Computers in Biology and Medicine 146 (2022) 105547

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs☆

Buse Yaren Tekin [a,**], Caner Ozcan [b,*], Adem Pekince [c], Yasin Yasa [d]

[a] Department of Computer Technologies, Kastamonu University, Kastamonu, Turkey
[b] Department of Software Engineering, Karabuk University, Karabuk, Turkey
[c] Department of Oral and Maxillofacial Radiology, Karabuk University, Karabuk, Turkey
[d] Department of Oral and Maxillofacial Radiology, Ordu University, Ordu, Turkey

ARTICLE INFO

Keywords:
Dental bitewing radiograph
Convolutional neural networks
Fdi notation
Tooth numbering
Instance segmentation

ABSTRACT

Bitewing radiographic imaging is an excellent diagnostic tool for detecting caries and restorations that are difficult to view in the mouth, particularly at the molar surfaces. Labeling radiological images by an expert is a labor-intensive, time-consuming, and meticulous process. A deep learning-based approach has been applied in this study so that experts can perform dental analyzes successfully, quickly, and efficiently. Computer-aided applications can now detect teeth and number classes in bitewing radiographic images automatically. In the deep learning-based approach of the study, the neural network has a structure that works according to regions. A region-based automatic segmentation system that segments each tooth using masks to help to assist analysis as given to lessen the effort of experts. To acquire precision and recall on a test dataset, Intersection Over Union value is determined by comparing the model's classified and ground-truth boxes. The chosen IOU value was set to 0.9 to allocate bounding boxes to the class scores. Mask R–CNN is a method that serves as instance segmentation and predicts a pixel-to-pixel segmentation mask when applied to each Region of Interest. The tooth numbering module uses the FDI notation, which is widely used by dentists, to classify and number dental items found as a result of segmentation. According to the experimental results were reached 100% precision and 97.49% mAP value. In the tooth numbering, were obtained 94.35% precision and 91.51% as an mAP value. The performance of the Mask R–CNN method used has been proven by comparing it with other state-of-the-art methods.

1. Introduction

Medical imaging tools for disease diagnosis and therapy are visible representations of the body's internal functioning for medical examination. Computed tomography medical imaging technology has aided in the treatment and diagnosis of a variety of ailments in recent years [1]. Dentistry is a profession that has observed the growth of innovative methods in radiological imaging. Radiological imaging has played an crucial task in effective diagnosis and therapy with such developments [2]. One of the most difficult aspects of dentistry is that the maxillofacial radiologist must interpret a large number of images in a short amount of time and provide a treatment plan quickly. Artificial intelligence's ability to perform as a diagnostic tool in medicine has been highlighted in recent years as a result of breakthroughs in the field [3]. Artificial intelligence (AI) systems that support the abilities of healthcare experts assure the beginning a new age in which can automate repetitive, time-consuming activities [4]. The time spent is reduced and it predicts potential bad situations, with innovative advanced technologies in the diagnosis and treatment methods applied to health problems [5]. Deep learning models are benefit for image analysis, anomaly detection, image segmentation, and classification in medical images. To summarize, AI systems have the ability to improve health data results, lower healthcare expenditures, and advance medical studies [6].

Radiological analysis is a method mostly used in intraoral imaging, which has an significantly position in dentistry. In clinics, periapical and bitewing films are commonly utilized for this purpose [7,8]. Experts

☆ This work is the results of the research project supported by the Scientific and Technological Research Council of Turkey (TUBİTAK).
* Corresponding author.
** Corresponding author.
E-mail addresses: bytekin@kastamonu.edu.tr (B. Yaren Tekin), canerozcan@karabuk.edu.tr (C. Ozcan).

https://doi.org/10.1016/j.compbiomed.2022.105547
Received 13 January 2022; Received in revised form 29 March 2022; Accepted 17 April 2022
Available online 27 April 2022

must put down their dental records precisely in hospital information systems, where all data kept in examinations are digitized [9]. In crowded clinics, misdiagnosis or underdiagnosis may occur depending on the experience and attentiveness of experts. The teeth must be appropriately identified and numbered to avoid this case [10]. Radiological imaging is divided into intraoral radiography and extraoral radiography for dental images. Intraoral radiographic images are a type of imaging commonly used by dentists. Intraoral radiography provides a radiological image in which the film or sensor is placed in the mouth [11].

Bitewing radiography is a type of imaging that shows the crown of teeth above and below the jaw. Film holders and bite papers are mostly used for imaging bitewing radiographs [12]. Bitewing radiography operates with an sensitivity of about 0.4 to identify lesions proximate to the tooth and bone with specificity. Bitewing radiographs on occlusal surfaces are not useful for detecting enamel lesions in an early stage, but are useful for identifying moderate lesions and determining proximity to the pulp with overall sensitivity/specific values at 0.4/0.8, respectively. Bitewing radiography has a low sensitivity for detecting early proximal lesions that only extend into the enamel, which aids in the early detection of proximal carious lesions. This leads to the assumption that very early lesions (histologically develop into outer enamel) often remain undetected. Furthermore, bitewing radiography is beneficial for detecting proximal secondary lesions (0.5/0.8), but not on smooth (buccal, lingual) surfaces [13].

During the taking of the bitewing radiograph, unique strip films with a wing in the middle are prepared before are placed in the patient's mouth in order for the patient to bite. Bitewing radiography images gathered for the study were obtained from Ordu University Faculty of Dentistry, Department of Oral, Dental and Maxillofacial Radiology, and necessary legal permissions were obtained. These images were obtained with a Kodak CS 2200 220–240-V bitewing dental imaging unit (Carestream Dental, Atlanta, GA) with 60 kVp, 7 mA, and 0.1 s parameters. Fig. 1 shows the samples of bitewing radiography images.The bitewing radiographs used in the study were annotated by the dentists after clinical confirmation for the presence of teeth. The data can be read by the machine learning model after it has been annotated. These explanations were used as ground truth (GT) for both training and testing. We used a total of 1000 radiographs to train our system. The remaining 200 radiographs were used for testing. The labels of our testing system were compared with the ground truth throughout the study. Detection or learning useful features that accurately characterize the regularities or patterns inherent in data is crucial in many tasks in medical image analysis when using machine learning. Deep learning is quickly becoming the industry standard, resulting in improved performance in a variety of medical applications. When there are a lot of samples available during the training stage, deep learning approaches are quite successful [1]. Artificial neural networks used in deep learning are network models in which the input that contains more than one layer is given to the model as a dataset and the prediction results are obtained as output. A Convolutional Neural Network (CNN) is an artificial neural network used to analyze visual images in deep learning [14]. As a result, information is supplied into deep learning as enormous data sets. Various scientists who performed tooth detection and segmentation by considering various neural networks derived from CNN and CNN also proposed their own models, and we summarize their work in Table 1. Most of the experiments summarized in table contain CNN algorithms, which are currently commonly used approaches. In addition, there are various methods for segmentation and detection for the purpose of identifying teeth.

In 2004, Said et al. [15] proposed the segmentation procedure for the image enhancement technique by combining different methods. The methods are combined, the first model according to morphological filtering and the second model according to the new improved 2D wavelet transform. In the study, better performance were obtained as a result of segmentation. However, it was concluded that it was not very successful or meaningful when dealing with dental images. Due to the multi-resolution characteristic that morphological filters cannot discriminate, wavelet-based segmentation was required for specific image types to capture the small edges between various teeth.

In 2010, Lin et al. [16] developed an algorithm that enables the numbering and classification of teeth using tooth region and contour information in bitewing radiographs. The forms of the teeth and pulps were shown to have a vital part in the classification for correct tooth

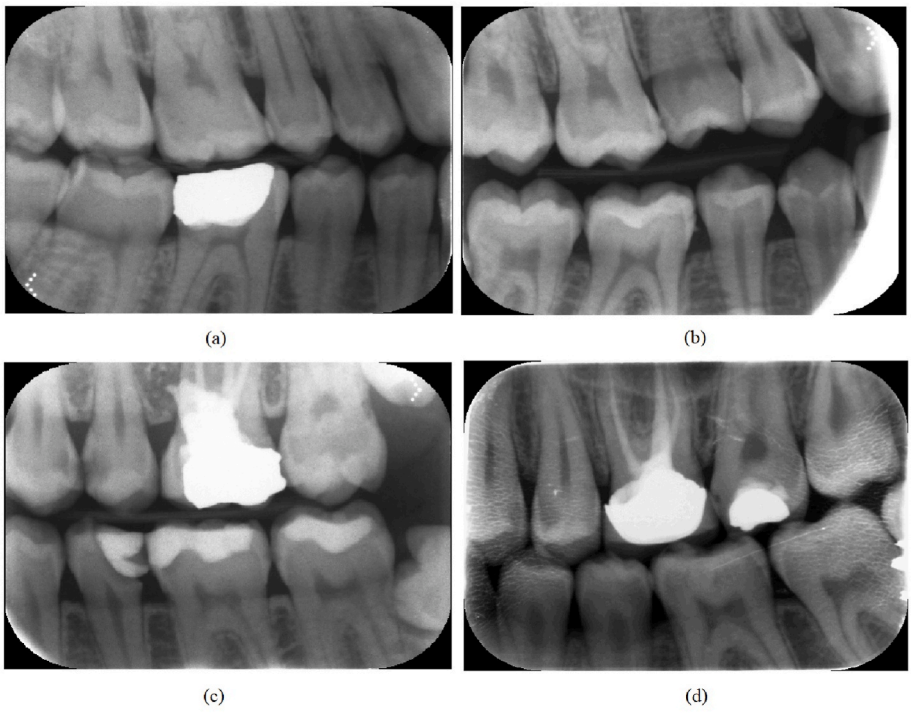


Fig. 1. Sample bitewing radiography images from the dataset.(a) and (b) are left radiograph, (c) and (d) are right radiograph of the intraoral region.

Table 1
Related work on tooth segmentation and numbering.

Literature	Proposed method	Limitations
Said et al. [15]	A tooth identification system has been proposed for missing and unknown persons using dental segmentation techniques on bitewing and periapical images.	To capture the small edges between teeth that morphological filters cannot discriminate, wavelet based method was required.
Lin et al. [16]	In order to classify teeth in bitewing radiography images, a dental classification system that divides them into segments is proposed.	In both maxilla and mandible tooth alignments, the numbering approach produces inaccurate results.
Aeini et al. [17]	In this study, an algorithm is proposed for tooth numbering and classification using the universal number system, using mesiodistal neck information.	As a result of experiments performed, it was presented that the teeth of the upper jaw were incorrectly numbered.
Chen et al. [18]	In this study, filtering algorithm improvement is proposed to delete the overlapping bounding boxes detected by Faster R–CNN.	The cases in which two "half teeth" were faulty for a tooth were deemed insufficient, despite the improvement.
Silva et al. [19]	A segmentation and numbering methods based on 4 various network architectures are proposed.	PANet outperformed other architectures by 70% and above in the evaluation.
Kim et al. [20]	According to the proposed model, a heuristic algorithm has been designed to classification of teeth type.	An IOU of 0.7 indicates that implant fixtures and crowns are less accurate.
Yasa et al. [21]	Faster *R*–CNN, a deep neural network type, shows potential for tooth identification and numbering while analyzing bitewing data.	The exact boundaries of the teeth could not be determined with the proposed solution according to the methods.

numbering in the study, and as a result, the classification was enhanced with segmentation. In dental radiographs, morphological transformations are mostly considered due to problems for instance noise, low contrast and uneven exposure. In the experimental results, the accuracy score was enhanced by using 47 bitewing images for classification and segmentation. In 2010, another study was developed by Aeini and Mahmoudi [17]. The use of the proposed methodology for forensic studies in the automatic tooth identification system is their main purpose. In this way, the benefits of classifying molars and premolars, as well as the spatial relationships between teeth in the jaws, with the proposed approach were utilized in the study. The dataset of 476 tooth images demonstrates that the suggested methodology is more advanced than several previously published studies on tooth numbering, fractured tooth detection, and the number and kind of missing teeth. According to the results of the trials, classification based on contour information should be used in circumstances when the difference between two various types of teeth is less than 0.3%.

In 2019, Chen et al. [18] proposed the use of the CNN network to classify and numbering teeth in periapical X-ray images. In order to enhance the detection accuracy in the study, post-processing is proposed to support the R–CNN network from certain prior domain information. The accuracy and recall of the classification study, which provided an FDI number to each detected tooth, was not satisfactory until certain post-treatment procedures were applied. Finally, the performances of our proposed automated system are improved in this study.

In 2020, Silva et al. [19] presented a work on teeth classification and segmentation in radiographic data through convolutional networks. In this study, the performance of different network architectures such as Mask R–CNN, PANet, HTC and ResNeSt were analyzed and concluded. Among the network architectures, PANet achieved the best result with 71.3% mean average precision (mAP) in segmentation and 74.0% in numbering. As a result of the model's testing, it was determined that the rare misclassifications were caused by the teeth being numbered as one of their neighbors. Mouths with deformed teeth, mislabeling of teeth, and vulgar descriptions yielded the lowest outcomes.

In 2020, Kim et al. [20] applied a new technique combining R–CNN, Single Shot Multi Box Detector (SSD) and optimization algorithms to identify and classify tooth and implant boxes in dental panoramic radiographic data. To identify the numbering of individual teeth, tooth classification and numbering algorithms were trained once dental items were detected. R–CNN was made to classify objects into each number using extracted dental information and location data. Then, utilizing the model as a guide, an algorithm is created to classify each tooth type. According to this approach, each tooth was identified and classified by assigning a number. As a consequence of the model's classification, the searched digits 1 and 7 were entered as the tooth's second digit and trained to classify them as incisors, canines, or molars based on tooth morphology.

In 2021, Yasa et al. [21] proposed an AI model approach for tooth detection and numbering using the Faster R–CNN algorithm, a convolutional neural network. It was utilized, which was an improved method to detect object. A convolutional neural network was used to determine the weighting elements on the dataset consisting of training and validation datasets. Of the 715 dental objects found in the 109 bitewing radiographs used in the study, 697 were accurately numbered and classified in test data of the dataset. In order to measure the accuracy of the neural network, the values obtained from the confusion matrix were used in the metric calculation. Crowns, bridges, implants, and implant-supported restorations were not present in the samples, and 12 radiographs showed a missing tooth. Primary teeth were not included in the study, and only adult radiographs were used. These limits, it is suggested, will be the subject of future research.

2. Tooth segmentation

The technique of separating objects from the image backdrop is known as image segmentation. It is commonly obtained by subtracting the detected objects from the background. A measure of density is typically used to locate an object, moreover a target identification technique as a different technique [15]. The precise identification of the shapes and edges of the teeth plays a vital part in classification when the tooth structures in dental bitewing images are analyzed. For this reason, the full shapes of the teeth are masked by providing segmentation with the detection stage.

The review for the use of CNN concluded that only bounding boxes and class scores were assigned in the section up to Faster R–CNN. For this reason, it was decided to use the Mask R–CNN, which is frequently used in the literature, for the segmentation process. Thus, while achieving high performance at a high threshold, the tooth and pulp masks in the images will be obtained at the same time. CNNs are extensively used for image classification [22].

2.1. Mask R–CNN

Mask R–CNN estimates segmented masks within every pixel region in addition to classification and regression boxes, using segmentation masking and the Faster R–CNN method. A mask branch is a tiny fully connected layer which is given to each ROI and shows a pixel-by-pixel masking method. Mask R–CNN is used to obtain high quality segmentation masks as well as extending other neural networks. Considering the complex setup steps and structure of the Faster R–CNN method, Mask R–CNN network is much easier to operate and deploy. Furthermore, the mask component adds even just a modest amount of computational overhead, allowing for a quick experimentation system [23].

Mask R–CNN includes instance segmentation function by painting each object's pixels in a different color. It is a feature extractor that uses a conventional convolutional neural network. Feature Pyramid Network (FPN) is utilized in the backbone network to enable multi-scale detection. FPN enhances the extracting features pyramid through gathering

features in the bottom layer before the first pyramid and adding them to a second pyramid [24]. As the backbone network, ResNet-101, which is a network that extracts image properties, is preferred from deep residual networks [25].

As shown in Fig. 2, in bitewing radiography images tested using Mask R–CNN, each object is painted on a separate color mask. When a sample bitewing image data is given to the neural network, ROI alignment is accomplished. As shown in figure, class objects are determined by extracting feature maps and running them through convolution filters [26]. In addition to the class object name and score, the output contains high-quality segmentation masks. Mask R–CNN is a straightforward modification of Faster R–CNN in theory, but it is critical to design the mask branch appropriately for decent results [24]. The most important advantage is that Faster R–CNN cannot match neural network inputs and outputs per pixel exactly. The ROI pooling methodology performs coarse spatial quantification and has become a de facto method for dealing with data [26]. Mask R–CNN uses pre-trained MS COCO [27] weights in the backbone structure. Region Proposal Network (RPN) is a neural network that estimates the object's boundaries and detection value at each position at the same time. It is trained state-of-the-art to generate higher level region proposals during object detection [28]. Feature pyramids are an important part of identification systems that detect objects of different size [16]. After the features are extracted during convolution processes with the used backbone network, ResNet-101, feature maps are created FPN and anchors are detected in the regions [24].

The neural network model was gradually trained for the head layers up to the head 400 epochs, respectively. It was done using the Adam optimization technique with a 0.001 learning rate. In the segmentation module, the Adam optimizer was employed since it is quick to adjust the weights in the neural network layers and performs better in practice on empirical results [29].

$$L = L_{cls} + L_{bbox} + L_{mask} \tag{1}$$

As seen in Eq. (1), during training the multitasking loss in each generated region of interest is specified with L. Loss value of the classifier L_{cls} and the losses in the bounding box are the same as those specified in L_{box}. L_{cls} and bounding box loss are the same as specified in L_{box}. The part of the segmentation mask has a $Km2$ output per region that codes mxm resolution K masks consisting of 0 and 1, one from each K class. This is supplied one sigmoid per pixel, and the L_{mask} is specified as the cross binary entropy loss. For the determination of the relevant region belonging to the ground truth class k, L_{mask} is just specified on the k -th segmentation mask [23].

The ground truth image labeled with expert opinion is shown in Fig. 3. As a consequence, the region of interest's bounding box and a high-quality mask in the images are created. Therefore, the specific objects to be specified are in their class's segmentation masks and bounding boxes, as shown in figure. After the model is created, the tooth objects are colored with tooth label in Fig. 3, where the real reference value objects are visualized. Non-toothed areas are left blank as they qualify as background. Masking of tooth objects in bitewing images as a result of 400 epoch training is as shown in figure. In addition to tooth segmentation masks, tooth probabilities are shown next to the bounding boxes. In the image provided, dental masks are included in addition to bounding boxes and matching scores.

3. Tooth numbering

During a dental examination, dentists usually offer their assistants numbers that are confusing. These are the numbers that dentists, assistants, and specialists use to communicate in a universal setting. The use of numerical terms throughout the examination when discussing a common tooth speeds up the practical breakdown. As a result, dentists and assistants can agree on a common expression. Dentists and experts benefit from the numbering of teeth since it helps them keep record of their patients clinical data. Various notation approaches are widely used for tooth numbering. Federation Dentaire Internationale (FDI) numbering approach, Palmer notation approach, and Universal numbering approach are the three most frequently used systems today [30]. Afterward, each of the teeth segmented by the Mask R–CNN method was painted with a various mask color. Thus, tooth segmentation, which is the first step in bitewing radiographs, has been completed. After the positions and shapes of the teeth were determined, the next step, tooth numbering, was started.

As the information given to the network, the "teeth" object is assigned as super class. During tooth segmentation and numbering, ROI extraction was also conducted using the requisite ROI align and ROI pooling procedures in Mask R–CNN. In the last step, teeth labeled according to FDI numbering, tooth objects passed through Mask R–CNN network are numbered according to FDI notation. Consequently, teeth were classified as having number information. Our tooth segmentation and numbering system is given in Fig. 4. The technique was launched with the labeling of 1200 bitewing radiographs. Bitewing radiographs are labeled by dentists specialized in Oral and Maxillofacial Radiology. Intact teeth (including crown and root) found on radiographs are labeled as polygon, which is a closed form object surrounded by successive lines. Labeled radiographs represent the tooth numbers corresponding to the GT to be given to the neural network.

3.1. FDI notation system

In this study, FDI two-digit numbering notation, which is a universal system, is used. The teeth numbering system of the FDI (ISO-3950) was applied. According to the numbering system, the numbers in the tens place represent regions and whether the teeth are permanent. The scientific names for teeth such as incisors, canines, molars, and premolars are represented by the numbers in the ones digit. As seen in Fig. 5, the explanations of the teeth according to their codes are given as follows: In digits, number 1 corresponds to central incisor, number 2 lateral incisors, number 3 canine teeth, while number 4 first premolars are represented as number 5 s premolars. In addition, 6 first molars, 7 s molars and 8 numbered third molars are defined as molars.

In the tooth numbering step, the bitewing images given to the Mask R–CNN are reduced to 1024x1024 as a common shape size. Accordingly, the image with the dimensions of 1308x1700 in Fig. 6 has undergone size reduction just before being transmitted to the neural network. Thus,

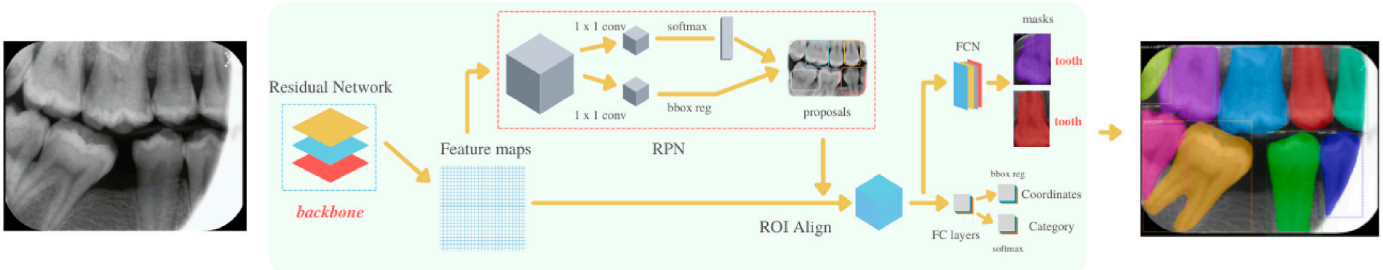
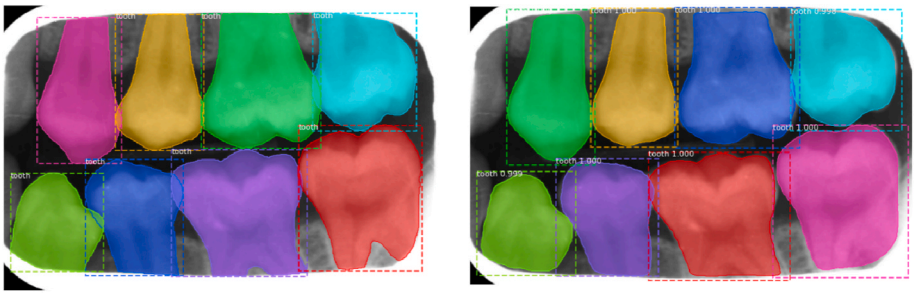


Fig. 2. System architecture and pipeline workflow.



(a) (b)

Fig. 3. Comparison of ground truth and prediction result of a sample bitewing dental image.(a) ground truth (b) prediction result.

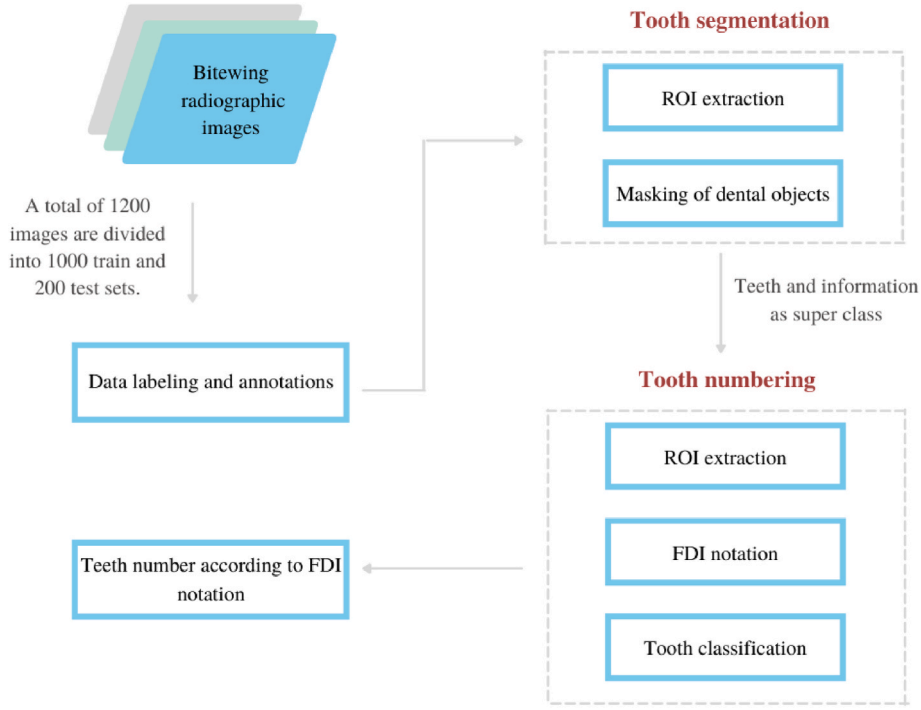


Fig. 4. The working architecture of our system.

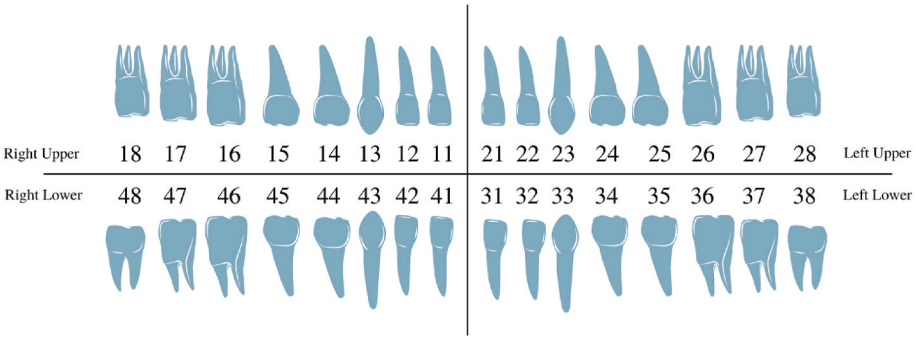


Fig. 5. Tooth numbering system according to FDI notation. 11–18 = right upper, 21–28 = left upper, 31–38 = left lower, 41–48 = right lower.

it is ensured that the model can detect all images in the form of 1024x1024. The molar teeth (numbers 46–47 and 48) shown in the figure are visualized according to the polygon points in the ground truth data labeled with the visualization function. Accordingly, while the background objects are painted in dark blue, the tooth objects are painted white. In this way, the teeth to be detected in the bitewing radiography are masked and their shapes become prominent.

The configuration parameter values for Mask R–CNN utilized for numbering teeth were calculated for the same bitewing radiographic dataset up to 400 epochs. The pre-trained backbone network used with

Fig. 6. Visualization of tooth named by FDI notation.

the model is the residual network. Compared to other network models, ResNet backbone network, which is formed by adding residual values to the next model, has thus ceased to be a classical model. Masking of tooth objects in bitewing images as a result of 400 iteration training is as shown in Fig. 7. In addition to tooth segmentation masks, tooth probabilities are shown next to the bounding boxes. The image in the figure includes dental masks in addition to bounding boxes and matching scores. These objects were painted in a various mask color for each ROI with instance segmentation as a result of training.

As shown in Fig. 7, it was concluded that the tooth number that was unlabeled in the ground truth data of the model was found to be correct as a result of the prediction. As can be seen in the figure, the prediction scores of molar and premolar tooth classes are given. When the prediction scores are compared, it is clear that molars and premolars were predicted with 100% accuracy. However, it can be shown that the unlabeled maxillary first premolar (14) is predicted with 99% accuracy.

4. Experimental results

In this part of the study, segmentation and classification results are presented, which Mask R–CNN method in tooth recognition. The suggested technique was performed on a NVIDIA Quadro RTX5000 using Jupyter Notebook software, one of the Anaconda frameworks. The total number of epochs for tooth segmentation and numbering was determined as 400. The total number of iterations was 400 000 as each training image corresponded to each iteration. The instance segmentation strategy we provided was tested with 200 bitewing radiographs randomly selected from the dataset. During the experiments, the dimensions for all bitewing radiographs were set to 1024*x*1024 so that images of various sizes could be used with Mask R–CNN. Transfer learning is provided with the residual network, which is the backbone network used while training the neural network. Transfer learning algorithms or backbones use trained models, which are ready-to-use neural network models with weights trained on labeled data [31]. Weights to be given to the model should be obtained from a pre-trained MS COCO [27] dataset.

The overlap boxes needed to measure this state are shown in Fig. 8.

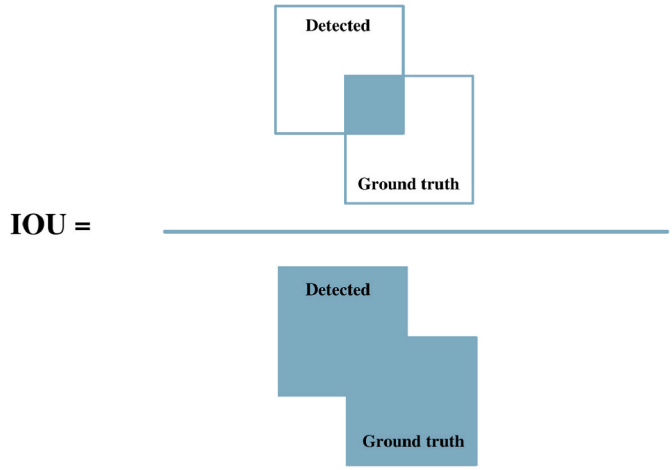
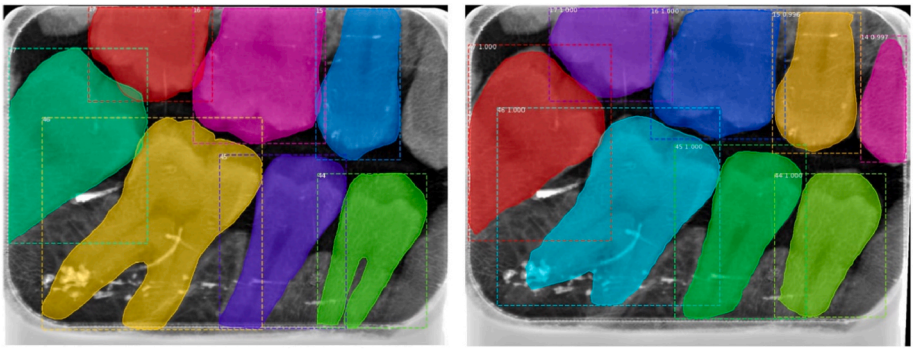


Fig. 8. Intersection over union (IOU) visualization for targeted object and detected masks.

The threshold value is used to determine the amount of overlap of 2 samples based on tile sizes. In order to measure the value, the overlap of the targeted and detected masks is checked. It is given in Eq. (2). If it overlaps more than the specified thresholds, only the highest confidence score is retained. Therefore, certain hyperparameter settings are made as well as neural network configurations. The specified value was determined as 0.5. This means that if the overlapping values are greater than or equivalent to 50%, the detected box is selected as true positive. Otherwise, it is considered a false positive.

$$IOU = \frac{Area_{Detected_{box}} \bigcap Area_{Groundtruth_{box}}}{Area_{Detected_{box}} \bigcup Area_{Groundtruth_{box}}} \quad (2)$$

Adam optimization algorithm for 0.001 learning rate and 400 periods was used to finish the training. At the end of the training, mAP, precision, recall and f1-score were utilized as shown in the equation to measure the accuracy of the existing test images. The most common



(a) (b)

Fig. 7. Segmentation masks of ground truth and prediction results in the sample bitewing image.(a) ground truth (b) prediction result.

accuracy metrics used in object recognition to evaluate model performance are precision and mAP. The calculation of the precision values is shown in Eq. (3). The mAP value is calculated using the average precision values as in Eq. (4).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{4}$$

The area under the precision and recall values graphic gives the AP value for the class. The calculation of recall values, which are among the other calculation metrics, is given in Eq. (5). In addition, precision and recall values are generally inversely related to each other. The f1-score is a metric for evaluating potential optimization methods. is obtained by taking the harmonic average of the recall and precision values and is calculated using the formula [32]:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

4.1. Tooth segmentation and numbering results

For tooth segmentation, overlapping boxes are scanned to non-maximum suppression (NMS). Anchor-based approach of the model divides regions containing objects into anchors with a specific threshold value. The maximum values in the pooling layer are selected to the CNN. According to the FDI notation numbering system, given to each tooths to one-of-a-kind number. This notation method is first segmented and assigns the number value to which the classified tooth result belongs. As a result of the experimental studies, a total of 200 bitewing radiographs taken randomly were tested separately based on Eq. (3), Eq. (4), Eq. (5) and Eq. (6). The performance metrics in Table 2 and Table 3 summarizes to comparison findings of ResNet-101 backbones.

Considering these performance metrics, it has been determined that the model with the highest number of epochs produces more successful results. When the values in Tables 2 and 3 are compared, it is clear that Table 2 has better results. The main reason is because the neural network only recognizes it as teeth/background. The model's ability to forecast becomes increasingly challenging as the number of classes grows. According to the experimental results obtained for tooth segmentation, 100% precision and 97.49% mAP value were reached. In the 24-class tooth numbering module, 94.35% precision and 91.51% mAP value were obtained. Although the R–CNN–based Mask R–CNN algorithm used in the study makes an important contribution to the literature, the examination and testing of other state of the art algorithms for the classification of teeth and tooth numbers are also given in the experimental studies section. In order to prove the scientific adequacy of the study, the performance of the algorithms was compared using different deep learning networks in the literature. Because retraining the weights would be a waste of time, it was ensured that high-performance weights learned using data from many fields were employed more conveniently [33]. For this reason, pre-trained convolutional neural network algorithms, which are frequently used for detection purposes in the literature were used.

Table 2
Experimental results table of tooth segmentation.

Epoch	Precision	mAP	Recall	F1-Score
50	89.99%	92.54%	83.45%	85.27%
100	90.90%	94.81%	89.25%	90.86%
200	92.70%	95.51%	93.57%	94.19%
400	100.0%	97.49%	97.24%	97.36%

Table 3
Experimental results table of tooth numbering.

Epoch	Precision	mAP	Recall	F1-Score
50	64.37%	62.85%	53.15%	57.59%
100	89.99%	82.40%	75.13%	64.92%
200	91.23%	90.00%	86.03%	88.14%
400	94.35%	91.51%	95.20%	93.33%

In this study, 1000 train and 200 test images, which were determined as the dataset, were cropped according to the bounding boxes at the labeling points to be classified by deep learning classifiers. As shown in Fig. 9, the dataset containing 1200 data was cropped to cover a specific tooth according to the points labeled by the experts and set as the input image. Then, it was reduced to 224*x*224 dimensions. Thus, uniformity was obtained in the data set to be given to the model. All ConvNets, including Mask R–CNN, which was the first approach used in the study, were trained on ImageNet [34] with well-labeled data instead of being trained from scratch [35]. As a result, experimental analysis was shown in the study by applying ConvNets as a comparison approach at the stage of detecting and classifying tooth numbers.

In the study, it is aimed to give successful results by transferring the weights of pre-trained ConvNets(ResNet [25], AlexNet [36], VGGNet [37], DenseNet [38], SqueezeNet [39], GhostNet [40], MobileNet [41], and GoogleNet [42]) with the improvement step, also known as fine-tuning, in the transfer learning stage. In a classification task with a total of 24 classes, the dataset is divided into 24 separate classes and divided into train and test sets. In this step, the size of the output layer has been reduced to 24, as there are 24 classes in the classification task. ConvNet models were prepared for training using Adam optimizer, which is the proposed gradient descent algorithm by combining the advantageous aspects of RMSProp [43]. And it is also very suitable for problems with large parameters [29].

When Table 4 is examined, the gamma value is 0.1, the momentum coefficient is 0.9, and the step size value is 1954 for the accuracy results given. VGGNets, which are very deep convolutional networks for classification of large-scale images, are a ConvNet model with drastically increased depth [37]. As a result of the observations when the VGG-16 model was trained, the train loss for the latest epoch reached 0.8852, while the best test loss reached 1.1757. DenseNet is a ResNet-based neural network that includes extra inputs from previous levels and uses dense blocks to convey information to the following layers [44]. In addition, 16 batch sizes were selected for DenseNet's 121-layer neural network, and when the training was completed, the train accuracy and test accuracy values were much higher than other networks. However, when the training results in which the DenseNet-121 was made by choosing 64 batch size were examined, it was observed that there was a very high difference between train accuracy and test accuracy. This means that the neural network is overfitting. Overfitting adversely affects the results by causing the network to memorize excessively during training. For this reason, it has been decided to select 16 as the batch size hyperparameter to be used in other neural networks.

Completing the training by dividing the dataset into 16 batches rather than dividing it into 64 batches means that it produces more successful results. Furthermore, when the step size hyperparameter was investigated, it was concluded that the number of data in the test set would affect the performance positively. The neural network classifiers in the table were trained with the determined parameters and compared with the Mask R–CNN. SqueezeNet, another neural network frequently used in classification, was used because it provides 50 times less parameters at the Alexnet level [39]. Thus, AlexNet and SqueezeNet classifiers should be compared. The common point of SqueezeNet and AlexNet networks is that both reach approximately the same level of accuracy when evaluated on the ImageNet image classification test dataset. When the data in Table 4 is examined, it has been concluded that the accuracy values are almost closer to each other than the other

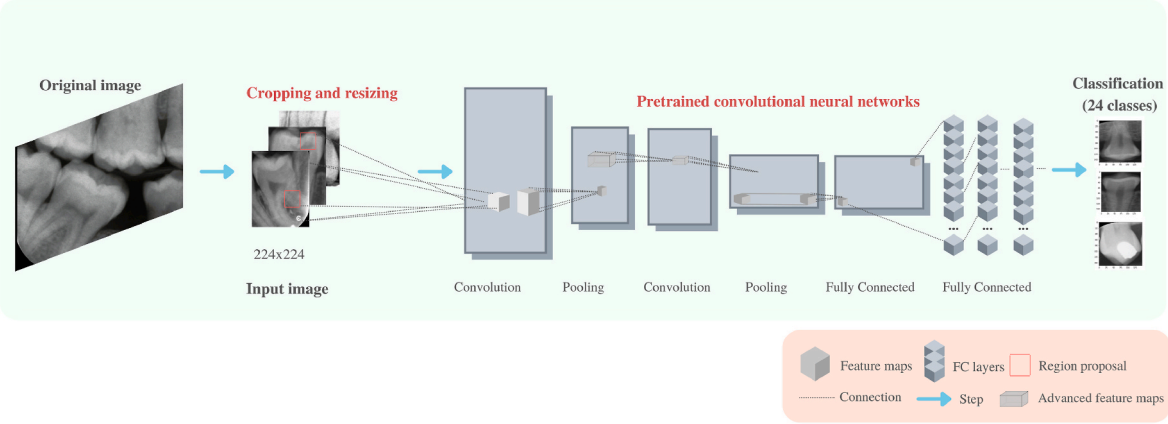


Fig. 9. A visualization of transfer learning architecture for comparison of classification algorithms using PyTorch framework in the study. One Convolutional Neural Network (ConvNet) is fine-tuned for classification of tooth numbers using weights from another ConvNet pre-trained on the ImageNet dataset.

Table 4
Comparison table of different deep learning architectures for tooth numbering according to FDI notation. This table is based on a comparison of individual findings of different network architectures for tooth numbering study.

Neural network	Batch size	Optimizer	Learning rate	Accuracy	Precision	Recall	F1-Score
SqueezeNet	16	Adam	0.001	58.68%	59.06%	59.06%	59.06%
AlexNet	16	Adam	0.001	64.67%	64.66%	65.07%	64.87%
VGGNet-16	16	Adam	0.001	65.62%	65.61%	65.82%	65.71%
WideResNet-101	16	Adam	0.001	70.98%	70.75%	71.42%	71.08%
GhostNet	16	Adam	0.001	72.87%	74.27%	72.87%	73.56%
DenseNet-121	16	Adam	0.001	78.23%	78.23%	78.23%	78.23%
HarDNet-85	16	Adam	0.001	78.55%	78.54%	78.79%	78.67%
ResNet-50	16	Adam	0.001	79.49%	79.75%	80.00%	79.87%
MobileNet-v2	16	Adam	0.001	80.44%	80.45%	80.44%	80.44%
DenseNet-201	16	Adam	0.001	81.39%	81.38%	81.38%	81.37%
GoogleNet	16	Adam	0.001	82.33%	82.85%	82.33%	82.59%
DenseNet-169	16	Adam	0.001	82.97%	82.70%	82.96%	82.83%
Mask R–CNN	**2**	**Adam**	**0.001**	**91.51%**	**94.35%**	**95.20%**	**93.33%**

networks. The GhostNet architecture, on the other hand, creates several ghost feature maps based-on a set of internal feature maps that can fully reveal the underlying information of the internal features [40]. In contrast to conventional residual models that use extended representations, MobileNet-v2 architecture is based on an inverted redundancy structure, where the input and output of residual blocks are thin layers of bottlenecks [41]. Harmonic DenseNet (HarDNet) was used to compare with DenseNet. For comparison, it runs 35% faster than ResNet running on the graphical processing unit compared to models with the same achievement [45].

ResNet-50 [25] and WideResNet-101 neural networks were also added to the study to compare the ResNet used as the backbone of Mask R–CNN. WideResNet has now been enhanced by introducing a "bottleneck" block that makes the ResNet blocks even thinner, with the authors now trying to make them as thin as possible in order to increase their depth and have fewer parameters [46]. Although there are great differences between WideResNet-101 and ResNet-50 in terms of both layers and depth, it is not always preferred to go deeper into neural networks in the literature. Because there are often limitations in going deep into neurons, one of the biggest reasons for this is the parts that are never mentioned because of going deep in the neurons. As a result, fine tuning rather than diving deeper will yield better outcomes.

Although Mask R–CNN seems to produce segmentation results, it produces the class name and accuracy result for classification purposes in addition to the masks of each tooth. And to train the Mask R–CNN model, 1 or 2 batch sizes with a tiny quantity per graphical processing unit is required. For this reason, batch size 2 of Mask R–CNN was chosen in the comparison study. Due to graphical processing unit memory, a large batch size causes a GPU out of memory error. Compared to other highly developed classification algorithms, Mask R–CNN has obtained superior results from all classifiers.

4.2. Forecasted failures

Although tooth segmentation and numbering has a high performance, it can be misclassified due to the anatomically symmetrical mouth anatomy and the alignment of the maxilla and mandible teeth. The bitewing radiographs used in the study allowed only certain cropped mouth regions to be used, not the entire mouth region. Furthermore, bitewing radiographs were shown to be more challenging to detect than panoramic radiographs because they do not show the inside of the mouth in as much detail.

The most disadvantageous characteristic of bitewing radiography analysis is that the X-ray images can occasionally reflect completely various tooth areas. For example, as can be seen in Fig. 10, the canine tooth of number 13 is present in a small area in some bitewing radiographs, while in others it is displayed larger. This situation causes the mask dimensions to change and therefore the detection ability of the model decreases. Thus, in line with the resulting algorithm, it was foreseen that the tooth class should be determined with a high accuracy score, but it would decrease to lower values such as 89%. Although it does not appear exactly on many bitewing radiographs, tooth number 13 has been identified, despite being of lower accuracy.

5. Conclusions

In this study, tooth segmentation and numbering were performed using Mask R–CNN, a convolutional neural network type, on bitewing

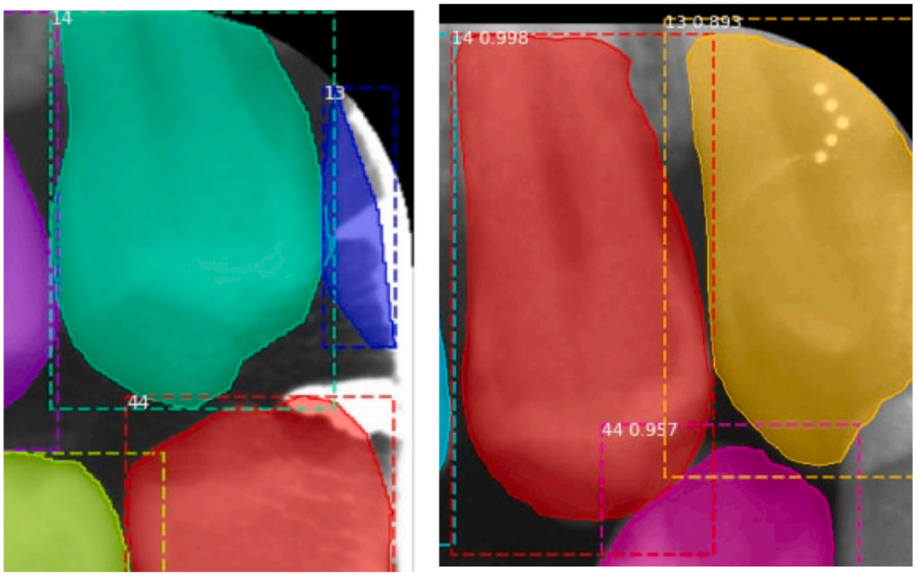


Fig. 10. A comparison of bitewing radiographs of the number 13 canine teeth.

radiographic images. As a result of Mask R–CNN, high quality segmentation masks were obtained in addition to the bounding box and class scores compared to other convolutional neural networks. For the classifiers used in the study and the Mask R–CNN, almost many hyperparameters were selected and training was provided on equal terms. This is very important in order to in classifiers.

The accuracy, precision, recall, and f1-measurement values for the most often used and most recommended classifiers are presented in Table 4. Among the classifier networks, GoogleNet architecture produced the highest accuracy and precision scores, while SqueezeNet produced the lowest. One of the reasons why the classical ResNet-50 classifier was especially used in the comparison step in the study was that the Mask R–CNN used ResNet as a backbone in its own architecture. However, when the ResNet-50 and Mask R–CNN classifiers are compared, it is concluded that there is a remarkable difference in terms of performance metrics.

The confusion of other classification networks is relatively high compared to Mask R–CNN results. Experts in clinical and practical use find this condition to be unsatisfactory. The highest accuracy and sensitivity scores are 82.33% and 82.97%, respectively, for DenseNet-169 and GoogleNet classifiers. When the different architectures of 201 layer and 169 layer of DenseNet network are compared, it is observed that DenseNet-169 gives better results. The main reason for this is that, as in other classifiers, some information is always lost when going deep. Moreover, high precision and accuracy can provide precise medical information on tooth numbering in FDI notation. In addition to classification, the most important reason for segmentation in the study is to distinguish and number the teeth from the background and other teeth by masking them in case of overlapping of more than one tooth anatomically on radiographs. In comparison to the other classifiers shown in Table 4, Mask R–CNN had the best classification performance. The major reason for this is that tooth objects are supported by segmentation masks in addition to their positions. It is aimed with the solution proposed in the study to provide support in the decision-making processes of dentists during radiographic analysis. High precision performance is provided with the system designed for the analysis of teeth. The artificial intelligence-based automatic teeth classification and segmentation of the system has been validated by dentists in clinical medical examinations and applications.

In the future studies, the segmentation and numbering of teeth in more complex data can be provided, so that the neural network model can achieve more efficient results. In the study, it is aimed to eliminate the mentioned possibilities with an intuitive approach in future studies, assuming that there are missing teeth, incorrect data labeling, the presence of anatomical asymmetry in the mouth and high overlap ratios of the teeth. However, bitewing radiographic images can detect not only teeth but also dental caries, implants and fillings.

Acknowledgement

This study is funded by the Scientific and Technological Research Council of Turkey (TUBİTAK) as part of the DentiAssist project numbered 2200272.

References

[1] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248.
[2] F.G. Zanjani, A. Pourtaherian, S. Zinger, D.A. Moin, F. Claessen, T. Cherici, S. Parinussa, P.H. de With, Mask-mcnet: tooth instance segmentation in 3d point clouds of intra-oral scans, Neurocomputing 453 (2021) 286–298.
[3] S. Abbasi, M. Hajabdollahi, P. Khadivi, N. Karimi, R. Roshandel, S. Shirani, S. Samavi, Classification of diabetic retinopathy using unlabeled data and knowledge distillation, Artif. Intell. Med. 121 (2021), 102176, https://doi.org/10.1016/j.artmed.2021.102176. URL, https://www.sciencedirect.com/science/article/pii/S093336572100169X.
[4] M.J. Cardoso, N. Houssami, G. Pozzi, B. Séroussi, Artificial intelligence (ai) in breast cancer care - leveraging multidisciplinary skills to improve care, Artif. Intell. Med. (2020), 102000, https://doi.org/10.1016/j.artmed.2020.102000. URL, https://www.sciencedirect.com/science/article/pii/S0933365720312653.
[5] E. Kröger, M. Dekiff, D. Dirksen, 3d printed simulation models based on real patient situations for hands-on practice, Eur. J. Dent. Educ. 21 (2017) e119–e125.
[6] J. Waring, C. Lindvall, R. Umeton, Automated machine learning: review of the state-of-the-art and opportunities for healthcare, Artif. Intell. Med. 104 (2020), 101822, https://doi.org/10.1016/j.artmed.2020.101822. URL, https://www.sciencedirect.com/science/article/pii/S0933365719310437.
[7] M. Chan, T. Dadul, R. Langlais, D. Russell, M. Ahmad, Accuracy of extraoral bitewing radiography in detecting proximal caries and crestal bone loss, J. Am. Dent. Assoc. 149 (2018) 51–58.
[8] B. Vandenberghe, R. Jacobs, H. Bosmans, Modern dental imaging: a review of the current technology and clinical applications in dental practice, Eur. Radiol. 20 (2010) 2637–2655.
[9] K. Suzuki, Overview of deep learning in medical imaging, Radiological physics and technology 10 (2017) 257–273.
[10] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, H. Fujita, Classification of teeth in cone-beam ct using deep convolutional neural network, Comput. Biol. Med. 80 (2017) 24–29.
[11] J. Aps, Extraoral Radiography in Pediatric Dental Practice, 2019, pp. 31–49.
[12] S.A. Prativi, S. Chairani, T. Hestingsih, Silicone loop alternative for posterior bitewing radiography, Dent. J. 54 (2021) 35–38.
[13] F. Schwendicke, G. Gostemeyer, Conventional bitewing radiography, Clin. Dentisy Rev. 4 (2020) 1–8.

[14] M.V. Valueva, N. Nagornov, P.A. Lyakhov, G.V. Valuev, N.I. Chervyakov, Application of the residue number system to reduce hardware costs of the convolutional neural network implementation, Math. Comput. Simulat. 177 (2020) 232–243.
[15] E. Said, G.F. Fahmy, D. Nassar, H. Ammar, Dental x-ray image segmentation, in: Biometric Technology for Human Identification, vol. 5404, 2004, pp. 409–417. International Society for Optics and Photonics.
[16] P.-L. Lin, Y.-H. Lai, P.-W. Huang, An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information, Pattern Recogn. 43 (2010) 1380–1392.
[17] F. Aeini, F. Mahmoudi, Classification and numbering of posterior teeth in bitewing dental images, in: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vol. 6, 2010, pp. V6–V66. IEEE.
[18] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, C.-H. Lee, A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films, Sci. Rep. 9 (2019) 1–11.
[19] B. Silva, L. Pinheiro, L. Oliveira, M. Pithon, A study on tooth segmentation and numbering using end-to-end deep neural networks, in: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI), 2020, pp. 164–171. IEEE.
[20] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, S. Youm, Automatic tooth detection and numbering using a combination of a cnn and heuristic algorithm, Appl. Sci. 10 (2020) 5624.
[21] Y. Yasa, Ö. Çelik, I.S. Bayrakdar, A. Pekince, K. Orhan, S. Akarsu, S. Atasoy, E. Bilgir, A. Odabaş, A.F. Aslan, An artificial intelligence proposal to automatic teeth detection and numbering in dental bite-wing radiographs, Acta Odontol. Scand. 79 (2021) 275–281.
[22] M. Momeny, A.M. Latif, M.A. Sarram, R. Sheikhpour, Y.D. Zhang, A noise robust convolutional neural network for image classification, Results Eng. 10 (2021), 100225.
[23] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[24] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira, Deep instance segmentation of teeth in panoramic x-ray images, in: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2018, pp. 400–407.
[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[26] Z. Cai, N. Vasconcelos, Cascade R-Cnn: High Quality Object Detection and Instance Segmentation, 2019. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[27] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: common objects in context, in: European Conference on Computer Vision, Springer, 2014, pp. 740–755.
[28] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S. I. Nikolenko, M.M. Sveshnikov, G.B. Bednenko, Tooth detection and numbering in panoramic radiographs using convolutional neural networks, Dentomaxillofacial Radiol. 48 (2019), 20180051.
[29] D.P. Kingma, J. Ba, Adam, A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.
[30] S. Peck, L. Peck, A time for change of tooth numbering systems, J. Dent. Educ. 57 (1993) 643–647.
[31] N.S. Shaik, T.K. Cherukuri, Transfer learning based novel ensemble classifier for covid-19 detection from chest ct-scans, Comput. Biol. Med. 141 (2022), 105127.
[32] F.P. Mahdi, K. Motoki, S. Kobashi, Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs, Sci. Rep. 10 (2020) 1–12.
[33] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big data 3 (2016) 1–40.
[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. Ieee.
[35] E. Baykal, H. Dogan, M.E. Ercin, S. Ersoz, M. Ekinci, Transfer learning with pre-trained deep convolutional neural networks for serous cell classification, Multimed. Tool. Appl. 79 (2020) 15593–15611.
[36] A. Krizhevsky, One Weird Trick for Parallelizing Convolutional Neural Networks, 2014 arXiv preprint arXiv:1404.5997.
[37] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.
[38] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[39] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, Squeezenet: Alexnet-Level Accuracy with 50x Fewer Parameters and¡ 0.5 Mb Model Size, 2016 arXiv preprint arXiv:1602.07360.
[40] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: more features from cheap operations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1580–1589.
[41] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets, Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017 arXiv preprint arXiv:1704.04861.
[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[43] E. Seyyarer, F. Ayata, T. Uçkan, A. Karcï, Derin öğrenmede kullanilan optimizasyon algoritmalarinin uygulanmasi ve kiyaslanmasi, Computer Science 5 (2020) 90–98.
[44] A. Gurses, A.B. Oktay, Tooth restoration and dental work detection on panoramic dental images via cnn, in: 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1–4. IEEE.
[45] P. Chao, C.-Y. Kao, Y.-S. Ruan, C.-H. Huang, Y.-L. Lin, Hardnet: a low memory traffic network, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3552–3561.
[46] S. Zagoruyko, N. Komodakis, Wide Residual Networks, 2016 arXiv preprint arXiv:1605.07146.