



Tooth Detection from Panoramic Radiographs Using Deep Learning

Shweta Shirsat¹(✉) and Siby Abraham²

¹ University Department of Computer Science, University of Mumbai, Mumbai 400098, India

² Center of Excellence in Analytics and Data Science, NMIMS Deemed to be University, Mumbai 400056, India

siby.abraham@nmims.edu

Abstract. The proposed work aims at implementing a Deep Convolutional Neural Network algorithm specialized in object detection. It was trained to perform tooth detection, segmentation, classification and labelling on panoramic dental radiographs. A dataset of dental panoramic radiographs was annotated according to the FDI tooth numbering system. Mask R-CNN Inception ResNet V2 object detection algorithm was able to give excellent results in terms of tooth segmentation and numbering. The experimental results were validated using standard performance metrics. The method could not only give comparable results to that of similar works but could detect even missing teeth, unlike similar works.

Keywords: Tooth detection · Deep Learning · Neural network · Transfer learning · Radiographs

1 Introduction

Deep Learning has become a buzzword in the technological community in recent years. It is a branch of Machine Learning. It is influenced by the functioning of the human brain in designing patterns and processing data for decision making. Deep neural networks are suitable for learning from unlabelled or unstructured data. Some of the key advantages of using deep neural networks are their ability to deliver high-quality results, eliminating the need for feature engineering and optimum utilization of unstructured data [1]. These benefits of deep learning have given a huge boost to the rapidly developing field of computer vision. Various applications of deep learning in computer vision are image classification, object detection, face recognition and image segmentation. An area that has achieved the most progress is object detection.

The goal of object detection is to determine which category each object belongs to and where these objects are located. The four main tasks in object detection include classification, labelling, detection and segmentation. Fast Convolutional Neural Networks (Fast-RCNN), Faster Convolutional Neural Network (Faster-RCNN) and Region-based Convolutional Neural Networks(R-CNN) are the most widely used deep learning-based object detection algorithms in computer vision. Mask R-CNN, an extension to faster-RCNN [2] is far superior to others in terms of detecting objects and generating high-quality masks. Mask-RCNN architecture is represented in Fig. 1

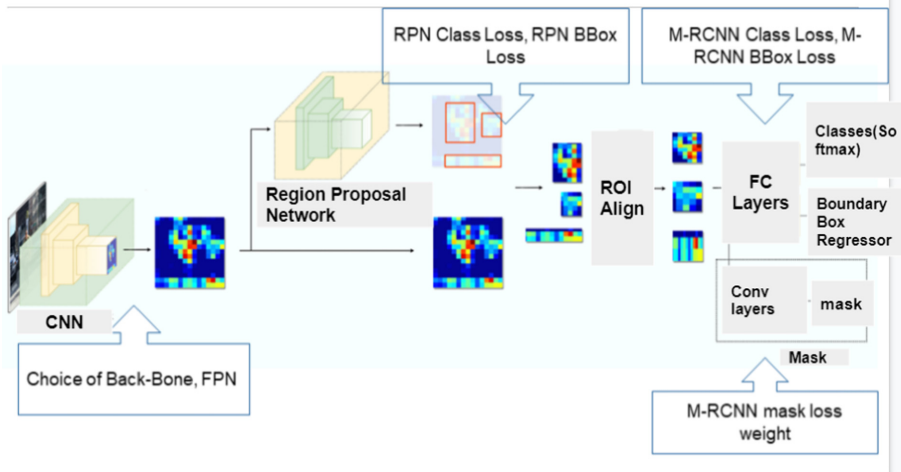


Fig. 1. Mask R-CNN architecture [3]

Mask-RCNN is also used for medical image diagnosis in locating tumours, measuring tissue volumes, studying anatomical structures, lesion detection, planning surgery, etc. Mask-RCNN uses the concept of transfer learning to drastically reduce the training time of a model and lower generalization results. Transfer learning is a method where a neural network model is trained on a problem that is similar to the problem being solved. These layers of the trained model are then used in a new model to train on the problem of interest. Several high-performing models can be used for image recognition and other similar tasks in computer vision. Some of the pre-trained transfer learning models include VGG, Inception and MobileNet [4].

Deep Learning is steadily finding its way to offer innovative solutions in the radiographic analysis of dental X-ray images. In dental radiology, the tooth numbering system is the format used by the dentist for recognizing and specifying information linked with a particular tooth. A tooth numbering system helps dental radiologists identify and classify the condition associated with a concerned tooth. The most frequently used tooth numbering methods are the Universal Numbering System, Zsigmondy-Palmer system, and the FDI numbering system [4] (Fig. 2).

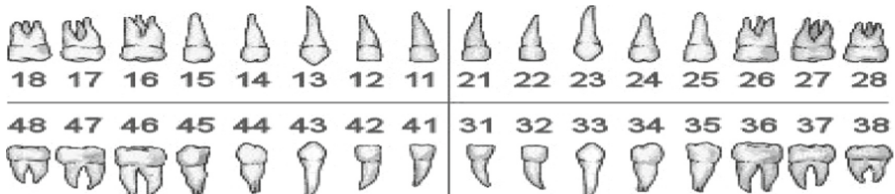


Fig. 2. FDI tooth numbering system chart [10]

The annotation method used in this research work primarily focuses on the FDI(Federation Dentaire Internationale) notation system(ISO 3950) [9]. The FDI tooth numbering system is an internationally recognized tooth numbering system where there are 4 quadrants. Maxillary right quadrant is quadrant 1, Maxillary left quadrant is quadrant 2, the Mandibular left quadrant is quadrant 3 and the Mandibular right quadrant is quadrant 4. Each quadrant is recognized from number 1 to 8. For example, 21 indicates maxillary left quadrant(quadrant 2) third teeth known as a central incisor.

Deep neural networks can be used with these panoramic radiographs for tooth detection and segmentation using different variations of Convolutional Neural Networks. Till date, most of the research work tackled the problem of tooth segmentation on panoramic radiographs using Fully Convolutional Neural Network and its variations. To the best of our knowledge, this is the first work to use Mask R-CNN Inception ResNet V2 trained on the COCO 2017 capable of working on Tensorflow version 2 object detection API. It guarantees to give better results in terms of performance metrics used to check the credibility of the Deep Convolutional Network algorithm used.

2 Related Works

There were a few attempts to apply deep learning techniques for teeth detection and segmentation.

Thorbjorn Louring Koch et al. [1] implemented a Deep Convolutional Neural Network (CNN) developed on the U-Net architecture for segmentation of individual teeth from dental panoramic X-rays. Here CNN reached the dice score of 0.934(AP). 1201 radiographic images were used for training, forming an ensemble that increased the score to 0.936.

Minyoung Chung et al. [3] demonstrated a CNN-based individual tooth identification and detection algorithm using direct regression of object points. The proposed method was able to recognize each tooth by labelling all 32 possible regions of the teeth including missing ones. The experimental results illustrated that the proposed algorithm was best among the state-of-the-art approaches by 15.71% in the precision parameter of teeth detection.

Dmitry V Tuzoff et al. [2] used the state-of-the-art Faster R-CNN architecture. The FDI tooth numbering system was used for teeth detection and localization. A classical VGG-16 CNN along with a heuristic algorithm was used to improve results according to the rules for the spatial arrangement of teeth.

Shuxu Zhao et al. [4] proposed the Mask R-CNN, for classification and segmentation of tooth. The results showed that the method achieved more than 90% accuracy in both tasks.

Gil Jader et al. [7] proposed Mask R-CNN for teeth instance segmentation by training the system with only 193 dental panoramic images of containing 32 teeth on average, they achieved an accuracy of 98%, F1-score of 88%, precision of 94%, recall of 84%, and 99% specificity.

Gil Jader et al. [8] performed a study of tooth segmentation and numbering on panoramic radiographs using an end-to-end deep neural network. The proposed work

used Mask R-CNN for teeth localization on radiographic images. The calculated accuracy was 98% out of which F1-score was 88%, precision was 94%, recall was 84% and specificity of 99% of over 1224 radiographs.

Gil Jader et al. [9] also proposed a segmentation system based on the Mask R-CNN and transfer learning to perform an instance segmentation on dental radiographs. The system was trained with 193 dental radiographs having a maximum of 32 teeth. Accuracy achieved was 98%.

3 Methodology

Figure 3 represents Mask R-CNN applied on a set of dental radiographs to perform tooth identification and numbering.

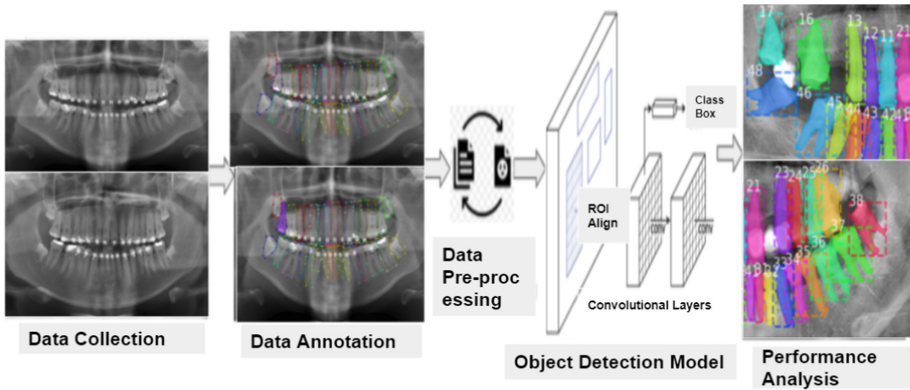


Fig. 3. Mask R-CNN architecture for tooth segmentation and numbering

3.1 Data Collection

To train a robust model we needed a lot of images that should vary as much as possible. The dataset of panoramic dental radiographs was collected from Ivison dental labs(UFBA_UESC_DENTAL_IMAGES_DEEP) [10]. The height and width of each panoramic dental radiograph ranged from 1014–1504 pixels and 2094 to 3432 pixels respectively. These radiographs were then resized to a fixed resolution of 800 * 600. The suitable format to store the radiographs was JPG.

3.2 Data Annotation

After collecting the required data these radiographs had to be annotated as per the FDI tooth numbering system. Rather than only annotating existing teeth in a radiograph, we annotated all 32 teeth including missing teeth. A JSON/XML file was created for each radiograph representing manually defined bounding boxes, and a ground truth label set

for each bounding box. Though there were a variety of annotation tools available such as the VGG Image annotation tool, labelme, and Pixel Annotation Tool [8]. The proposed work uses labelme software because of its efficiency and simplicity. All annotations were verified by a clinical expert in the field.

3.3 Data Preprocessing

In addition to the labelled radiographs a TFRecord file needed to be created that could be used as input for training the model. Before creating TFRecord files, we had to convert the labelme labels into COCO format, as we had used the same as the pre-trained model. Once the data is in COCO format it was easy to create TFRecord files [9].

3.4 Object Detection Model

The below steps illustrate how the Mask R-CNN object detection model works:-

- A set of radiographs was passed to a Convolutional Neural Network.
- The results of the Convolutional Neural Network were passed through to a Region Proposal Network (RPN) which produces different anchor boxes known as ROI(Regions of Interest) based on each occurrence of tooth objects being detected.
- The anchor boxes were then transmitted to the ROI Align stage. It is essential to convert ROI's to a fixed size for future processing.
- A set of fully connected layers will receive this output which will result in the generating class of the object in that specific region and defining coordinates of the bounding box for the object.
- The output of the ROI Align stage is simultaneously sent to CNN's to create a mask according to the pixels of the object.

Hyper Parameter Tuning: The training was performed on dental radiographic images having 32 different objects that were identified and localized. The hyperparameter values of the object detection models were: Number of classes = 32; image_resolution = 512 * 512; mask_height * width = 33 * 33; standard_deviation = 0.01; IOU_threshold = 0.5; Score_converter = Softmax;batch_size = 8; No. of steps = 50,000; learning_rate_base = 0.008. The parameters like standard deviation, score_converter, batch_size, no.of epochs and fine_tune_checkpoint_type were optimised.

3.5 Performance Analysis

Test results after training the model for 50K epochs are shown in Fig. 4. To measure the performance accuracy of object detection models some predefined metrics such as Precision, Recall and Intersection Over Union(IoU) are required [8].

Precision: Precision is the capability of a model to identify only the relevant objects. It is the percentage of correct positive predictions and is given by $\text{Precision} = \frac{TP}{(TP + FP)}$ [9].

Recall: Recall is the capability of a model to find all the relevant cases (all ground truth bounding boxes). It is the percentage of true positives detected among all relevant ground truths and is given by: $\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$ [10].

Where,

TP = **true positive** is observed when a prediction-target mask pair has an IoU score that exceeds some predefined threshold;

FP = **false positive** indicates a predicted object mask has no associated ground truth object mask.

FN = **false negative** indicates a ground truth object mask has no associated predicted object mask.

IoU: Intersection over Union is an evaluation metric used to measure the accuracy of an object detector on a specific dataset [10].

$\text{IoU} = \text{Area of Overlap}/\text{Area of Union}$

4 Experimental Results

The experiment was executed on a GPU ($1 \times$ Tesla K80), with 1664 CUDA cores and 16 GB memory. The algorithm was running on TensorFlow version 2.4.1 having python version 3.7.3. In general the tooth detection numbering module demonstrated results for detecting each tooth from dental radiographs. Then it was also able to provide tooth numbers for each detected tooth as per the FDI tooth numbering system. The sample results are shown in Fig. 4. When the training process was successfully completed, a precision of 0.98 and recall of 0.97 was recorded as seen in Table 1.

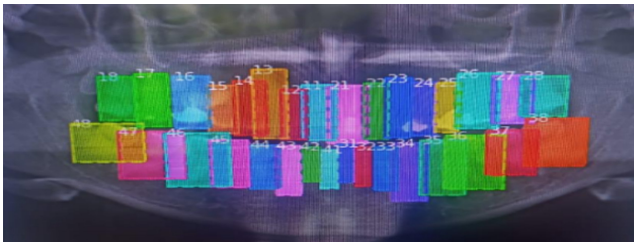


Fig. 4. Dental radiograph with results

Table 1. Evaluation metrics

Evaluation metric	Value
Precision	0.98
Recall	0.97
IoU	0.5

Along with evaluation metrics, a graphical representation of various loss functions presented on the tensorboard is given in Figs. 5, 6, 7 and 8.

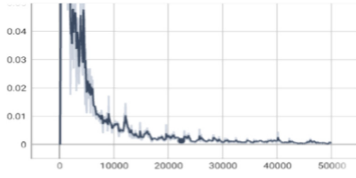


Fig. 5. Localization loss

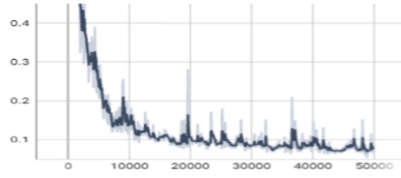


Fig. 6. Total loss

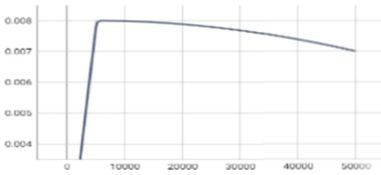


Fig. 7. Learning rate

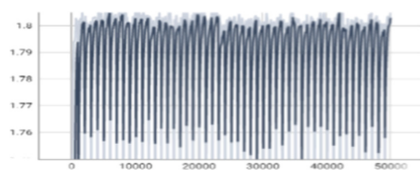


Fig. 8. Steps_per_epoch

4.1 Localization Loss

Localization loss is used to demonstrate loss between a predicted bounding box and ground truth [10]. As the training progresses, the localization loss decreases gradually and then remains stable as illustrated in Fig. 5.

4.2 Total Loss

The total loss is a summation of localization loss and classification loss as represented in Fig. 6. The optimisation model reduces these loss values until the loss sum reaches a point where the network can be considered as fully trained.

4.3 Learning Rate

Learning Rate is the most important hyperparameter for this model which is shown in Fig. 7. Here we can see there is a gradual increase in the learning rate after every batch recording the loss at every increment. When entering the optimal learning rate zone it is observed that there is a sudden drop in the loss function.

4.4 Steps Per Epoch

This hyperparameter is useful if there is a huge dataset with considerable batches of samples to train. It defines how many batches of samples are used in one epoch. In our tooth detection model, the total number of epochs was 50,000 with an interval of 100 as represented in Fig. 8.

5 Comparative Study

5.1 Comparison with Clinical Experts

The results provided by the proposed model were asked to be verified by a clinical expert for detected and undetected bounding boxes, confusion with similar teeth or missing tooth labels, failure in complicated cases and objects detected more than ground tooth.

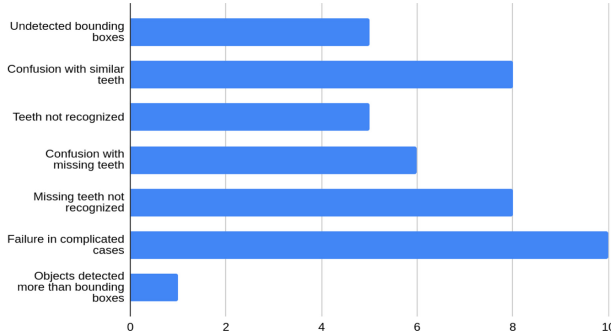


Fig. 9. Expert analysis

Figure 9 demonstrates the percentage difference between the expert reviews and predictions from the system.

Out of the total sample data of 25 radiographs there exists 8% confusion with similar kinds of teeth in a single radiograph. 5% of teeth were not correctly recognized by the algorithm. This algorithm was also able to identify tooth numbers for missing teeth but there exists a confusion of 6%. Under some complicated scenarios around 8% of missing teeth were not recognized. There exists around 10% failure in complicated cases such as tooth decay, impacted teeth, cavities or because of partial or full dental implantation. As per the graphical representation inaccuracies, though small in number, are attributed to a large extent because of the poor quality of the data and not the performance deficiency of the model.

5.2 Comparison with Other Works

Table 2 provides a comparative study of the effectiveness of the proposed model with similar works. Comparison has been done based on seven criterias as shown in Table 2. The proposed work is different from others as it demonstrates the implementation of the Mask R-CNN Inception ResNet V2 object detection model. Only one Mask R-CNN Model is supported with TensorFlow 2 object detection API at the time of writing of this paper [10]. The transfer learning technique was successfully implemented using a pre-trained COCO dataset. Along with tooth detection, tooth segmentation, tooth numbering we were also able to predict missing teeth. These missing teeth are known as edentulous spaces from maxilla or mandible. M-RCNN was able to correctly recognize edentulous

spaces on a radiograph. This study demonstrates that the proposed method is far superior to the other state of the art models pertaining to tooth detection and localization. Also, there is a clear understanding of poor detection, wherever it occurred though small in number, and its verification is done by a clinical expert.

Table 2. Comparative study

Author	Minyoung Chung [3]	Shuxu Zhao [4]	Guohua Zhu [5]	Gil Jader [7]	Proposed work
Type of radiograph	Panoramic	Panoramic	Digital X-rays	Panoramic	Panoramic
Deep learning method	Fast R-CNN	Mask R-CNN	Mask R-CNN	Mask R-CNN, PANet, HTC and ResNeSt	Mask R-CNN Inception V2
Transfer learning	No	Yes	Yes	Yes	Yes
Tooth detection	Yes	Yes	Yes	Yes	Yes
Tooth segmentation	Yes	Yes	Yes	Yes	Yes
Tooth numbering	Yes	Yes	Yes	Yes	Yes
Missing tooth detection	No	Yes	No	No	Yes
Comparison with experts	No	No	No	Yes	Yes

6 Conclusion

The Mask R-CNN Inception ResNet V2 object detection model was used to train a dataset of dental radiographs for tooth detection and numbering. It was observed that the training results were exceptionally good especially in tooth identification and numbering with high IOU, precision and recall. The visualization results were considerably better than Fast R-CNN. The performance of our selected model was very close to the level of the clinical expert who was selected as a referee in this study. In future studies, we will consider working with more advanced models for periodontal bone loss, early caries diagnosis, and various periapical diseases.

References

1. Koch, T.L., Perslev, M., Igel, C., Brandt, S.S.: Accurate Segmentation of Dental Panoramic Radiographs with U-NETS. In: Published in Proceedings of 16th IEEE International Symposium on Biomedical Imaging, Venice, Italy (2019)

2. Tuzoff, D.V., et al.: Tooth detection and numbering in panoramic radiographs using convolutional neural networks. In: Published in Conference of Dental Maxillofacial Radiology in medicine, Russia, March-2019 (2019)
3. Chung, M., et al.: Individual Tooth Detection and Identification from Dental Panoramic X-Ray Images via Point-wise Localization and Distance Regularization. Elsevier, Artificial Intelligence in medicine, South Korea (2020)
4. Zhao, S., Luo, Q., Liu, C.: Automatic Tooth Segmentation and Classification in Dental Panoramic X-ray Images. In: IEEE International Conference on Pattern Analysis and Machine Intelligence, China (2020)
5. Zhu, G., Piao, Z., Kim, S.C.: Tooth detection and segmentation with mask R-CNN. In: Published in International Conference on Artificial Intelligence in Information and Communication (2020)
6. Jader, G., Oliveira, L., Pithon, M.: Automatic segmentation teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. In: Elsevier, Expert systems with applications volume 10, March-2018 (2018)
7. Gil Jader et al.: Deep instance segmentation of teeth in panoramic x-ray images. In: SIBGRAPI 31st International Conference on Graphics, Patterns and Images (SIBGRAPI), Brazil, May 2018 (2018)
8. Silva, B., Pinheiro, L., Oliveira, L., Pithon, M.: A study on tooth segmentation and numbering using end-to-end deep neural networks. In: Published in 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Brazil (2020)
9. Jader, G. et al.: Deep instance segmentation of teeth in panoramic X-ray images (2018). http://sibgrapi.sid.inpe.br/col/sid.inpe.br/sibgrapi/2018/08.29.19.07/doc/tooth_segmentation.pdf
10. Freny, R.K.: Essentials of Oral and Maxillofacial Radiology. Jaypee Brothers Medical Publishers(P) Ltd., New Delhi. MDS (Mumbai University)