H.Q. and R.C.; funding acquisition, H.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hubei Natural Science and Research Project under Grant 2020418, Hubei Provincial Natural Science Foundation (2024AFB851), 2021 Light of Taihu Science and Technology Project, and 2022 Wuxi Science and Technology Innovation and Entrepreneurship Program.

Institutional Review Board Statement: The study was conducted following the Declaration of Helsinki and approved by the Ethics Committee of the School of Computer Science, Yangtze University (117/10.10.2022).

Informed Consent Statement: Informed consent was obtained from all participants involved in this study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, H. Qin, upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

C2f	Channel to Feature Map
BiFPN	Bidirectional Feature Pyramid Network
EMCA	Enhanced Efficient Multi-Channel Attention
IOU	Intersection over Union
mAP	mean Average Precision
FPS	Frames Per Second
WHO	World Health Organization
CNNs	Convolutional Neural Networks
GPUs	Graphics Processing Units
ROI	Region of Interest
RPN	Region Proposal Network
YOLO	You Only Look Once
CSPDarknet	Cross-Stage Partial Darknet
ELAN	Efficient Layer Aggregation Network
PAN	Path Aggregation Network
FPN	Feature Pyramid Network
YOLOACT	You Only Look At CoefficienTs
PConv	Partial Convolution
CBS	Convolution Layers, Batch Normalization Layers, and Activation Functions
Avgpool	Global Average Pooling
MaxPool	Global Max Pooling
CIOU	Complete Intersection over Union
GFLOPs	Giga Floating-Point Operations Per Second

Appendix A

Algorithm A1 Workflow of Real-Time Dental Detection and Segmentation in Panoramic Radiographs

Input: Oral panoramic radiograph dataset $D = \{x_1, x_2, \cdots, x_n\}$, corresponding oral dental disease labels $Y = \{y_1, y_2, \cdots, y_n\}$, training epochs ($E$), batch size ($B$), learning rate ($\eta$), and weight decay ($\beta$).
Output: Detection and segmentation results on the oral panoramic radiograph dataset.

Initialize model parameters θ;
Set training parameters;
    Prepare enhanced dataset D and corresponding labels Y;
for epoch = 1 to E **do**
    Train the model (refer to training process below);
    if performance does not improve **then**
        Apply an early stopping strategy;
    end if
end for
return detection and segmentation results.

Algorithm A2 Training Process for Real-Time Dental Pathology Detection and Segmentation in Panoramic Radiographs

Input: Batch of oral panoramic radiograph images $\{x_1, x_2, \cdots, x_B\}$, corresponding oral dental disease labels $\{y_1, y_2, \cdots, y_B\}$.
Output: Model updated with optimized parameters.

for each batch **do**
  Forward Propagation:
  Replace standard convolution with lightweight partial convolution (PConv) in the C2f module to create the C2f-Faster module for efficient feature extraction;
  $PConv,\ Output = PConv(Input) = W * Input + b$;
  Utilize the BiFPN structure to enhance multiscale feature fusion, enabling better information flow among both lower-order features and higher-order features;
  Feature Fusion: $F_{out} = \sum_i a_i \cdot F_i$;
  Introduce the EMCA attention mechanism to strengthen attention on disease-related features through weighted feature maps;
  Attention Mechanism: $F_{att} = F \cdot \sigma(W_{att} * F)$;
  Apply the segmentation head to execute forward propagation and obtain predictions:
  $\hat{y}_i = \text{Segmentation head}(x_i; \theta),\ \forall i \in \{1, \cdots, B\}$
  Loss Calculation:
  Localization Loss (L_{box}): Compute using Powerful-IOU to evaluate the gap between estimated and actual bounding boxes used to assess oral disease;
  $L_{box} = u(\lambda q) \cdot L_{PIoU} = 3 \cdot (\lambda q) \cdot e^{-(\lambda q)^2} \cdot L_{PIoU}$;
  Classification Loss (L_{cls}): Compute using BCE (binary cross-entropy) to evaluate the gap between actual and estimated classes;
  $L_{cls} = -\sum_i \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right)$;
  Segmentation Mask Loss (L_{seg}): Compute using BCE to evaluate the gap between actual and estimated masks;
  $L_{seg} = -\sum_i \left(M_i \log(\hat{M}_i) + (1 - M_i) \log(1 - \hat{M}_i) \right)$;
  Total Loss Calculation:
  $L(\theta) = \lambda_{box} L_{box} + \lambda_{cls} L_{cls} + \lambda_{seg} L_{seg} + \beta \|\theta\|_2^2$;
  Backpropagation:
  Use the SGD optimizer to update model parameters;
  $\theta = \theta - \eta \frac{\partial L(\theta)}{\partial \theta}$;
end for
return updated model parameters θ.

References

1. World Health Organization. *Global Oral Health Status Report: Towards Universal Health Coverage for Oral Health by 2030*; World Health Organization: Geneva, Switzerland, 2022.
2. Li, S.; Fevens, T.; Krzyżak, A.; Li, S. An automatic variational level set segmentation framework for computer aided dental X-rays analysis in clinical environments. *Comput. Med. Imaging Graph.* **2006**, *30*, 65–74. [CrossRef]
3. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [CrossRef] [PubMed]
4. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef] [PubMed]
5. Choy, G.; Khalilzadeh, O.; Michalski, M.; Do, S.; Samir, A.E.; Pianykh, O.S.; Geis, J.R.; Pandharipande, P.V.; Brink, J.A.; Dreyer, K.J. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* **2018**, *288*, 318–328. [CrossRef]
6. Kumar, V.; Gu, Y.; Basu, S.; Berglund, A.; Eschrich, S.A.; Schabath, M.B.; Forster, K.; Aerts, H.J.W.L.; Dekker, A.; Fenstermacher, D.; et al. Radiomics: The Process and the Challenges. *Magn. Reson. Imaging* **2012**, *30*, 1234–1248. [CrossRef]
7. Kim, J.U.; Kim, H.G.; Ro, Y.M. Iterative deep convolutional encoder-decoder network for medical image segmentation. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Republic of Korea, 11–15 July 2017; pp. 685–688.
8. Majanga, V.; Viriri, S. Dental Images' Segmentation Using Threshold Connected Component Analysis. *Comput. Intell. Neurosci.* **2021**, *2021*, 2921508.. [CrossRef]

9. Li, S.; Fevens, T.; Krzyżak, A.; Jin, C.; Li, S. Semi-Automatic Computer Aided Lesion Detection in Dental X-Rays Using Variational Level Set. *Pattern Recognit.* **2007**, *40*, 2861–2873. [CrossRef]
10. Huang, C.-H.; Hsu, C.-Y. Computer-Assisted Orientation of Dental Periapical Radiographs to the Occlusal Plane. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endodontology* **2008**, *105*, 649–653. [CrossRef] [PubMed]
11. Mohamed razali, M.R.; Ahmad, N.S.; Mohd Zaki, Z.; Ismail, W. Region of Adaptive Threshold Segmentation between Mean, Median and Otsu Threshold for Dental Age Assessment. In Proceedings of the 2014 International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, Malaysia, 2–4 September 2014; pp. 353–356.
12. Subramanyam, R.B.; Prasad, K.P.; Anuradha, B. Different Image Segmentation Techniques for Dental Image Extraction. *Eng. Res. Appl.* **2014**, *4*, 173–177.
13. Kim, J.; Lee, S. Extracting Major Lines by Recruiting Zero-Threshold Canny Edge Links along Sobel Highlights. *IEEE Signal Process. Lett.* **2015**, *22*, 1689–1692. [CrossRef]
14. Razali, M.R.M.; Ahmad, N.S.; Hassan, R.; Zaki, Z.M.; Ismail, W. Sobel and Canny Edges Segmentations for the Dental Age Assessment. In Proceedings of the 2014 International Conference on Computer Assisted System in Health, Kuala Lumpur, Malaysia, 19–21 December 2014; pp. 62–66.
15. Abimannan, S.; El-Alfy, E.-S.M.; Chang, Y.-S.; Hussain, S.; Shukla, S.; Satheesh, D. Ensemble Multifeatured Deep Learning Models and Applications: A Survey. *IEEE Access* **2023**, *11*, 107194–107217. [CrossRef]
16. AbuSalim, S.; Zakaria, N.; Islam, M.R.; Kumar, G.; Mokhtar, N.; Abdulkadir, S.J. Analysis of Deep Learning Techniques for Dental Informatics: A Systematic Literature Review. *Healthcare* **2022**, *10*, 1892. [CrossRef]
17. Thompson, N.; Greenewald, K.; Lee, K.; Manso, G.F. The Computational Limits of Deep Learning. In Proceedings of the Ninth Computing Within Limits 2023, Online, 14 June 2023.
18. Tikhe, S.V.; Naik, A.M.; Bhide, S.D.; Saravanan, T.; Kaliyamurthie, K. Algorithm to identify enamel caries and interproximal caries using dental digital radiographs. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 27–28 February 2016; pp. 225–228.
19. Lin, P.L.; Huang, P.Y.; Huang, P.W.; Hsu, H.C.; Chen, C.C. Teeth Segmentation of Dental Periapical Radiographs Based on Local Singularity Analysis. *Comput. Methods Programs Biomed.* **2014**, *113*, 433–445. [CrossRef] [PubMed]
20. Lin, P.L.; Huang, P.W.; Huang, P.Y.; Hsu, H.C. Alveolar Bone-Loss Area Localization in Periodontitis Radiographs Based on Threshold Segmentation with a Hybrid Feature Fused of Intensity and the H-Value of Fractional Brownian Motion Model. *Comput. Methods Programs Biomed.* **2015**, *121*, 117–126. [CrossRef] [PubMed]
21. Lurie, A.; Tosoni, G.M.; Tsimikas, J.; Walker, F. Recursive Hierarchic Segmentation Analysis of Bone Mineral Density Changes on Digital Panoramic Images. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2012**, *113*, 549–558. [CrossRef]
22. Modi, C.K.; Desai, N.P. A Simple and Novel Algorithm for Automatic Selection of Roi for Dental Radiograph Segmentation. In Proceedings of the 24th Canadian Conference on Electrical and Computer Engineering (CCECE), Niagara Falls, ON, Canada, 8–11 May 2011.
23. Indraswari, R.; Kurita, T.; Arifin, A.Z.; Suciati, N.; Astuti, E.R.; Navastara, D.A. 3D Region Merging for Segmentation of Teeth on Cone-Beam Computed Tomography Images. In Proceedings of the Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS)/19th International Symposium on Advanced Intelligent Systems (ISIS), Toyama, Japan, 5–8 December 2018.
24. Gan, Y.; Xia, Z.; Xiong, J.; Zhao, Q.; Hu, Y.; Zhang, J. Toward Accurate Tooth Segmentation from Computed Tomography Images Using a Hybrid Level Set Model. *Med. Phys.* **2015**, *42*, 14–27. [CrossRef]
25. Wang, Y.; Liu, S.; Wang, G.; Liu, Y. Accurate Tooth Segmentation with Improved Hybrid Active Contour Model. *Phys. Med. Biol.* **2018**, *64*, 015012. [CrossRef]
26. Fernandez, K.; Chang, C. Teeth/palate and interdental segmentation using artificial neural networks. In Proceedings of the Artificial Neural Networks in Pattern Recognition: 5th INNS IAPR TC 3 GIRPR Workshop, ANNPR 2012, Trento, Italy, 17–19 September 2012; Proceedings 5; Springer: Berlin/Heidelberg, Germany, 2012.
27. Prakash, M.; Gowsika, U.; Sathiyapriya, S. An Identification of Abnormalities in Dental with Support Vector Machine Using Image Processing. In *Emerging Research in Computing, Information, Communication and Applications, New Delhi*; Shetty, N., Prasad, N., Nalini, N., Eds.; Springer: Berlin, Germany, 2015; pp. 29–40.
28. Mortaheb, P.; Rezaeian, M. Metal artifact reduction and segmentation of dental computerized tomography images using least square support vector machine and mean shift algorithm. *J. Med. Signals Sens.* **2016**, *6*, 1–11. [CrossRef]
29. Son, L.H.; Tuan, T.M. A Cooperative Semi-Supervised Fuzzy Clustering Framework for Dental X-Ray Image Segmentation. *Expert Syst. Appl.* **2016**, *46*, 380–393. [CrossRef]
30. Geetha, V.; Aprameya, K.S.; Hinduja, D.M. Dental Caries Diagnosis in Digital Radiographs Using Back-Propagation Neural Network. *Health Inf. Sci. Syst.* **2020**, *8*, 8. [CrossRef]

31. Lo Casto, A.; Spartivento, G.; Benfante, V.; Di Raimondo, R.; Ali, M.; Di Raimondo, D.; Tuttolomondo, A.; Stefano, A.; Yezzi, A.; Comelli, A. Artificial Intelligence for Classifying the Relationship between Impacted Third Molar and Mandibular Canal on Panoramic Radiographs. *Life* **2023**, *13*, 1441. [CrossRef] [PubMed]
32. Mureșanu, S.; Hedeșiu, M.; Iacob, L.; Eftimie, R.; Olariu, E.; Dinu, C.; Jacobs, R.; Group, T.P. Automating Dental Condition Detection on Panoramic Radiographs: Challenges, Pitfalls, and Opportunities. *Diagnostics* **2024**, *14*, 2336. [CrossRef] [PubMed]
33. Lakshmi, M.M.; Chitra, P. Tooth decay prediction and classification from X-ray images using deep CNN. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 28–30 July 2020; pp. 1349–1355.
34. Banar, N.; Bertels, J.; Laurent, F.; Boedi, R.M.; De Tobel, J.; Thevissen, P.; Vandermeulen, D. Towards Fully Automated Third Molar Development Staging in Panoramic Radiographs. *Int. J. Leg. Med.* **2020**, *134*, 1831–1841. [CrossRef] [PubMed]
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
36. Wang, K.; Zhang, S.; Wei, Z.; Fang, X.; Liu, F.; Han, M.; Du, M. Deep Learning-Based Efficient Diagnosis of Periapical Diseases with Dental X-Rays. *Image Vis. Comput.* **2024**, *147*, 105061. [CrossRef]
37. Patil, S.; Kulkarni, V.; Bhise, A. Caries Detection Using Multidimensional Projection and Neural Network. *Int. J.-Knowl.-Based Intell. Eng. Syst.* **2018**, *22*, 155–166. [CrossRef]
38. Zhu, H.; Cao, Z.; Lian, L.; Ye, G.; Gao, H.; Wu, J. CariesNet: A Deep Learning Approach for Segmentation of Multi-Stage Caries Lesion from Oral Panoramic X-Ray Image. *Neural Comput. Appl.* **2018**, *22*, 155–166. [CrossRef]
39. Ma, T.; Zhou, X.; Yang, J.; Meng, B.; Qian, J.; Zhang, J.; Ge, G. Dental Lesion Segmentation Using an Improved ICNet Network with Attention. *Micromachines* **2022**, *13*, 1920. [CrossRef]
40. Bağ, İ.; Bilgir, E.; Bayrakdar, İ.Ş.; Baydar, O.; Atak, F.M.; Çelik, Ö.; Orhan, K. An Artificial Intelligence Study: Automatic Description of Anatomic Landmarks on Panoramic Radiographs in the Pediatric Population. *BMC Oral Health* **2023**, *23*, 764. [CrossRef] [PubMed]
41. Beser, B.; Reis, T.; Berber, M.N.; Topaloglu, E.; Gungor, E.; Kılıc, M.C.; Duman, S.; Çelik, Ö.; Kuran, A.; Bayrakdar, I.S. YOLO-V5 Based Deep Learning Approach for Tooth Detection and Segmentation on Pediatric Panoramic Radiographs in Mixed Dentition. *BMC Med. Imaging* **2024**, *24*, 172.
42. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
43. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
44. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
45. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-Time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9156–9165.
46. Chen, J.; Kao, S.-H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, Don't walk: Chasing higher FLOPS for faster neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023.
47. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
48. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Supplementary material for "ECA-Net: Efficient channel attention for deep convolutional neural networks". In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
49. Liu, C.; Wang, K.; Li, Q.; Zhao, F.; Zhao, K.; Ma, H. Powerful-IoU: More Straightforward and Faster Bounding Box Regression Loss with a Nonmonotonic Focusing Mechanism. *Neural Netw.* **2024**, *170*, 276–284. [CrossRef] [PubMed]
50. Hamamci, I.E.; Er, S.; Simsar, E.; Yuksel, A.E.; Gultekin, S.; Ozdemir, S.D.; Yang, K.; Li, H.B.; Pati, S.; Stadlinger, B.; et al. DENTEX: An Abnormal Tooth Detection with Dental Enumeration and Diagnosis Benchmark for Panoramic X-Rays. *arXiv* **2023**, arXiv:2305.19112.
51. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
52. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
53. Qin, X.; Li, N.; Weng, C.; Su, D.; Li, M. Simple attention module based speaker verification with iterative noisy label detection. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 6722–6726.

54. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3139–3148.
55. Park, S.; Kim, H.; Shim, E.; Hwang, B.-Y.; Kim, Y.; Lee, J.-W.; Seo, H. Deep Learning-Based Automatic Segmentation of Mandible and Maxilla in Multi-Center CT Images. *Appl. Sci.* **2022**, *12*, 1358. [CrossRef]