RESEARCH ARTICLE

Tooth detection and numbering in panoramic radiographs using convolutional neural networks

[1]Dmitry V. Tuzoff, [2]Lyudmila N. Tuzova, [3]Michael M. Bornstein, [4]Alexey S. Krasnov,
[2]Max A. Kharchenko, [1]Sergey I. Nikolenko, [2]Mikhail M. Sveshnikov and [5]Georgiy B. Bednenko

[1]*Steklov Institute of Mathematics, Saint Petersburg, Russia;* [2]*Denti. AI Technology Inc., Toronto, Canada;* [3]*Applied Oral Sciences, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China;* [4]*Dmitry Rogachev National Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia;* [5]*Reutov Stomatological Clinic, Reutov, Russia*

Objectives: Analysis of dental radiographs is an important part of the diagnostic process in daily clinical practice. Interpretation by an expert includes teeth detection and numbering. In this project, a novel solution based on convolutional neural networks (CNNs) is proposed that performs this task automatically for panoramic radiographs.

Methods: A data set of 1352 randomly chosen panoramic radiographs of adults was used to train the system. The CNN-based architectures for both teeth detection and numbering tasks were analyzed. The teeth detection module processes the radiograph to define the boundaries of each tooth. It is based on the state-of-the-art Faster R-CNN architecture. The teeth numbering module classifies detected teeth images according to the FDI notation. It utilizes the classical VGG-16 CNN together with the heuristic algorithm to improve results according to the rules for spatial arrangement of teeth. A separate testing set of 222 images was used to evaluate the performance of the system and to compare it to the expert level.

Results: For the teeth detection task, the system achieves the following performance metrics: a sensitivity of 0.9941 and a precision of 0.9945. For teeth numbering, its sensitivity is 0.9800 and specificity is 0.9994. Experts detect teeth with a sensitivity of 0.9980 and a precision of 0.9998. Their sensitivity for tooth numbering is 0.9893 and specificity is 0.9997. The detailed error analysis showed that the developed software system makes errors caused by similar factors as those for experts.

Conclusions: The performance of the proposed computer-aided diagnosis solution is comparable to the level of experts. Based on these findings, the method has the potential for practical application and further evaluation for automated dental radiograph analysis. Computer-aided teeth detection and numbering simplifies the process of filling out digital dental charts. Automation could help to save time and improve the completeness of electronic dental records.

Dentomaxillofacial Radiology (2019) **48**, 20180051. doi: 10.1259/dmfr.20180051

Introduction

In the last two decades, computer-aided diagnosis (CAD) has developed significantly due to the growing accessibility of digital medical data, rising computational power, and progress in artificial intelligence. CAD systems assisting physicians and radiologists in decision-making have been applied to various medical problems, such as breast and colon cancer detection, classification of lung diseases, and localization of brain lesions.[1,2] The increasing popularity of digital radiography stimulates

further research in the area. In dentistry, processing of radiographic images became an important subject of automation too, as radiographic image interpretation is an essential part of the diagnosis, monitoring of dental health, and treatment planning.

Although CAD systems have been used in the clinical environment for decades, in most cases they do not aim to substitute the medical experts but rather to assist them.[1] Automated solutions might help dentists in clinical decision-making, save time, and reduce the negative effects of stress and fatigue in daily practice.[1] In this paper, the problem of teeth detection and numbering in dental radiographs according to the FDI two-digit notation was studied.[3] An algorithmic solution may be used to automatically fill in digital patients' records for dental history taking and treatment planning. It may also serve as a pre-processing step for further pathology detection.

During the last decade, a number of studies addressed this problem. For teeth detection, Lin et. al[4] and Hosntalab et al[5] proposed pixel-level segmentation methods based on traditional computer vision techniques, such as thresholding, histogram-based, and level set methods. They detected teeth with the recall (sensitivity) of 0.94 and 0.88 respectively. Miki et al[6] used a manual approach to place bounding boxes enclosing each tooth on CT images.

For teeth numbering, Lin et. al[4] and Hosntalab et al[5] methods consisted of two stages: feature extraction and classification. To extract features from segmented teeth, Lin et. al[4] utilized the parameters such as width/height teeth ratio and crown size, when Hosntalab et al[5] used wavelet-Fourier descriptor to represent the shape of the teeth. To classify teeth, support vector machines (SVM), sequence alignment algorithm[4] and feedforward neural networks (NNs)[5] were used. These papers reported the classification accuracy results of 0.98[4] and above 0.94.[5] Miki et al[6] presented a convolutional neural network (CNN) model based on the AlexNet network[7] to classify manually isolated teeth on CT achieving a classification accuracy of 0.89.

In the present study, CNNs for both detection and numbering of teeth are applied. CNNs are a standard class of architectures for deep feedforward neural networks, and they are typically applied for image recognition tasks. CNNs were first introduced more than two decades ago,[8] but in 2012, when the AlexNet architecture significantly outperformed other teams on ImageNet Large Scale Visual Recognition Competition challenge,[7] the deep learning revolution came to computer vision, and CNNs have enjoyed rapid development ever since. Currently, CNNs are used in numerous applications and represent a state-of-the-art approach for various computer vision tasks.[9,10]

In dentistry, the application of CNNs has been studied for cephalometric landmark detection,[11,12] teeth structures segmentation,[13] and teeth classification.[6] These works demonstrated promising results, but it is still an underdeveloped area of research. In particular,

CNN architectures for object detection have not been used yet in dentistry, while this approach was successfully used in other types of medical applications, *e.g.* detection of colitis on abdominal CT[14] or brain lesions detection on brain MRI.[2]

This study is focused on the analysis of panoramic radiographs that depict the upper and lower jaws in one single image to apply CNNs. A panoramic radiograph is one of the most common dental radiographic examinations as it allows to screen a broad anatomical region and at the same time requires a relatively low radiation dose. Panoramic radiographs have not been discussed for teeth detection and numbering before, but similar approaches have been studied for other types of radiographs such as medical computed tomography (CT)[5,6] and bitewings.[4] Furthermore, CNN architectures for both teeth detection and numbering have not been described in the available literature. The current study aims to verify the hypothesis that CNN-based models can be trained to detect and number teeth on panoramic radiographs.

Methods and Materials

Deep learning and CNNs
The proposed solution is based on deep learning techniques. Deep learning is a class of learnable artificial intelligence (AI) algorithms that allows a computer program to automatically extract and learn important features of input data for further interpretation of previously unseen samples. The key distinction of deep learning methods is that they can learn from a raw data input, *e.g.* pixels of images, with no hand-crafted feature engineering required. Deep CNNs is one of the most popular sets of deep learning methods that is commonly applied for image recognition tasks. CNN architectures exploit specific characteristics of an image data input, such as spatial relationships between objects, to effectively represent and learn hierarchical features using multiple levels of abstraction; see a detailed overview of deep learning techniques, including CNNs, in LeCun et al.[10]

Basic workflow of a computer-aided diagnostic system
Panoramic radiographs are used as an input for the system presented here. The teeth detection module processes the radiograph to define the boundaries of each tooth. The system then crops the panoramic radiograph based on the predicted bounding boxes. The teeth numbering module classifies each cropped region according to the FDI notation,[3] combines all teeth, and applies the heuristics producing the final teeth numbers. The system outputs the bounding boxes coordinates and corresponding teeth numbers for all detected teeth on the image. The overall architecture and workflow are shown in Figure 1.
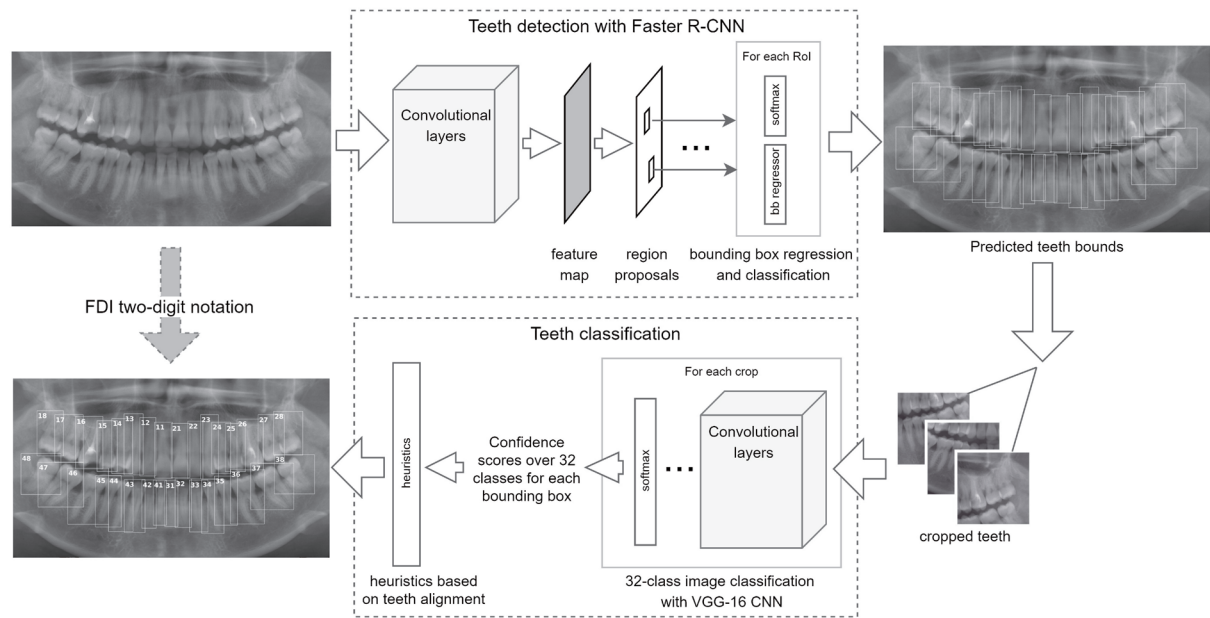
Figure 1 System architecture and pipeline: the system consists of two modules for teeth detection and teeth classification. The teeth detection module finds teeth on the original panoramic radiograph outputting the bounding boxes. The teeth classification module classifies each tooth to assign a number according to the dental notation and applies a heuristic method to ensure arrangement consistency among the detected teeth. CNNs, convolutional neural networks.

Radiographic data set

The data set included 1574 anonymized panoramic radiographs of adults randomly chosen from the X-ray images archive provided by the Reutov Stomatological Clinic in Russia from January 2016 to March 2017. No additional information such as gender, age, or time of image taking was used to for the database. The images were randomly distributed into:

- Training group: 1352 images
- Testing group: 222 images

The training group was used to train teeth detection and numbering models, and the testing group was used for evaluation of the performance of the software.

All panoramic radiographs were captured with the Sirona Orthophos XG-3 X-ray unit (Sirona Dental Systems GmbH, Bensheim, Germany). Five radiology experts of varying experience provided ground truth annotations for the images. The following method was used to collect annotations: experts were presented with high-resolution panoramic radiographs and asked to draw bounding boxes around all teeth and, at the same time, to provide a class label for each box with the tooth number (FDI system). The model was trained only to detect teeth with natural roots, excluding dental implants and fixed bridges.

For this study, the use of the radiographic material was exempt from an approval by an ethical committee or IRB according to the official decision of Steklov Institute of Mathematics in St. Petersburg, Russia due to the retrospective nature of the data collection, complete anonymization of the data used, and the subject of the study being not related to clinical interventions.

Teeth detection

Teeth detection method uses the state-of-the-art Faster R-CNN model.[15] Faster R-CNN was evolved from Fast R-CNN architecture, which was, in turn, based on R-CNN method (Region-based CNN). The challenging task of object detection is to define the regions of interest where the objects can be located. R-CNN proposed a combined solution for both the region of interest proposal generation and object localization. Fast R-CNN improved the performance of R-CNN by simplifying pipeline and optimizing computation. Finally, Faster R-CNN proposed even more advanced solution fully based on CNNs.

Faster R-CNN is a single unified network consisted of two modules: the regional proposal network (RPN) and object detector. RPN proposes regions where the objects of interest might be located. The object detector uses these proposals for further object localization and classification. Both modules share the convolution layers of the base CNN that provides a compact representation of the source image, known as a feature map. The features are learned during the training phase, which is a key difference compared to classical computer vision algorithms in which the features are engineered by hand.

To generate regional proposals, RPN slides the window over the feature map, and, at each window location, produces the potential bounding boxes named "anchors". For each anchor, the RPN estimates the

probability of an anchor to contain an object or a background, and tightens the bounding box with the special regressor. The top N-ranked region proposals are then used as an input for the object detection network. The object detector refines the class score of a region to be a tooth or a background for two-class detection task and generates the final bounding box coordinates.

The VGG-16 Net[16] was used as a base CNN for both RPN and object detection. VGG-16 is a 16-layer CNN architecture named by the research group that designed this network (Visual Geometry Group, Department of Engineering Science, University of Oxford). The hyperparameters that define the anchor properties were tuned to reflect the potential boundaries of teeth. These parameters include base anchor size, anchor scales, and anchor ratios. To minimize the false-positives rate of teeth detection, the Intersection-over Union threshold for non-maximum suppression algorithm used in the system and the prediction score threshold were also tuned.

During the training, model weights pretrained on the ImageNet data set were used for the basic CNN.[17] All layers of CNN were fine-tuned since the data set is large enough and differs significantly from ImageNet. The initial learning rate was chosen as 0.001 with further exponential decay. Teeth detection model was implemented as a customized version of the Faster R-CNN python implementation[18] with the TensorFlow backend.[19]

Teeth numbering

Teeth numbering method is based on the VGG-16 convolutional architecture.[16] The model was trained to predict the number of a tooth according to the FDI two-digit notation.

To classify the teeth by numbers, this module uses the output of the teeth detection module. It crops the teeth based on the predicted bounding boxes. Then VGG-16 CNN classifies each cropped image to predict a two-digit tooth number. The classifier outputs a set of confidence scores over all 32 classes for each bounding box estimating the probability of the tooth to be any of possible 32 teeth numbers. These data are then post-processed by a heuristic method to improve prediction results. Post-processing is based on the natural assumption that each tooth can occur at most once in the image in the specific order; the algorithm operates as follows.

Step 1. Sort predicted teeth bounding boxes by coordinates within each jaw.

Step 2. Count the number of missed teeth based on the known maximum teeth count.

Step 3. Iterate over all possible valid combinations of teeth and calculate the total confidence score.

Step 4. Choose the combination with the highest total confidence score.

As with teeth detection, model weights pretrained on the ImageNet data set were used to initialize the CNN. For training, cropped images were produced based on ground truth annotations of full panoramic X-rays, and the cropping method was tuned to include neighbouring structures, which allowed to improve the prediction quality of CNN because of additional context. The images were also augmented to increase the variety of data set. The batch size of 64 was used to train the CNN. Teeth numbering module is written in Python programming language using the Keras library[20] with TensorFlow backend.[19]

Performance analysis

The data set of the testing group of 222 images was used to evaluate the performance of the system, and to compare it to the expert. Each image was analyzed by the system and an experienced radiologist independently. The testing data set was not seen by the system during the training phase.

The annotations made by the system and the experts were compared to evaluate the performance. A detailed analysis of all cases where expert and system annotations were not in agreement was performed by another experienced expert in dentomaxillofacial radiology to review possible reasons of incorrect image interpretation. In such cases, the verifying expert had the final say to determine the ground truth. In the cases where the system and the expert provided the same annotations, both were considered correct.

For the detection task, the expert and system annotations were deemed to agree if they intersected substantially. The remaining unmatched boxes were composed of two error types: false positive results, where redundant boxes were annotated, and false negative results, where existent teeth were missed.

For the numbering task, expert and system annotations were deemed to agree, if the class labels provided by experts and the system for the same bounding boxes were identical. Since numbering task is a multiclass problem (32 teeth number), the metrics were evaluated using the one-against-all strategy and then were aggregated. For each tooth number (C):

- "True positives of C" are all C instances that are classified as C;
- "True negatives of C" are all non-C instances that are not classified as C;
- "False positives of C" are all non-C instances that are classified as C;
- "False negatives of C" are all C instances that are not classified as C.

Table 1 Results of tooth detection

	System	Expert
True-positives	5023	5043
False-negatives	30	10
False-positives	28	1
Precision	0.9945	0.9998
Sensitivity	0.9941	0.9980

The metrics depict true positives, false negatives, false positives, sensitivity and precision.
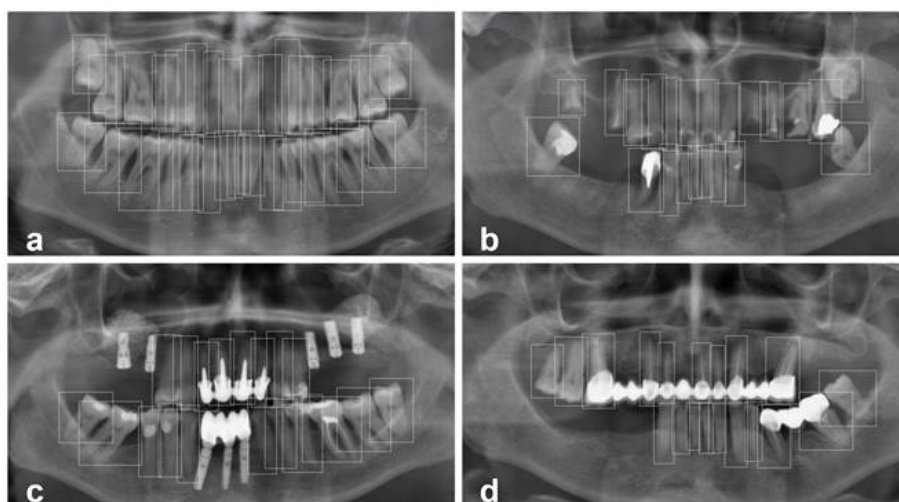
Figure 2 Teeth detection results: (a) all 32 teeth were detected, (b) severely decayed and impacted teeth were detected, (c) implants were excluded and dental crowns were detected, (d) cantilever elements of fixed bridges were excluded.

Based on the results for detection and numbering tasks, the metrics were calculated to evaluate the performance of the system and the expert.

For teeth detection, the following metrics were used: sensitivity = $\frac{TP}{TP+FN}$ and precision = $\frac{TP}{TP+FP}$, where TP, FP, FN represent true-positive, false-positive, and false-negative results, respectively.

For teeth numbering, the following metrics were used: sensitivity = $\frac{TP}{TP+FN}$ and specificity = $\frac{TN}{TN+FP}$, where TP, FP, FN, TN represent true-positive, false-positive, false-negative, and true negative results, respectively.

## Results

Teeth detection results
The method achieved a sensitivity of 0.9941 and a precision of 0.9945. The expert achieved a sensitivity

of 0.9980 and a precision of 0.9998. The detailed data are presented in Table 1. In Figure 2, the sample detection results are shown. In general, the teeth detection module demonstrated excellent results, both for high-quality images with normal teeth arrangement and more challenging cases such as overlapped or impacted teeth, images of a poor quality with blurred contours of teeth, or teeth with crowns. In most cases, the detector correctly excluded bridges and implants from detection results.

In Figure 3, the system error samples are presented. Errors analysis shows that for false-negatives, the main reasons for the faults include the following: root remnants, the presence of orthopaedic appliances, highly impacted and overlapped teeth. The system produced false-positive results in the form of incorrectly detected implants and bridges, extra boxes for teeth with orthopaedic constructions and multiple-rooted teeth,
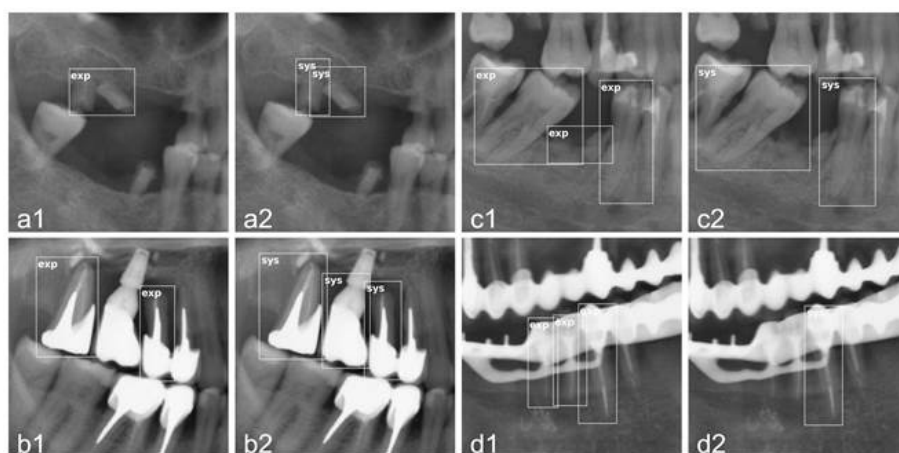
Figure 3 Teeth detection errors produced by the system: for each case, the left image shows the boxes annotated by the experts, the right image shows the boxes detected by the system. False positives: (a) an extra box for the multiple-root tooth was detected, (b) an implant was classified as a tooth. False negatives: (c) a root remnant was missed, (d) teeth obstructed by a prosthetic construction were missed.
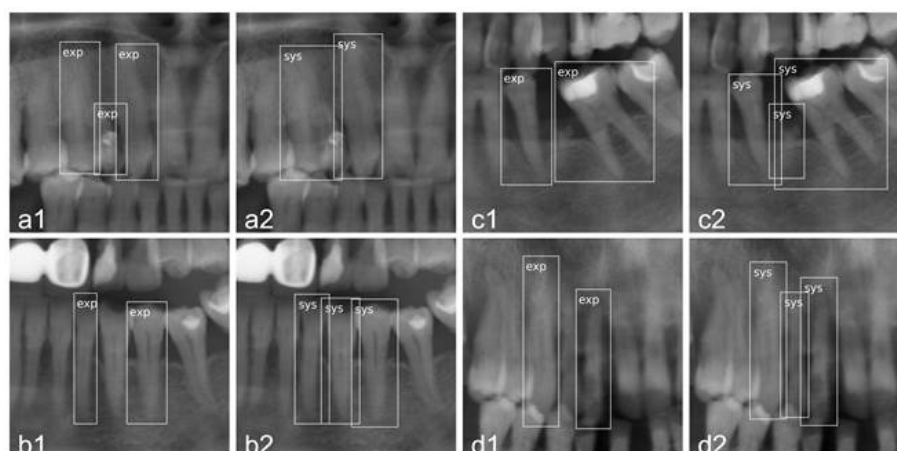
Figure 4 Teeth detection errors produced by the experts: for each case, the left image shows the boxes annotated by the experts, the right image shows the boxes detected by the system. False positives: (a) persistent deciduous tooth was annotated. False negatives: (b) a whole tooth was missed, (c) a root remnant was missed, (d) a tooth obstructed by another one was missed.

and detected fragments outside of the jaw. In Figure 4, errors of the expert teeth detection are presented. Most of the expert errors are false negatives caused by missed root remnants probably as a result of a lack of concentration.

Teeth numbering results

The method achieved a sensitivity of 0.9800 and a specificity of 0.9994, while the expert achieved a sensitivity of 0.9893 and a specificity of 0.9997. The detailed data are presented in Table 2. In Figure 5, the sample numbering results are presented.

Extending the region of cropped teeth to include additional context and augmenting the images resulted in approximately 6 and 2 pp increase of sensitivity respectively. The heuristic method based on spatial teeth number arrangement rules increased the sensitivity by 0.5 pp.

In Figure 6, the system errors samples are shown. The main reasons for numbering errors included lack of nearby teeth near the target tooth, too small remaining tooth fragments (root remnants or severely decayed teeth), and evidence of extensive dental works. In most errors, the system confused a tooth with a missing adjacent one. The molars were mainly misclassified. The

Table 2 Results of tooth numbering

	System	*Experts*
True-positives	4938	4985
True negatives	15,6108	15,6155
False negatives	101	54
False positives	101	54
Specificity	0.9994	0.9997
Sensitivity	0.9800	0.9893

The metrics depict aggregated values of true-positives, true-negatives, false-negatives,and false-positives for all 32 teeth numbers. It also shows sensitivity and specificity.

same cases are reported by experts to be challenging. In Figure 7, examples of expert errors are presented.

In Figure 8, the sensitivity and specificity results for all teeth numbers are presented for both the system and the expert. This figure demonstrates the similarity between the numbering patterns produced by the system and the expert.

Discussion

In the present study, the potential of modern CNN architectures is demonstrated for automated dental X-ray interpretation and diagnostic tasks using panoramic radiographs, specifically for teeth detection and numbering. The system achieved high-performance results for both detection and numbering of teeth that are close to the expert level. The detailed error analysis showed that experts made errors caused by the similar problems in the images. Based on the final results, it can be concluded that the proposed approach shows high enough quality to be integrated into a software used for real-life problems and introduced into daily practice.
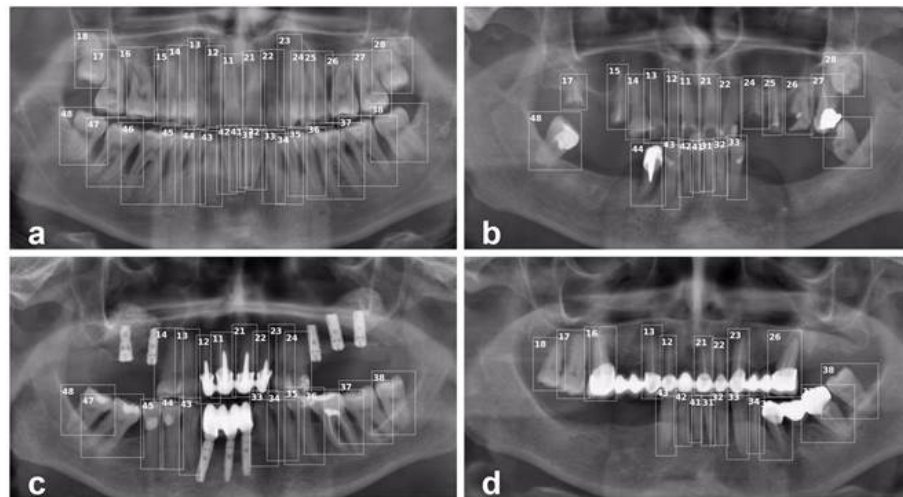
Compared with the Hosntalab et al[5] and Lin et al[4] methods, the present study has an important advantage: the classification performance of the proposed solution does not rely on the accuracy of hand-crafted feature extraction algorithms. Another limitation was that Lin et al[4] analyzed bitewing images that capture only the posterior teeth of one side of the mouth at one given moment. As another CNN-based approach for teeth numbering, Miki et al[6] demonstrated promising results of numbering teeth on CT; however, this study had some limitations: the process of teeth isolation using bounding boxes was manual, the third molars were excluded from the data set.

The analysis of the errors produced by the proposed system showed that the problem of teeth numbering is more challenging for both the system and the experts.
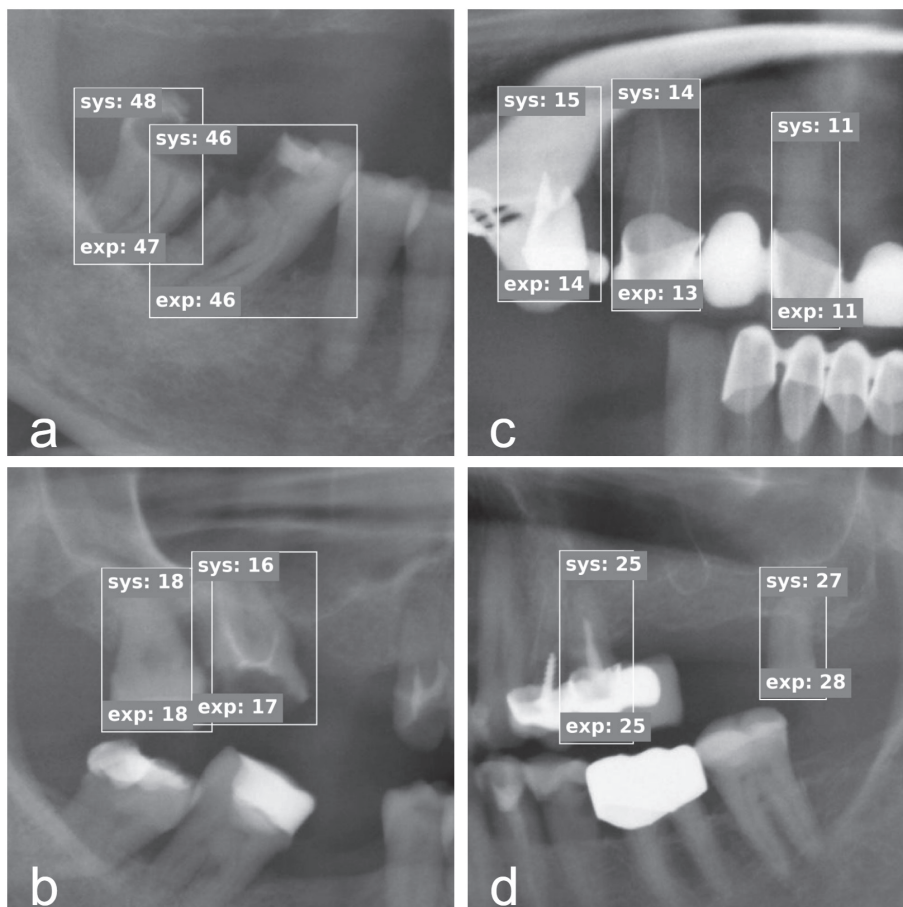
Figure 5 Teeth numbering results: (a) all 32 teeth were correctly classified, (b) severely decayed and impacted teeth were correctly classified, (c) teeth with dental crowns were correctly classified, (d) teeth were correctly classified considering the missed teeth and lack of context.

Figure 6 Teeth numbering errors produced by the system: for each case, the classification provided by the software is at the top, the expert annotation is at the bottom. (a) decayed tooth 47 was misclassified, (b) tooth 17 (severely decayed) was misclassified, (c) teeth 13, 14 obstructed by a prosthetic device were misclassified, (d) tooth 28 was misclassified probably due to the lack of context (missing neighbouring teeth).
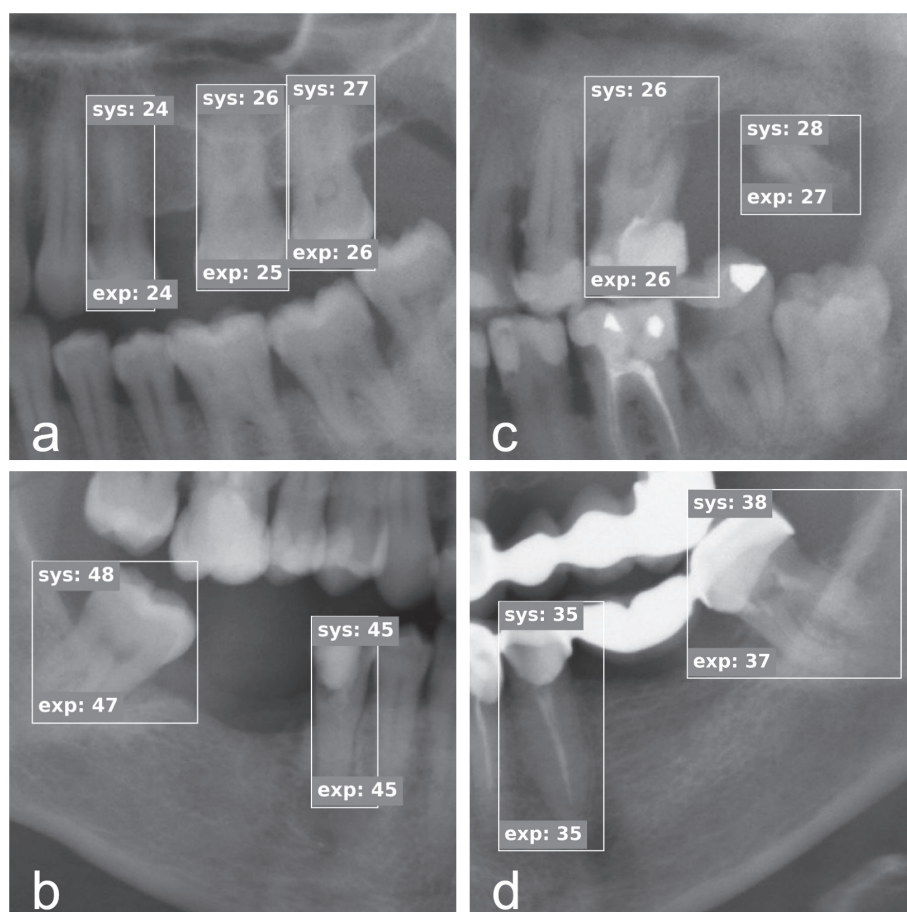
Figure 7 Teeth numbering errors produced by the experts: for each case, the system classification result is at the top, the expert annotation is at the bottom. (a) teeth 26, 27 were misclassified, (b) tooth 48 was misclassified, (c) a root remnant of 28 was misclassified, (d) tooth 38 was misclassified.

The most misclassifications occurred among the neighbouring teeth, especially in case of missed teeth. Despite the fact, that many errors are explainable, the system's performance in both detection and numbering tasks is still lower than the one of the experts. There are more techniques to study that can further improve the system's output, including an application of more advanced augmentation techniques,[21] extending the data set, and the use of more recent CNN architectures.

Since CNNs do not rely on hand-crafted features, application of deep learning techniques can be studied for other tasks as well. First, in further works, the present model can be extended to interpret other types of X-ray images, such as cephalograms, bitewings, or even 3D images such as cone beam CT. Even more exciting prospects open up in generalizing the model to detect and interpret other dental structures and even pathologies. For dental and oral structures, it is useful to locate implants, bridges, and crowns. For pathologies, the primary goal is to detect caries, periodontitis, and dental cysts. Admittedly, to achieve expert level results for such tasks, these new systems will most likely require much larger samples for training than used in this work.

These directions of research might require extending the number of classes for Faster R-CNN, implementing segmentation techniques for more accurate pathology localization, experimenting with new architectures and networks. One important advantage of the CNN approach is that these improvement steps can be gradual, and results of previous steps can be reused in the form of transfer learning: fine-tuning of existing models, training new models on already annotated data sets, segmentation or localization of objects within previously detected boundaries.

Conclusions

This study verifies the hypothesis that a CNN-based system can be trained to detect and number teeth on panoramic radiographs for the automated dental charting purposes. The proposed solution aims to assist dentists in their decision-making process rather than substitute them. The systems performance level is close to the experts' level, which means that the radiologist can use the output of the system for automated charting
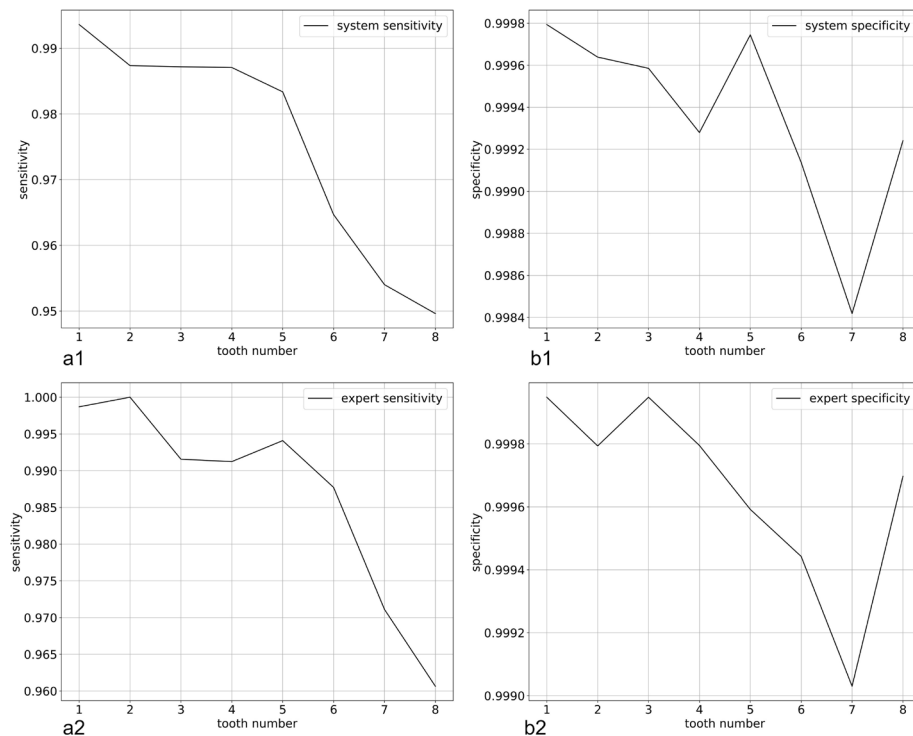
Figure 8 Sensitivity and specificity plots for all teeth numbers. The plots show sensitivity and specificity for each of eight teeth numbers averaged by four quadrants: (a1–b1) sensitivity and specificity for the system, (a2–b2) sensitivity and specificity for the expert. These plots demonstrate similarity in the numbering patterns produced by the system and the expert.

when only evaluation and minor corrections are required instead of manual input of information.

As this is a proof-of-concept study, there is a potential to increase the system performance. The results of the teeth detection and numbering could be improved or made more robust by implementing additional techniques, such as advanced image augmentation,[21] and using more recent CNN architectures for feature extraction and classification.

Based on the results achieved, it can be concluded that AI deep learning algorithms have a potential for further investigation of their applications and implementation

in a clinical dental setting. This approach to CADs has an important advantage over conventional computer vision and machine learning techniques, as it does not rely on hand-crafted features or special-purpose programming, instead it is actually learning only from initial image representations such as pixels of dental panoramic radiographs.

Funding

References

1. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 2007; **31**(4-5): 198–211p.. doi: https://doi.org/10.1016/j.compmedimag.2007.02.002
2. Rezaei M, Yang H, Meinel C. Deep Neural Network with l2-norm Unit for Brain Lesions Detection. In: *International Conference on Neural Information Processing*; 2017. pp. 798–807.
3. ISO ISO 3950:2016 Dentistry-Designation system for teeth and areas of the oral cavity. 2016;.
4. Lin PL, Lai YH, Huang PW. An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information. *Pattern Recognition* 2010; **43**: 1380–92. doi: https://doi.org/10.1016/j.patcog.2009.10.005
5. Hosntalab M, Aghaeizadeh Zoroofi R, Abbaspour Tehrani-Fard A, Shirani G, Zoroofi RA, Tehrani-Fard AA. Classification and

numbering of teeth in multi-slice CT images using wavelet-Fourier descriptor. *Int J CARS* 2010; **5**: 237–49. doi: https://doi.org/10.1007/s11548-009-0389-8
6. Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med* 2017; **80**: 24–9. doi: https://doi.org/10.1016/j.compbiomed.2016.11.003
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: NIPS, ed. *Annual Conference on Neural Information Processing Systems*; 2012. pp. 1097–105.
8. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc. IEEE* 1998; **86**: 2278–324. doi: https://doi.org/10.1109/5.726791

9. Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi Aet al. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In: CVPR, ed. *The IEEE Conference on Computer Vision and Pattern Recognition*; 2017. pp. 3296–7.

10. LeCun Y, Bengio Y, Hinton G, learning D. Deep learning. *Nature* 2015; **521**: 436–44. doi: https://doi.org/10.1038/nature14539

11. Lee H, Park M, Kim J. Cephalometric landmark detection in dental X-ray images using Convolutional neural networks. in SPIE medical imaging. 2017; 1–6.

12. Arik Sercan Ö., Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imag* 2017; **4**: 014501. doi: https://doi.org/10.1117/1.JMI.4.1.014501

13. Wang C-W, Huang C-T, Lee J-H, Li C-H, Chang S-W, Siao M-J, et al. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal* 2016; **31**: 63–76. doi: https://doi.org/10.1016/j.media.2016.02.004

14. Liu J, Wang D, Lu L, Wei Z, Kim L, Turkbey EB, et al. Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks. *Med Phys* 2017; **44**: 4630–42. doi: https://doi.org/10.1002/mp.12399

15. 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* 2017; **39**: 1137–49. doi: https://doi.org/10.1109/TPAMI.2016.2577031

16. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: ICLR, ed. *International Conference on Learning Representations*; 2015.

17. Deng J, Dong W, Socher R, LJ L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: CVPR, ed. *The Conference on Computer Vision and Pattern Recognition*; 2009. pp. 248–55.

18. Hosang J. Faster RCNN TF. 2016. Available from: https://github.com/smallcorgi/Faster-RCNN_TF.

19. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available from: tensorflow.org.

20. Chollet F. keras. 2015. Available from: https://github.com/fchollet/keras.

21. Jung A. Image augmentation for machine learning experiments. 2015. Available from: https://github.com/aleju/imgaug.