scientific reports

OPEN

Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs

Fahad Parvez Mahdi✉, Kota Motoki & Syoji Kobashi

Computer-assisted analysis of dental radiograph in dentistry is getting increasing attention from the researchers in recent years. This is mainly because it can successfully reduce human-made error due to stress, fatigue or lack of experience. Furthermore, it reduces diagnosis time and thus, improves overall efficiency and accuracy of dental care system. An automatic teeth recognition model is proposed here using residual network-based faster R-CNN technique. The detection result obtained from faster R-CNN is further refined by using a candidate optimization technique that evaluates both positional relationship and confidence score of the candidates. It achieves 0.974 and 0.981 mAPs for ResNet-50 and ResNet-101, respectively with faster R-CNN technique. The optimization technique further improves the results i.e. F_1 score improves from 0.978 to 0.982 for ResNet-101. These results verify the proposed method's ability to recognize teeth with high degree of accuracy. To test the feasibility and robustness of the model, a tenfold cross validation (CV) is presented in this paper. The result of tenfold CV effectively verifies the robustness of the model as the average F_1 score obtained is more than 0.970. Thus, the proposed model can be used as a useful and reliable tool to assist dental care professionals in dentistry.

Visual examination by dental care experts during dental treatment alone cannot provide sufficient information to diagnose a number of dental anomalies. It is because of their location in the mineralized tissues (bone and teeth). Thus, it is indispensable to use digital radiographs during dental treatment. Immediate availability of digital images, limited radiation dose and the possibility of applying image processing techniques (such as image enhancement and image registration) are some of the advantages of dental radiographs. However, the possibility of computer aided analysis of digital radiographs is one of the most important aspects of using dental radiographs. Dental panoramic radiograph or simply pantomograph is a type of dental radiographs that provides the dentist an unobstructed view of the whole dentition (both upper and lower jaws). It contains detail information of the dentomaxillofacial anatomy[1].

The usage of dental radiographs is growing day by day and therefore, it is highly desirable to assist dentist with computer-aided analysis. Automatic recognition of teeth from dental radiographs can be a great way to aid dentists in dental treatment. It will not only reduce workload from dental professionals but also reduce interpretation error and diagnosis time, and eventually will increase the efficiency of dental treatment. Usage of machine learning (ML) and computer vision techniques is not new in dental radiographs. Nomir et al.[2] proposed an automatic system that can identify people from dental radiographs. The proposed system can segment the radiographs into individual teeth automatically, represent and match teeth contours and finally provides matching scores between antemortem (AM) and postmortem (PM) teeth. Nassar et al.[3] created a prototype architecture of automated dental identification system (ADIS) to address the problem of postmortem identification through matching image feature. This matching problem was tackled by high level feature extraction in the primary step to expedite retrieval of potential matches followed by image comparison using inherent features of dental images. To detect areas of lesions in dental radiographs, a semiautomatic framework is proposed by Li et al.[4] by using level set method. The framework, at first, segments the radiograph into three meaningful regions. It was done by using two coupled level set functions. Then, an analysis scheme influenced by a color emphasis scheme prioritizes radiolucent areas automatically. After that the scheme employed average intensity profile based method to isolate and locate lesions in teeth. The framework improved the interpretation in a clinical settings and enables dentist

Graduate School of Engineering, University of Hyogo, 2167 Shosha, Himeji, Hyogo 671-2280, Japan. ✉email: fahadapecedu@gmail.com

to focus their attention on critical areas. Local singularity analysis based teeth segmentation was proposed by Lin et al.[5] in periapical radiographs in order to detect periapical lesion or periodontitis. This method works on four different stages that include adaptive power law transformation (as image enhancement technique), Hölder exponent (for local singularity analysis), Otsu's thresholding and connected component analysis (as tooth recognition) and finally, snake boundary and morphological operations (for tooth delineation). The overall accuracy (considering true positive) was found to be near 90 percent.

Kavitha et al.[6] employed a new support vector machine (SVM) method to diagnose osteoporosis (a disease that increases the risk of fractures in bone) at early stage to reduce the risk of fractures. They utilized dental panoramic radiographs to measure the thin inferior cortices of mandible which is very useful to identify osteoporosis in women.

The limitation of the conventional ML techniques in processing raw natural data requires careful engineering to construct feature extractor in order to transform the raw data into a suitable representation for detecting or classifying input patterns. This limitation was overcome effectively by the introduction of deep learning techniques. Deep learning techniques are representation learning based techniques that allow a machine to be fed with raw data and then process the data in different layers to automatically discover necessary representations to detect or classify input data. The main advantage of deep learning is that these layers of features are learned directly from the raw data by using a general purpose learning procedure instead of design constructed by the human engineers[7]. It has thus caused remarkable improvements in artificial intelligence. Deep learning techniques beat records in image[8] and speech recognitions[9], supersedes other machine learning techniques in analyzing particle accelerator data[10], predicting activity in potential drug molecules[11], reconstructing brain circuits[12], and produced promising results in natural language understanding[13].

Transfer learning based convolutional neural network (CNN) was utilized by Prajapati et al.[14] to classify three kinds of dental diseases from dental radiographs. They have utilized a pretrained VGG16[15] as feature detector. Lin et al.[16] proposed an algorithm based on CNN to automatically detect teeth and classify their conditions in panoramic dental radiographs. In order to increase the amount of data, different data augmentation techniques such as flipping and random cropping are used. They claimed to achieve accuracy around 90% using different image enhancement techniques along with CNN. Chen et al.[17] proposed faster R-CNN technique that included three post processing steps on dental periapical films. The post processing steps included a filtering system, a neural network model and a rule-base module to refine and supplement faster R-CNN. Although, the detection rate was exceptionally well, there classification result was only very close to the level of a junior scientist even after applying three post-processing steps. Tuzoff et al.[18] proposed a model that used faster R-CNN for teeth detection, VGG-16 based convolutional network for classification and heuristic-based algorithm for result refinement. Although, their heuristic algorithm heavily depended on the confidence scores produced by the convolution network, adequate performance analysis of the convolutional network for teeth classification was not present. Muramatsu et al.[19] proposed a fully convolutional network (FCN) based on GoogleNet to detect teeth. A ResNet-50 based pretrained network was then used to classify tooth by its type i.e. incisors, canines, premolars, and molars as well as three different tooth conditions. They tried to improve the classification result by introducing double input layers with multisized image data. However, their final classification result i.e. 93.2% for teeth classification was fallen short to the required accuracy needed for clinical implementation.

This paper proposes a residual network based faster R-CNN algorithm in panoramic radiographs for automatic teeth recognition. Faster R-CNN object detector is the modified and upgraded version of R-CNN[20] and fast R-CNN[21]. The main advantage of faster R-CNN is that it does not need a separate algorithm for region proposals; rather the same convolution network is used for region proposal generation and object detection and hence much faster than its predecessors are. Two variants of trained residual network i.e. ResNet-50 and ResNet-101 are utilized in this paper to increase the effectiveness of the proposed system. Residual network is widely known for mitigating infamous vanishing gradients problem in deep network[22]. This paper proposes a candidate optimization algorithm based on prior knowledge of the dataset to further refine the detection results obtained by residual network based faster R-CNN. The proposed candidate optimization method considers both the position patterns of detected boxes as well as the confidence scores of the candidates given by the faster R-CNN algorithm to refine the detected boxes. The proposed method, thus, combines an optimization algorithm with deep learning technique for teeth recognition in dental panoramic radiographs.

The rest of the paper is structured as follows; "Materials" section describes about the data i.e. dental radiographs used in this research. It also describes the tooth numbering systems used for testing the performance of the proposed method. "Proposed method" section presents the method and the architecture of the proposed model as well as explanation of training and test datasets. It also includes description of the candidate optimization method proposed in this paper. "Results and discussions" section comprises of results and discussions include simulation results. Finally, the paper is concluded with the "conclusion" section, providing the gist of the paper and possible future work.

Materials

A total of 1000 panoramic radiographs were collected for this research. The dimension of the images was around $(1400–3100) \times (800–1536)$ pixels and stored as a jpeg format. The images were collected by Narcohm Co. Ltd. from multiple dental clinics under the approval of each clinic. The authors obtained the images from Narcohm Co. Ltd. with permission. Also, the images were anonymously collected so that no additional information, like age, gender or height was revealed. Figure 1 shows an example of a collected dental panoramic radiograph. For the sake of training and validation, all the images were labeled by putting a rectangular bounding box around each tooth with proper roots and shape. The panoramic radiographs consisted of normal teeth, missing teeth, residual roots and dental implants. This paper followed universal tooth numbering (UTN) system. In universal

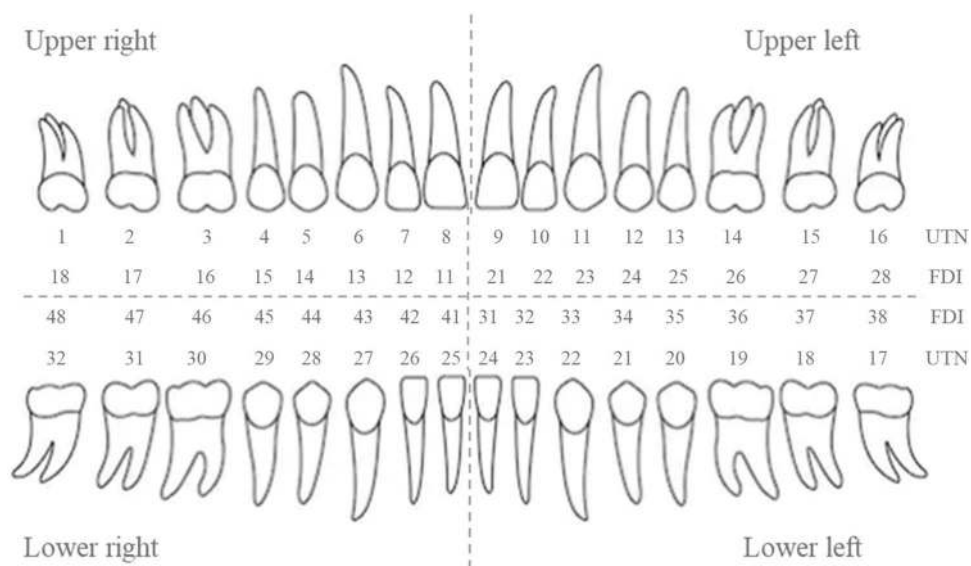**Figure 1.** Dental panoramic radiograph or pantomograph.

Figure 2. Universal tooth numbering (UTN) system and Fédération Dentaire Internationale (FDI) system.

tooth numbering system, teeth count starts from upper right part to upper left part as 1 to 16 and then lower left to lower right as 17 to 32. Fédération Dentaire Internationale (FDI)[23] and Palmer notation (PN)[24] are other two notable tooth numbering systems. Figure 2 illustrates the UTN and FDI system simultaneously.

Annotated 1000 radiographs were split into total tenfolds, each containing 100 radiographs; (1) training dataset consisted of total ninefolds and (2) test dataset consisted of onefold. Therefore, training and test datasets were consisted of 900 radiographs and 100 radiographs, respectively. Training dataset was used to train faster R-CNN, whereas test dataset was used to analyze and validate the performance of the proposed method.

Proposed method

The proposed method consists of two steps; (i) candidate detection and (ii) optimization. The first step detects candidates from the panoramic radiographs using faster R-CNN and in the second step, the detected candidates are refined using an optimization method. Image-based CNN detectors are used in this research and therefore, a brief overview of these detectors are given at first. Next, deep learning-based faster R-CNN technique, its architecture and criteria of introducing transfer learning techniques are presented and discussed. Finally, the proposed optimization method based on prior knowledge is explained along with brief descriptions of performance evaluation metrics.

Layer name	Output size	ResNet-50	ResNet-101
conv1	112×112	7×7, 64	
conv2_x	56×56	3×3 max pool $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times4$
conv4_x	14×14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$
conv4_x	7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
avg_pool, fc1000, fc1000_softmax	1×1	Average pool, classification, softmax	

Table 1. Architectures of ResNet-50 and ResNet-101.

Image-based CNN detectors and candidate detection using faster R-CNN. Mainly two types of image-based CNN detectors have been developed, they are (i) single-stage methods and (ii) two-stage methods. Both types of method utilize multiple feature maps of different resolutions for object detection. These feature maps are generated by a backbone network e.g. AlexNet[8] or ResNet[22].

The single-stage name came from the fact that these kind of methods performed directly into these multi-scale feature maps for object detection. In contrast, two-stage methods at first work on the feature maps to generate region proposals from the anchor boxes. Anchor boxes are a set of predefined bounding boxes with different aspect ratio. The network that generates region proposals is known as region proposal network (RPN) in faster R-CNN. RPN produces region proposals by predicting whether the anchor boxes contain an object or not (without classifying which object). Region proposals with best confidence scores are then processed into the second stage for further classification and regression. Thus, the region proposals are classified and regressed twice and that is why usually these kinds of methods achieve higher accuracy. However, the second stage computation adds an extra computational burden and the system, thus, tends to be less efficient and slow[25].

In this research, faster R-CNN[26] technique is used for automatic teeth recognition. Faster R-CNN is modified and updated version of fast R-CNN[21]. It utilizes two different modules; one is deep convolutional network also known as region proposal network (RPN) for region proposals and second is fast R-CNN object detector that utilizes proposed regions. Two modules, however, work as single unified network for object detection. The main advantage of faster R-CNN from its predecessors R-CNN[20] and fast R-CNN is that it successfully alleviate the problem of needing a separate algorithm for region proposals and thus enabling a cost effective region proposals. A single unified network is used for both region proposals and object detection. The technique won 1st place in several tracks e.g. ImageNet detection, ImageNet localization, COCO detection and COCO segmentation of ILSVRC and COCO competitions[26].

Transfer learning. Transfer learning is a technique used in both machine learning and deep learning problems to improve the learning performance of a particular task through transferring the knowledge gained from a different task that has already been learned. Usually it works best when the tasks are quite similar. However, it has been found that this technique works well even though the tasks are completely different. There are many pre-trained architectures that are trained on huge datasets such as AlexNet[8], VGG-16[15], VGG-19[15], Inception-V3[27], ResNet-50[22] and ResNet-101[22]. Probably the most popular such dataset is ImageNet. It contains millions of data sample to classify 1000 different categories. Transfer learning technique enables the researchers to train models with minimal training data by fetching architecture and weights from some popular pre-trained model. Furthermore, it drastically reduces the computational cost and therefore the training time.

ResNet-50 & ResNet-101. In order to construct the proposed model, the ResNet-50 and the ResNet-101 architectures are adopted by the faster R-CNN framework separately, i.e. the proposed model utilizes both the architectures separately for teeth recognition task. He et al.[22] presented a framework based on residual learning to overcome the difficulty of training deeper neural network. The degradation problem was addressed by introducing a deep residual learning framework. It showed that optimizing residual mapping is easier than to optimize the original and therefore, gained accuracy easily from network with greater depth. Residual network won the 1st place in classification task of ILSVRC 2015 competition[22]. ResNet-50 consists of four stages with total 50 layers and hence its name. ResNet-101 is the deeper version of ResNet-50, consisting of additional 17 blocks (3-layer block) in the third stage that made it total 101 layers. The architecture of ResNet-50 and ResNet-101 is summarized in Table 1.

Activation layer selection. In order to use a pre-trained model for a completely different task, few pre-processing steps should be considered. The steps include the removal of the original classifier, add a new classifier according to the task and fine tune the model[28]. There are three strategies to fine tune the model.

(i) The first strategy is to train the entire model, i.e. use only the architecture of the pre-trained model and train the model according to the available dataset. In short, training from the scratch. In order to achieve sufficient accuracy, large dataset is required for this strategy. It also involves huge computational cost.

(ii) The second option is to train some layers of the model, while leaving other frozen. In general, lower layers keep information about general features, whereas the higher layers keep information about specific features. As general features are problem independent, lower layers can be left frozen in case of small dataset. The training then only be done in the higher layers (problem dependent). However, when large dataset is available, overfitting does not become an issue and lower layers can also be train with the higher layers.

(iii) Third option is to freeze all the convolutional layers and thus, use only the classifier. This option can should only be considered where dataset is small and sufficient computational power is unavailable.

Based on the above strategies, activation layer 40 (activation_40_relu) is used as a feature extraction layer for ResNet-50 and activation layer res4b22 (res4b22_relu) for ResNet-101.

Candidate optimization. The residual network based faster R-CNN together with careful selection of parameters can provide very good recognition performance. However, there may still be a good number of false positives including double detections for a single tooth. In order to cope up with this problem, a candidate optimization algorithm based on prior knowledge is proposed in this research. This model selects the best combination of candidates in order to filter out the false positives and thus improving overall efficiency of the model.

Assume that tooth x ($1 \leq x \leq 32$) has $N(x)$ candidates detected by faster R-CNN. And, in some case, all candidates are false positives. Therefore, the selection is to find the best combination of candidates in $\prod_{x=1}^{32}(N(x)+1)$ combinations. The selection is done by optimizing Eq. (1). In this equation, the first term evaluates the confidence score, and the second term evaluates the relational position from other teeth.

$$f(\boldsymbol{P}) = \frac{1}{32}\sum_{x=1}^{32}\omega_c\mu_c(\boldsymbol{p_x}) + \omega_p\mu_p(\boldsymbol{p_x}) \tag{1}$$

where, $\boldsymbol{P} = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \ldots, \boldsymbol{p_{32}}\}$ is the combination pattern (T1 to T32), ω_c and ω_p are weights of confidence score and coordinate score, respectively. $\mu_c(\boldsymbol{p_x})$ is confidence score of candidate $\boldsymbol{p_x}$ obtained from faster R-CNN in the range of [0–1], whereas $\mu_p(\boldsymbol{p_x})$ is the positional relationship score of candidate $\boldsymbol{p_x}$ calculated by using the following equation:

$$\mu_p(\boldsymbol{p_x}) = \frac{1}{|\Omega|}\sum_{\boldsymbol{p_y}\in\Omega}\delta(\boldsymbol{p_x}, \boldsymbol{p_y}) \tag{2}$$

$$\Omega = \{\boldsymbol{p_{x+i}}|i = \{-2, -1, 1, 2\}, 1 \leq x+i \leq 32\}$$

where, $\boldsymbol{p_x}$ is the tooth candidate under consideration and $\delta(\boldsymbol{p_x}, \boldsymbol{p_y})$ is a customized function created using the prior knowledge of the dataset. The function evaluates the horizontal distance between the tooth $\boldsymbol{p_x}$ with its neighboring teeth $\boldsymbol{p_y}$ [where, $\boldsymbol{p_y} = \boldsymbol{p_{x-2}}, \boldsymbol{p_{x-1}}, \boldsymbol{p_{x+1}}, \boldsymbol{p_{x+2}}$]. It then calculates and assigns score for each tooth defined as coordinate score following the equation below:

$$\delta(\boldsymbol{p_x}, \boldsymbol{p_y}) = \begin{cases} 0 & \left|p_x^x - p_y^x\right| < a \\ k + (1-k)\frac{\left|p_x^x - p_y^x\right| - a}{b-a} & a < \left|p_x^x - p_y^x\right| < b \\ 1 & b < \left|p_x^x - p_y^x\right| < c \\ 1 - (1-k)\frac{\left|p_x^x - p_y^x\right| - c}{d-c} & c < \left|p_x^x - p_y^x\right| < d \\ 0 & d < \left|p_x^x - p_y^x\right| \\ k & p_x = \emptyset \text{ or } p_y = \emptyset \end{cases} \tag{3}$$

p_x^x and p_y^x are x-coordinate values of center point of candidate $\boldsymbol{p_x}$ and $\boldsymbol{p_y}$, respectively. The parameters, a, b, c and d determine the shape of the function. The values are determined experimentally and set as $(a, b, c, d) = (30, 47, 114, 130)$ when $\boldsymbol{p_y} = \boldsymbol{p_{x-1}}$ or $\boldsymbol{p_{x+1}}$, and $(a, b, c, d) = (80, 97, 164, 180)$ when $\boldsymbol{p_y} = \boldsymbol{p_{x-2}}$ or $\boldsymbol{p_{x+2}}$. The value of k is also determined experimentally and set as 0.35 for this experiment. Figure 3 shows the mechanism of calculating the coordinate score. The different colors refer to the different teeth number. The weights of confidence score and coordinate score are two of the parameters of this algorithm and they should be selected carefully. There may have multiple candidates for a single tooth that the optimization algorithm should fix. In that case, the candidates are numbered in accordance with its confidence value, i.e. candidate tooth with higher confidence value will be numbered first. For example, tooth T1 has two candidates with confidence value 0.95 and 0.7, they will be denoted as $C_{1,1}$ and $C_{1,2}$, respectively. The first subscript refers to the tooth number. And, $C_{x,0}$ represents the missing tooth candidate. The candidate optimization processes in three steps.

Step 1: Initialize pattern $\boldsymbol{P_{max}} = \{C_{1,1}, C_{2,1}, \ldots, C_{32,1}\}$ by choosing the candidate with the highest confidence value for each tooth. When there are no candidate at tooth x, $C_{x,0}$ is used as the candidate. The score of pattern $\boldsymbol{P_{max}}$ is calculated using Eq. (8), and let the calculated score be S_{max}.
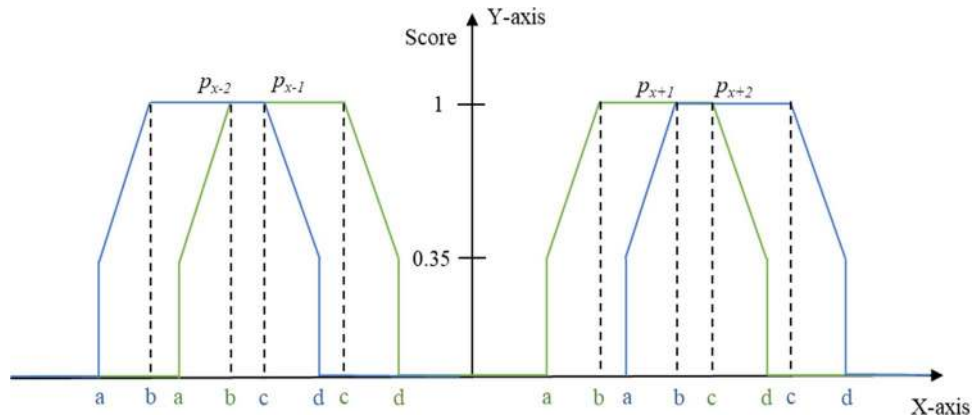
5

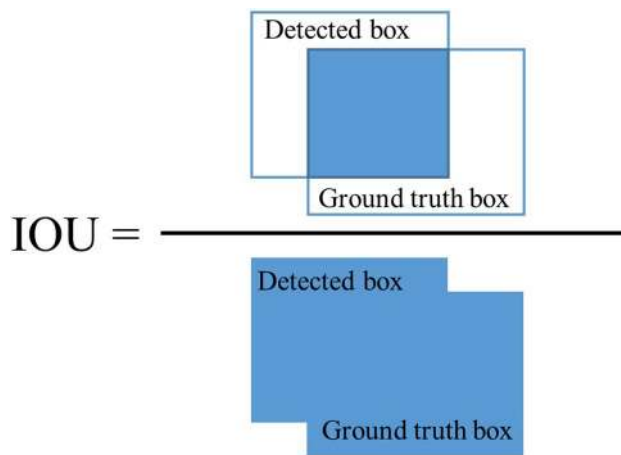**Figure 3.** Mechanism of calculating the coordinate score (μ_p).

Figure 4. Illustration of intersection-over-union (IOU).

Step 2: For every tooth x, try all candidates of C_x, where $C_x = (C_{x,0}, C_{x,1}, C_{x,2} \ldots)$ and calculate the score using Eq. (8). Update S_{max}, if new best combination is found, and set the candidate to P_{tmp}.

Step 3: If there are update for S_{max}, P_{max} is replaced by P_{tmp} and return to *Step 2*. Else, the algorithm finds the best combination of all.

Performance analysis. Average precision (AP)[29] is calculated for each category of tooth to evaluate the candidate detection performance of the proposed method. At first, the detected boxes are compared with the ground truth boxes by calculating the intersection-over-union (IOU) as shown in Fig. 4 and defined as below

$$IOU = \frac{Area_{DetectedBox} \cap Area_{GroundTruthBox}}{Area_{DetectedBox} \cup Area_{GroundTruthBox}} \tag{4}$$

The IOU threshold value is set as 0.5 i.e. if the IOU value is greater or equal to 0.5 then the detected box is considered as true positive, otherwise is considered as false positive. To calculate the evaluation index, i.e. AP, precision and recall are calculated using the equations as follows.

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

where, *TP* is defined as the number of ground truth boxes that overlap with the detected boxes with $IOU \geq 0.5$; *FP* is defined as the number of detected boxes that overlap with the ground truth boxes with $IOU < 0.5$, and *FN* is defined as the number of teeth that are not detected or detected with $IOU < 0.5$. Finally, the model is tested with a test dataset of 100 images. The above mentioned metrics are used to evaluate the detected boxes.
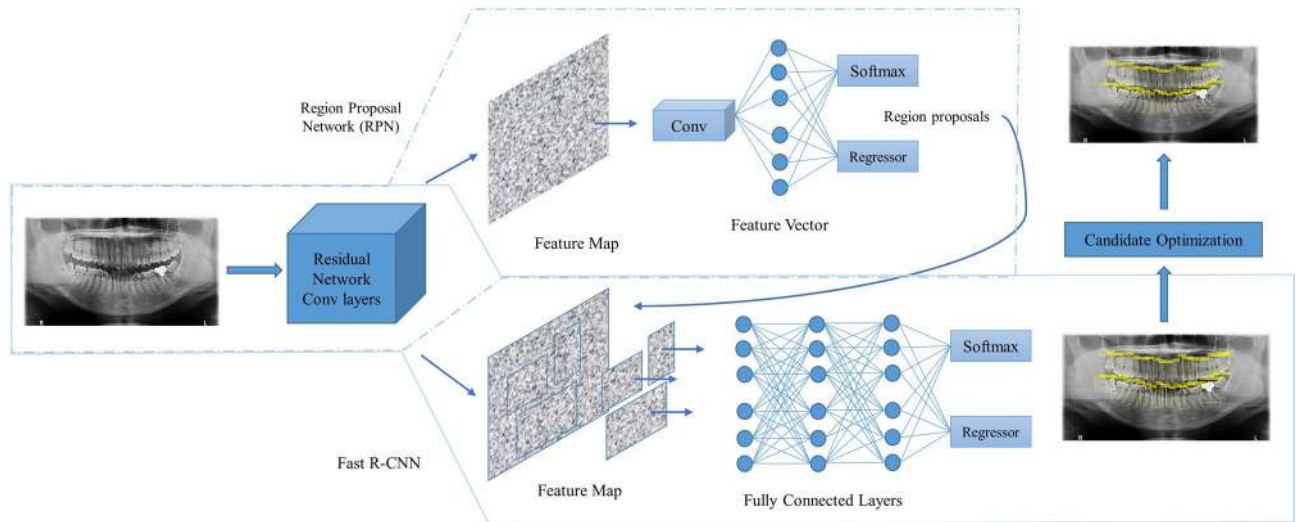
6

Figure 5. Illustration of proposed teeth recognition model.

Parameter	Value
Epoch	10
Iteration	9000
Initial learning rate	0.001
Mini batch size	1
Number of regions to sample	256
Number of strongest regions	2000
Negative overlap range	[0–0.3]
Positive overlap range	[0.6–1]

Table 2. Parameter settings of faster R-CNN.

The overall proposed model with candidate optimization algorithm is depicted in Fig. 5. To evaluate the performance of candidate optimization algorithm, F_1 score is calculated. F_1 score is the harmonic mean of precision and recall and is defined by the following equation:

$$F_1 score = \frac{2 \times (precision \times recall)}{(precision + recall)} \tag{7}$$

Results and discussion

This section presents the overall simulation results using residual network based faster R-CNN for teeth recognition. The proposed method was implemented in MATLAB 2019a software and executed with Ryzen 7 2700 Eight-Core Processors (16 CPUs) with clock speed ~ 3.2 GHz. The training and testing were done with TITAN RTX 24 GB display memory (VRAM). Table 2 refers to the parameter settings of faster R-CNN for teeth recognition task. Total number of epoch was set as 10. As each training image corresponds to each iteration, thus total number of iteration was 9000. Number of regions to sample from each training image was set as 256, whereas number of strongest regions to be used for generating training samples was set as 2000. Negative overlap range and positive overlap range were set as [0–0.3] and [0.6–1], respectively. To better explain the results of the proposed model, total three cases were considered.

Test case 1: Total 900 panoramic radiographs were used to train the network, while 100 images were used for testing. Both, ResNet-50 and ResNet-101 networks were implemented separately as the base networks of faster R-CNN for the evaluation purpose. Figure 6 is presenting the comparison of candidate detection results between ResNet-50 and ResNet-101 for each tooth category. It can be seen from the Fig. 7 that only T1 achieves less than 0.900 AP while using ResNet-50. Other than that the AP of other teeth is above 0.900 and the mAP is 0.974. On the other hand, results obtained using ResNet-101 shows extremely good detection performance as total seven teeth categories achieve maximum average precision and mAP is 0.981, which is better than the results obtained by ResNet-50. Table 3 presents precision and recall value for each tooth category. Some of the teeth categories achieved perfect recall values for ResNet-101 based faster R-CNN i.e. there were no false negative for those teeth categories. Figure 7 shows recall-precision curves for different teeth categories using ResNet-101. For better visualization, the curves are shown in four different figures, and each figure shows recall-precision curves for eight
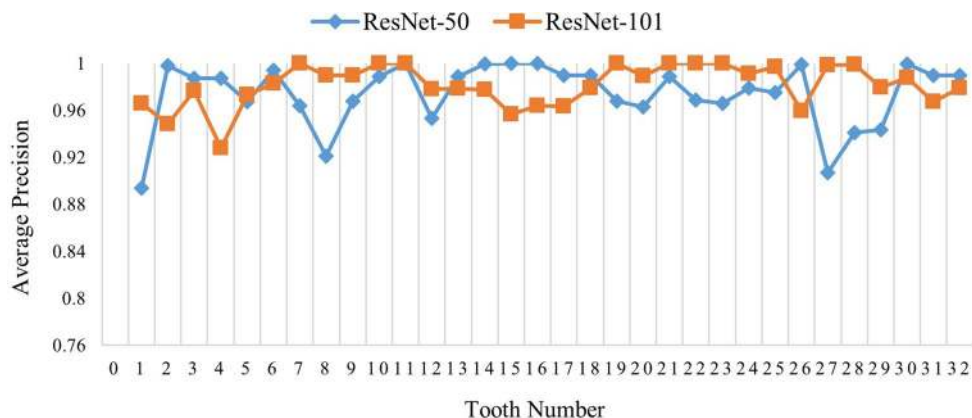
Figure 6. Average precision for different tooth category.

Tooth number	ResNet-50		ResNet-101		Tooth number	ResNet-50		ResNet-101	
	Precision	Recall	Precision	Recall		Precision	Recall	Precision	Recall
T1	0.8876	0.9186	0.9231	0.9796	T17	0.9474	0.9643	0.9565	0.9706
T2	0.8585	0.9529	0.9400	0.9592	T18	0.9381	0.9529	0.9798	0.9798
T3	0.9272	0.9745	0.9412	0.9796	T19	0.9458	0.9846	1.0000	1.0000
T4	0.9785	0.9333	0.9479	0.9286	T20	0.9234	0.9847	0.9899	0.9899
T5	0.8788	0.9667	0.8788	0.9775	T21	0.9641	0.9947	0.9495	1.0000
T6	0.9598	0.9695	0.9700	0.9898	T22	0.9802	1.0000	1.0000	1.0000
T7	0.9469	0.9899	1.0000	1.0000	T23	0.9369	0.9747	0.9900	1.0000
T8	0.9327	0.9848	0.9706	0.9900	T24	0.8977	0.9650	0.9174	1.0000
T9	0.9557	0.9848	0.9703	0.9899	T25	0.8122	0.9300	0.9901	1.0000
T10	0.9423	0.9849	0.9802	1.0000	T26	0.9234	0.9650	0.9600	0.9600
T11	0.9897	0.9747	0.9802	1.0000	T27	0.9557	0.9848	0.9604	1.0000
T12	0.9267	0.9779	0.9375	0.9783	T28	0.9444	0.9791	0.9505	1.0000
T13	0.9439	0.9536	0.9604	0.9798	T29	0.9200	0.9583	0.9798	0.9798
T14	0.8720	0.9787	0.9314	0.9794	T30	0.9444	0.9791	0.9697	0.9897
T15	0.8488	0.9305	0.8727	0.9697	T31	0.9109	0.9684	0.9691	0.9691
T16	0.9036	0.9036	0.8393	0.9792	T32	0.9211	0.9722	0.9688	0.9841

Table 3. Recall and precision value for each tooth category.

categories of teeth. Almost all of the curves show ideal behavior and visibly it can be seen that the convergence performance is close to that of an ideal one. The recognition performance obtained by the proposed model is close to the level of an expert dentist.

Test case 2: Second test case examines the compatibility and the feasibility of using candidate optimization algorithm along with faster R-CNN technique to improve overall recognition result. The candidate optimization algorithm based on prior knowledge was implemented along with faster R-CNN technique to refine the obtained detected boxes by faster R-CNN. In this paper, two sets of weights were considered. In set A, weight of confidence score (ω_c) was set as 0.8, whereas the weight of positional relationship score (ω_p) was set as 0.2. In set B, both weights were selected as 0.5. Table 4 shows the recognition results after applying the candidate optimization algorithm. From the table, it is clear that for both the networks, the overall F_1 scores improves. For ResNet-50 and ResNet-101, F_1 score improves from 0.965 to maximum 0.976 and 0.978 to 0.983, respectively. The optimization technique also effectively balances the difference between precision and recall, which indicates that the algorithm is fully compatible with the model and successfully improves its robustness. Figure 8 visualizes the result given in Table 3. In terms of F_1 score, set A performed better than set B for ResNet-50, whereas it remained in balance for ResNet-101.

Test case 3: In order to check the robustness of the proposed method, K-fold cross validation (CV) technique was utilized in this experiment. In this paper, K is equal to 10 i.e. the whole dataset is divided into 10 different folds. Total 10 runs were required to perform tenfold CV. Both, ResNet-50 and ResNet-101 were used with faster R-CNN, separately to perform the recognition task. The overall results consisting of AP of all teeth categories in all 10 test datasets for ResNet-50 and ResNet-101 are summarized in Fig. 9. The obtained results of 10-CV presented in Fig. 9 shows that the residual network based faster R-CNN performed quite strongly and consistently with the average of mAP is 0.958 for ResNet-50 and 0.960 for ResNet-101. Furthermore, the robustness
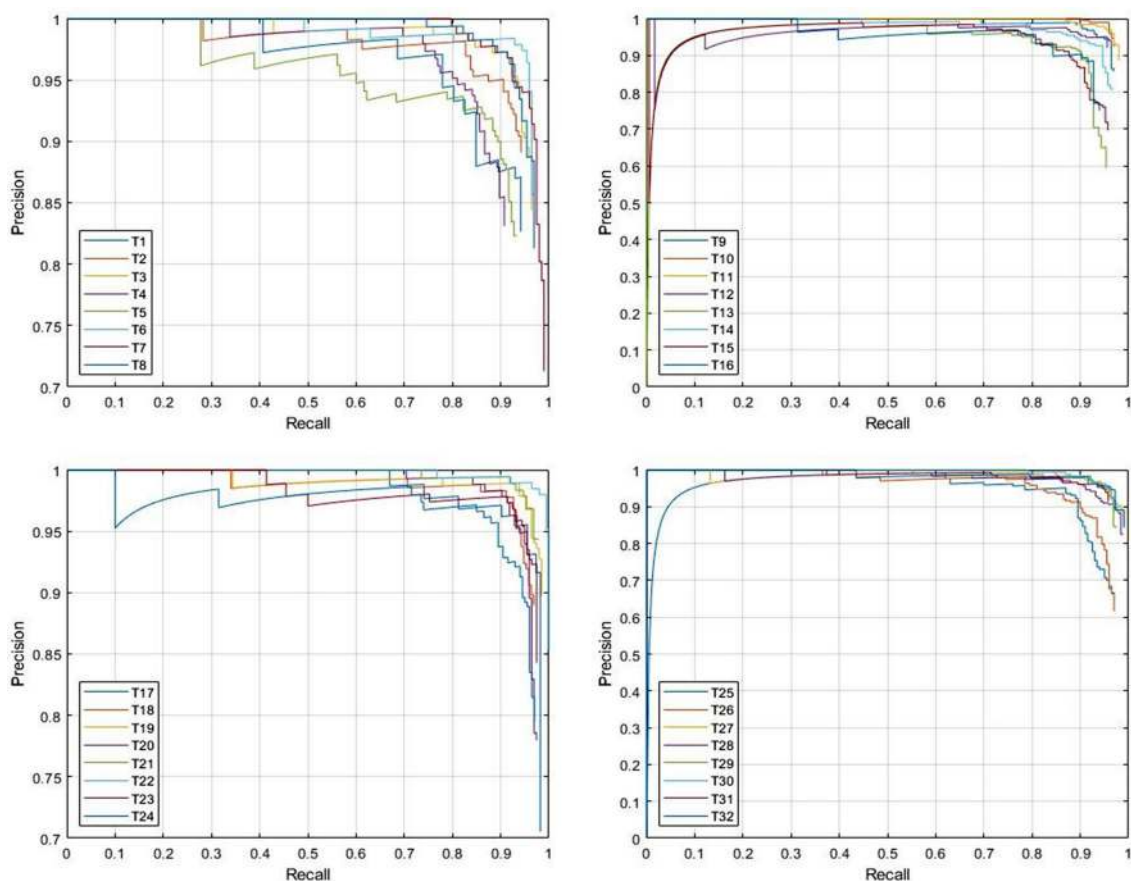
Figure 7. Recall-precision curve of different teeth categories for ResNet-101. The curves were generated by MATLAB 2019a software.

	ResNet-50			ResNet-101		
	Original	$\omega_c = 0.8$ $\omega_p = 0.2$	$\omega_c = 0.5$ $\omega_p = 0.5$	Original	$\omega_c = 0.8$ $\omega_p = 0.2$	$\omega_c = 0.5$ $\omega_p = 0.5$
Precision	0.942	**0.975**	0.971	0.964	**0.988**	0.977
Recall	**0.990**	0.978	0.980	**0.993**	0.978	0.989
F_1 Score	0.965	**0.976**	0.975	0.978	**0.983**	**0.983**

Table 4. Recognition results after applying candidate optimization algorithm. Bold values indicate the best results in that particular row (particular section).

of the method in different test data shows that it is clinically applicable. The comparison of different residual networks presented in Fig. 9 shows ResNet-101 performs better for 7 folds, whereas ResNet-50 performs better in 3 other folds. The lowest AP achieved by ResNet-50 is 0.800 for T27 in K7 fold and 0.840 for T1 in K7 fold by ResNet-101. However, most of the lower detections came from the eighth number fold (i.e. K8) for both of the residual networks. To assess the feasibility and robustness of the proposed model after applying candidate optimization algorithm, tenfold CV technique was performed and the results are presented in Table 5. After applying candidate optimization, the average F_1 score improves from 0.962 to 0.971 for ResNet-50 and 0.975 to 0.976 for ResNet-101, respectively. Furthermore, for all cases, the candidate optimization algorithm refined the detected boxes successfully and thus, improved the overall recognition performance.

The stand-alone residual network based faster R-CNN performed exceedingly well in recognizing tooth by its number. Two kinds of residual networks i.e. ResNet-50 and ResNet-101 were used as base networks of faster R-CNN. ResNet-101 is deeper network than ResNet-50 and it performs marginally better than its shallower counterpart. Although, ResNet-101 performs better in terms of recognition, it is computationally costlier than ResNet-50. As their recognition performance is not much different, authors recommend using ResNet-50 as a base network when computational cost is a concern. The inclusion of candidate optimization algorithm further improves the recognition performance of the proposed model. However, the optimization parameters should be chosen carefully based on the dataset in order to have a good impact on the overall recognition results. Figure 10
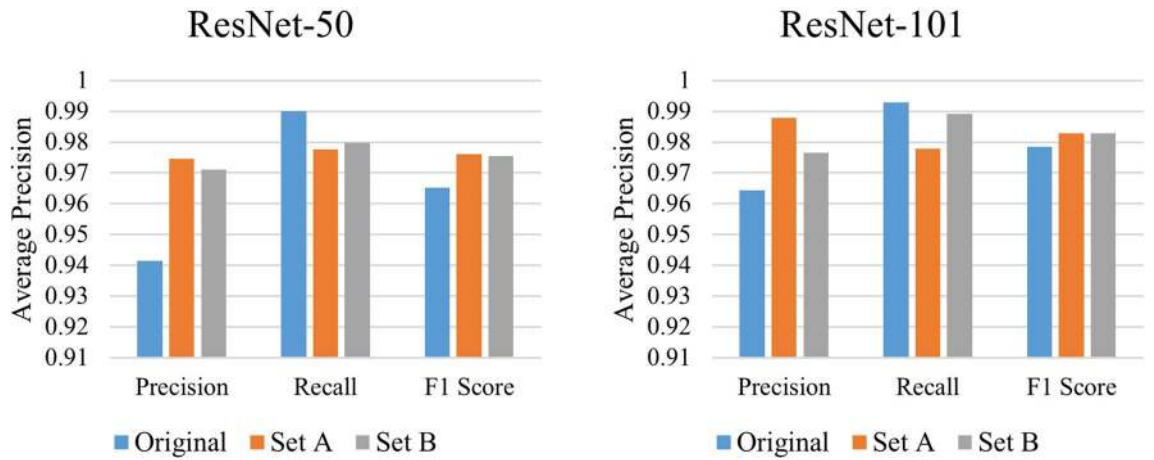
Figure 8. Comparison of results before and after applying candidate optimization algorithm (Set A: $\omega_c = 0.8$, $\omega_p = 0.2$; Set B: $\omega_c = 0.5$, $\omega_p = 0.5$).
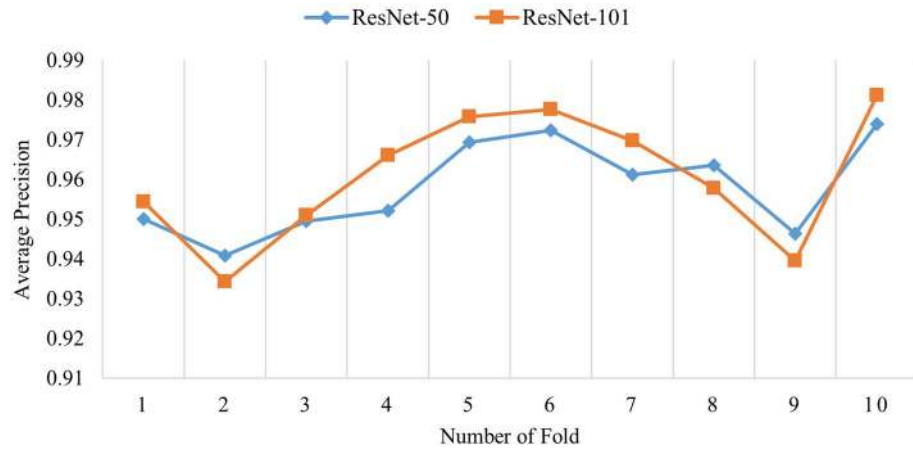

Figure 9. Mean average precision (mAP) for different folds in tenfold cross validation.

		ResNet-50			ResNet-101		
	Number of teeth in test dataset	Original	$\omega_c = 0.8$ $\omega_p = 0.2$	$\omega_c = 0.5$ $\omega_p = 0.5$	Original	$\omega_c = 0.8$ $\omega_p = 0.2$	$\omega_c = 0.5$ $\omega_p = 0.5$
K1	2947	0.957	0.969	**0.971**	0.975	0.976	**0.977**
K2	2890	0.962	0.967	**0.967**	0.967	0.969	**0.970**
K3	2904	0.961	0.969	**0.969**	0.970	0.973	**0.974**
K4	2909	0.962	**0.969**	**0.969**	0.967	**0.972**	0.972
K5	2958	0.974	**0.980**	0.979	0.982	**0.984**	0.984
K6	2847	0.958	**0.969**	0.969	0.971	**0.975**	0.974
K7	2862	0.974	0.983	**0.984**	0.984	0.987	**0.987**
K8	2796	0.956	0.961	**0.963**	0.964	**0.967**	0.967
K9	2846	0.951	0.962	**0.964**	0.962	0.968	**0.970**
K10	2969	0.965	**0.976**	0.975	0.978	**0.983**	0.983
Average	2893	0.962	0.971	**0.971**	0.972	0.975	**0.976**

Table 5. Results (in F_1 score) of teeth recognition after applying candidate optimization algorithm for tenfold cross validation. Bold values indicate the best results in that particular row (particular section).
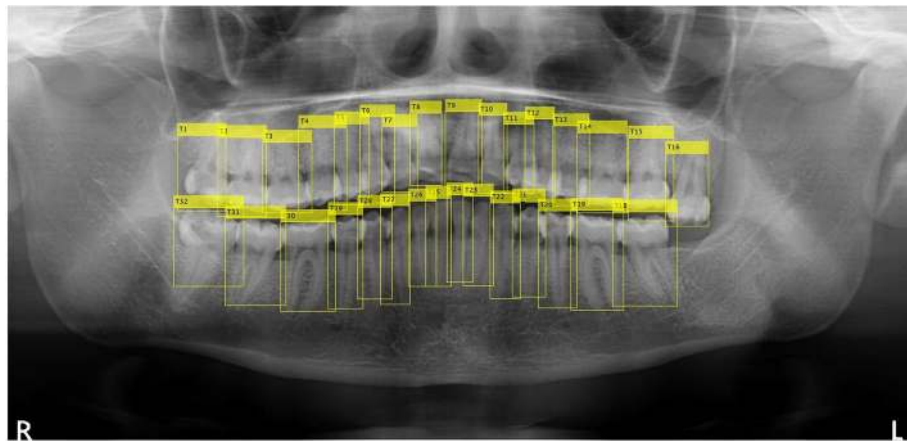
Figure 10. Successful teeth detection in noisy panoramic radiograph. The detected boxes were generated by MATLAB 2019a software.

shows detected teeth in a noisy panoramic radiograph. This research excludes severely broken teeth from the experiment.

Conclusion

This research proposes a method for automatic teeth recognition in dental panoramic radiographs. The method is based on candidate detection with residual network based faster R-CNN and candidate optimization using a prior knowledge. Two versions of residual network i.e. ResNet-50 and ResNet-101 are used as base networks for faster R-CNN separately. The combination of residual network with faster R-CNN method successfully performs the recognition task with a high degree of accuracy. It achieves maximum 0.980 mAP. A prior knowledge based candidate optimization technique is also incorporated to improve the overall recognition performance. The introduction of the optimization method improves the F_1 score from 0.965 to maximum 0.976 and 0.978 to 0.983 for ResNet-50 and ResNet-101, respectively. The K-fold cross validation technique is also implemented with and without candidate optimization technique that effectively verifies the feasibility and the robustness of the proposed method. The level of performance achieved by the proposed model is close to an expert dentist and thus, clinically implementable. Finally, it can be said that the proposed model can be used as a reliable and useful tool to assist dental care professionals in dentistry. In future, we plan to extend the current model to include automatic dental condition evaluation and prosthetic detection features.

References

1. Hallikainen, D. History of panoramic radiography. *Acta Radiol.* **37**, 441–445. https://doi.org/10.1177/02841851960373p207 (1996).
2. Nomir, O. & Abdel-Mottaleb, M. A system for human identification from X-ray dental radiographs. *Pattern Recogn.* **38**, 1295–1305. https://doi.org/10.1016/j.patcog.2004.12.010 (2005).
3. Nassar, D. E. M. & Ammar, H. H. A neural network system for matching dental radiographs. *Pattern Recogn.* **40**, 65–79. https://doi.org/10.1016/j.patcog.2006.04.046 (2007).
4. Li, S., Fevens, T., Krzyżak, A., Jin, C. & Li, S. Semi-automatic computer aided lesion detection in dental X-rays using variational level set. *Pattern Recogn.* **40**, 2861–2873. https://doi.org/10.1016/j.patcog.2007.01.012 (2007).
5. Lin, P. L., Huang, P. Y., Huang, P. W., Hsu, H. C. & Chen, C. C. Teeth segmentation of dental periapical radiographs based on local singularity analysis. *Comput. Methods Programs Biomed.* **113**, 433–445. https://doi.org/10.1016/j.cmpb.2013.10.015 (2014).
6. Kavitha, M. S., Asano, A., Taguchi, A., Kurita, T. & Sanada, M. Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC Med. Imaging* **12**, 1–11. https://doi.org/10.1186/1471-2342-12-1 (2012).
7. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436. https://doi.org/10.1038/nature14539 (2015).
8. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90. https://doi.org/10.1145/3065386 (2017).
9. Hinton, G. *et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **29**, 82–97. https://doi.org/10.1109/MSP.2012.2205597 (2012).
10. Ciodaro, T., Deva, D., de Seixas, J. M. & Damazio, D. Online particle detection with neural networks based on topological calorimetry information. *J. Phys.: Conf. Ser.* **368**, 012030. https://doi.org/10.1088/1742-6596/368/1/012030 (2012).
11. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **55**, 263–274. https://doi.org/10.1021/ci500747n (2015).
12. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168. https://doi.org/10.1038/nature12346 (2013).
13. Collobert, R. *et al.* Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **12**, 2493–2537 (2011).
14. Prajapati, S. A., Nagaraj, R. & Mitra, S. In *5th International Symposium on Computational and Business Intelligence (ISCBI)* 70–74.
15. Simonyan, K. & Zisserman, A. In *International Conference on Learning Representations* (2014).
16. Lin, N.-H. *et al.* Teeth detection algorithm and teeth condition classification based on convolutional neural networks for dental panoramic radiographs. *J. Med. Imaging Health Inform.* **8**, 507–515. https://doi.org/10.1166/jmihi.2018.2354 (2018).

11

17. Chen, H. *et al.* A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci. Rep.* **9**, 3840. https://doi.org/10.1038/s41598-019-40414-y (2019).
18. Tuzoff, D. V. *et al.* Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dento Maxillo Fac. Radiol.* **48**, 20180051. https://doi.org/10.1259/dmfr.20180051 (2019).
19. Muramatsu, C. *et al.* Tooth detection and classification on panoramic radiographs for automatic dental chart filing: Improved classification by multi-sized input data. *Oral Radiol.* https://doi.org/10.1007/s11282-019-00418-w (2020).
20. Girshick, R., Donahue, J., Darrell, T. & Malik, J. in *2014 IEEE Conference on Computer Vision and Pattern Recognition.* 580–587.
21. Girshick, R. in *2015 IEEE International Conference on Computer Vision (ICCV).* 1440–1448.
22. He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 770–778.
23. Keiser-Nielsen, S. Federation Dentaire Internationale. Two-Digit System of designating teeth. *DP. Dental practice* **3**, 6 passim (1971).
24. Palmer, C. Palmer's dental notation. *Dent. Cosmos.* **33**, 194–198 (1891).
25. Liu, W., Liao, S. & Hu, W. Perceiving motion from dynamic memory for vehicle detection in surveillance videos. *IEEE Trans. Circuits Syst. Video Technol.* **29**, 3558–3567. https://doi.org/10.1109/TCSVT.2019.2906195 (2019).
26. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149. https://doi.org/10.1109/tpami.2016.2577031 (2017).
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2818–2826.
28. Marcelino, P. *Transfer Learning from Pre-Trained Models,* <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751> (2018).
29. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **88**, 303–338. https://doi.org/10.1007/s11263-009-0275-4 (2010).

Acknowledgements

Author contributions

F.P.M. conducted the full experiment involving the deep learning method, compiled the results and wrote the full manuscript. K.M. designed the optimization part and execute the experiment involving the optimization technique. S.K. supervised the whole experiment with important instructions and advices. All authors reviewed the manuscript.

Competing interests

This research was sponsored by Narcohm Co. Ltd. and may lead to the development of products which may be licensed to University of Hyogo, in which the authors are affiliated to and have financial interest. The authors have disclosed those interests fully to Nature, and have in place an approved plan for managing any potential conflicts arising from this arrangement.

Additional information

Correspondence and requests for materials should be addressed to F.P.M.

Reprints and permissions information is available at www.nature.com/reprints.

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