



# A novel deep learning-based perspective for tooth numbering and caries detection

Baturalp Ayhan<sup>1</sup> · Enes Ayan<sup>2</sup> · Yusuf Bayraktar<sup>1</sup>

Received: 15 November 2023 / Accepted: 17 February 2024 / Published online: 27 February 2024  
© The Author(s) 2024

## Abstract

**Objectives** The aim of this study was automatically detecting and numbering teeth in digital bitewing radiographs obtained from patients, and evaluating the diagnostic efficiency of decayed teeth in real time, using deep learning algorithms.

**Methods** The dataset consisted of 1170 anonymized digital bitewing radiographs randomly obtained from faculty archives. After image evaluation and labeling process, the dataset was split into training and test datasets. This study proposed an end-to-end pipeline architecture consisting of three stages for matching tooth numbers and caries lesions to enhance treatment outcomes and prevent potential issues. Initially, a pre-trained convolutional neural network (CNN) utilized to determine the side of the bitewing images. Then, an improved CNN model YOLOv7 was proposed for tooth numbering and caries detection. In the final stage, our developed algorithm assessed which teeth have caries by comparing the numbered teeth with the detected caries, using the intersection over union value for the matching process.

**Results** According to test results, the recall, precision, and F1-score values were 0.994, 0.987 and 0.99 for teeth detection, 0.974, 0.985 and 0.979 for teeth numbering, and 0.833, 0.866 and 0.822 for caries detection, respectively. For teeth numbering and caries detection matching performance; the accuracy, recall, specificity, precision and F1—Score values were 0.934, 0.834, 0.961, 0.851 and 0.842, respectively.

**Conclusions** The proposed model exhibited good achievement, highlighting the potential use of CNNs for tooth detection, numbering, and caries detection, concurrently.

**Clinical significance** CNNs can provide valuable support to clinicians by automating the detection and numbering of teeth, as well as the detection of caries on bitewing radiographs. By enhancing overall performance, these algorithms have the capacity to efficiently save time and play a significant role in the assessment process.

**Keywords** Artificial intelligence · Digital bitewing radiography · Deep learning · Detection · Numbering · Dental caries

## Introduction

Dental caries is a preventable, chronic, multifactorial, infectious disease that is common in many populations worldwide [1]. Carious lesions, which usually progress slowly, can be a serious health problem that can lead to pain, infection and tooth loss if not treated in early stages [2]. While the perspective on caries treatment was more invasive in the past, today in modern dentistry, this perspective has been replaced by early diagnosis with preventive and minimally invasive treatment methods [3]. Early and accurate detection of caries lesions is a very important step that guides clinical planning by determining the treatment to be applied. Therefore, timely and accurate diagnosis has a key role in a successful treatment [4]. Although there are different methods in the diagnosis of caries lesions, the visual-tactile examination

✉ Baturalp Ayhan  
baturalpayhan@kku.edu.tr

Enes Ayan  
enesayan@kku.edu.tr

Yusuf Bayraktar  
yusufbayraktar@kku.edu.tr

<sup>1</sup> Department of Restorative Dentistry, Faculty of Dentistry, Kırıkkale University, Kırıkkale, Turkey

<sup>2</sup> Department of Computer Engineering, Faculty of Engineering and Architecture, Kırıkkale University, Kırıkkale, Turkey

method which is insufficient in the diagnosis of early approximal caries detection is one of the most basic of them [5, 6]. In addition to the visual-tactile examination method, radiological examination which is commonly employed in intraoral imaging plays a crucial role in dental practice. In this regard, panoramic, periapical and bitewing radiography are regularly utilized in practice. Visualization and evaluation of dental status on radiographs are one of the most important steps in disease diagnosis [7, 8]. Today, with the help of digital bitewing radiographs, which have high recall and specificity in the detection of approximal caries, the detection of caries lesions is done easily, cheaply and successfully compared to many methods [9, 10].

As can be understood from all these, radiological imaging and visual-tactile examination methods are essential for detecting carious lesions. In some situations, depending on the experience and attention of the dentist, misdiagnosis or insufficient diagnosis can be made in bustling clinics. The use of computer aided systems can help to dentists during diagnostic process. Also, it can provide the physician to achieve success by assisting with treatment planning and implementation [11, 12].

Artificial intelligence term (AI), which has become very popular in our lives and affects many industries and promises to be a real turning point [13]. AI is a general term that refers to systems, machines or computers that mimic human behaviors and intelligence [14]. It includes different areas of research fields and algorithms. Deep learning (DL) is one of these research areas that has gained popularity in recent years [15]. DL algorithms consist of various artificial neural networks according to the different types of problems such as natural language processing, large language models, computer vision [16]. Convolutional neural networks (CNNs) are supervised deep learning algorithms that are artificial neural networks inspired by the mammalian visual system [17]. CNNs are actively used in many computer vision problems in dentistry and medicine for improving health services and increasing efficiency [13, 18].

The adoption of CNNs in the diagnosis of caries in dentistry is a successful method as can be understood from the studies carried out. There are many studies on CNN and similar neural networks [19–24]. Bitewing radiographs are often used by dentists as an assistant tool to diagnose dental caries [25]. In some of the researches, caries diagnosis studies were performed on bitewing radiographs [19, 20, 26, 27]. Correct detection and numbering of teeth during the examination can reduced the examination time and provide a better diagnosis [28]. Yasa et al. [29] researched teeth detection and numbering in bitewing radiographs. Also, there are similar studies have been searched with in panoramic or bitewing radiographs [30–34]. These studies showed that the developments based on DL in the health industry can be an assistant and a second opinion provider in the daily clinical

practice of the physician [35]. To the best of our knowledge, no study has been investigated that teeth detection, numbering, and caries detection at the same time in digital bitewing radiographs utilizing a CNN approach based on DL. The proposed method makes this study unique compared to existing literature.

The aim of this study was to automatically detecting and numbering teeth in digital bitewing radiographs obtained from patients, and to evaluate the diagnostic efficiency of decayed teeth in real time, with a pipeline architecture using CNNs. Thus, it is planned to reduce the chair time, to assist the physician and to increase the effectiveness of the treatment by making the diagnosis more effective during the examination.

A successful proposal introduces a novel DL-based pipeline architecture for teeth numbering and caries detection. The architecture enhances the preservation of crucial information in bitewing images by incorporating convolutional block attention modules (CBAMs) and average pooling layers into the YOLOv7 model. This implementation results in a more efficient performance in teeth numbering and decay detection compared to the original YOLOv7 model. Additionally, an algorithm has been developed to assess teeth with caries by correlating the numbered teeth with the detected caries.

## Material and methods

### Sample size determination

The determination of the sample size for the test dataset was performed with Fisher's exact test A design incorporating clustering was employed because of every digital bitewing radiograph contained averagely 16 surfaces which was calculated by design effect (DE). Design effect (DE), which assesses the performance of a particular sampling method compared to that of a simple random sample, is a valuable tool in considering the precision of a cluster sample. Equation of DE is  $1 + (m-1) \times ICC$ , in where  $m$  represents size of cluster, and ICC is the intraclass correlation coefficient. ICC of 0.2 was assumed based on Masood et al. [36]. According to previous study's power analysis, required test set units multiplied with DE [37]. When  $DE = 1 + (16 - 1) \times 0.2 = 4$ , the total surfaces required should be 2136 or 134 bitewing radiographs. In the present study, test dataset contained 170 bitewing radiographs.

### Image dataset

A total of 1170 digital bitewing radiographs of adult patients aged over 18 were collected from radiology archive of Kırıkkale University, Faculty of Dentistry for this study.

Bitewing radiographs which obtained from patients who came for examination and/or treatment between January-2018 and January-2023 included by examining. No further details, such as gender, age or acquisition date of images was revealed, since the radiographs were collected anonymously. In the study, bitewing radiographs containing patient positioning errors, technical errors, fixed prosthetic restoration and fixed orthodontic braces that would complicate or completely prevent the diagnosis did not include in the dataset. Approval for this study was obtained from the non-interventional Clinical Research Ethical Committee of Kırkkale University. (Decision Number: 2022.11.14) The study adhered to the ethical guidelines stated in the Declaration of Helsinki.

All digital bitewing radiographs were taken by a Gendex Expert DC (Gendex Dental Systems, IL, USA) device that is located in the clinics with a focal spot value of 0.04 mm, tube voltage 65 kVp, tube current 7 mA. Exposure time, distance, and quality of radiographs can be variable, because of the digital bitewing radiographs were captured by various personnel across different departments. Each digital bitewing radiograph was anonymized and saved as.jpg format. Maxillary and mandibular dental arches were distinctly

visible on digital bitewing images. The assessment included all visible teeth, excluding restorations present on the proximal surfaces of the teeth. Because of the aim of this study was evaluating the proposed model performance for detecting primary approximal caries and numbering teeth. The collected digital bitewing radiographs displayed heterogeneous resolutions in terms of their sizes. For this reason, all resolutions were adjusted to  $640 \times 480$  pixels, such as a previous study [19]. After this process, the dataset was randomly split as follows: 85% for training and validation (1000 bitewings) and 15% for testing (170 bitewings). Thus, 1000 digital bitewing radiographs that obtained with random distribution were used for training the proposed CNN model. The remaining 170 digital bitewing radiographs were used for evaluating the model. In the training dataset, 9344 teeth were numbered. There are 16674 approximal surfaces and 2289 of them are caries. The lesion prevalence of the training data is 24%. Details of the dataset are given in Table 1 and image samples from the dataset are given in Fig. 1.

## Study design

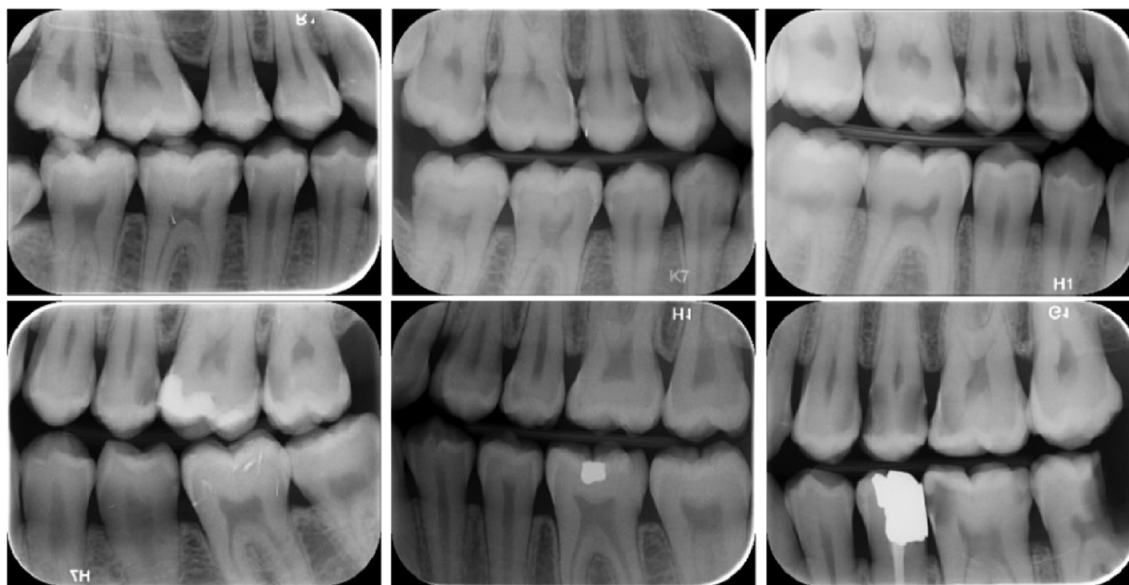
In this study, a pipeline architecture that can simultaneously detect teeth numbering and caries in bitewing images is proposed. The proposed architecture consists of 3 basic steps. A diagram illustrating the application of the proposed architecture is given in Fig. 2.

## Image evaluation and data labeling

A restorative specialist (author 3) with 13 years of experience and a research assistant (author 1) with 4 years of

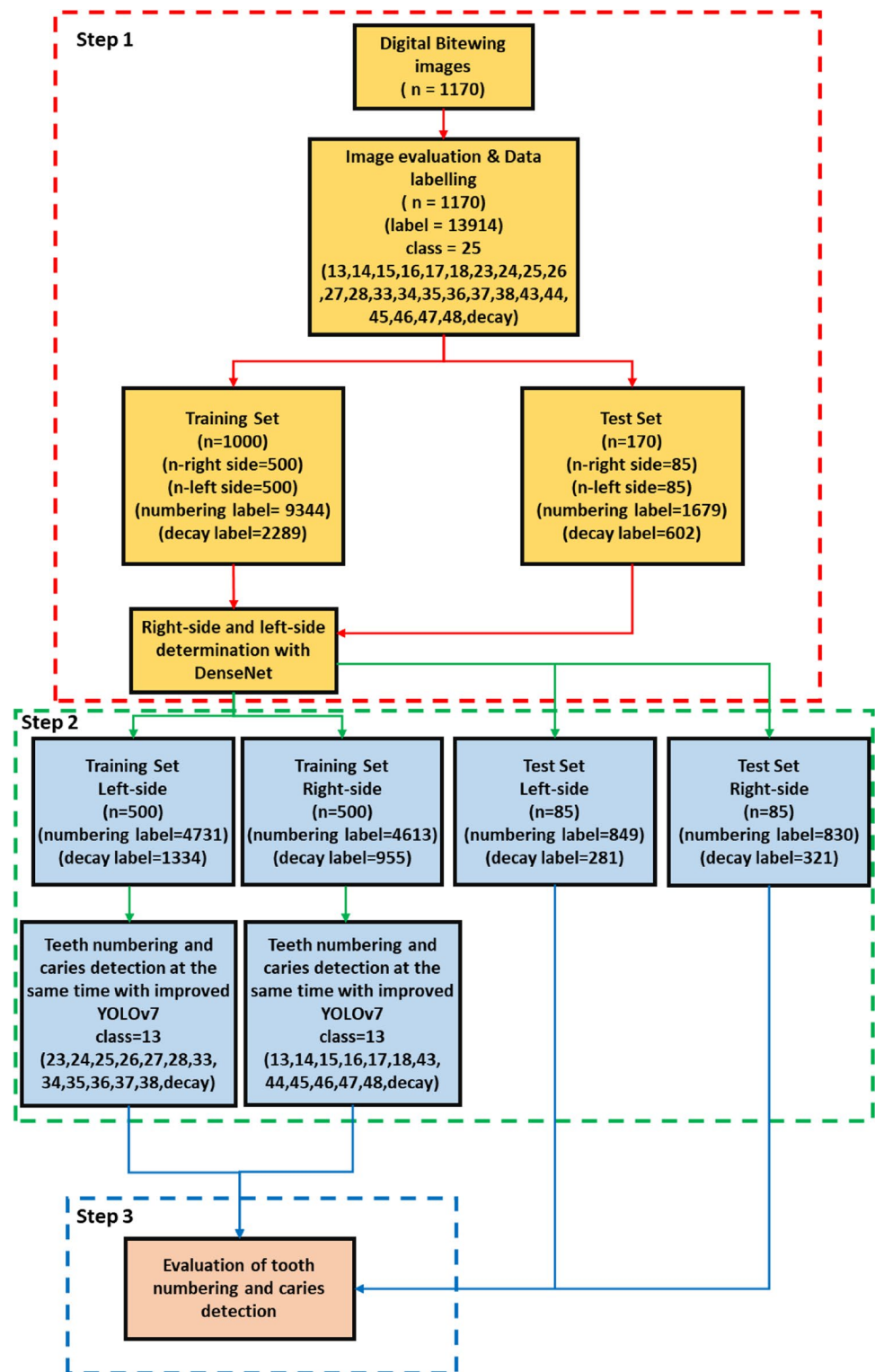
**Table 1** Summary of whole dataset

	Train	Test
Decay	2289	602
Numbering	9344	1679
Total surfaces	16,674	2842
Lesion prevalence	24%	21.16%
Total image	1000	170



**Fig. 1** Image samples from the dataset

**Fig. 2** Diagram of the implementation of the proposed architecture



experience generated ground truth annotations for the images using software program which developed in Python programming language. For that, annotators labeled bounding boxes around all teeth while simultaneously providing a class label for each box with the tooth number according

to Federation Dentaire Internationale (FDI) tooth notation system (16–17–18–26–27–28–36–37–38–46–47–48 (molars), 14–15–24–25–34–35–44–45 (premolars) and 13–23–33–43 (canines)). Also, caries lesions labeled as decay in digital bitewing images. While in the data labeling procedure,



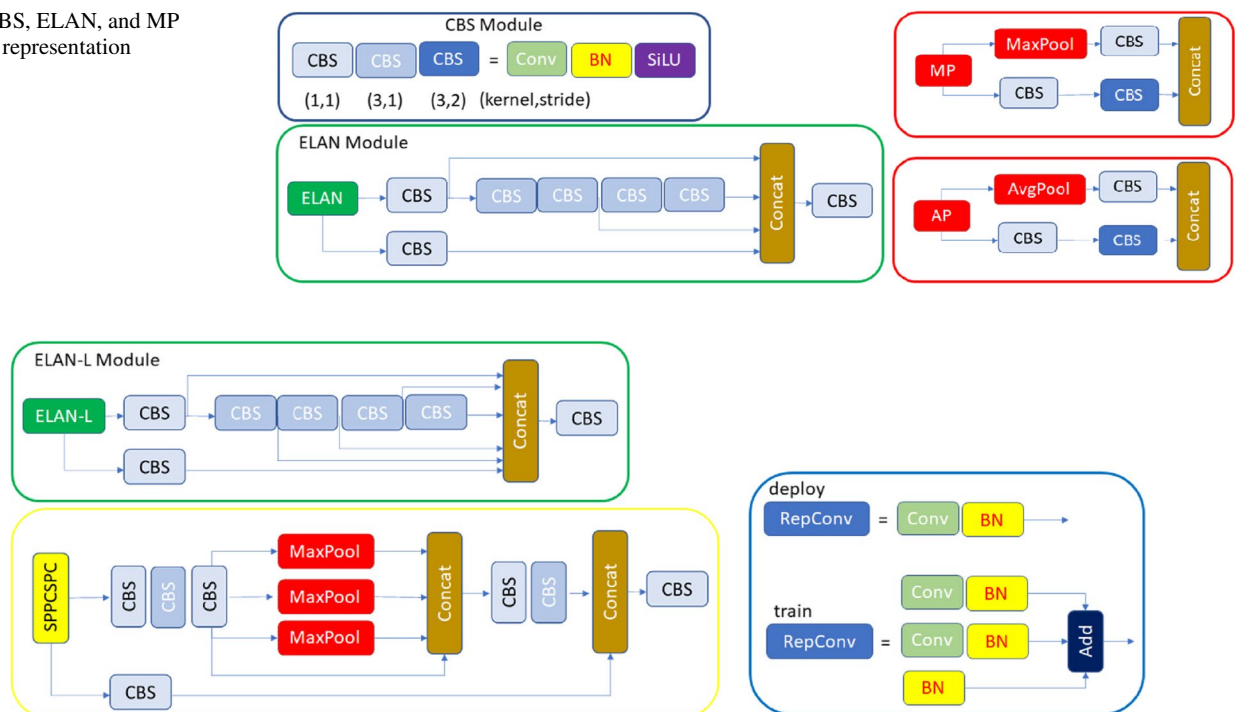


In YOLOv7, one of the latest versions of the YOLO series, some improvements have been made in the loss function, backbone network and activation function. In this way, the model has achieved improved detection accuracy and accelerated detection speed. For this reason, the YOLOv7 algorithm was chosen to be used in the study. The original YOLOv7 architecture consists of four modules: input, backbone, neck and head. The input module resizes the input image to a standardized pixel size ( $640 \times 640$ ) to fulfill the input size specifications of the backbone module. The backbone of the original architecture consists of convolutional blocks, batch normalization process, and SiLU activation function (CBS), extended efficient long-range attention network (E-ELAN) and MaxPool-1 (MP1) modules. The task of the backbone part is to extract feature maps from the input image. The CBS module consists of three components: convolution blocks, batch normalization process (BN), and SiLU activation function (sigmoid-weighted linear unit activation function) (Fig. 5). The ELAN module comprises multiple CBS modules that ensure the input and output feature sizes remain unchanged. By guiding the computational blocks of different feature groups to learn various features, the network's learning capacity is enhanced while preserving the integrity of the original gradient path (Fig. 5). The MP1 module is divided into two branches specifically an upper branch

and a lower branch and consisting of CBS modules and the max pooling (MaxPool) layer (Fig. 4). In this module, the upper branch utilizes MaxPool to capture the maximum value information from small local areas, while the CBS modules in the lower branch extract all valuable information from small local areas. This dual approach enhances the network's feature extraction capability.

The neck module consists of ELAN-L, Upsample, MP2 and SPPCSPC modules using the PAFPN Path Aggregation Feature Pyramid Network (PAFPN) architecture. The neck part combines the feature maps obtained from the different layers of the backbone module and sends them as input to the head module. The SPPCSPC module consists of CBS modules and MaxPool layers with different sizes of parallel max pooling (Fig. 5). The ELAN-L module has a similar structure to the ELAN module, but it combines five different outputs (Fig. 5). The Up-sample module consists of the CBS module and is used for upsampling. The Head module includes of a Rep block to set the number of image channels for the features output from the neck part. This is followed by the application of  $1 \times 1$  convolution module, which predicts the category, confidence, and bounding box. The REP module is inspired by RepVGG [44] and employs a simplified convolution process during the testing phase (Fig. 5). This simplification effectively reduces network complexity without compromising performance.

**Fig. 4** CBS, ELAN, and MP modules representation



**Fig. 5** ELAN-L, Rep and SPPCSPC modules representation

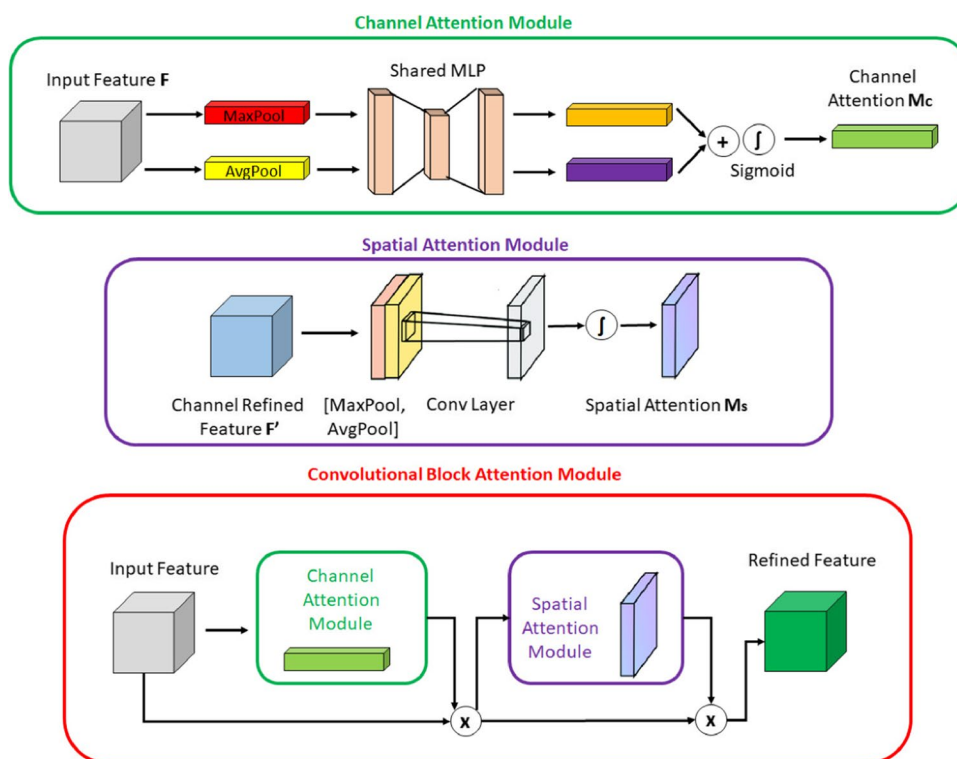
## Improved YOLOv7 with convolutional block attention module

The scientific community widely acknowledges the significant role of attentional mechanisms in human vision. A fundamental characteristic of the human visual system is its ability to selectively focus on salient aspects of a scene rather than processing the entire visual field simultaneously. This selective attention enables enhanced capture of visual details [45]. In recent years, researchers have developed various selective attention mechanisms for neural networks. These mechanisms enable neural networks to learn "where" or "what" to focus more on, using explicit dependencies between channels and weighted spatial regions. In this context, the convolutional block attention module (CBAM) was introduced by Woo et al. [46] as an efficient attention mechanism designed for CNNs. Within this context, Woo et al. [46] proposed the convolutional block attention module (CBAM), which is an effective attention mechanism for CNNs. The CBAM module includes of a channel attention module (CAM) and a spatial attention module (SAM). The CAM is focused on the relevant aspects of an input image by exploiting the inter-channel relationships of features. Two spatial context descriptors are generated in the module using two separate pooling methods (average pooling and max pooling). Average pooling (AP) is mainly used to gather spatial information, whereas max pooling retains significantly more contextual details pertaining to

the object within the image, thereby enabling more precise channel attention. These generated context descriptors are forwarded to a shared multi-layer network. After the vector obtained from the multilayer network is collected, the channel attention map is generated by applying the sigmoid function. Within the CBAM module, the spatial attention module (SAM) specifically targets the localization of the most informative regions within the input image. Spatial attention computation within the SAM module involves the application of average pooling and max pooling techniques along the channel axis of the input image. The obtained two-channel image is subjected to convolutional processing to convert it into a single-channel feature map, which is then merged and forwarded to a standard convolutional layer to create an efficient feature descriptor. The output from the convolutional layer is passed through a sigmoid activation layer. The Sigmoid function, which is a probability-based activation, transforms all values to a range between 0 and 1, thereby revealing the spatial attention map. In summary, CBAM combines the two attention modules described above to determine what is important in the input image and to locate the descriptive regions within the image. Figures of CBAM, CAM and SAM modules are given in Fig. 6.

When reviewing the studies in the literature, it is seen that there are successful combinations of the CBAM module in CNNs in the solution of computer vision problems [47, 48]. However, creating a neural network is a complex process that requires consideration of the utilized dataset,

**Fig. 6** CBAM, CAM and SAM modules representation



the selection of optimized metrics, and the availability of computational resources.

In this study, we injected CBAM layers in the YOLOv7 model as a result of our experience with CNNs and various tests we performed. After conducting various experiments, three CBAM modules have been placed at three specific locations where the intersections of the neck and head in the YOLOv7 architecture. The selection of the number of integrated CBAM blocks during the incorporation of CBAM modules was done meticulously to minimize any significant increase in the computational overhead on the network. The integration of CBAM modules into the YOLOv7 architecture by researchers to develop novel designs for solving various problems has emerged as a highly popular research area [49, 50]. Therefore, the determination of position, number of blocks, and their integration with other network components are design decisions that can vary depending on the specific problem under consideration. To the best of our knowledge, there is currently no known study that specifically addresses the region we have incorporated in our research. Figure 7 provides an illustration of the improved model and details about the location where CBAM modules are positioned.

Another improvement we have made in our proposed YOLOv7 model is to use average pooling (AP) instead of max pooling (MP) in subsampling layers. In the original YOLOv7, subsampling is done at four different locations. In these subsamples, the average pooling

method was used instead of the max pooling. Max pooling returns the maximum value in each window, while average pooling returns the average value in each window. The benefit of using average pooling instead of max pooling is that it can help reduce overfitting. Max pooling is more likely to memorize as it preserves only the strongest feature in each window that may not represent the entire window. On the other hand, average pooling considers all features in each window, which can help preserve more information. Additionally, average pooling can help smooth out the output, which can be useful when the input is noisy or volatile. In this context, the changes we made are given in Fig. 7. It has been observed that this improvement has a positive effect on detection performance.

The architecture of YOLOv7-AP-CBAM obtained with the updates made in the YOLOv7 architecture was used in the numbering of teeth and in the detection of caries. In this context, two different models were trained for the right and left parts of the mouth. Since the numbering performance of a single model remained low in the tests, two different model training methods were used. The models for both sides were trained with the following hyperparameters: batch size 16, epochs 20, input image size (640,640), initial learning rate 0.01, optimizer SGD algorithm, momentum 0.937, weight decay 0.005. A flowchart summarizing this stage is given in Fig. 7.

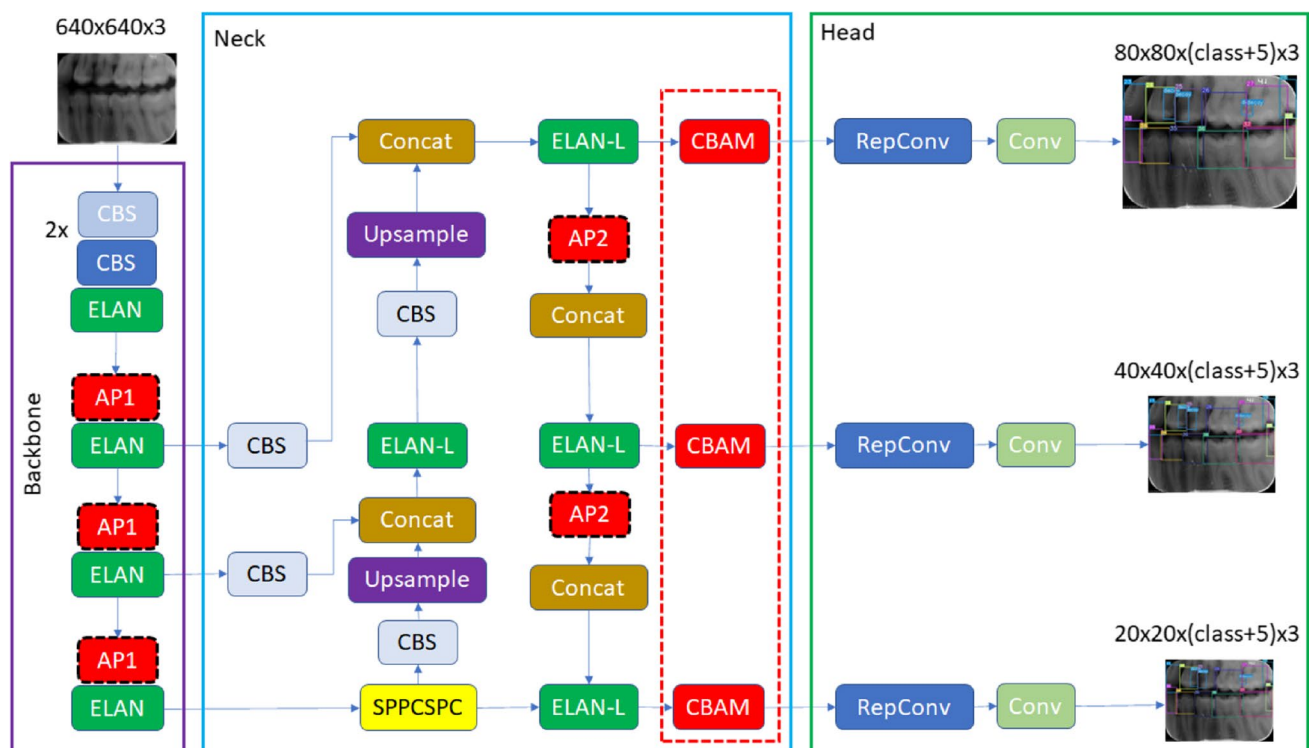


Fig. 7 Improved YOLOv7 with CBAM, and AP modules



## Evaluation of teeth numbering and caries detection matching

In the last stage of the proposed architecture, it was aimed to give the output of the system which tooth had caries by matching the numbered teeth and the detected caries. It has been determined whether there was a bounding box drawn for caries in the bounding boxes drawn for tooth numbering. If there was caries in the numbered bounding box, it was reported which tooth had caries. The Intersection Over Union (IoU) value was used to detect caries in the numbered bounding box (see Fig. 8).

If the IoU value was above 90%, it was accepted that the number of the tooth with caries label was determined correctly. The flow chart of the algorithm we developed for this purpose is given in Fig. 9.

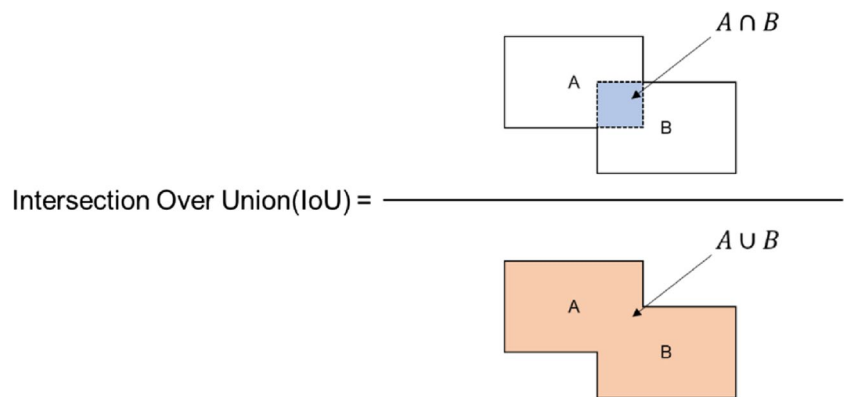
## Experimental setup

All experiments in this study are executed on NVIDIA 1080 Ti GPU with 12G video memory size and two Intel Xenon CPU processors with 64 GB ram on the Ubuntu operating system. The experimental running environment includes Python programming language, Pytorch [51], Keras [52] DL frameworks, and OpenCV computer vision library [53].

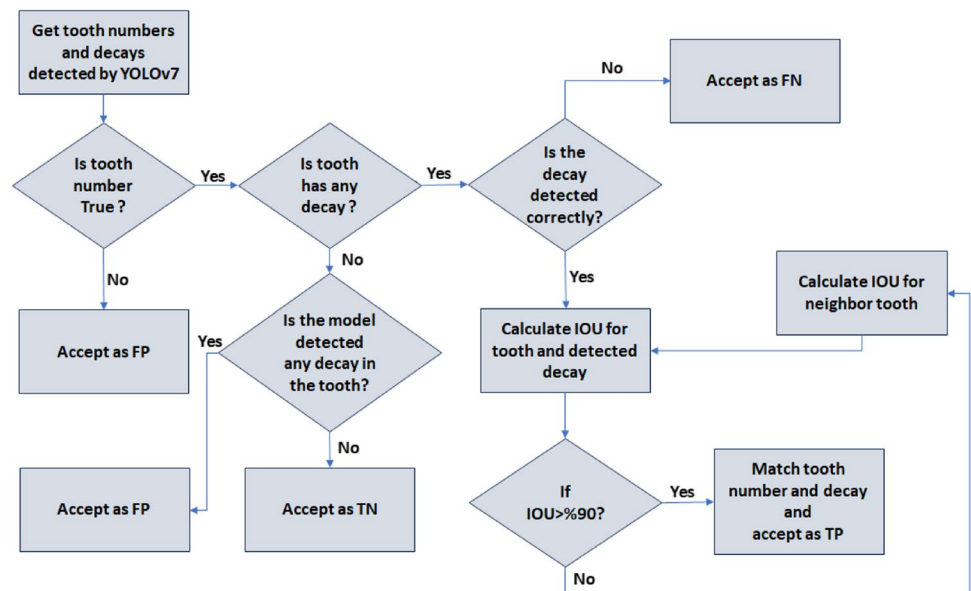
## Performance evaluation analysis

To evaluate the success of the proposed model, a confusion matrix, that summarizes the actual and predicted situations in a table, was used together with the following procedures and metrics. The confusion matrix that we adapted for evaluating performance of teeth numbering and caries detection

**Fig. 8** The formula of IoU. A represents the numbering box, while B represents the caries box



**Fig. 9** Diagram of teeth numbering and caries detection matching decision



matching is given below. Also, the performances of teeth detection, numbering, and caries detection were evaluated individually.

Adaptation of the confusion matrix for evaluating performance of teeth numbering and caries detection matching

TP: the output in which the model correctly predicts the positive class (teeth correctly detected, correctly numbered, caries correctly detected on bitewing radiographs)

FP: the output in which the model incorrectly predicts the positive class (teeth correctly detected, correctly numbered, but caries prediction on sound surface; teeth correctly detected, but regardless of caries incorrectly numbered)

FN: the output in which the model incorrectly predicts the negative class (teeth correctly detected, correctly numbered, but no caries prediction on decayed surface)

TN: the output in which the model correctly predicts the negative class (teeth correctly detected, correctly numbered, no caries prediction when there was no caries)

Initially, True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) rates were determined. Then, Accuracy (Eq. 1), Recall (Eq. 2), Specificity (Eq. 3), Precision (Eq. 4), F1 – Score (Eq. 5) and IoU (Eq. 6) scores were calculated. IoU is defined as the measure of overlap between the actual bounding box labeled by the annotator for an object and the bounding box predicted by the model. By comparing the IoU value with the threshold value determined prior to model training, it is possible to interpret the success of the trained model in object detection. The F1 score, which is considered the harmonic average of precision and recall, is commonly used for providing a balanced evaluation of a classification model's performance.

The equations are shown below (Eqs. 1–6).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{F1 – Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

## Results

In the testing dataset which include 170 digital bitewing radiographs, there were 1679 teeth, 2842 approximal surfaces, 1679 numbering labels, 602 caries labels. The lesion prevalence of the testing data is 21.18%.

Without splitting into right and left sides, the Recall, precision, and F1-score values were 0.409, 0.236, and 0.299 for teeth numbering, and 0.676, 0.707, and 0.691 for caries detection, respectively (Table 5).

After splitting into right and left sides, the Recall, precision, and F1-score values were 0.994, 0.987 and 0.99 for teeth detection, 0.974, 0.985 and 0.979 for teeth numbering, and 0.833, 0.866 and 0.822 for caries detection, respectively (Table 2 and 3).

For teeth numbering and caries detection matching performance; the Accuracy, recall, specificity, precision and F1—Score values were 0.934, 0.834, 0.961, 0.851 and 0.842, respectively (Table 4).

The improved YOLOv7-AP-CBAM models performed well in detecting and numbering teeth, as well as detecting caries at the same time on digital bitewing radiographs (Fig. 10).

## Ablation Experiments

This study performed ablation experiments to evaluate the impact of various enhancements on the model's performance. Initially, the original YOLOv7 model was employed to extract features without splitting into right and left sides. Subsequently, the original YOLOv7 model was modified to include the split into right and left sides, as well as the integration of AP and MP with CBAM. Finally, AP with CBAM was applied. The teeth numbering and caries detection recall, precision, and F1-score of the model with used AP and CBAM modules significantly improved compared to initial of this study. The experimental results are presented in Table 5.

In CNNs, complexity is described by the total number of parameters of the model and the number of Giga Floating Point Operations (GFLOPs) [54]. We compared basic version of YOLOv7 with improved version of YOLOv7-AP-CBAM. The original YOLOv7 has 314

**Table 2** Evaluation metrics of the teeth detection model

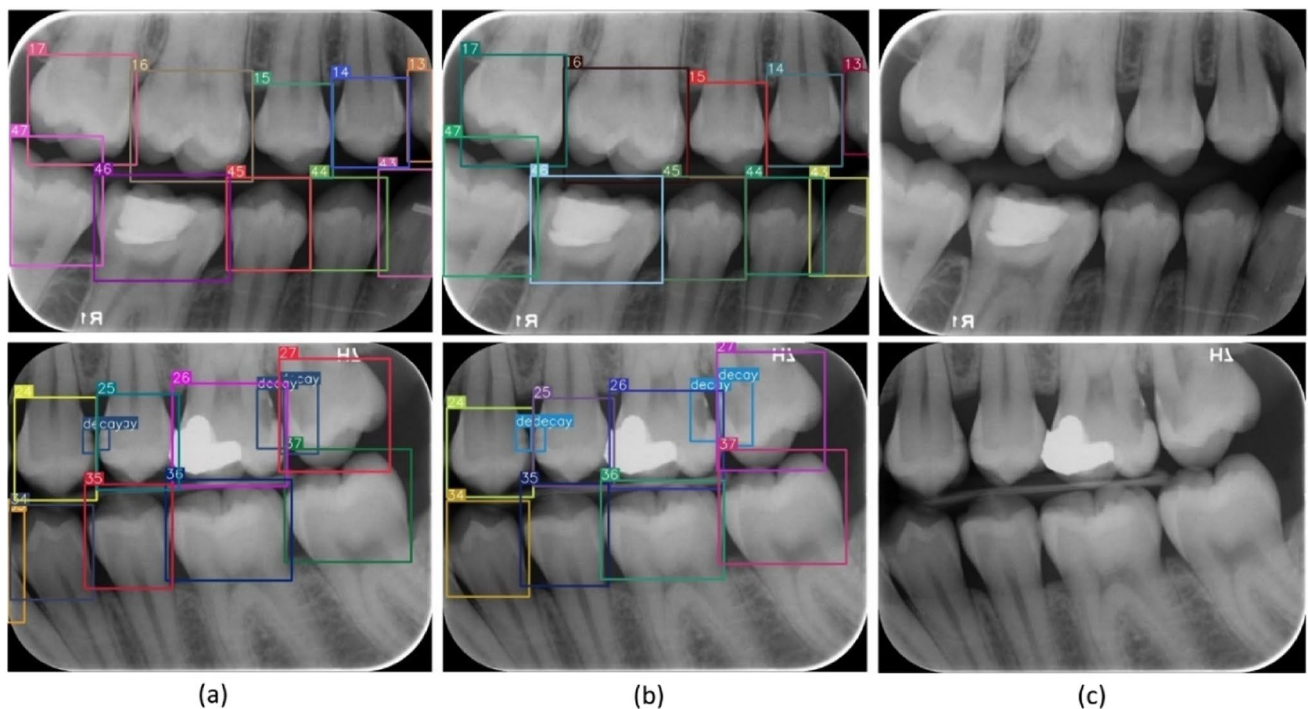
	Recall	Precision	F1-score
Left side	0.995	0.987	0.99
Right side	0.993	0.988	0.99
All	0.994	0.987	0.99

**Table 3** Evaluation metrics of the teeth numbering and caries detection model

	Recall	Precision	F1—score
Left side teeth numbering	0.976	0.978	0.977
Right side teeth numbering	0.972	0.993	0.982
Left side caries detection	0.857	0.833	0.844
Right side caries detection	0.810	0.900	0.800
Left side—numbering and caries detection	0.966	0.966	0.966
Right side—numbering and caries detection	0.959	0.985	0.968
Left and right—side numbering and caries detection—All	0.962	0.975	0.967
Teeth numbering—All	0.974	0.985	0.979
Caries detection—All	0.833	0.866	0.849

**Table 4** Evaluation metrics of the teeth numbering and caries detection matching model

	TP	FP	FN	TN	Accuracy	Recall	Specificity	Precision	F1-score
Left side	240	48	47	1102	0.933	0.836	0.958	0.833	0.834
Right side	260	30	66	1049	0.931	0.797	0.972	0.896	0.843
All	500	78	113	2151	0.932	0.815	0.965	0.865	0.839

**Fig. 10** Image samples for detecting and numbering teeth, as well as detecting caries with the proposed YOLOv7-AP-CBAM model on digital bitewing radiographs. **a** labels of improved YOLOv7-AP-CBAM model, **b** labels of ground truth, **c** raw image

layers 36,611,228 million parameters, 103.7 GFLOPS, 75.1 MB size in memory. On the other hand, the proposed YOLOv7-AP-CBAM has 350 layers 36,655,485 million

parameters 103.8 GFLOPS and 75.2 MB size in memory. A satisfactory performance increase was achieved with a small increase in complexity in the model.

**Table 5** Ablation comparison of performance enhancements on the model

	Recall		Precision		F1-score	
Original YOLOv7-MP without splitting into right and left	Decay all	0.676	Decay all	0.707	Decay all	0.691
	Numbering all	0.409	Numbering all	0.236	Numbering all	0.299
Original YOLOv7-MP with splitting into right and left	Right side decay	0.659	Right side decay	0.803		
	Left side decay	0.696	Left side decay	0.848		
	Decay all	0.677	Decay all	0.825	Decay all	0.743
	Right side numbering	0.971	Right side numbering	0.974		
	Left side numbering	0.979	Left side numbering	0.978		
YOLOv7-AP	Numbering all	0.975	Numbering all	0.976	Numbering all	0.975
	Right side decay	0.626	Right side decay	0.773		
	Left side decay	0.797	Left side decay	0.775		
	Decay all	0.711	Decay all	0.774	Decay all	0.741
	Right side numbering	0.967	Right side numbering	0.977		
YOLOv7-MP-CBAM	Left side numbering	0.981	Left side numbering	0.983		
	Numbering all	0.974	Numbering all	0.980	Numbering all	0.976
	Right side decay	0.579	Right side decay	0.838		
	Left side decay	0.754	Left side decay	0.832		
	Decay all	0.666	Decay all	0.836	Decay all	0.741
YOLOv7-AP-CBAM (ours)	Right side numbering	0.974	Right side numbering	0.983		
	Left side numbering	0.981	Left side numbering	0.978		
	Numbering all	0.977	Numbering all	0.980	Numbering all	0.978
	Right side decay	0.810	Right side decay	0.900	All decay	0.849
	Left side decay	0.857	Left side decay	0.833		
	All decay	0.833	All decay	0.866		
	Right side numbering	0.972	Right side numbering	0.993		
	Left side numbering	0.976	Left side numbering	0.978		
	Numbering all	0.974	Numbering all	0.985	Numbering all	0.979

## Discussion

In dentistry, achieving an accurate and precise diagnosis of dental caries is essential for the primary step towards successful treatment implementation. Dental caries diagnosis methods aim to improve the diagnostic process by facilitating the early detection of caries or enabling its objective identification [55]. Although tooth detection and numbering may seem straightforward, the accuracy of tooth numbering and caries detection can be affected by the limited and diverse knowledge of novice clinicians, potentially resulting in inconsistent diagnoses and inadequate treatment approaches. In addition to experience, practitioners' inattentiveness or demanding workloads can also have an impact on the diagnostic results. To prevent all these challenges and facilitate the diagnostic process, a computer-aided system based on DL methods can be employed. DL has gained significant popularity in the medical field due

to its outstanding problem-solving capabilities. Within this context, numerous DL architectures have been developed, with especially CNNs standing out for their impressive performance in image recognition tasks [56]. By utilizing CNN models, practitioners can effectively enhance work efficiency and achieve precise outcomes in the accurate diagnostic process [34, 57].

In the current study, the efficiency of the improved CNN model YOLOv7-AP-CBAM is presented for automatically detecting and numbering teeth, as well as detecting caries in digital bitewing radiographs, concurrently. So far, various studies have been conducted in the field of dentistry, utilizing CNNs for the detection, classification, and segmentation of various entities such as objects, tooth, caries, numbering, root fracture, periapical lesions and periodontal bone loss [34, 58–63]. According to these studies, CNN-based systems showed promising results in overcoming diagnosing difficulties and assisting practitioners.



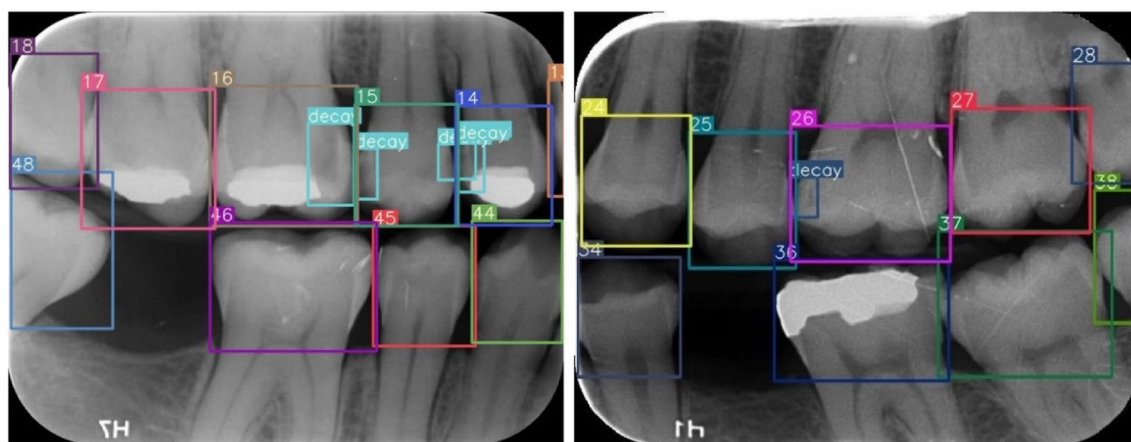
Various dental imaging systems have been used for these studies. Digital bitewing radiography technique is widely employed due to its practicality and advantages. Less radiation doses requirement, time efficiency, image storing, simplicity of image enhancing and processing are some of the advantages and reason for preference of digital bitewing radiography technique [64]. In the research papers, there were numbers of study utilizing digital bitewing radiography [19, 20, 65–69]. It is a more beneficial technique than other x-ray containing methods for the simultaneous evaluation of the crowns and alveolar bone levels of posterior teeth in both upper and lower dental arch, caries detection that especially in approximal caries, and teeth detection and numbering [10, 70, 71].

Also, there have been studies of the implementation of CNN to bitewing, panoramic, periapical, cone beam computed tomography radiographs and various diagnostic methods [34, 72, 73]. Zhang et al. [73] applied a DL method based on CNN for teeth detection and classification in periapical radiographs. Both precision and recall scores were high and, 95.8% and 96.1%, respectively. In our study, the improved model achieved recall of 99.4% and precision of 98.75%, in digital bitewing radiographs. In another study on CBCT radiographs reported that implementing a 3D U-Net network model showed satisfactory results for the segmentation and classification of teeth. Furthermore, an AI-based segmentation of all teeth in a single scan resulted in a speed improvement of 1800 times compared to that of an expert [72]. Tuzoff et al. [34] utilized the Faster R-CNN model for tooth detection and the VGG-16 Net model for tooth numbering on panoramic radiographs. The CNN-based models demonstrated high recall and specificity values for teeth detection and numbering tasks, comparable to experts. However, their model exhibited misclassifications in cases where a tooth was missing, whereas our improved model accurately detected these cases. Image samples for these cases are given in Fig. 11.

A DL method Faster R-CNN proposed for tooth detection and numbering in periapical radiographs by Chen et al. [57]. In order to improve the detection accuracy, they proposed three post-processing techniques to complement the baseline Faster R-CNN. Their model achieved precision of 98.8% and recall of 98.5% for detection, and precision of 91.7% and recall of 91.4% for numbering. In their study, the model was unable to accurately predict teeth located at the corners of the bitewing image. However, our improved model demonstrated a precision of 98.75% and a recall of 99.4% for detection, as well as a precision of 98.5% and a recall of 97.4% for numbering.

In literature, researchers generally benefit image enhancement and cropping techniques to improve accuracy. Mao et al. [74] improved the new Alexnet model, a CNN-based approach for caries and restoration detection on bitewing images, by utilizing image enhancement and cropping techniques. The model showed accuracy of 90.30% for caries detection and 95.56% for restoration detection. However, it is important to note that high accuracy scores alone may not indicate the reliability of a method. If a CNN model labels every surface as "decayed," it can achieve perfect accuracy but lack meaningful results. To address this, recall and specificity scores are reported in our study, following the precedent set by previous research, to provide a more comprehensive evaluation without sacrificing interpretability. Besides, we did not use any image enhancement or cropping methods.

In a previous study, an automatic assistance system developed for detecting caries lesions in bitewing radiographs using a U-Net model. The system's performance was evaluated through a comparative analysis with dentists. According to the findings of this study, the model demonstrated an overall accuracy of 80%, while the mean accuracy of practitioners was 71%. The recall, specificity, and F1-score values were also high, at 75%, 83%, and 73%, respectively



**Fig. 11** Sample of correct teeth numbering in cases where a tooth was missing

[20]. Although the results of the model in this study were not directly compared with the performance of human experts, it showed promising results in the detection of caries lesions, achieving recall, precision and F1 score values of 83.3%, 86.6% and 82.2%, respectively. These results are very similar to those reported in the study by Cantu et al. In addition, the present study evaluated the performance of tooth numbering simultaneously with the detection of caries lesions.

Yasa et al. [29] proposed a Faster R-CNN model for the detection and numbering of teeth in bitewing radiographs. They conducted a comparison between two different models for tooth numbering. Initially, the teeth were not separated into right and left sides. However, after splitting and training the model in this manner, promising results were obtained, with recall, precision, and F1-score values of 97.48%, 92.93%, and 95.15%, respectively. In our study, we adopted the approach of separating teeth into right and left sides, following a similar methodology. Through the adoption of this approach, our improved model exhibited promising performance, as indicated by notable recall, precision, and F1-score values of 97.4%, 98.5%, and 97.9%, respectively.

In another study, Bayraktar and Ayan [19] reported detection of interproximal caries lesions in digital bitewing radiographs with YOLOv3 model. In that study, the recall was 72.26%, precision was 86.58%. According to these promising results, the use of CNN-based system could be useful in diagnosing caries lesions. In the current study, our improved model is YOLOv7-AP-CBAM model with implementation of CBAM module. Through to CBAM module, the model could focus more precisely. In this way, the model reached recall of 83.3% which is higher than the previous study. However, the precision values were nearly same which are 86.58% and 86.6%.

The present study, YOLOv7-AP-CBAM model exhibited promising results in terms of the evaluated parameters. In the confusion matrix that we created, while the specificity score is higher, contrary effect was noted in the recall score. This result showed that the improved model can accurately predict tooth numbering while also detecting sound surfaces, similar to the study by Lee et al. [22]. The overall accuracy, recall, specificity, precision and F1-score values were 0.934, 0.834, 0.961, 0.851 and 0.842, respectively.

The performance of a CNN model is directly dependent on the amount of training data. The use of large and diverse data sets in the training process improves the consistency and accuracy of the CNN models. Therefore, the quality and quantity of training data play a vital role in ensuring the effectiveness of CNNs [75]. In this study, training dataset was relatively small. Although the training dataset contains an equal number of right and left bitewing images, the model was more successful on the left side in detecting caries in the final results. This can be explained by the higher amount of caries in the bitewing radiographs

on the left side in the training group. To avoid this problem, new bitewing images including caries may be added to the right-side training set. Dataset bias is another problem that occurs when specific samples in a dataset are over or under-represented. In order to avoid this problem, the dataset was generated images with taken from the same devices in various clinics with different resolutions and exposure times. However, the images obtained from different x-ray machines and from people with different demographic characteristics may affect the results. The proposed model's parameters were determined using validation data during the training phase, and the model's performance on unseen data was evaluated using different test data from the training and validation data. CNN models require high hardware requirements, especially in the training phase. In our study, we performed our training and tests on 1080 Ti graphic card. Computational performance will be higher on more advanced graphics cards. The dental caries lesions that were not evaluated using the gold standard method of histological assessment detected and labeled by the annotators on the radiographic images [76]. The model encountered challenges in accurately predicting teeth positioned at the corners of the bitewing image, as opposed to other teeth. This could be related to the reduced clarity of these specific teeth and the limited amount of data in the training set. Also, the improved YOLOv7-AP-CBAM model was not compared with an expert. Despite the limitations of the study, it is worth noting that there are several significant strengths present as well. The present study offered a comprehensive analysis by integrating the results of tooth detection, numbering, and caries detection. This comprehensive approach facilitated a more cohesive comprehension of the outcomes. Additionally, the experimental design included the split of data into right and left sides, enabling a more detailed analysis. This approach enhanced the performance metrics of our results. Furthermore, by adopting an interdisciplinary approach, our study introduced a novel perspective and makes an innovative contribution to previous papers.

## Conclusion

In this study, an improved YOLOv7-AP-CBAM model was proposed for accurate teeth detection, numbering and caries detection, concurrently assessing with which tooth had caries by matching the numbered teeth and the detected caries. For teeth numbering and caries detection matching performance; the accuracy, recall, specificity, precision and F1—Score values 0.934, 0.834, 0.961, 0.851 and 0.842, respectively. The model exhibited high recall, specificity, precision, and F1-score, highlighting the potential of CNNs. This result

revealed that CNNs can provide valuable support to clinicians by automating the detection and numbering of teeth, as well as detection of caries on bitewing radiographs. This can potentially contribute for assessment, improve overall performance, and ultimately save precious time.

**Acknowledgements** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author contribution** B.A: Data acquisition and labeling, Tests, Formal analysis, Writing – original draft, Writing – review & editing.

E.A: Conceptualization, Software, Visualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing.

Y.B: Conceptualization, Methodology, Data acquisition and labeling, Writing – review & editing.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). No funding was obtained for this study.

## Declarations

**Ethics approval** All procedures performed in studies involving human participants were in accordance with ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by Ethical Committee of Kirikkale University (Date: 09/11/2022, Decision Number: 2022.11.14).

**Consent to participate** For this type of study, formal consent is not required.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Featherstone JDB (2000) The science and practice of caries prevention. *J Am Dent Assoc* 131:887–899. <https://doi.org/10.14219/jada.archive.2000.0307>
2. Mortensen D, Dannemand K, Twetman S, Keller MK (2014) Detection of non-cavitated occlusal caries with impedance spectroscopy and laser fluorescence: an in vitro study. *Open Dent J* 8:28–32. <https://doi.org/10.2174/1874210601408010028>
3. Pitts NB (2004) Are we ready to move from operative to non-operative/preventive treatment of dental caries in clinical practice? *Caries Res* 38:294–304. <https://doi.org/10.1159/000077769>
4. Baelum V, Heidmann J, Nyvad B (2006) Dental caries paradigms in diagnosis and diagnostic research. *Eur J Oral Sci* 114:263–277. <https://doi.org/10.1111/j.1600-0722.2006.00383.x>
5. Pitts NB, Stamm JW (2004) International consensus workshop on caries clinical trials (ICW-CCT)—final consensus statements: agreeing where the evidence leads. *J Dent Res* 83:125–128. <https://doi.org/10.1177/154405910408301s27>
6. Selwitz RH, Ismail AI, Pitts NB (2007) Dental caries. *The Lancet* 369:51–59. [https://doi.org/10.1016/S0140-6736\(07\)60031-2](https://doi.org/10.1016/S0140-6736(07)60031-2)
7. Chan M, Dadul T, Langlais R, Russell D, Ahmad M (2018) Accuracy of extraoral bite-wing radiography in detecting proximal caries and crestal bone loss. *J Am Dent Assoc* 149:51–58. <https://doi.org/10.1016/j.adaj.2017.08.032>
8. Vandenbergh B, Jacobs R, Bosmans H (2010) Modern dental imaging: a review of the current technology and clinical applications in dental practice. *Eur Radiol* 20:2637–2655. <https://doi.org/10.1007/s00330-010-1836-1>
9. Baelum V (2010) What is an appropriate caries diagnosis? *Acta Odontol Scand* 68:65–79. <https://doi.org/10.3109/00016350903530786>
10. Kamburoğlu K, Kolsuz E, Murat S, Yüksel S, Özen T (2012) Proximal caries detection accuracy using intraoral bitewing radiography, extraoral bitewing radiography and panoramic radiography. *Dentomaxillofacial Radiol* 41:450–459. <https://doi.org/10.1259/dmfr/30526171>
11. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM (2020) The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofacial Radiol* 49:20190107. <https://doi.org/10.1259/dmfr.20190107>
12. Suzuki K (2017) Overview of deep learning in medical imaging. *Radiol Phys Technol* 10:257–273. <https://doi.org/10.1007/s12194-017-0406-5>
13. Rodrigues JA, Krois J, Schwendicke F (2021) Demystifying artificial intelligence and deep learning in dentistry. *Braz Oral Res* 35. <https://doi.org/10.1590/1807-3107bor-2021.vol35.0094>
14. Barr A, Feigenbaum EA, Cohen PR (1981) The handbook of artificial intelligence. In: *Artificial Intelligence*, William Kaufman Inc, California, pp 3–11
15. Lee J-G, Jun S, Cho Y-W, Lee H, Kim GB, Seo JB, Kim N (2017) Deep learning in medical imaging: general overview. *Korean J Radiol* 18:570. <https://doi.org/10.3348/kjr.2017.18.4.570>
16. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
17. Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243. <https://doi.org/10.1113/jphysiol.1968.sp008455>
18. Schwendicke F, Samek W, Krois J (2020) Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 99:769–774. <https://doi.org/10.1177/0022034520915714>
19. Bayraktar Y, Ayan E (2022) Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin Oral Investig* 26:623–632. <https://doi.org/10.1007/s00784-021-04040-1>
20. Cantu AG, Gehring S, Krois J, Chaurasia A, Rossi JG, Gaudin R, Elhennawy K, Schwendicke F (2020) Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent* 100:103425. <https://doi.org/10.1016/j.jdent.2020.103425>
21. Choi J, Eun H, Kim C (2018) Boosting proximal dental caries detection via combination of variational methods and convolutional neural network. *J Signal Process Syst* 90:87–97. <https://doi.org/10.1007/s11265-016-1214-6>
22. Lee J-H, Kim D-H, Jeong S-N, Choi S-H (2018) Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 77:106–111. <https://doi.org/10.1016/j.jdent.2018.07.015>



23. Moutselos K, Berdouses E, Oulis C, Maglogiannis I (2019) Recognizing occlusal caries in dental intraoral images using deep learning. In: 2019 41st Annu Int Conf IEEE Eng Med Biol Soc EMBC, IEEE, pp. 1617–1620. <https://doi.org/10.1109/EMBC.2019.8856553>
24. Prajapati SA, Nagaraj R, Mitra S (2017) Classification of dental diseases using CNN and transfer learning. In: 2017 5th Int Symp Comput Bus Intell ISCBI, IEEE, pp. 70–74. <https://doi.org/10.1109/ISCBI.2017.8053547>
25. Vidnes-Kopperud S, Tveit AB, Espelid I (2011) Changes in the treatment concept for approximal caries from 1983 to 2009 in Norway. *Caries Res* 45:113–120. <https://doi.org/10.1159/000324810>
26. Moran M, Faria M, Giraldo G, Bastos L, Oliveira L, Conci A (2021) Classification of approximal caries in bitewing radiographs using convolutional neural networks. *Sensors* 21:5192. <https://doi.org/10.3390/s21155192>
27. Srivastava MM, Kumar P, Pradhan L, Varadarajan S (2017) Detection of tooth caries in bitewing radiographs using deep learning. <http://arxiv.org/abs/1711.07312>. Accessed 7 July 2023
28. Leite AF, Gerven AV, Willems H, Beznik T, Lahoud P, Gaêta-Araujo H, Vranckx M, Jacobs R (2021) Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clin Oral Investig* 25:2257–2267. <https://doi.org/10.1007/s00784-020-03544-6>
29. Yasa Y, Çelik Ö, Bayrakdar IS, Pekince A, Orhan K, Akarsu S, Atasoy S, Bilgir E, Odabaş A, Aslan AF (2021) An artificial intelligence proposal to automatic teeth detection and numbering in dental bite-wing radiographs. *Acta Odontol Scand* 79:275–281. <https://doi.org/10.1080/00016357.2020.1840624>
30. Bilgir E, Bayrakdar İŞ, Çelik Ö, Orhan K, Akkoca F, Sağlam H, Odabaş A, Aslan AF, Özçetin C, Kılılı M, Rozylo-Kalinowska I (2021) An artificial intelligence approach to automatic tooth detection and numbering in panoramic radiographs. *BMC Med Imaging* 21:124. <https://doi.org/10.1186/s12880-021-00656-7>
31. Kaya E, Gunec HG, Gokyay SS, Kutal S, Gulum S, Ates HF (2022) Proposing a CNN method for primary and permanent tooth detection and enumeration on pediatric dental radiographs. *J Clin Pediatr Dent* 46:293. <https://doi.org/10.22514/1053-4625-46.4.6>
32. Kılıc MC, Bayrakdar IS, Çelik Ö, Bilgir E, Orhan K, Aydın OB, Kaplan FA, Sağlam H, Odabaş A, Aslan AF, Yılmaz AB (2021) Artificial intelligence system for automatic deciduous tooth detection and numbering in panoramic radiographs. *Dentomaxillofacial Radiol* 50:20200172. <https://doi.org/10.1259/dmfr.20200172>
33. Tekin BY, Özcan C, Pekince A, Yasa Y (2022) An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs. *Comput Biol Med* 146:105547. <https://doi.org/10.1016/j.compbiomed.2022.105547>
34. Tuzoff DV, Tuzova LN, Bornstein MM, Krasnov AS, Kharchenko MA, Nikolenko SI, Svishnikov MM, Bednenko GB (2019) Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofacial Radiol* 48:20180051. <https://doi.org/10.1259/dmfr.20180051>
35. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25:44–56. <https://doi.org/10.1038/s41591-018-0300-7>
36. Masood M, Masood Y, Newton JT (2015) The clustering effects of surfaces within the tooth and teeth within individuals. *J Dent Res* 94:281–288. <https://doi.org/10.1177/0022034514559408>
37. Chen X, Guo J, Ye J, Zhang M, Liang Y (2022) Detection of proximal caries lesions on bitewing radiographs using deep learning method. *Caries Res* 56:455–463. <https://doi.org/10.1159/000527418>
38. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2016) densely connected convolutional networks. <https://doi.org/10.48550/ARXIV.1608.06993>
39. Girshick R (2015) Fast R-CNN. In: 2015 IEEE Int. Conf. Comput. Vis. ICCV, IEEE, Santiago, Chile, pp. 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
40. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39:1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
41. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. In: 2017 IEEE Int Conf Comput Vis ICCV, IEEE, Venice, pp. 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>
42. Terven J, Cordova-Esparza D (2023) A Comprehensive Review of YOLO: From YOLOv1 and Beyond. <http://arxiv.org/abs/2304.00501> (accessed July 7, 2023)
43. Wang C-Y, Bochkovskiy A, Liao H-YM (2022) YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. <https://doi.org/10.48550/ARXIV.2207.02696>
44. Ding X, Zhang X, Ma N, Han J, Ding G, Sun J (2021) RepVGG: Making VGG-style ConvNets Great Again. In: 2021 IEEE/CVF Conf Comput Vis Pattern Recognit CVPR, IEEE, Nashville, TN, USA, pp. 13728–13737. <https://doi.org/10.1109/CVPR46437.2021.01352>
45. Larochelle H, Hinton G (2010) Learning to combine foveal glimpses with a third-order Boltzmann machine. *Adv Neural Inform Process Syst* 23:1243–1251
46. Woo S, Park J, Lee J-Y, Kweon IS (2018) CBAM: Convolutional Block Attention Module. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (Eds.). *Comput Vis—ECCV 2018*, Springer International Publishing, Cham, pp. 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
47. Shi J, Yang J, Zhang Y (2022) Research on steel surface defect detection based on YOLOv5 with attention mechanism. *Electronics* 11:3735. <https://doi.org/10.3390/electronics11223735>
48. Xue Z, Xu R, Bai D, Lin H (2023) YOLO-tea: a tea disease detection model improved by YOLOv5. *Forests* 14:415. <https://doi.org/10.3390/f14020415>
49. De Moraes JL, De Oliveira Neto J, Badue C, Oliveira-Santos T, De Souza AF (2023) Yolo-papaya: a papaya fruit disease detector and classifier using cnns and convolutional block attention modules. *Electronics* 12:2202. <https://doi.org/10.3390/electronics12102202>
50. Yan J, Zhou Z, Zhou D, Su B, Xuanyuan Z, Tang J, Lai Y, Chen J, Liang W (2022) Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling. *Front Mar Sci* 9:1056300. <https://doi.org/10.3389/fmars.2022.1056300>
51. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 32:8024–8035
52. Chollet F et al (2015) Keras. <https://github.com/fchollet/keras>. Accessed 7 July 2023
53. Bradski G (2000) The OpenCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25:120–123
54. Zhang Y, Li J, Fu W, Ma J, Wang G (2023) A lightweight YOLOv7 insulator defect detection algorithm based on DSC-SE. *PLoS One* 18:e0289162. <https://doi.org/10.1371/journal.pone.0289162>
55. Gomez J (2015) Detection and diagnosis of the early caries lesion. *BMC Oral Health* 15:S3. <https://doi.org/10.1186/1472-6831-15-S1-S3>
56. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O (2018) Deep learning with convolutional neural network in radiology. *Jpn J Radiol* 36:257–272. <https://doi.org/10.1007/s11604-018-0726-3>



57. Chen H, Zhang K, Lyu P, Li H, Zhang L, Wu J, Lee C-H (2019) A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep* 9:3840. <https://doi.org/10.1038/s41598-019-40414-y>
58. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, Krejci I, Markram H (2019) Caries detection with near-infrared transillumination using deep learning. *J Dent Res* 98:1227–1233. <https://doi.org/10.1177/0022034519871884>
59. Fukuda M, Inamoto K, Shibata N, Arijji Y, Yanashita Y, Kutsuna S, Nakata K, Katsumata A, Fujita H, Arijji E (2020) Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol* 36:337–343. <https://doi.org/10.1007/s11282-019-00409-x>
60. Lee J-H, Kim D, Jeong S-N, Choi S-H (2018) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontol Implant Sci* 48:114. <https://doi.org/10.5051/jpis.2018.48.2.114>
61. Murata M, Arijji Y, Ohashi Y, Kawai T, Fukuda M, Funakoshi T, Kise Y, Nozawa M, Katsumata A, Fujita H, Arijji E (2019) Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiol* 35:301–307. <https://doi.org/10.1007/s11282-018-0363-7>
62. Park J-H, Hwang H-W, Moon J-H, Yu Y, Kim H, Her S-B, Srinivasan G, Aljanabi MNA, Donatelli RE, Lee S-J (2019) Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod* 89:903–909. <https://doi.org/10.2319/022019-127.1>
63. Pauwels R, Brasil DM, Yamasaki MC, Jacobs R, Bosmans H, Freitas DQ, Haite-Neto F (2021) Artificial intelligence for detection of periapical lesions on intraoral radiographs: Comparison between convolutional neural networks and human observers. *Oral Surg. Oral Med Oral Pathol Oral Radiol* 131:610–616. <https://doi.org/10.1016/j.oooo.2021.01.018>
64. Rockenbach MI, Veeck EB, da Costa NP (2008) Detection of proximal caries in conventional and digital radiographs: an in vitro study. *Stomatologija* 10:115–120
65. Abdinian M, Razavi SM, Faghihian R, Samety AA, Faghihian E (2015) Accuracy of digital bitewing radiography versus different views of digital panoramic radiography for detection of proximal caries. *J Dent Tehran Iran* 12:290–297
66. Hintze H (1993) Screening with conventional and digital bitewing radiography compared to clinical examination alone for caries detection in low-risk children. *Caries Res* 27:499–504. <https://doi.org/10.1159/000261588>
67. Peker İ, ToramanAlkurt M, Altunkaynak B (2007) Film tomography compared with film and digital bitewing radiography for proximal caries detection. *Dentomaxillofacial Radiol* 36:495–499. <https://doi.org/10.1259/dmfr/13319800>
68. Schaefer G, Pitchika V, Litzenburger F, Hickel R, Kühnisch J (2018) Evaluation of occlusal caries detection and assessment by visual inspection, digital bitewing radiography and near-infrared light transillumination. *Clin Oral Investig* 22:2431–2438. <https://doi.org/10.1007/s00784-018-2512-0>
69. Wenzel A (2004) Bitewing and digital bitewing radiography for detection of caries lesions. *J Dent Res* 83:72–75. <https://doi.org/10.1177/154405910408301s14>
70. Mahoor MH, Abdel-Mottaleb M (2005) Classification and numbering of teeth in dental bitewing images. *Pattern Recognit* 38:577–586. <https://doi.org/10.1016/j.patcog.2004.08.012>
71. Reed BE, Polson AM (1984) Relationships between bitewing and periapical radiographs in assessing Crestal Alveolar Bone Levels. *J Periodontol* 55:22–27. <https://doi.org/10.1902/jop.1984.55.1.22>
72. Shaheen E, Leite A, Alqahtani KA, Smolders A, Van Gerven A, Willems H, Jacobs R (2021) A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study. *J Dent* 115:103865. <https://doi.org/10.1016/j.jdent.2021.103865>
73. Zhang K, Wu J, Chen H, Lyu P (2018) An effective teeth recognition method using label tree with cascade network structure. *Comput Med Imaging Graph* 68:61–70. <https://doi.org/10.1016/j.compmedimag.2018.07.001>
74. Mao Y-C, Chen T-Y, Chou H-S, Lin S-Y, Liu S-Y, Chen Y-A, Liu Y-L, Chen C-A, Huang Y-C, Chen S-L, Li C-W, Abu PAR, Chiang W-Y (2021) Caries and restoration detection using bitewing film based on transfer learning with CNNs. *Sensors* 21:4613. <https://doi.org/10.3390/s21134613>
75. Koppányi Z, Iwaszczuk D, Zha B, Saul CJ, Toth CK, Yilmaz A (2019) Multimodal semantic segmentation: fusion of rgb and depth data in convolutional neural networks, in: multimodal scene underst., Elsevier, pp. 41–64. <https://doi.org/10.1016/B978-0-12-817358-9.00009-3>
76. Subka S, Rodd H, Nugent Z, Deery C (2019) In vivo validity of proximal caries detection in primary teeth, with histological validation. *Int J Paediatr Dent* 29:429–438. <https://doi.org/10.1111/ipd.12478>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.