

PAPER • OPEN ACCESS

## Object detection on dental x-ray images using deep learning method

To cite this article: D Suryani *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1073** 012058

View the [article online](#) for updates and enhancements.

### You may also like

- [Synergistic effect of ion-releasing fillers on the remineralization and mechanical properties of resin–dentin bonding interfaces](#)  
Yimeng Xie, Ruhua Chen, Wei Yao et al.
- [Defects detection in dentistry: designing a graphene multi-layered based plasmonic sensor](#)  
Jitendra Singh Tamang, Somenath Chatterjee and Rudra Sankar Dhar
- [The effect of thermal aging on flexural strength of CAD/CAM hybrid and polymeric materials](#)  
Hanin E Yeslam, Shadia Alharbi, Waad Albalawi et al.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research



# Object detection on dental x-ray images using deep learning method

**D Suryani, M N Shoumi\*, and R Wakhidah**

Department of Information Technology, State Polytechnic of Malang, Jl. Soekarno-Hatta No. 9, Malang 65141, East Java, Indonesia

\*milyun.nima.shoumi@polinema.ac.id

**Abstract.** Radiological examination has an important role in determining the diagnosis of dental problems and making decisions about the right type of treatment according to the case indications. Dental x-ray is a medical procedure for taking pictures of the inside of the mouth using radiation fluid, where the results are used diagnostically to help the dentist see the entire structure of the jaw bone and teeth, and dental problems that cannot be seen directly. Dental radiographic interpretation, which is generally performed by dentists, is a time-consuming and error-prone process due to high variations in tooth structure, low experience levels, and fatigue factors experienced by dentists. The workload of a dentist and the occurrence of misdiagnosis can be reduced by the existence of a system that can automatically interpret the x-ray results. To overcome these problems, a model will be developed to be able to detect objects in the dental panoramic x-ray images using Mask R-CNN, one of the methods in Deep Learning. Deep Learning is an artificial intelligence function that modelled the workings of human brain in processing data and creating patterns for use in decision making. With the detection of objects in panoramic x-ray image automatically, it is expected to save time, improve the quality of dental care, and also the quality of diagnosis made by dentists.

## 1. Introduction

In the world of dentistry, radiological examination has an especially important role in determining the diagnosis and making the right type of treatment decision according to the indication of the case. Dental x-rays or dental radiography are medical procedures to take pictures of the inside of the mouth using radiation fluids, where the results are used diagnostically to help dentists see the entire structure of the jawbone and teeth, as well as dental problems that cannot be seen directly by the eye. Without dental X-rays, many dental problems go undiagnosed. One of the most performed radiographic techniques is the panoramic radiograph technique. A panoramic x-ray can provide a complete picture of the upper and lower jaw at once, as well as various adjacent anatomical structures. Doctors perform this procedure to find out disorders in the mouth. This procedure can also be used to plan treatment for dentures, braces, tooth extractions, or dental implants.

Dental radiographic interpretation, which is generally performed by dentists, is a time-consuming and error-prone process due to high variation in tooth structure, low level of experience, and the large number of dental X-rays that need to be checked every day by the dentist, which can cause dentists to become exhausted. leading to a misdiagnosis or an inaccurate diagnosis, which in turn can hinder treatment. A dentist's workload and the occurrence of misdiagnoses can be reduced by having a system that can automatically interpret x-ray results. To overcome this problem, a model that can detect objects in a panoramic x-ray image of teeth will be developed using the Mask R-CNN method, which



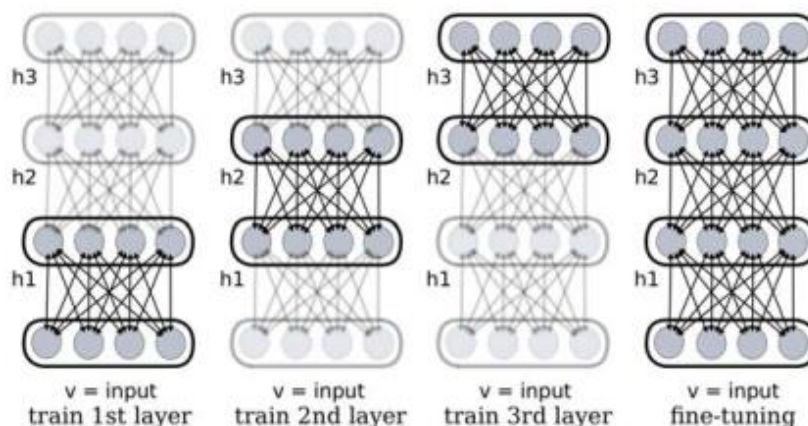
Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

is one of the methods in Deep Learning. The Mask R-CNN method is proven to produce high accuracy in recognizing objects in an image with input data in the form of an annotated image. The object to be detected is a restoration object. Dental restoration is a treatment to repair cavities or damaged teeth, as well as dental repair treatment after endodontic treatment, to restore the normal function and appearance of teeth. Tooth restoration after endodontic treatment is the key to successful treatment. With the automatic detection of objects on panoramic x-rays, it is expected save time, improve the quality of dental care, and the quality of diagnosis made by dentists.

## 2. Background

### 2.1. Deep learning

The term deep learning was first introduced to the machine learning community by Rina Dechter in 1986 [1] and was introduced to the realm of artificial neural networks by Igor Aizenberg in 2000 [2], in the context of Boolean threshold neurons. Deep learning is a subset of machine learning and it is a function of artificial intelligence that imitate the way of human brain processes data and creates patterns for use in decision making. Deep Learning is a network that can learn unsupervised from unstructured or unlabelled data. In deep learning, a learning model performs classification tasks from images, text, or sounds. The model is trained using large labelled datasets and a neural network architecture that contains multiple layers. The deep learning scheme can be seen in Figure 1. Each layer is treated separately and trained sequentially. When the previous layer has been trained, the new layer will be trained using the encoding from the previous layer's input.



**Figure 1.** The deep learning scheme [3].

Most deep learning models are based on Artificial Neural Networks, and the most popular type is Convolutional Neural Network (CNN). CNN combines features learned from input data and uses 2D convolutional layers to make this architecture particularly suitable for processing 2D data, such as images. CNN eliminates the need for manual feature extraction, eliminating the need to identify the features used to classify images. CNN works by extracting features directly from images. This automated feature extraction makes deep learning models highly accurate for tasks such as object classification and detection. One type of CNN that has high accuracy in detecting an object is Mask R-CNN.

### 2.2. Mask R-CNN

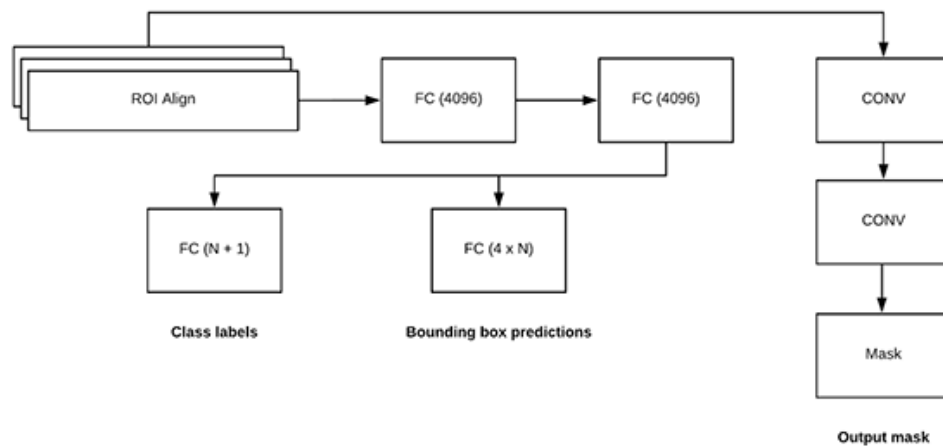
Region-Based Convolutional Neural Networks (R-CNN), is a family of convolutional neural network models designed for object detection, developed by Girshick [4]. There are four main variations of the R-CNN approach. The salient aspects of each variation can be summarized as follows:

- R-CNN: The bounding box is proposed by a "selective tracing" algorithm, which is each stretched, and features extracted via a deep convolutional neural network, before the final set of object classifications is created with linear SVM.
- Fast R-CNN: Simplified design with single model, bounding box is still defined as input, but the merge area of interest layer is used after CNN in to combine regions and the model predicts class and region labels of interest directly.
- Faster R-CNN: Added Regional Proposal Network that interprets extracted features from deep CNN and learns to propose areas of interest directly.
- Mask R-CNN: Extension of Faster R-CNN which adds an output model to predict the mask for each detected object.

The Mask R-CNN Model introduced in 2017 [5] is the latest variation of the family model and supports object detection and object segmentation. The R-CNN Mask algorithm builds on the Faster R-CNN architecture with two main contributions:

- Replacing the ROI Pooling module with a more accurate ROI Align module
- Incorporate additional branches from the ROI Align module

This additional branch receives the ROI Align output and then feeds it into the two CONV layers. The output of the CONV layer is the mask itself. The R-CNN Mask architecture can be visualized in Figure 2.

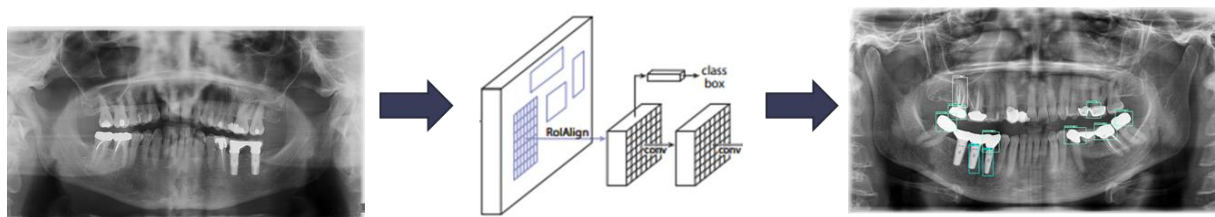


**Figure 2.** Mask R-CNN architecture [6].

### 3. Research methodology

#### 3.1. System architecture

The architecture of object detection system in dental images using mask R-CNN can be seen in Figure 3.



**Figure 3.** System architecture of object detection in dental images.

The first step is to take the images and extract the features using the ResNet 101 architecture [7]. These features act as an input for the next layer. Then take the feature map obtained in the previous step and apply the Region Proposal Network (RPN). This stage basically predicts whether an object is present in that region (or not) and get a map of the region or feature map that the model predicts contains multiple objects. The next step is to apply a pooling layer and convert all the regions to the same shape. Then this region is passed through a fully connected network so that the class labels and bounding boxes are predictable. For all predicted regions, Intersection over Union (IoU) will be calculated with the ground truth box. The IoU calculation is as follows:

$$IoU = \text{Area of intersection} / \text{Area of unity} \quad (1)$$

If the IoU is greater than or equal to 0.5, it is considered as a region of interest. If not, the region will be ignored. It is implemented for all regions and then selects only one set of regions which the IoU is greater than 0.5. After having a RoI based on the IoU value, the next process is to add a mask branch to the existing architecture. This returns the segmentation mask for each region containing the object. This returns a mask with a size of 28x28 for each region which is then increased for inference. The model has segmented all objects in the image. This is the final step in Masks R-CNN where it performs mask prediction for all objects in the image.

## 4. Results and discussion

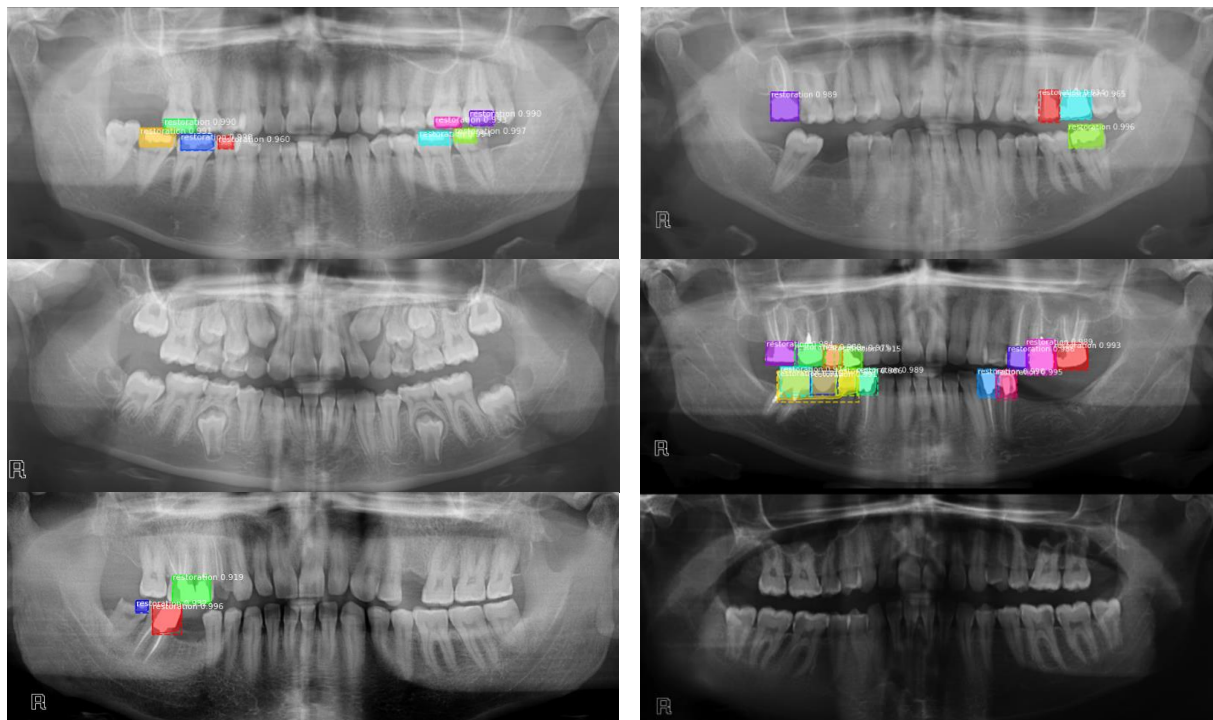
### 4.1. Dataset

The data used in this study are panoramic x-ray images of teeth that have been annotated using the open-source Visual Object Tagging Tool (VoTT) software. This data set consisted of anonymous and de-identified panoramic dental X-rays from 116 patients, taken at the Noor Medical Imaging Centre, Qom, Iran [8]. Subjects covered a wide variety of dental conditions from healthy, to partial and complete toothless cases. The dataset is annotated with one class, which is restoration, and it is divided into two parts, 110 images for the training process and 6 images for the testing process.

### 4.2. Detection results

Experiments were carried out using a GPU with a number of training steps per epoch of 100, the learning rate was 0.006, and the minimum value for detection confidence was 0.9. The pre-fit R-CNN Mask model on the MS COCO object detection dataset can be used as a starting point and then adjusted to a specific dataset, in this case, a dental x-ray image. The model is trained on the output layer (heads) with a layer on top to speed up the learning process, using 5 epochs. Detection results for 6 testing data can be seen in Figure 4.





**Figure 4.** Experiments results on testing image.

#### 4.3. Analysis

Model performance for object detection is evaluated using mean absolute precision, or mAP. The bounding box is predictable so it can be determined whether the bounding box prediction is good or not based on how well the predicted and actual bounding boxes overlap. It can be calculated by dividing the area of the overlap by the total area of the two bounding squares, or the intersection divided by the union, it is known as the "intersection over union," or IoU. The perfect bounding box will have IoU of 1.

IoU is a standard for assuming positive predictions from the bounding box. Precision refers to the percentage of bounding boxes predicted correctly ( $\text{IoU} > 0.5$ ) of all predicted bounding boxes. Meanwhile, recall is the percentage of bounding boxes that are predicted correctly ( $\text{IoU} > 0.5$ ) of all objects in the photo.

#### 5. Conclusion

The object detection model on panoramic x-ray images of teeth has been developed and has successfully detected the location of the restoration object in the data. Only one type of object in the panoramic tooth image was detected, namely tooth restoration. This is due to the difficulty in finding a panoramic dental image dataset that contains several objects, such as dental implants and endodontic treatments. The results of object detection in the panoramic tooth image show confidence level values in the range of 0.91 to 0.96. This research can be developed for several different objects in one image (multiple object). So, it is necessary to search for deeper datasets for further research development.

#### References

- [1] Dechter R 1986 Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory
- [2] Aizenberg I, Aizenberg N N, and Vandewalle J P 2000 *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*. Springer Science & Business Media
- [3] Ludovic A, Sébastien R, Sylvain C, and Hélène P 2011 An Introduction to Deep Learning.

- European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium
- [4] Girshick R, Donahue J, Darrell T, and Malik J 2014 Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*
  - [5] He K, Gkioxari G, Dollár P, and Girshick R 2018 Mask R-CNN. arXiv:1703.06870v3 [cs.CV]
  - [6] Rosebrock A 2018 Deep Learning For Computer Vision With Python. Pyimagesearch.
  - [7] He K, Zhang X, Ren S, and Sun J 2016 Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770–778.
  - [8] Abdi A and Kasaei S 2020 Panoramic Dental X-rays With Segmented Mandibles. Mendeley Data, v2. DOI: 10.17632/hxt48yk462.2.