IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 February 2025, accepted 18 March 2025, date of publication 20 March 2025, date of current version 28 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3553228

RESEARCH ARTICLE

Optimized Multi-Scale Detection and Numbering of Teeth in Panoramic Radiographs Using DentifyNet

SALIH TAHA ALPEREN ÖZÇELIK[1], HÜSEYİN ÜZEN[2], ABDULKADIR SENGUR[3], MUAMMER TÜRKOĞLU[4], ADALET ÇELEBI[5], AND NEBRAS M. SOBAHI[6]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Bingöl University, 12000 Bingöl, Türkiye
[2]Department of Computer Engineering, Faculty of Engineering, Bingöl University, 12000 Bingöl, Türkiye
[3]Department of Electrical and Electronics Engineering, Faculty of Technology, Fırat University, 23000 Elâzığ, Türkiye
[4]Department of Software Engineering, Samsun University, 55000 Samsun, Türkiye
[5]Oral and Maxillofacial Surgery Department, Faculty of Dentistry, Mersin University, 33000 Mersin, Türkiye
[6]Department of Electrical and Electronics Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Abdulkadir Sengur (ksengur@firat.edu.tr)

This work was supported by Fırat University, Scientific Research Project Committee, under Grant TEKF.24.46.
This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Fırat University Non-Interventional Research Ethics Committee under Application BSA8797E1H, 2023/13-37.

ABSTRACT Manual tooth detection and numbering in panoramic radiographs are time-consuming and prone to human errors, negatively impacting diagnostic accuracy and treatment outcomes. These challenges necessitate robust automated solutions to improve efficiency and precision in dental imaging. This study introduces DentifyNet, a novel deep learning architecture designed for automatic tooth detection and numbering in panoramic radiography images. DentifyNet integrates a customized Faster R-CNN framework with Feature Pyramid Networks (FPN), flexible anchor structures, and RoI Align to enhance detection precision. The model was trained and evaluated on 468 panoramic radiographs annotated by dental experts using the FDI numbering system. Experimental results demonstrate that DentifyNet achieved state-of-the-art performance with an Intersection over Union (IoU) of 86.42%, precision of 97.52%, recall of 97.49%, F1-score of 97.51%, and accuracy of 97.50%. The architecture effectively detects challenging cases, such as adjacent similar teeth and missing teeth. These findings suggest that DentifyNet surpasses standard Faster R-CNN architectures, offering a reliable solution for automated tooth detection and numbering. Future research will focus on utilizing broader datasets and architectural advancements to address current limitations and expand clinical applications.

INDEX TERMS Computer-aided diagnostics, convolutional neural networks, tooth detection, tooth numbering, panoramic radiography.

I. INTRODUCTION

Tooth identification and numbering play a critical role in dentistry practice, in terms of diagnosis and treatment planning. When correct tooth identification is not made, serious medical errors, especially incorrect tooth extraction, can occur, leading to consequences that negatively affect the quality of patient care [1], [2]. In addition, tooth detection is widely used in identification studies in forensic medicine, since the morphological structure of the teeth is unique to the individual, it also has an important place in archaeological and forensic studies [3], [4]. In recent years, advances in imaging technologies have enabled the more widespread use of panoramic radiography in dentistry [5], [6]. Panoramic radiography is a method that shows both the upper and lower jaw in a single image and offers significant advantages such as low radiation dose, short imaging time, and minimal burden for the patient [7]. Thanks to these features, panoramic radiography has become a routine diagnostic tool in dentistry. The ability to see both teeth and their surrounding structures

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

together provides clinicians with a wide range of evaluation opportunities during the diagnosis and treatment process.

Deep learning and artificial intelligence technologies have made great progress in areas such as tooth detection and numbering [8], [9]. Deep learning algorithms help clinicians make faster and more accurate diagnoses, especially by automating complex and time-consuming processes. Deep learning techniques such as Convolutional Neural Networks (CNN) extract features at various levels from raw image data, providing high accuracy in tasks such as object detection, classification, and segmentation [10]. CNN architectures are widely used in medical image analysis, and successful results have been achieved in sensitive tasks such as tooth detection and numbering [11].

Tooth detection and numbering in panoramic radiography images are usually done manually and this process is time-consuming for experts. However, human errors are also common during these manual processes; errors such as incorrect numbering or missing some teeth can occur, especially due to excessive workload, lack of experience, or distraction [12], [13]. Such errors can negatively affect the treatment process and may also lead to wrong treatment decisions for patients. CNN and similar deep learning algorithms automate this process, allowing teeth to be detected and numbered correctly. In addition, the rapid results of these technologies save time in dentistry practices and accelerate the treatment process of patients.

A. LITERATURE REVIEW

Recent advancements in tooth identification and numbering have been significantly enhanced by deep learning methodologies. Models for object recognition that stand out in this field include Faster Region-Based CNN (R-CNN) [14] and You Only Look Once (YOLO) [15] These models have provided critical improvements in terms of speed and accuracy in tooth detection tasks and have been widely accepted in the field of dental imaging. Faster R-CNN is derived from R-CNN and Fast R-CNN and has a structure that makes the object detection process more efficient. This model has achieved high accuracy rates by combining region proposals and detection processes in sensitive tasks such as tooth detection. Especially in dental panoramic radiography images, modified versions of Faster R-CNN are widely used in tooth detection and numbering tasks. The YOLO model stands out with its speed in the object detection process. YOLO's architecture significantly speeds up the detection process by analyzing an image in a single pass, making the model ideal for real-time applications. Newer versions such as YOLOv4 and YOLOv5 have increased accuracy as well as speed, allowing effective results to be obtained on large data sets in dentistry applications. YOLO's success in tooth detection and numbering is due to the model's ability to maintain accuracy despite operating at high speed. These features make this model stand out, especially in the analysis of large-scale dental images.

In this literature review, we will examine the results of studies in the field of tooth detection and numbering and the success of these studies with deep learning models. Laishram et al. [16] proposed an improved Faster R-CNN-based model for accurate tooth segmentation in dental images. The proposed modified model was used on 145 Orthopantomogram (OPG) images and a 92% accuracy and 91% mAP evaluation scores were obtained. Top-level results such as 99.8% in AP50 and 92.7% in AP75 were obtained. Kaya et al. [17] proposed a Yolov4-based model for the detection of permanent tooth germs. This model strengthened with the CSPDarknet53 backbone, was tested on 4518 panoramic radiograph images and reached a 90% F1 score with 94.16% average precision. Bumann et al. [18] utilized the Mask R-CNN model for segmenting teeth and dental fillings. The study achieved a mean Average Precision (mAP) of 95.32% and an F1 score of 92.5%. These results were obtained from experiments conducted on a dataset comprising 368 publicly available panoramic radiographs, supplemented by 80 additional private images sourced from the University of Missouri-Kansas City, School of Dentistry, for more complex segmentation tasks. The use of these additional private images aimed to enhance the model's robustness in handling diverse dental conditions, contributing to its high segmentation accuracy. Zhang et al. [19] proposed a modified Faster R-CNN model to detect and classify tooth locations. The authors reported a sensitivity of 95.1%, a recall of 95.5%, and an F1 score of 95.3%, based on trials conducted on 1,250 dental periapical X-ray images. Privado et al. [20] investigated the presence or lack of teeth in their study. 8000 dental panoramic radiograph images were used for object recognition using the Matterport Mask R-CNN and classification with the CNN model. Consequently, 99.24% total accuracy was achieved. Bilgir et al. [8] used automated teeth detection and numbering in their investigation. Experiments with 2482 panoramic radiograph images using the Faster R-CNN Inception V2 model yielded 95.59% sensitivity, 96.52% accuracy, and 96.06% F1 score. Zhang et al. [21] employed a VGG16-based model to recognize and classify teeth in periapical radiography images. The authors reported 98% accuracy, 98.3% recall, and 98.1% F1 score values in their experiments. Putra et al. [22] performed automatic tooth detection and numbering in panoramic radiograph images. The authors employed the CSPDarknet53 backbone of the YOLOv4 model, which was evaluated on 500 panoramic radiographs. In the studies, 88.5% accuracy, 87.7% precision, 100% sensitivity, and a 93.51% F1 score were achieved. Choi et al. [23] investigated tooth detection and dental treatment patterns. The EfficientDet D3 model was considered by the authors to assess 1683 panoramic radiograph images. During the studies, 99.1% sensitivity and 97.4% recall were achieved. Chen et al. [24] proposed a model for detecting tooth location, shape, and bone level in periapical radiographs. The model was built on a hybrid deep-learning architecture that included YOLOv5, VGG-16, and U-Net.

In experimental studies, 8000 periapical radiograph images were employed, with 88.8% accuracy for tooth location recognition, 90.4% for tooth form detection, and 85.7% for bone level detection. Etsai et al. [25] also performed automatic tooth detection and numbering in OPG images. ROI detection and segmentation were performed with U-Net, and detection was performed with Faster R-CNN and tested on 591 OPG images. The results showed that ROI detection had a 71% IoU, 99.2% recall, and 99.4% precision. Yuksel et al. [26] performed tooth detection and numbering on panoramic radiograph images. The authors developed an innovative model, Dentect, based on YOLO, and tested it on 1005 panoramic radiographs. For the tooth numbering model, 89.1% AP (0.5 thresholds) and 47.4% AP (0.5:0.95) were achieved. Yilmaz et al. [27] used panoramic radiographs to detect and classify teeth. 1200 panoramic radiographs were tested using the YOLOv4 and Faster R-CNN models. The YOLOv4 model obtained 99.9% accuracy, 99.18% recall, and 99.54% F1 score, whereas the Faster R-CNN model achieved 93.67% accuracy, 90.79% recall, and 92.21% F1 score. Gülüm et al. [28] carried out teeth detection and numbering in panoramic radiographs. The YOLOv4 model was used for experiments on 3000 panoramic radiographs. The F1 score was 83%, while the mAP was 81%. This study investigated the impact of dataset size on tooth numbering performance. Recent advancements in deep learning, particularly the integration of transformer-based architectures and hybrid models, have significantly improved the accuracy and efficiency of tooth detection and numbering tasks. A recent study [29] by applied DETR to detect, segment, and number teeth in panoramic radiographs, achieving a mAP of 0.69 without data augmentation. When data augmentation techniques were applied, the mAP improved to 0.82. Additionally, the study utilized inpainting techniques with stable diffusion to generate synthetic panoramic images, further enhancing the model's performance mAP of 0.76. Hybrid models combining CNNs and Transformers have shown promise in dental imaging. For instance, a study by [30] proposed a robust hybrid learning framework for automatic teeth segmentation and labeling on 3D dental models. The model leverages the spatial feature extraction capabilities of CNNs and the global context understanding of transformers, achieving high precision in tooth detection and numbering. A study by [31] proposed an improved object detection method for dental X-ray images using a hybrid architecture that combines YOLOv5, VGG-16, and U-Net. The model achieved high accuracy in tooth location recognition (88.8%), tooth shape detection (90.4%), and bone level detection (85.7%). The hybrid architecture leverages the strengths of each component, with YOLOv5 providing real-time detection capabilities and U-Net offering precise segmentation.

B. GAPS IN THE LITERATURE AND CONTRIBUTIONS OF THE STUDY

One of the most significant gaps in the literature is the prevalent practice of labeling all teeth under a single ''tooth'' class, without distinguishing individual teeth. This approach oversimplifies the detection process and fails to address the nuances of identifying and numbering all 32 individual teeth, as required in dental practice. Since each tooth exhibits distinct anatomical features and spatial relationships, a model capable of individual tooth-level classification is essential for accurate detection and numbering. However, most existing studies focus on global tooth detection rather than a detailed classification approach.

Standard models such as YOLO and Faster R-CNN rely on single-scale feature extraction, limiting their ability to capture fine-grained variations in tooth morphology. These models often struggle with detecting small or overlapping teeth due to their fixed-scale anchor structures and lack of adaptive multi-level feature representation. In contrast, DentifyNet integrates multi-scale feature extraction through a Feature Pyramid Network (FPN), allowing the model to preserve structural details across different tooth sizes and positions.

Another limitation of existing methods is the reliance on fixed anchor structures, which may not be well-suited for the irregular shapes and diverse orientations of teeth in panoramic radiographs. DentifyNet overcomes this issue by employing flexible-sized and proportionally adaptive anchor structures, which enhance detection accuracy, particularly for small and closely spaced teeth.

Additionally, while many previous studies have used a limited range of backbone architectures, comparative performance evaluations across various backbones remain sparse. Since backbone selection directly influences feature extraction quality, it is critical to assess different architectures to determine the most effective design for dental imaging applications.

In this study, considering the deficiencies in the literature, some important novelties in tooth detection and numbering processes are presented:

1. DentifyNet, was developed to overcome the limitations of standard models. This customized architecture enables more precise analysis of tooth structures, addressing challenges unique to dental imaging.
2. A Feature Pyramid Network (FPN) structure enhances multi-scale feature extraction, improving the model's ability to handle varying tooth sizes and positions.
3. Flexible-sized and proportionally adaptive anchor structures replace traditional fixed anchors, leading to more robust detection of small and closely positioned teeth.
4. A dataset of 468 panoramic radiographs, meticulously labeled by expert dentists using the FDI numbering system, was employed. Unlike prior studies that labeled teeth under a single class, this dataset includes labels for all 32 individual teeth, enabling comprehensive detection and numbering. This original dataset allowed the evaluation of the model's generalization capacity and applicability to a wide range of clinical scenarios.

In subsequent sections, this study provides a detailed evaluation of DentifyNet's performance. The Materials and Methods section describes the model architecture, dataset preparation, and training processes. The Results section presents a comprehensive analysis of the model's performance across various backbone architectures and datasets, highlighting its accuracy and overall success. Finally, the Discussion section compares the findings with prior studies, emphasizing the innovations introduced by DentifyNet in tooth detection and numbering tasks.

II. MATERIALS AND METHODS

In this study, a novel dataset consisting of panoramic radiographic labeled by expert dentists was used to evaluate the performance of the DentifyNet model developed in tooth detection and numbering processes. In the training and testing stages of the model, different backbone architectures and techniques were compared on this dataset.

A. PREPARING THE DATASET

The dataset used in this study consists of 500 panoramic radiographs, and all images were collected anonymously to protect patient confidentiality. The images were collected by expert dentists without distinguishing between the gender, age, and identity information of the patients. However, patients between the ages of 18-65 and patients with medical conditions such as cysts or jaw fractures were excluded, and a total of 468 panoramic radiographs were selected. These selected images were recorded in JPEG format, each with an average resolution of 2700 × 1316 pixels and 8-bit depth. All teeth were carefully labeled by an oral and maxillofacial surgeon using the FDI tooth numbering system. The FDI numbering system was selected for this study due to its widespread adoption and standardization in dental practices globally. This system provides a consistent framework for tooth identification and is commonly used by clinicians and researchers for diagnostic and treatment purposes. While alternative systems such as the Universal Numbering System are prevalent in certain regions, the FDI system's compatibility with international standards makes it a preferred choice for developing models aimed at broader applicability. The FDI system divides the oral cavity into four anatomical regions and numbers the teeth with a two-digit coding system. In this system, the teeth in the upper right jaw are represented as "1", the upper left jaw as "2", the lower left jaw as "3" and the lower right jaw as "4". Each tooth receives a second number according to its order in the region it is located in. For example, the second molar in the upper right jaw is numbered as "17", while the first incisor in the upper left jaw is labeled as "21". During the labeling process, each tooth is framed with a bounding box and recorded in the COCO [32] format. The COCO format specified the bounding boxes and categories of the teeth for each image, allowing the model to be trained with this data. These labels in the COCO format provided easy access and use of the dataset during the training and testing stages of the model. An image labeled according to the FDI numbering system by a specialist dentist is given in Figure 1.

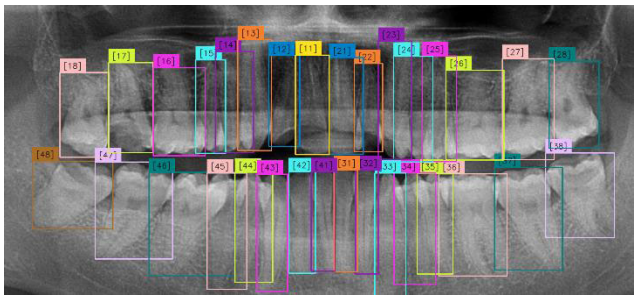

FIGURE 1. A panoramic radiography image labeled by an expert dentist.

B. PREPROCESSING

The labels of each image were processed in COCO format and converted to tensors to become the target structure to be used by the model. During the tensor conversion process, each pixel value was normalized between 0-1.

C. PROPOSED METHOD: DENTIFYNET

The developed DentifyNet model is a Faster R-CNN-based deep learning architecture specifically designed for tooth detection and numbering tasks. The model differs from the standard Faster R-CNN with several aspects such as multi-scale feature extraction, customized anchor structures, and integrated RoI Align mechanism. A detailed illustration of DentifyNet is given in Figure 2.

As seen in Figure 2, DentifyNet uses a ResNet-50-based FPN for multi-scale feature extraction. ResNet-50 was selected as the backbone for DentifyNet due to its strong performance in feature extraction for object detection tasks. Its architecture provides a balance between computational efficiency and depth, allowing it to effectively handle complex radiographic images. The use of skip connections ensures efficient gradient propagation, which is critical for stable training. While alternative backbones such as DenseNet and EfficientNet were considered, ResNet-50's proven reliability and consistent performance in similar imaging tasks made it the most suitable choice. FPN extracts feature maps from different layers of the network from C2 to C5 and combines these features to obtain richer and more diverse features. This structure facilitates the simultaneous detection of smaller and larger teeth and increases the overall sensitivity of the model. Especially in layers C2, C3, C4 and C5, the channel numbers are reduced to 256 with 1 × 1 convolutions to make them suitable for FPN. The obtained features are named P2 from C2, P3 from C3, P4 from C4, and P5 from C5 and each is combined to be transferred to RPN. DentifyNet uses customized anchor structures to detect teeth of different sizes and proportions in dental panoramic images. 15 anchors are used at each level in the model, resulting in a total of 60 anchors. These anchors are defined in different sizes and ratios. Anchor boxes are created with pixel sizes

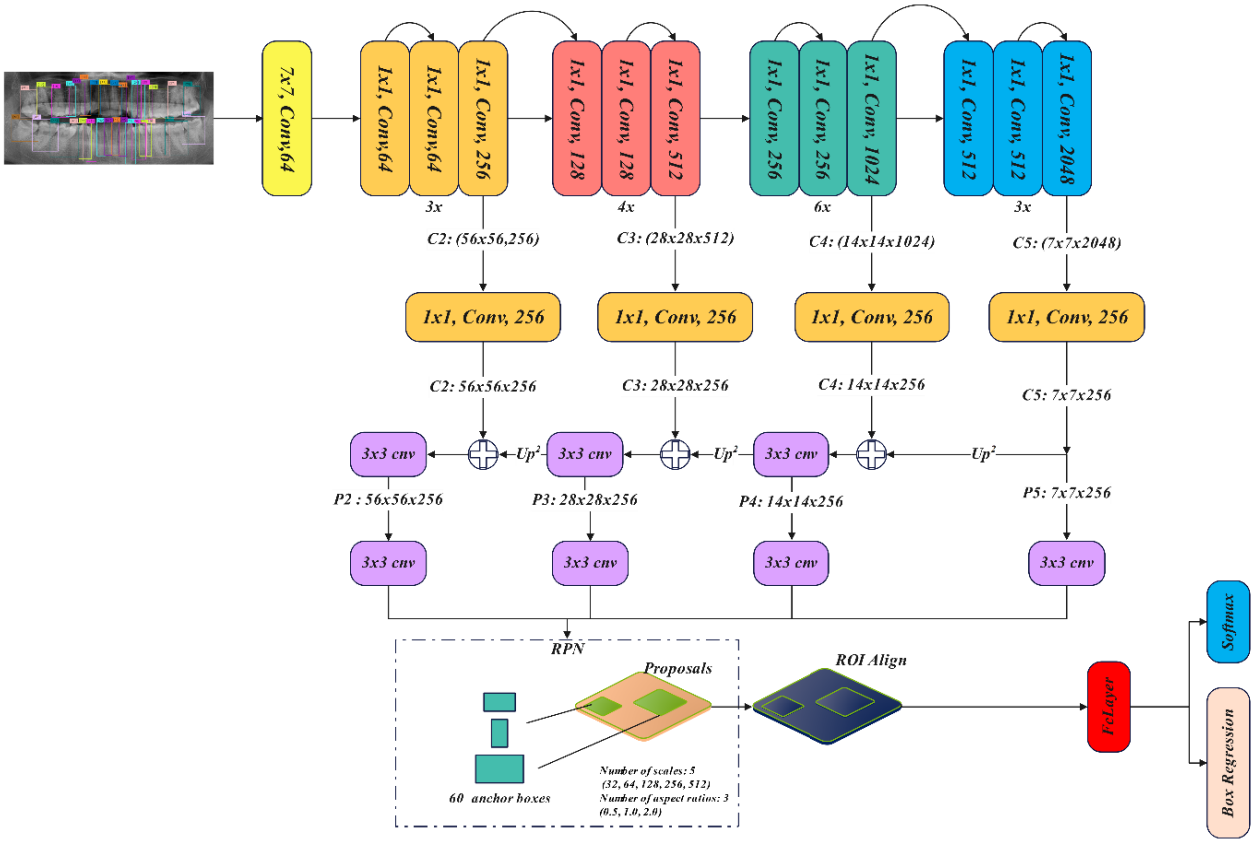

FIGURE 2. Dentifynet model architecture.

of [32, 64, 128, 256, 512] and ratios of [0.5, 1.0, 2.0]. Thus, the model can effectively cover both small and large teeth. RPN detects potential tooth regions using these anchors and provides the most suitable region suggestions. The innovation of this structure is that it provides better accuracy in large and complex images such as panoramic radiography. RoI Align is an important part of DentifyNet and improves detection performance by minimizing alignment errors. RoI Align resizes the feature maps from each proposed region, allowing them to be processed more accurately for classification and bounding box estimation. This integration provides more precise tooth detection and numbering. During the RoI Align process, the features obtained from the regions proposed using each feature map (P2-P5) are transferred to the FC layers and finalized in two different heads: classification (softmax) and bounding box coordinates (box regression). DentifyNet estimates the classification and bounding box coordinates using two separate heads. First, the features from RoI Align are processed by a fully connected layer. This fully connected layer obtains higher-level representations of the features and is then divided into two heads namely classification and box regression heads. While classification head uses the softmax activation function to determine the type of teeth. Box regression head uses box regression to estimate the positions of the teeth. This double-head structure performs the tooth detection and numbering tasks precisely and enables DentifyNet to perform successfully on panoramic radiograph images. Table 1 provides the layers, feature dimensions and transfer details of the features used in the architecture of the DentifyNet model:

DentifyNet incorporates several modifications to the standard Faster R-CNN framework to optimize its performance for panoramic dental radiographs. A key enhancement is the integration of Feature Pyramid Networks (FPN), which enables multi-scale feature extraction by processing features at different resolution levels. Unlike conventional single-scale feature extraction methods, FPN constructs a top-down pathway that refines

high-level semantic information while preserving spatial details from lower layers. This approach enhances the model's ability to detect teeth of varying sizes and positions, addressing common challenges such as partially erupted teeth and variations in occlusion patterns. Another critical customization is the implementation of flexible anchor structures. Standard Faster R-CNN models rely on fixed-scale anchor boxes, which may not be well-suited for detecting objects with irregular shapes and varying aspect ratios, such as teeth. In DentifyNet, anchor sizes and aspect ratios are adaptively scaled to ensure better localization of small, closely positioned, or missing teeth. By employing

TABLE 1. DentifyNet model architecture and feature flow.

Layer Name	Layer Type	Input Size	Output Size	Feature Extraction Location
Conv1	Convolutional Layer	(3, 224, 224)	(64, 112, 112)	-
MaxPool	Max Pooling	(64, 112, 112)	(64, 56, 56)	-
Layer1	Bottleneck Blocks (3 blocks)	(64, 56, 56)	(256, 56, 56)	Provided as C2 feature to FPN
Layer2	Bottleneck Blocks (4 blocks)	(256, 56, 56)	(512, 28, 28)	Provided as C3 feature to FPN
Layer3	Bottleneck Blocks (6 blocks)	(512, 28, 28)	(1024, 14, 14)	Provided as C4 feature to FPN
Layer4	Bottleneck Blocks (3 blocks)	(1024, 14, 14)	(2048, 7, 7)	Provided as C5 feature to FPN
FPN	Feature Pyramid Network	(C2, C3, C4, C5)	(P2, P3, P4, P5: 256 channels)	Transferred to RPN from P2, P3, P4, and P5
RPN	Region Proposal Network	(P2, P3, P4, P5)	Proposed regions	Works with features from FPN
ROI Align	Region of Interest Align	Proposed regions	Processed regions	Processes regions from RPN
Box Predictor	Fast R-CNN Head	Processed regions	Class and BBox results	Final classification and bounding box prediction after ROI Align

multiple anchor sizes (32, 64, 128, 256, 512 pixels) and three aspect ratios (0.5, 1.0, 2.0) at each feature level, the model improves detection performance across different anatomical structures in panoramic images. Additionally, DentifyNet refines the region proposal alignment process by replacing the conventional RoI Pooling operation with RoI Align. This modification eliminates spatial misalignment caused by quantization errors, allowing for more precise bounding box predictions. RoI Align ensures that feature maps are interpolated more accurately, preserving structural details critical for differentiating adjacent teeth. This enhancement is particularly beneficial for cases where teeth are closely spaced or overlap, improving the model's classification and localization accuracy. These architectural refinements enable DentifyNet to achieve higher detection accuracy and robustness in complex clinical scenarios, distinguishing it from standard Faster R-CNN implementations.

D. MODEL TRAINING AND HYPERPARAMETERS

The Stochastic Gradient Descent (SGD) [33] optimization algorithm was preferred for model training. The main reason for choosing SGD is its faster convergence and potential to improve overall performance. The learning rate was initially set to 0.005, based on values commonly used in similar object detection tasks. To ensure stable convergence and prevent abrupt weight updates, a StepLR scheduler was employed, reducing the learning rate by a factor of 0.1 every 5 epochs. This approach allows the model to start with a relatively higher learning rate for faster initial convergence and then gradually refine its weight adjustments as training progresses. The impact of this learning rate scheduling was validated empirically, demonstrating improved training stability and overall performance on the dataset. This value was further validated through iterative experimentation to ensure optimal performance on the dataset. In this way, the weights of the

model were optimized in each epoch and the accuracy rate of the model was tried to be increased. Momentum was set to 0.9 and weight decay to 0.0005, values chosen based on both literature recommendations and fine-tuning during preliminary experiments, thus aiming to reduce the risk of over-learning of the model. During the training process, the model was trained for a total of 30 epochs. Loss values were calculated and monitored in each epoch. To ensure the model's generalization capability, several measures were taken during the training process. First, the dataset used for training included diverse panoramic radiographs labeled by dental experts, exposing the model to various anatomical and imaging variations. This diversity helped the model adapt to different scenarios and reduced the risk of overfitting. Additionally, an 80:20 train-validation split was employed to evaluate the model's performance on unseen data during training, ensuring it could generalize well beyond the training set. The training of the model was performed on a computer equipped with an NVIDIA RTX 2080 Ti GPU, an Intel i5-12400F processor, and 32 GB of RAM. The 11 GB GDDR6 memory of the GPU provided a great advantage in terms of speed and performance in training the model. During training, CUDA acceleration was used and the efficiency of the model was increased through the PyTorch [34] deep learning library. The visual of the training process is given in Figure 3.

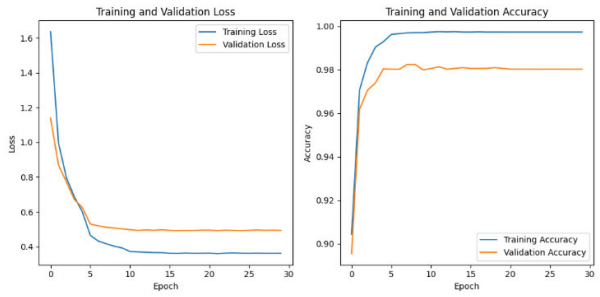

FIGURE 3. The process of training Dentifynet.

The training loss and accuracy curves illustrate the performance of the DentifyNet model during the training process. The training loss decreased steadily and stabilized after approximately 10 epochs, demonstrating proper convergence. The validation accuracy curve remained consistent and closely aligned with the training accuracy, indicating the model's ability to generalize well to unseen data.

1) OPTIMIZATION ALGORITHM DETAILS

Optimization algorithms play a critical role in training deep learning models by minimizing the loss function and improving model accuracy.

2) STOCHASTIC GRADIENT DESCENT (SGD)

SGD updates the model parameters iteratively based on the gradient of the loss function with respect to the weights:

$$w_{t+1} = w_t - \eta \nabla L(w_t)$$

where w_t represents the weights at iteration t, η is the learning rate, and $\nabla L(w_t)$ denotes the gradient of the loss function. To accelerate convergence and reduce oscillations, SGD can incorporate a momentum term:

$$v_{t+1} = \beta v_t - \eta \nabla L\left(w_t\right), w_{t+1} = w_t + v_{t+1}$$

Here, v_t is the velocity term that accumulates the past gradients, and β is the momentum coefficient, typically set between 0.8 and 0.99. The momentum mechanism helps the optimizer traverse ravines more efficiently, avoiding oscillations and improving stability.

3) ADAM OPTIMIZER

The Adam optimizer combines the advantages of momentum-based optimization with adaptive learning rates. It computes the first moment (mean) and second moment (uncentered variance) of the gradient:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla L(w_t)$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_1)(\nabla L(w_t))^2$$

Bias-corrected estimates are used to prevent initialization bias:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

The weights are updated as follows:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\left(\hat{v}_t\right)} + \epsilon}\hat{m}_t$$

where ϵ is a small constant added for numerical stability.

Comparison and Influence on Convergence: The addition of momentum allows the optimizer to maintain a consistent direction, accelerating convergence in valleys and reducing oscillatory behavior. Adam's adaptive learning rates enable each parameter to have an individualized step size, allowing faster convergence on sparse gradients and improved stability during training.

The choice of optimizer significantly impacts model performance, particularly in complex datasets like panoramic radiographs with diverse tooth structures. This study evaluates both optimizers, as discussed in the Results section.

E. EVALUATION METRICS

The performance of the DentifyNet model has been evaluated with various metrics. These metrics provide important criteria to measure the success of the model in tooth detection and enumeration tasks:

Intersection over Union (IoU): IoU [35] measures the overlap ratio between the bounding boxes predicted by the model and the true bounding boxes. The IoU value determines how well the predicted box overlaps and has a value in the range [1, 0]. Higher IoU values indicate that the model detects objects more accurately. The IoU threshold used in the evaluation was set to 0.5, meaning that detections with an IoU value above 0.5 were considered correct.

TABLE 2. Comparative performance results of DentifyNet and standard faster R-CNN with different backbones.

Model	IoU	Precision	Recall	F1 Score	Accuracy
ResNet50 Faster R-CNN	0.8059	0.9441	0.9432	0.9432	0.9432
DenseNet121 Faster R-CNN	0.8130	0.9578	0.9576	0.9575	0.9576
MobileNet Faster R-CNN	0.7778	0.9440	0.9438	0.9438	0.9438
EfficientNet-B0 Faster R-CNN	0.7534	0.9331	0.9326	0.9326	0.9326
RegNet400xMF Faster R-CNN	0.7592	0.9302	0.9393	0.9394	0.9393
DenseNet201Faster R-CNN	0.8222	0.9522	0.9519	0.9518	0.9519
DentifyNet	**0.8642**	**0.9752**	**0.9749**	**0.9751**	**0.9750**

Accuracy: Accuracy measures how well the model predicts among the total predictions. In the classification of teeth, it is calculated as the ratio of correctly labeled examples to the total examples. This metric reflects the overall success of the model and its performance over all classes. However, accuracy alone can be misleading in imbalanced datasets; therefore, it is evaluated together with other metrics.

Precision: Precision is calculated as the ratio of the number of teeth correctly detected by the model to the total number of teeth predicted. High precision indicates that the model avoids false positive detections and succeeds incorrect detections. This metric is an important performance indicator, especially when false detections are costly.

Recall: Recall is calculated as the ratio of the number of teeth correctly detected by the model to the total number of correct teeth. High recall indicates how successful the model is in detecting all correct teeth. This metric is critical when false negative detections need to be kept low.

F1 Score: The F1 score is calculated as the harmonic mean of precision and recall and measures the overall performance of the model. Balancing both precision and recall, this metric better reflects the performance of the model in imbalanced data sets. A high F1 score indicates that both the correct detection rate and the overall success of the model are high.

Confusion Matrix: The confusion matrix is an evaluation tool that visualizes the correct and incorrect detections of the model for each tooth class. This matrix is presented as a table containing true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. The confusion matrix is used to understand in which classes the model performs well or poorly.

III. EXPERIMENTAL RESULTS

In this section, the achievement of the DentifyNet model in tooth detection and numbering tasks will be evaluated in detail through numerical and visual results. The model was compared with standard Faster R-CNN models equipped with state-of-the-art backbone architectures and analyzed by considering numerical performance metrics (IoU, precision, recall, F1 score, accuracy). In addition, visual estimates obtained to better demonstrate the detection capabilities of DentifyNet are also presented.

The DentifyNet model was tested with a specially designed Faster R-CNN architecture and compared with different state-of-the-art backbones such as DenseNet121 [36], MobileNet [37], EfficientNet-B0 [38], RegNet400xMF [39], and DenseNet201 integrated into standard Faster R-CNN models. The IoU, precision, recall, F1 score, and accuracy values obtained for each backbone were used as the basic metrics to evaluate the overall accuracy and precision performance of the model. This comparison demonstrates the detection achievement of DentifyNet and its effectiveness against different backbones. The comparison of DentifyNet with other faster R-CNN models is given in Table 2.

The results given in Table 2 clearly show the superior performance of the DentifyNet model when compared with different backbones of the standard Faster R-CNN. DentifyNet outperformed other models by exhibiting higher accuracy, precision, and overall performance in tooth detection and numbering tasks. Especially, the best values were obtained by DentifyNet in basic metrics such as IoU, precision, recall, F1 score, and accuracy. Experiments conducted with commonly used backbones of the standard Faster R-CNN such as ResNet50 and DenseNet121 showed that DentifyNet offers higher performance due to multi-scale feature extraction and customized anchor structures. Although DenseNet121 provides better results than other backbones in terms of precision and recall, it could not reach the consistent accuracy and precision level offered by DentifyNet. Similarly, when lighter and faster backbones such as MobileNet and EfficientNet-B0 were used, a significant decrease in the performance of Faster R-CNN was observed, but DentifyNet provided better results even with these lightweight models. In particular, the lower IoU value and accuracy rate of EfficientNet-B0 show that the improved structure of DentifyNet is effective for better detection of complex structures. On the other hand, the results obtained with advanced backbones such as RegNet400xMF and DenseNet201 reveal that DentifyNet makes a significant difference in terms of overall performance. Although the standard Faster R-CNN reached the highest IoU and accuracy values with DenseNet201, DentifyNet exhibited superior metric values even against this backbone. This clearly shows that DentifyNet provides better overall performance and achieves successful results even with different backbones by providing high accuracy and precision in tooth detection and numbering. In addition, the confusion matrix obtained with the DentifyNet model is given in Figure 4.

The confusion matrix presented in Figure 4 shows the overall performance of the DentifyNet model in the tooth detection and numbering task. The high correct detection numbers along the diagonal line of the matrix indicate

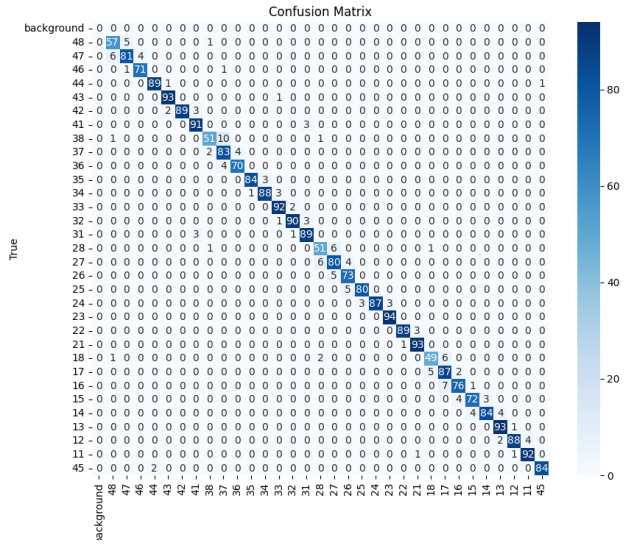

FIGURE 4. Confusion matrix obtained with the DentifyNet model.

the model's strong ability to classify most tooth classes accurately. Particularly, tooth numbers such as 11, 21, and 31 exhibit nearly perfect true positive rates, highlighting the model's effectiveness for distinct classes. In particular, the high true positive rates for distinct tooth numbers are striking, indicating that the model is effective in determining these classes. However, confusion occasionally occurs in the classification of adjacent and visually similar teeth. For instance, misclassifications are noticeable between teeth such as 27 and 28, or 47 and 48, where false positives and negatives are observed. These errors can be attributed to the anatomical similarities between these teeth in terms of shape, size, or location. The panoramic radiographs sometimes lack clear boundaries between such closely placed teeth, making the classification task more challenging. Despite these challenges, the confusion matrix demonstrates that the model maintains high accuracy for most tooth numbers, with misclassifications being exceptions rather than the norm. This performance underscores the robustness of DentifyNet and its potential for real-world applications, while also identifying areas where further architectural improvements can enhance precision for challenging cases.

A. TIME COMPLEXITY AND PERFORMANCE ANALYSIS

The computational efficiency of DentifyNet was thoroughly evaluated through comprehensive performance testing across different hardware configurations and input resolutions. Our analysis encompassed both theoretical complexity assessment and practical performance measurements to provide a complete understanding of the model's capabilities. Performance testing was conducted using both CPU implementations across various input resolutions (640 × 640, 800 × 800, and 1024 × 1024 pixels). The results obtained are presented in table 3.

TABLE 3. FPS comparison on CPU.

Resolution	Standard Faster R-CNN	DentifyNet	Difference
640 x 640	1.17 FPS	2.00 FPS	+0.83 FPS
800 x 800	1.33 FPS	2.01 FPS	+0.68 FPS
1024 x 1024	1.18 FPS	1.99 FPS	+0.81 FPS

Inference speed was also evaluated using GPU acceleration (CUDA). The custom architecture demonstrated superior performance across all resolutions, as shown in Table 4.

TABLE 4. FPS comparison on GPU.

Resolution	Standard Faster R-CNN	DentifyNet	Difference
640 x 640	53.37 FPS	66.14 FPS	+12.77 FPS
800 x 800	57.25 FPS	66.61 FPS	+9.36 FPS
1024 x 1024	56.33 FPS	66.90 FPS	+10.57 FPS

Our CPU-based performance analysis revealed remarkable improvements across all tested resolutions. DentifyNet achieved approximately 2.00 FPS consistently across different input sizes, representing a substantial improvement over the standard Faster R-CNN implementation. The most significant enhancement was observed at 640 × 640 resolution, where DentifyNet demonstrated a 70.9% improvement in processing speed. Notably, the performance remained stable even as the input resolution increased, with the model maintaining consistent FPS rates at higher resolutions (2.01 FPS at 800 × 800 and 1.99 FPS at 1024 × 1024).

The GPU-accelerated implementation showed even more promising results. DentifyNet achieved 66.14 FPS at 640 × 640 resolution, representing a 23.9% improvement over the standard architecture. As the input resolution increased, the model maintained its superior performance, achieving 66.61 FPS at 800 × 800 resolution and 66.90 FPS at 1024 × 1024 resolution. This consistent performance across increasing resolutions demonstrates the architecture's efficient scaling capabilities and robust resource utilization.

A particularly noteworthy aspect of DentifyNet's performance is its stability across different input resolutions. While the standard Faster R-CNN showed variable performance with changing resolutions, DentifyNet maintained consistent FPS rates, especially in GPU implementations. This stability can be attributed to the efficient Feature Pyramid Network (FPN) implementation and optimized anchor generation mechanism, which better handles multi-scale feature extraction and processing.

The performance improvements are particularly significant in the context of real-time dental imaging applications. The enhanced processing speed, especially in GPU implementations, makes DentifyNet suitable for integration into clinical workflows where rapid analysis is crucial. The consistent performance across different resolutions also provides flexibility in handling various image qualities and sizes without compromising processing speed.

We further evaluated the obtained results visually by an expert dentist via the visual results. The visual performance

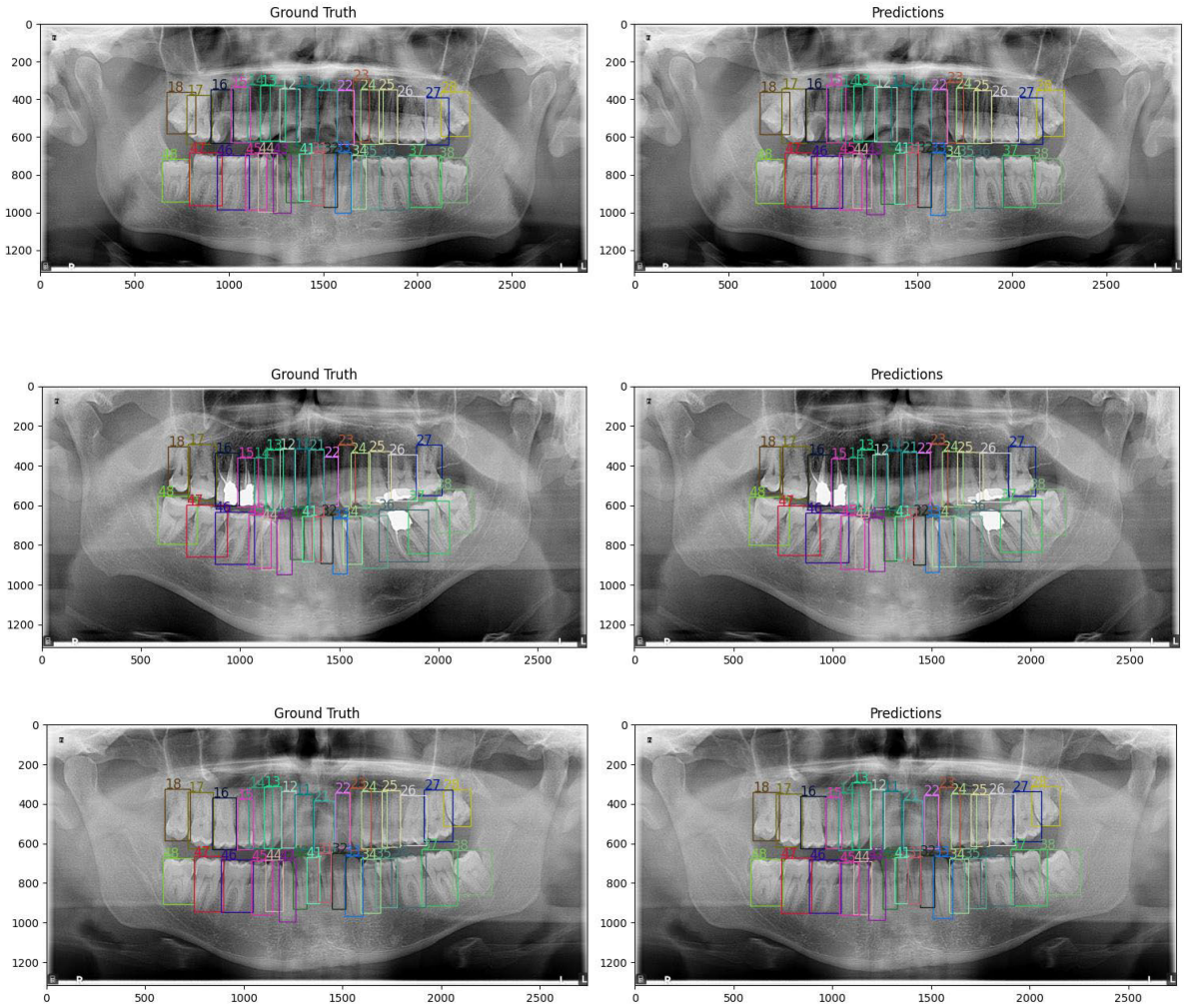

FIGURE 5. Visual prediction results of the model in mouths with a full set of teeth.

of the DentifyNet model in tooth detection and numbering tasks will be evaluated in detail using sample images. The images show the model's accuracy and precision by showing the model's ground truth and predicted bounding boxes side by side. Each prediction allows us to observe whether the model made correct detections and to understand the reasons for incorrect detections. The model's visual prediction results were filtered using a predetermined threshold value. Only predictions with a confidence score above a certain value (e.g. 0.3) were accepted. Thus, the model's most accurate predictions were visualized and analyzed. In addition, the Non-Maximum Suppression (NMS) [40] method was used on the predictions to reduce overlapping boxes. Thus, marking the same tooth with different boxes was prevented and more accurate results were obtained. The images are presented in two columns, each side by side: the ground truth is shown in the left column, and the model's predicted bounding boxes are shown in the right column. The performance of the model in mouths with full teeth or very close to 32 teeth is important to observe the overall detection capability of DentifyNet. In such images, the model is expected to number and detect the teeth correctly. Therefore, Figure 5 presents the predictions made by the model on teeth with full teeth or almost full teeth. This image represents a situation where the model correctly identifies and classifies all teeth and shows how well DentifyNet performs in mouths with full teeth.

This image represents DentifyNet's performance in mouths with full or nearly full teeth. Figure 5 shows that the model correctly identifies and numbers all teeth. Both upper and lower jaw teeth are identified, and the model's predictions are very close to the ground truth. The model performs with high accuracy on such full teeth, with no deviations or misidentifications in the placement of the teeth. This demonstrates how DentifyNet can perform under ideal conditions. Figure 6 shows how the DentifyNet model performs in cases with missing teeth or more irregularly structured mouths. This image examines in detail how the model identifies and numbers of teeth in more complex and missing teeth situations. These more challenging scenarios are important for evaluating the model's overall detection ability and resistance to errors.

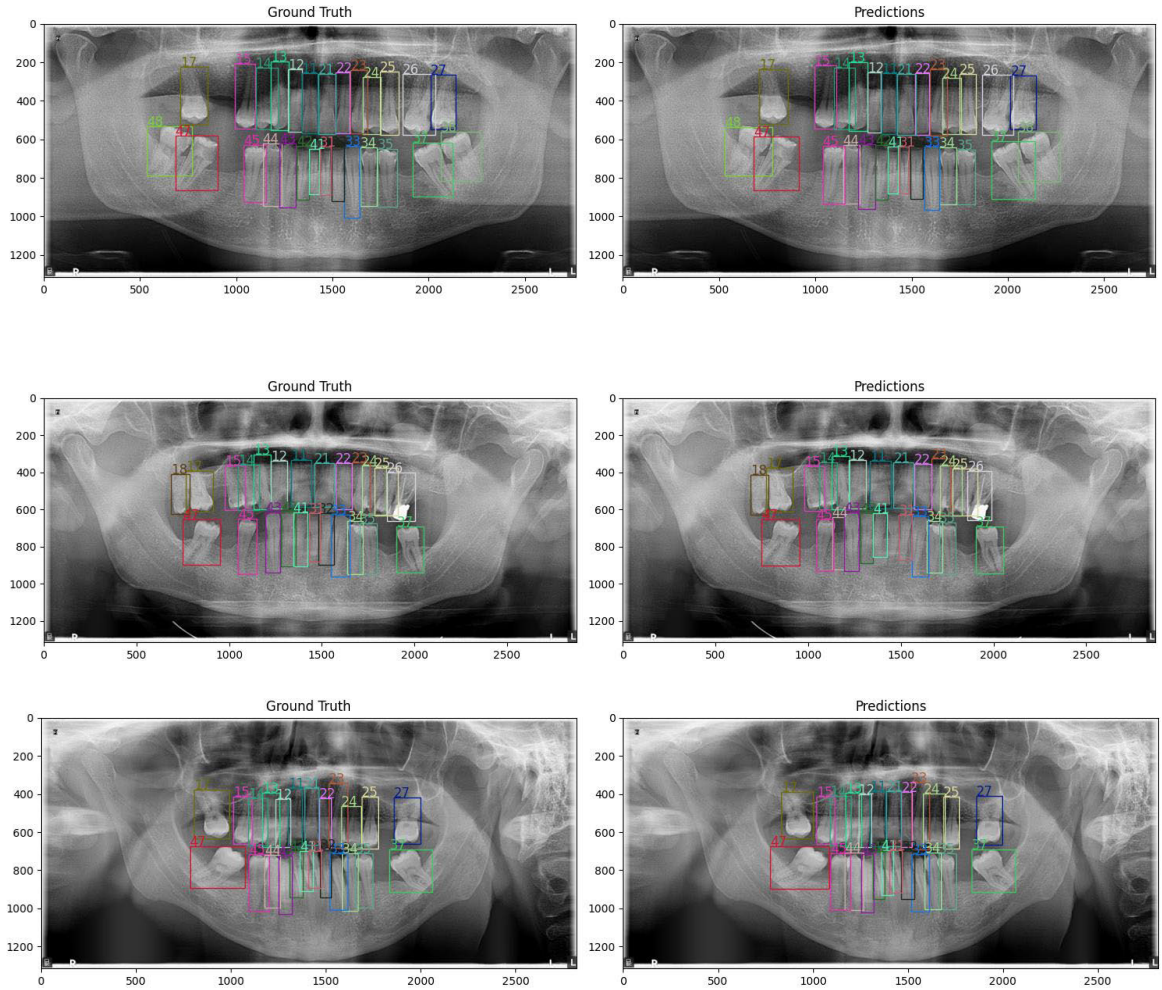

FIGURE 6. Visual prediction results of the model in mouth structures with missing teeth.

Figure 6 shows the performance of the DentifyNet model in more challenging oral environments where teeth are missing. In these images, the model is evaluated for both the detection of missing teeth and the correct numbering of existing teeth. As can be seen in the image, the model is generally able to cope with more complex environments where teeth are missing and make correct determinations. Figure 7 shows the incorrect predictions made by the DentifyNet model on some teeth. The model tends to make errors, especially when similar teeth are in proximity or when some teeth are missing. These images show that the model is generally successful in detecting and numbering teeth, but it can get confusing in certain scenarios.

Figure 7 visually presents the incorrect predictions of the DentifyNet model on some teeth. In these images, it is observed that the model incorrectly classified or confused some teeth, contrary to the previous accuracy results. For example, in the first image, the model predicted tooth number 14 on tooth number 13, whereas tooth number 14 is not present in the ground truth. Similarly, in another image, the model predicted that teeth number 44 and 15 were present even though they were not present in the ground truth. These errors show that the model is prone to errors, especially in areas where there are missing teeth and when detecting similar teeth in proximity. However, in most of the images, the model generally makes correct predictions, but it makes mistakes in some complex cases. However, these errors do not significantly affect the overall accuracy and precision of the model.

B. COMPARISON OF SGD AND ADAM OPTIMIZERS

The performance of the model was evaluated using both Stochastic Gradient Descent (SGD) and Adam optimizers. Key metrics such as training loss, training accuracy, validation accuracy, Intersection over Union (IoU), precision, recall, F1-score, and overall accuracy were compared. The results are summarized in Table 5.

SGD demonstrated superior training performance, achieving a lower training loss (0.3622) and higher training accuracy (0.9970) compared to Adam. This highlights the efficiency of SGD in minimizing the loss function during training.

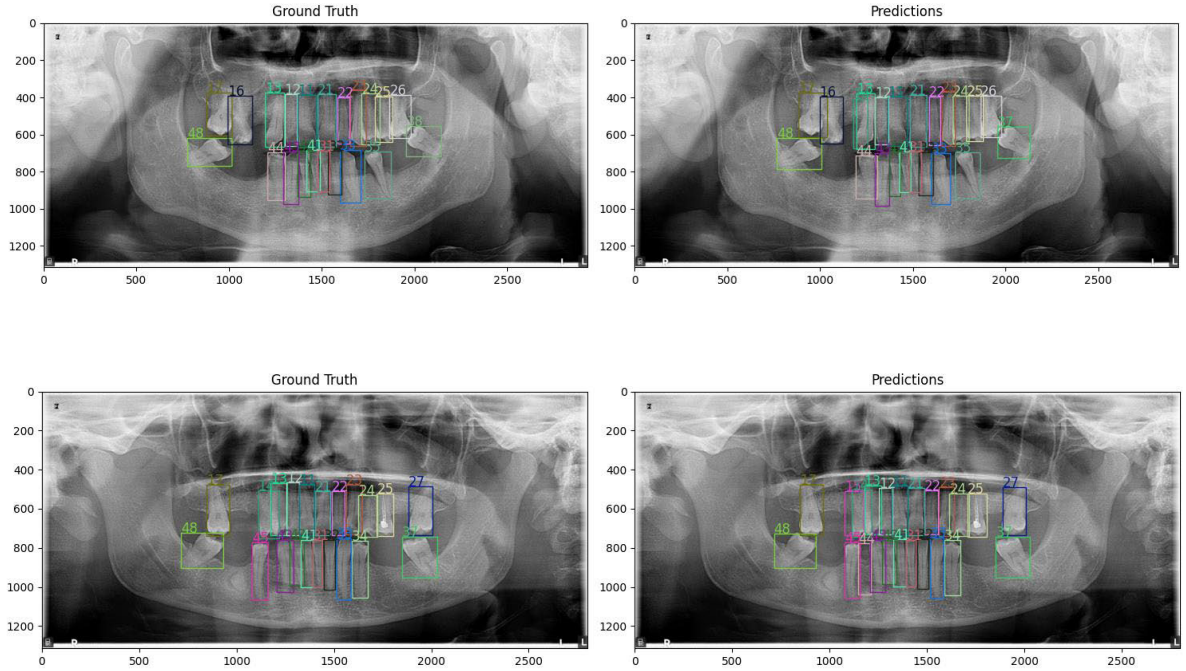

FIGURE 7. Visual errors in model predictions.

TABLE 5. Performance comparison of SGD and ADAM optimizers.

Metric	Adam Optimizer	SGD Optimizer
Training Loss	0.7110	0.3622
Training Accuracy	0.9500	0.9970
Validation Accuracy	0.8964	0.9750
IoU	0.7547	0.8642
Precision	0.7879	0.9752
Recall	0.7825	0.9749
F1 Score	0.7840	0.9751
Accuracy	0.7825	0.9750

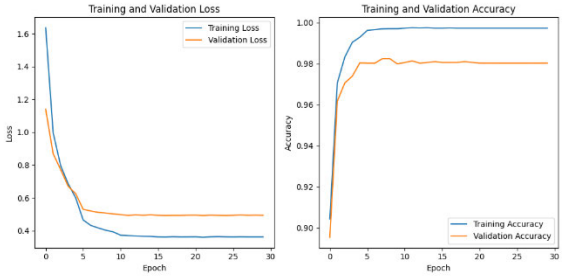

FIGURE 8. Training loss and accuracy curves for SGD optimizer, as shown in the uploaded visual. The results indicate a rapid and smooth convergence with minimal oscillations during training.

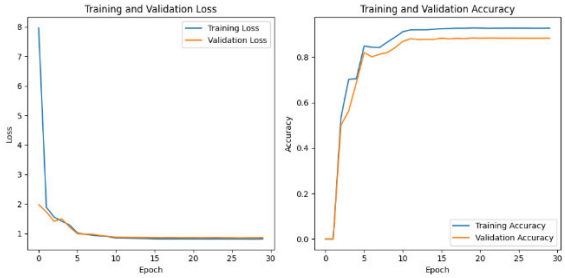

FIGURE 9. Training loss and accuracy curves for Adam optimizer, provided in the second uploaded visual. The results highlight slower initial convergence compared to SGD but demonstrate stable performance over time.

SGD outperformed Adam in validation accuracy, achieving 0.9794 compared to 0.8964 for Adam. This indicates better generalization performance of SGD on unseen data.

The SGD optimizer consistently achieved better IoU (0.8642) and classification metrics (precision, recall, and F1-score) compared to Adam. The precision and F1-score values for SGD were significantly higher (0.9752 and 0.9751, respectively), demonstrating its robustness in handling dental imaging tasks.

As shown in Figures 8 and 9, the training loss for both optimizers decreased over the training epochs, with SGD showing a steeper decline and smoother convergence. Adam, while initially showing higher loss, stabilized towards the later epochs.

IV. DISCUSSION

The developed DentifyNet model has provided significant performance improvements over previous approaches when evaluated based on numerical and visual results obtained in tooth detection and numbering tasks. Compared to the standard Faster R-CNN model, DentifyNet has shown a significant increase in metrics such as IoU, precision, recall, and F1 score. This success is due to the architectural innovations of DentifyNet and proves how effective the model is in a complex and sensitive task such as tooth detection. In particular, the integration of multi-scale feature extraction, customized anchor structures, and ROI Align mechanism were the key factors contributing to the overall success of the model. DentifyNet's use of FPN (Feature Pyramid Network) has been effective in reducing detection difficulties caused by teeth of different sizes and positions. FPN allows better detection of both large and small teeth by combining features at different resolutions. This feature supported the model to provide high accuracy in both mouths with full teeth and in cases with missing teeth. The high IoU and recall values of the model reflect the capacity of FPN to detect the boundaries of teeth and distinguish even small teeth accurately.

The customization of anchor structures is another critical component of DentifyNet's success. While the fixed anchors used in the standard Faster R-CNN may be insufficient to detect different sizes and shapes of teeth, the anchors designed in different scales and proportions in DentifyNet provided better adaptation to the dimensions of the teeth. This customization was effective in correctly distinguishing adjacent teeth and directly contributed to the improvements observed in Precision and F1 Score values. In this way, both the number of false positives and false negatives were reduced, and the overall accuracy and precision performance of the model increased.

The ROI Align mechanism further improved the performance of DentifyNet. Unlike standard ROI Pooling, ROI Align processes the information in the feature maps more precisely and increases the location accuracy. This feature allowed for clearer determination of the boundaries of the teeth and more accurate placement of the boxes. The effect of ROI Align was observed in the accuracy of the number of teeth detected by the model and the precision of the boxes. For example, even in challenging oral structures with missing teeth, the model provided an effective solution for tooth detection and numbering tasks by providing high accuracy rates. However, despite the success of DentifyNet, there are some limitations. Some errors were observed in the predictions made between similar teeth located next to each other. For example, the model sometimes had difficulty distinguishing the boundaries of similar teeth such as 13 and 14, which led to incorrect predictions. Similarly, problems such as incorrect numbering or missing detection of teeth in oral structures with missing teeth were also detected. To overcome these limitations, it is recommended to train with more data and develop the architecture to adapt to these special situations.

Table 6 evaluates the success of the DentifyNet model in tooth detection and numbering task in comparison with other studies in the literature. DentifyNet was tested on 468 panoramic radiographs and achieved high results of 86.42%, 97.52%, 97.49%, 97.51% and 97.50% in IoU, Precision, Recall, F1 score, and accuracy metrics, respectively. These values show that DentifyNet is one of the best models available in tooth detection and numbering. When the table is examined, it is seen that other studies in literature have achieved various success metrics with different dataset sizes and models. While studies such as Laishram et al. [16] and Zhang et al. [19] stand out with their high AP and accuracy rates, DentifyNet's success in F1 score and other metrics demonstrates the comprehensive detection capabilities of the model. For example, Privado et al. [20] and Choi et al. Although the accuracy values of 99.24% and 99.1% obtained in [23] studies are high, it is noteworthy that DentifyNet shows a more balanced performance, especially in other metrics such as F1 score and IoU.

On the other hand, while some studies (e.g., Etsai et al. [25]) are limited to lower IoU and AP values, the overall performance of DentifyNet is higher, which shows that the model is effective in difficult tooth detection and numbering scenarios. In addition, the special Faster R-CNN structure used in DentifyNet and FPN-based multi-scale feature extraction played an important role in increasing the sensitivity and accuracy of the model.

As a result, DentifyNet stands out as an innovative model that shows high performance in tooth detection and numbering tasks. However, it would be useful to retrain the model with a more diverse dataset and improve its architecture in future studies for it to perform better, especially in cases where there are similar teeth and missing teeth. In this regard, it can be said that the model is open to additional developments so that it can address a wider range of clinical applications and provide more successful results in different patient groups. The advantages and disadvantages of the proposed model are summarized as follows:

Advantages

- DentifyNet shows a significant performance increase over standard Faster R-CNN models according to both numerical and visual results.
- The flexible anchor mechanism of the model allows better detection of small and closely spaced teeth.
- Thanks to the multi-scale feature extraction performed with FPN integration, the model can effectively detect teeth of different sizes and provide high accuracy in tooth structures of different sizes.
- The model has demonstrated a wide generalization capacity by providing high accuracy rates even in difficult mouth structures with missing teeth.

TABLE 6. Comparison of DentifyNet model with tooth detection and numbering models in literature.

Study	Model	Dataset Size	Success Metrics
Laishram et al. [16]	Modified Faster R-CNN	145 OPG images	92% accuracy, 91% mAP, AP50: 99.8%, AP75: 92.7%
Kaya et al. [17]	YOLOv4 (CSPDarknet53 backbone)	4518 panoramic images	94.16% mean precision, 90% F1 score
Bumann et al. [18]	Mask R-CNN	368 panoramic + 80 private data	95.32% mAP, 92.5% F1 score
Zhang et al. [19]	Modified Faster R-CNN	1250 dental periapical X-rays	95.1% precision, 95.5% recall, 95.3% F1 score
Privado et al. [20]	Matterport Mask R-CNN & CNN	8000 panoramic images	99.24% accuracy
Bilgir et al. [8]	Faster R-CNN (Inception V2)	2482 panoramic images	95.59% sensitivity, 96.52% precision, 96.06% F1 score
Zhang et al. [21]	VGG16	Dental periapical radiography	98% accuracy, 98.3% recall, 98.1% F1 score
Putra et al. [22]	YOLOv4 (CSPDarknet53 backbone)	500 panoramic images	88.5% accuracy, 87.7% precision, 93.51% F1 score
Choi et al. [23]	EfficientDet D3	1683 panoramic images	99.1% precision, 97.4% recall
Chen et al. [24]	Hybrid Model (YOLOv5, VGG-16, U-Net)	8000 periapical images	Position: 88.8%, Shape: 90.4%, Bone Level: 85.7%
Etsai et al. [25]	U-Net & Faster R-CNN	591 OPG images	ROI IoU: 71%, Detection: 99.2% recall, 99.4% precision
Yuksel et al. [26]	YOLO-based Model (Dentect)	1005 panoramic images	Tooth AP (0.5): 89.1%, AP (0.5:0.95): 47.4%
Yilmaz et al. [27]	YOLOv4 & Faster R-CNN	1200 panoramic images	YOLOv4: 99.9% accuracy, 99.18% recall, 99.54% F1 score; Faster R-CNN: 93.67% accuracy, 90.79% recall, 92.21% F1 score
Gülüm et al. [28]	YOLOv4	3000 panoramic images	83% F1 score, 81% mAP
This Study	DentifyNet	468 panoramic images	IoU: 86.42%, Precision: 97.52%, Recall: 97.49%, F1: 97.51%, Accuracy: 97.50%

Disadvantages

• The model can sometimes confuse similar teeth located next to each other, such as teeth numbered 13 and 14, and this can cause incorrect classifications.

• It was observed that the model predicted incorrect teeth in mouths with missing teeth. This is related to the model's inability to generalize missing data well enough in some cases.

• Training and running the model requires high-performance GPUs, which can be a challenge especially for research teams with limited resources.

• The current performance of the model is limited by the dataset used. It is thought that the performance of the model can be further improved with larger and more diversified datasets.

V. CONCLUSION

In this study, the DentifyNet model, which was specifically developed for tooth detection and numbering in dental panoramic radiography images, was introduced and evaluated comprehensively. DentifyNet exhibited effective performance with higher accuracy, recall, precision, and F1 scores compared to standard Faster R-CNN models. DentifyNet, which has a specially designed feature pyramid network (FPN) and flexible anchor structures, better captured the multi-scale features of teeth and provided high accuracy rates even in challenging oral structures. The success of DentifyNet was demonstrated with high-performance metrics such as IoU (%86.42), precision (%97.52), recall (%97.49), F1 score (%97.49), and accuracy (%97.49). Compared to standard Faster R-CNN models, DentifyNet provided effective results, especially in challenging dental structures, in cases where there were missing teeth, and in distinguishing similar teeth. Visual evaluation results also showed that the model performed consistently in mouths with 32 teeth as well as in mouths with missing teeth. For future studies, it is recommended to train DentifyNet with different datasets, use transfer learning techniques, and perform optimizations for faster detection of the model. In addition, it may be useful to try additional algorithms and data enrichment techniques to accurately classify more complex dental structures.

REFERENCES

[1] *Reduce the Risk of Wrong-Site and Wrong-Tooth Dental Procedures*. Accessed: Oct. 11, 2024. [Online]. Available: https://www.thedoctors.com/articles/reduce-the-risk-of-wrong-site-and-wrong-tooth-dental-procedures/

[2] J. S. Lee, A. W. Curley, and R. A. Smith, "Prevention of wrong-site tooth extraction: Clinical guidelines," *J. Oral Maxillofacial Surg.*, vol. 65, no. 9, pp. 1793–1799, Sep. 2007, doi: 10.1016/j.joms.2007.04.012.

[3] N. Swain, S. Patel, J. Pathak, P. R. Sarkate, N. K. Sahu, and R. M. Hosalkar, "Role of dental hard tissue in human identification," *J. Contemp. Dentistry*, vol. 9, no. 3, pp. 130–134, Mar. 2021, doi: 10.5005/JP-JOURNALS-10031-1264.

[4] J. Pramod, A. Marya, and V. Sharma, "Role of forensic odontologist in post mortem person identification," *Dental Res. J. Isfahan*, vol. 9, no. 5, p. 522, 2012, doi: 10.4103/1735-3327.104868.

[5] R. Izzetti, M. Nisi, G. Aringhieri, L. Crocetti, F. Graziani, and C. Nardi, "Basic knowledge and new advances in panoramic radiography imaging techniques: A narrative review on what dentists and radiologists should know," *Appl. Sci.*, vol. 11, no. 17, p. 7858, Aug. 2021, doi: 10.3390/app11177858.

[6] *Optimal Panorex Imaging—Decisions in Dentistry*. Accessed: Oct. 11, 2024. [Online]. Available: https://decisionsindentistry.com/article/optimal-panorex-imaging/

[7] I. Kaffe, D. Fishel, and M. Gorsky, "Panoramic radiography in dentistry," *Refuat Hapeh Vehashinayim*, vol. 26, no. 2, pp. 25–30, Apr. 1977, doi: 10.1007/s41894-021-00111-4.

[8] E. Bilgir, I. Ş. Bayrakdar, Ö. Çelik, K. Orhan, F. Akkoca, H. Sağlam, A. Odabaş, A. F. Aslan, C. Ozcetin, M. Kıllı, and I. Rozylo-Kalinowska, "An artificial ıntelligence approach to automatic tooth detection and numbering in panoramic radiographs," *BMC Med. Imag.*, vol. 21, no. 1, pp. 1–9, Dec. 2021, doi: 10.1186/S12880-021-00656-7.

[9] B. Ayhan, E. Ayan, and Y. Bayraktar, "A novel deep learning-based perspective for tooth numbering and caries detection," *Clin. Oral Investig.*, vol. 28, no. 3, pp. 1–17, Feb. 2024, doi: 10.1007/s00784-024-05566-w.

[10] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, and S. Youm, "Automatic tooth detection and numbering using a combination of a CNN and heuristic algorithm," *Appl. Sci.*, vol. 10, no. 16, p. 5624, Aug. 2020, doi: 10.3390/app10165624.

[11] S. Sadr, R. Rokhshad, Y. Daghighi, M. Golkar, F. T. Kheybari, F. Gorjinejad, A. M. Kojori, P. Rahimirad, P. Shobeiri, M. Mahdian, and H. Mohammad-Rahimi, "Deep learning for tooth identification and numbering on dental radiography: A systematic review and meta-analysis," *Dentomaxillofacial Radiol.*, vol. 53, no. 1, pp. 5–21, Jan. 2024, doi: 10.1093/dmfr/twad001.

[12] I. Rozylo-Kalinowska, *Imaging Techniques in Dental Radiology: Acquisition, Anatomic Analysis and Interpretation of Radiographic Images*. Springer, 2020.

[13] N. R. Suparno, A. Faizah, and A. N. Nafisah, "Assessment of panoramic radiograph errors: An evaluation of patient preparation and positioning quality at Soelastri dental and oral hospital," *Open Dentistry J.*, vol. 17, no. 1, pp. 1–9, Oct. 2023, doi: 10.2174/0118742106261974230925073155.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. *You Only Look Once: Unified, Real-Time Object Detection*. Accessed: Oct. 18, 2024. [Online]. Available: http://pjreddie.com/yolo/

[16] A. Laishram and K. Thongam, "A deep learning approach based on faster R-CNN for automatic detection and classification of teeth in orthopantomogram radiography images," *IETE J. Res.*, vol. 70, no. 2, pp. 1316–1327, Feb. 2024, doi: 10.1080/03772063.2022.2154283.

[17] E. Kaya, H. G. Gunec, K. C. Aydin, E. S. Urkmez, R. Duranay, and H. F. Ates, "A deep learning approach to permanent tooth germ detection on pediatric panoramic radiographs," *Imag. Sci. Dentistry*, vol. 52, no. 3, p. 275, 2022, doi: 10.5624/isd.20220050.

[18] E. E. Bumann, S. Al-Qarni, G. Chandrashekar, R. Sabzian, B. Bohaty, and Y. Lee, "A novel collaborative learning model for mixed dentition and fillings segmentation in panoramic radiographs," *J. Dentistry*, vol. 140, Jan. 2024, Art. no. 104779, doi: 10.1016/j.jdent.2023.104779.

[19] K. Zhang, H. Chen, P. Lyu, and J. Wu, "A relation-based framework for effective teeth recognition on dental periapical X-rays," *Computerized Med. Imag. Graph.*, vol. 95, Jan. 2022, Art. no. 102022, doi: 10.1016/j.compmedimag.2021.102022.

[20] M. Prados-Privado, J. G. Villalón, A. B. Torres, C. H. Martínez-Martínez, and C. Ivorra, "A validation employing convolutional neural network for the radiographic detection of absence or presence of teeth," *J. Clin. Med.*, vol. 10, no. 6, p. 1186, Mar. 2021, doi: 10.3390/jcm10061186.

[21] K. Zhang, J. Wu, H. Chen, and P. Lyu, "An effective teeth recognition method using label tree with cascade network structure," *Computerized Med. Imag. Graph.*, vol. 68, pp. 61–70, Sep. 2018, doi: 10.1016/j.compmedimag.2018.07.001.

[22] R. H. Putra, E. R. Astuti, D. K. Putri, M. Widiasri, P. A. M. Laksanti, H. Majidah, and N. Yoda, "Automated permanent tooth detection and numbering on panoramic radiograph using a deep learning approach," *Oral Surgery, Oral Med., Oral Pathol. Oral Radiol.*, vol. 137, no. 5, pp. 537–544, May 2024, doi: 10.1016/j.oooo.2023.06.003.

[23] H.-R. Choi, T. S. Siadari, J.-E. Kim, K.-H. Huh, W.-J. Yi, S.-S. Lee, and M.-S. Heo, "Automatic detection of teeth and dental treatment patterns on dental panoramic radiographs using deep neural networks," *Forensic Sci. Res.*, vol. 7, no. 3, pp. 456–466, Jul. 2022, doi: 10.1080/20961790.2022.2034714.
[24] C.-C. Chen, Y.-F. Wu, L. M. Aung, J. C.-Y. Lin, S. T. Ngo, J.-N. Su, Y.-M. Lin, and W.-J. Chang, "Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence," *J. Dental Sci.*, vol. 18, no. 3, pp. 1301–1309, Jul. 2023, doi: 10.1016/j.jds.2023.03.020.
[25] M. Estai, M. Tennant, D. Gebauer, A. Brostek, J. Vignarajan, M. Mehdizadeh, and S. Saha, "Deep learning for automated detection and numbering of permanent teeth on panoramic images," *Dentomaxillofacial Radiol.*, vol. 51, no. 2, p. 51, Oct. 2021, doi: 10.1259/dmfr.20210296.
[26] A. E. Yüksel, S. Gültekin, E. Simsar, Ş. D. Özdemir, M. Gündoğar, S. B. Tokgöz, and İ. E. Hamamcı, "Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Jun. 2021, doi: 10.1038/s41598-021-90386-1.
[27] S. Yilmaz, M. Tasyurek, M. Amuk, M. Celik, and E. M. Canger, "Developing deep learning methods for classification of teeth in dental panoramic radiography," *Oral Surgery, Oral Med., Oral Pathol. Oral Radiol.*, vol. 138, no. 1, pp. 118–127, Jul. 2024, doi: 10.1016/j.oooo.2023.02.021.
[28] S. Gülüm, S. Kutal, K. C. Aydin, G. Akgün, and A. Akdağ, "Effect of data size on tooth numbering performance via artificial intelligence using panoramic radiographs," *Oral Radiol.*, vol. 39, no. 4, pp. 715–721, Oct. 2023, doi: 10.1007/s11282-023-00689-4.
[29] H. Kadi, T. Sourget, M. Kawczynski, S. Bendjama, B. Grollemund, and A. Bloch-Zupan, "Segmentation, and numbering in oral rare diseases: Focus on data augmentation and inpainting techniques," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2023, pp. 1358–1363, doi: 10.1109/CSCI62032.2023.00298.
[30] S. Zhuang, G. Wei, Z. Cui, and Y. Zhou, "Robust hybrid learning for automatic teeth segmentation and labeling on 3D dental models," *IEEE Trans. Multimedia*, vol. 27, pp. 792–803, 2023, doi: 10.1109/TMM.2023.3289760.
[31] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Mar. 2019, doi: 10.1038/s41598-019-40414-y.
[32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, vol. 8693, May 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1.
[33] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[34] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. NeurIPS*, 2019, p. 12. [Online]. Available: https://doi.org/10.48550/arXiv.1912.01703
[35] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666, doi: 10.1109/CVPR.2019.00075.
[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
[37] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[38] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, May 2019, pp. 10691–10700. Accessed: Oct. 24, 2024.
[39] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10425–10433.
[40] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6469–6477, doi: 10.1109/CVPR.2017.685.

SALIH TAHA ALPEREN ÖZÇELIK was born in Elâziğ, Türkiye, in 1995. He received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Fırat University, Elâziğ, in 2017 and 2021, respectively, where he is currently pursuing the Ph.D. degree. Since 2020, he has been a Research Assistant with the Department of Electrical and Electronics Engineering, Bingöl University, Bingöl, Türkiye. His research interests include deep learning, artificial intelligence, biomedical signal processing, and image processing.

HÜSEYİN ÜZEN received the B.Sc. degree in computer engineering from Süleyman Demirel University, Isparta, Türkiye, in June 2015, and the M.Sc. and Ph.D. degrees in computer engineering from İnönü University, Malatya, Türkiye, in July 2018 and October 2022, respectively. He is currently an Assistant Professor with the Department of Computer Engineering, Bingöl University, Bingöl, Türkiye. His research interests include artificial intelligence, image processing, and machine learning applications.

ABDULKADIR SENGUR received the B.Sc. degree in electronics and computer education, the M.Sc. degree in electronics education, and the Ph.D. degree in electrical and electronics engineering from Fırat University, Türkiye, in 1999, 2003, and 2006, respectively. He was a Research Assistant with the Faculty of Technical Education, Fırat University, in February 2001, where he is currently a Professor with the Faculty of Technical Education. His research interests include signal processing, image segmentation, pattern recognition, medical image processing, and computer vision.

MUAMMER TÜRKOĞLU received the B.Sc. and M.Sc. degrees in electronics and computer education from Fırat University, Türkiye, in 2011 and 2013, respectively, and the Ph.D. degree in computer engineering from Inonu University, Türkiye, in 2019. He was a Research Assistant with the Engineering Faculty, Bingöl University, in April 2013. His research interests include signal processing, machine learning, and image processing.

ADALET ÇELEBI received the B.Sc. degree from the Faculty of Dentistry, Istanbul University, İstanbul, in 2012, and the Specialty degree from the Department of Maxillofacial Surgery, Faculty of Dentistry, Dicle University, in 2018. She is currently an Assistant Professor with the Faculty of Dentistry, Mersin University.

NEBRAS M. SOBAHI is currently an Associate Professor with the Department of Electrical and Computer Engineering, King Abdulaziz University, Saudi Arabia. His research interests include nano/microfabrication, MEMS, microfluidics, BioMEMS, and signal and image processing.