



Object Detection on Dental X-ray Images Using Region Based Convolutional Neural Networks

Rakib Hossen, Minhazul Arefin and Mohammed Nasir Uddin

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 21, 2022

Object Detection on Dental X-ray Images using Region Based Convolutional Neural Networks

Rakib Hossen, Minhazul Arefin  and Mohammed Nasir Uddin

Abstract In dentistry, Dental X-ray systems help dentists by showing the basic structure of tooth bones to detect various kinds of dental problems. However, depending only on dentists can sometimes impede treatment since identifying things in X-ray pictures requires human effort, experience, and time, which can lead to delays in the process. In image classification, segmentation, object identification, and machine translation, recent improvements in deep learning have been effective. Deep learning may be used in X-ray systems to detect objects. Radiology and pathology have benefited greatly from the use of deep convolutional neural networks, which are a fast-growing new area of a medical study. Deep learning techniques for the identification of objects in dental X-ray systems are the focus of this study. As part of the study, Deep Neural Network algorithms were evaluated for their ability to identify dental cavities and a root canal on periapical radiographs. We used tensor flow packages to detect dental caries and root canals in X-rays. This method used faster R-CNN technology. For this reason, the proposed method is accurate 83.45% which is 10% greater than previous research.

Key words: Deep Learning . Faster R-CNN . Convolutional Neural Network . Dental X-ray Image

Rakib Hossen
Jagannath University, Dhaka, Bangladesh e-mail: rakibcse5003@gmail.com

Minhazul Arefin*
Jagannath University, Dhaka, Bangladesh e-mail: minhazularefin21@gmail.com

Mohammed Nasir Uddin
Jagannath University, Dhaka, Bangladesh e-mail: nasir@cse.jnu.ac.bd

1 Introduction

In recent decades, dentistry has improved. However, as dentistry has grown, so has the number of people with dental issues. Dentists are often required to see a large number of patients in a single day. A significant number of dental X-ray [1] films are taken every day as an important diagnostic tool for assisting dentists. A majority of dentists perform film reading work, which takes up important clinical time and can lead to mis- or under-diagnosis because of personal aspects like fatigue, emotions and poor skill levels. Intuitive dental disease detection technologies [2] may minimize the labor load of dentists and the incidence of misdiagnoses, therefore improving the quality of dental care. For smart health care in this context, automated object recognition in dental x-ray systems is a critical duty to be completed on time. It is a form of structural deterioration that causes teeth to develop cavities or holes. Cavities, on the other hand, are the consequence of tooth decay, which is caused by bacterial infections. Root canal is being due to inflammation or infection or external fracture in the roots of a tooth. They are common infectious oral diseases [3]. Many individuals, from teenagers to adults suffer from caries and root canal which leads to severe pain for a lifetime and even to tooth loss. Further, the severe problem may lead to oral cancer. In 2016, the WHO estimated that about 50% of the world population [4] is affected by dental caries and in some Asian-Pacific countries, the incidence of oral cancer is within the top 3 of all cancers. Caries and a root canal can be difficult to detect and diagnosis due to internal damage confused lesions that can't be identified with naked eyes. So far, for caries and root canal detection, several image segmentation and classification techniques such as level set method, morphological image processing, machine learning (ML) based classification methods have been improved. But they couldn't achieve any significant results.

Dental caries and root canals have been detected using a variety of techniques in recent decades. As a result of tooth anatomy and restoration shapes, there hasn't been a significant improvement. Particularly when deep fissures, strong interproximal contacts, dark areas, and subsequent lesions are present, identification becomes difficult. With the help of standard machine learning and image processing techniques, these methods concentrated on extracting hand-crafted features, building effective classifiers for the identification and recognition of objects and people. Such techniques are restricted by the fact that they typically require medical specialists to create useful features since vital visual characteristics for clinical choices [5] may be combined with irrelevant pixels in these approaches. Aside from that, each component of the detection pipeline is improved independently, resulting in a sub-optimal detection pipeline overall. As a side note, these characteristics are typically restricted and useless, resulting in incorrect ROI and edge detection. That means that both expert systems and medical science will gain from this research. Using this framework does not require a lot of experience on the part of the end-users.

2 Background Studies

Most teens and adults in the globe suffer from dental caries, a persistent infectious oral illness. For the identification of dental caries and root canals, a number of studies have been conducted. These attempts to identify tooth cavities and root canals on the basis of dental X-ray images [6] are discussed here. It was proposed by Rad et al. [7] that a method of automated segmentation and feature extraction be used for the detection of caries in dental radiographs. A wide range of image processing techniques were employed by the author, including picture enhancement and segmentation [8] as well as feature extraction, detection and classification. To improve the contrast of the structure of interest, picture enhancement methods are utilized. A k-mean clustering technique is then used to segment the picture in order to identify any errors. Individual tooth areas have been separated using an integrated projection method. If there are caries in the area of interest (ROI), characteristics such as entropy, intensity, mean and energy are retrieved from the ROI and compared to nearby areas. For caries detection, dental experts analyze the characteristics of the caries region. As a result of this method's effectiveness, segmentation is often challenging owing to changes in intensity area, and tooth extraction issues might lead to incorrect results.

To help in the diagnosis of proximal dental caries in periapical pictures, Choi and colleagues [9] developed an automated detection method. Proximal dental caries were investigated in regions near the crown borders, according to the authors. The crown areas were initially divided using a level set approach in order to identify them. Second, the authors used the identified crown regions to filter out areas outside of crowns. Then, crown borders were used to reduce the risk of dental pulp cavities. To decrease computing expenses, the scientists scaled all of the experimental pictures to the same size. Five folds were randomly assigned to the pictures. Four folds were used to train the CNN, and one fold was used to test the CNN. Test folds are altered one at a time to ensure that all periapical pictures have been checked. HAPT (horizontal alignment of photographed teeth), probability map creation, crown extraction, and refining are the four components of the system.

An improved caries detection system was suggested by Rad et al. [10] using a novel segmentation approach and detection methodology. It consists of three primary phases: pre-processing, segmentation, and analysis, which are described in detail below. Initial contour creation and intelligent level set segmentation finish the segmentation process. In this technique, features are extracted manually, which might lead to incorrect IC synthesis. Despite the author's attention on the segmentation process, a sufficient amount of research has not been done to identify caries. The BPNN also suffers from local minima, therefore this technique does not give good convergence. As the learning rate must be set carefully, it is possible that the detection of caries will be inaccurate. In this technique, features are extracted manually, which might lead to incorrect IC synthesis. Despite the author's attention on the segmentation process, a sufficient amount of research has not been done to identify caries.

The BPNN also suffers from local minima, therefore this technique does not give good convergence. As the learning rate must be set carefully, it is possible that the detection of caries will be inaccurate. According to Lee and colleagues [11] deep CNN algorithms may be used to identify and diagnose dental caries on periapical radiographs. Rotation, width and height shifting, zooming, shearing, and horizontal flipping were used to augment the training dataset ten times at random. The datasets were preprocessed before being trained with transfer learning using a GoogleNet Inception v3 CNN network. The v3 architecture's inception has learned roughly 1.28 million photos with 1000 object categories. To extract distinct scale properties, a convolutional filter with 22 deep layers can be utilized. It is possible to achieve a variety of scale features by combining convolutional filters of different sizes inside the same layer. The 9 inception modules used included an auxiliary classifier, two fully-connected layers, and softmax functions. With a learning rate of 0.01 and 1000 epochs of training, the dataset was randomly partitioned into 32 batches for each epoch. To improve the detection of dental caries, fine tuning was done to optimize weights and increase output power by modifying hyperparameters. This method uses a sophisticated and expensive algorithm to calculate the sliding window for caries detection. Detecting premolar caries may be challenging as a result of this procedure.

On the basis of dental radiograph image categorization, Yang et al. [12] presented an automated root canal treatment quality evaluation technique. For the root canal filling therapy, the scientists employed an automated apical foreman-area recognition technique based on dental scans. Researchers in this study employed a labeled dataset of periapical radiography pictures obtained before and after therapy to develop their findings. Root canal filling was identified using image subtraction techniques at first. The ROIs — the apical area in image preprocessing — were extracted here. It was first necessary to discover feature points between the two images using the SIFT and SURF algorithms. The authors then used the minimal grayscale difference approach to get the best ternary matching spots. These best ternary points were then used to construct an affine matrix, which was then applied to images in order to do an image subtraction. As a result of this, the authors were able to determine the apical foreman and its surrounding region using the filling area. The authors then input this portion of the pictures to the CNN training software in order to train their classification models on this component of the dataset. The authors employed six layers of deep neural network to tackle the over fitting problem. Two convolution layers were added to the Inception structure in order to minimize the amount of parameters while maintaining a high level of performance. Because ROI detection is so imprecise, this research has a major flaw. There are a few CNN layers used in this technique, which limits the performance. A similar method does not function well with molar pictures. The discrepancies between certain pictures are so great that even manual calibration, which relies solely on the affine matrix, does not work effectively. As a result, new ways for dealing with these picture pairings are needed.

3 Proposed Methodology

This study technique follows the comparable Faster RCNN, a state of the art deep learning system for object detection. This research method follows a similar deep learning framework. The first element is a Regional Proposal Network (RPN) that generates a list of areas that are likely to contain objects, or ROIs; the second part is a Fast RCNN that categorizes a section of the picture as objects (and background) and reinforces their borders. The convolution layers used for feature extraction in both components have identical values, allowing object detection tasks to be completed at a reasonable speed. This research paper proposed a new model for object detection that uses a deep learning technique. It is based on the Faster Region-based Convolution Neural Networks (Faster R-CNN), made up of the Regional Proposal Network (RPN) & (Fast R-CNN). We will describe the process flow with a high level view in the first place. The Figure 1 shows the top level view of the proposed model.

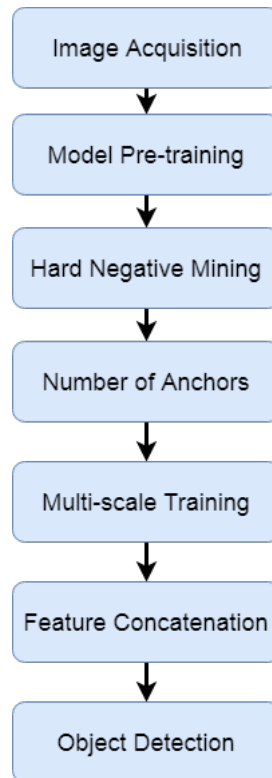


Fig. 1: Proposed Methodology of object detection on dental X-ray images using R-CNN

3.1 Image Acquisition

The Digital Dental Periapical X-Ray Dataset was utilized to evaluate the suggested model. The primary objective of this research is to recognize and mark each pixel for the item category designated (e.g implant & root canal etc.). Over the years, the Computer Vision Community has produced several benchmark data sets in order to evaluate academics and enthusiasts' models. A small segment of the dataset is shown in Figure 2.



Fig. 2: Segment of Digital Dental Periapical X-Ray Dataset

3.2 Model pre-training

We have decided to fine tune a pre-trained *faster_rcnn_inception_v2* model using a COCO dataset to adapt the Faster RCNN to object detection. The Digital Dental Periapical X-Ray Database is a widely known dataset featuring unconstrained objects. However, it may not be a good choice to merely refine this data set, as it is a very tiny dataset with just 2 items in 120 pictures. We have pre-trained our model in a COCO dataset, which is a lot bigger object dataset with many more challenging instances, before we finalize the digital dental periapical X-ray database. In order to manage those tough instances that might interrupt the process of convergence,

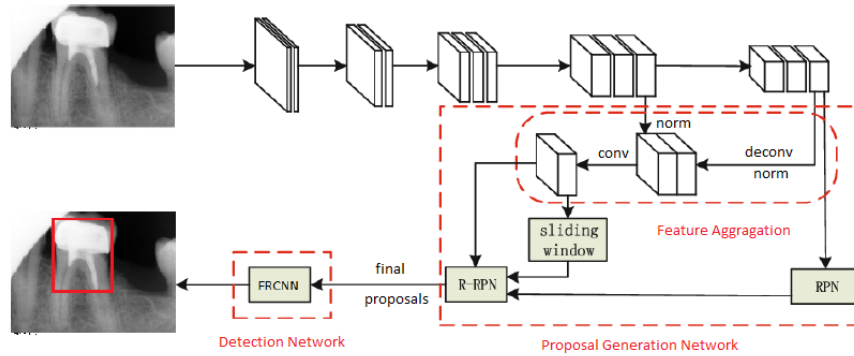


Fig. 3: Architecture of Faster RCNN

attention should be paid to the discovery of certain training data where the experimental section contains information. Hard negative mining is also required in this dataset before training in order to limit the amount of false positive results created. This technique shows the intricate architecture in the Figure 3.

3.3 Hard negative mining

Hard negative mining has been demonstrated as an efficient technique to increase the performance of deep education, particularly for the identification of objects. The concept behind this procedure is that the places in which the network has failed to forecast correctly are harsh negative. Hard negatives are therefore sent back into the network to enhance our trained model. The ensuing training procedure can then enhance our model to reduce false positive effects and increase classification performance. When the classifier incorrectly classifies a negative example, the model will automatically identify it as a negative example and place it in the training set for retraining. This phase is akin to having the improper set of questions. If the same question is answered incorrectly many times, the student is incorrect. During the first stage of our training process, hard negatives are extracted from the pre-trained model. Then we regard a region to be hardly negative, if it is less than 0.5 at its intersection between the union (IoU). We intentionally add these harsh negatives to RoI's throughout the hard negative training process in order to finalize the model. It also complete the ratio of background and foreground near about 1:3. It is the same ratio we used in the first step.

3.4 Number of anchors

In this research, the Faster RCNN architecture has set up numerous critical hyperfactors, in which the most important of these parameters seems to be the number of anchors in the RPN portion. Usually Faster RCNN utilizes 9 anchors that are able to remember tiny things. Smaller items such as implants seem to be quite frequent in object detection tasks. In the faster R-CNN we uses two conditions to assign a positive label to an anchor.

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

A mini-batch has an anchor i , and p_i is a projected probability that anchor i will be a real item. We assume the anchor in the figure is positive. For that reason, the ground-truth label p_i^* is 1, else it's 0. There are 4 parameterized coordinates in t_i , and t_i^* is the ground-truth bounding box. It is associated with a positive anchor in t_i . L_{cls} is a logarithm base loss across two classes. For the regression loss, we use $L_{reg}(t_i, t_i^*) = R(t_i, t_i^*)$ where, R is denoted as the loss function (smooth L_1). A regression loss $L_{reg}(t_i, t_i^*) = R(t_i, t_i^*)$ is only enabled for positive anchors ($p_i^* = 1$) and deactivated for negative anchors ($p_i^* = 0$). As a result, the outputs of cls and reg are p_i and t_i , respectively.

3.5 Multi-scale training

Technique for all the training pictures, the Faster RCNN architectures use a fixed scale. The detector can learn features over a broad variety of sizes, thereby enhancing its performance towards a level invariance, by scaling pictures to a random scale. We allocate one of three scales for each image randomly in this study prior to it being included to the network. Our experimental section provides details. Our empirical results demonstrate that our model is more resilient to diverse dimensions and improves the performance of detection on benchmarks. We compute the bounding box regression by parameterized the four co-ordinates of the detected object. It is shown in equation 2:

$$\begin{aligned} t_p &= (p - p_a)/r_a, & t_q &= (q - q_a)/s_a \\ t_r &= \log(r/r_a), & t_s &= \log(s/s_a) \\ t_p^* &= (p^* - p_a)/r_a, & t_q^* &= (q^* - q_a)/s_a \\ t_r^* &= \log(r^*/r_a), & t_s^* &= \log(s^*/s_a) \end{aligned} \quad (2)$$

where, p, q represents the width and r, s represents the height of the box. Variable p represents the predicted box. Variable p_a denotes the anchor box. Variable p^* is for ground truth box.

3.6 Feature Concatenation

The RoI pooling on the final functional map layer is used for the classic Fast RCNN networks to create regional characteristics that are subsequently examined by the classification component of your network. With this innovative architecture, the classification network may use the characteristics computed by RPN to save many needless calculations. This method is however not always ideal and occasionally certain crucial characteristics may be missed since features of the deeper convolution layer output have broader receptive fields and are more granular. The reverse function computed the partial derivative of the loss function. It worked in the RoI pooling layer. For calculating it works with respect of each input variable p_i . Using argmax quation we find that in equation 3:

$$\frac{\delta L}{\delta p_i} = \sum_r \sum_j [i = i^*(r, j)] \frac{\delta L}{\delta q_{rj}} \quad (3)$$

Here, r is denoted as each mini-batch RoI and y_{rj} represents each pooling output unit. The partial derivative $\frac{\delta L}{\delta q_{rj}}$ is accumulated if i is the argmax selected for q_{rj} by max pooling. Because there is no concept of a ground-truth in the bounding box for background RoIs, L_{loc} is ignored. We utilize the loss, for bounding-box regression equation 4:

$$L_{loc}(t^u, v) = \sum_{i \in p, q, r, s} smooth(t_i^u - v_i) \quad (4)$$

in which,

$$smooth_{L_1}(p) = \begin{cases} 0.5p_2 & \text{if } |p| < 1 \\ |p| - 0.5 & \text{otherwise,} \end{cases} \quad (5)$$

Here, L_1 is used, cause it is sensitive in the image outliers and L_2 used in R-CNN. They both are loss function. In this research, we want to enhance the RoI pooling by merging many convolution-layered characteristic maps, which include both low- and high-level functions, to capture more finely-grained RoI information. In order to get the final bonding features for detection tasks, we propose to concatenate the output of several convolution feature maps. In other words, combined with the final feature map in RPN, we use certain intermediate findings, which we combine to produce the final pooling functions. Specifically, features are RoI-pooled and L2-normalized from several lower-level convolution layers accordingly. These characteristics are then concatenated and rescaled as if the original scale of the features had not been adopted. A 1×1 convolution is done to match the original network's number of channels. This technique shows the intricate architecture in the Figure 4.

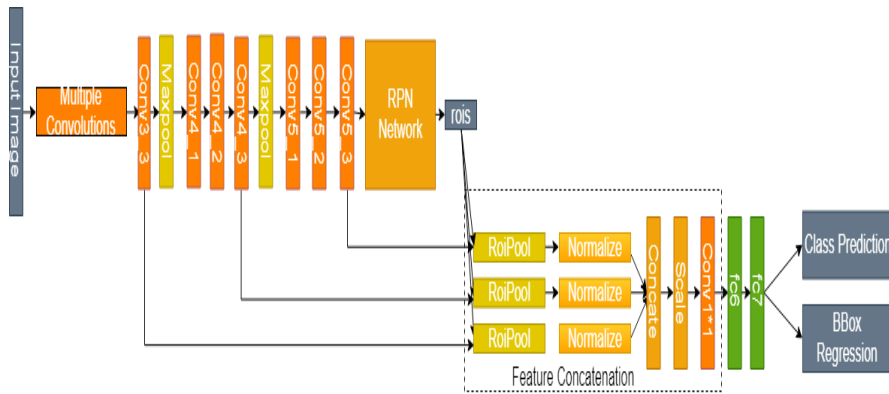


Fig. 4: Architecture of Feature Concatenation

3.7 Object Detection

Object detection is the last step of this process where usually we assign some levels to an object based on their features. In the last step of this methodology, using softmax classifier this method can finally detect objects in dental x-rays image successfully which is shown in Figure 5.

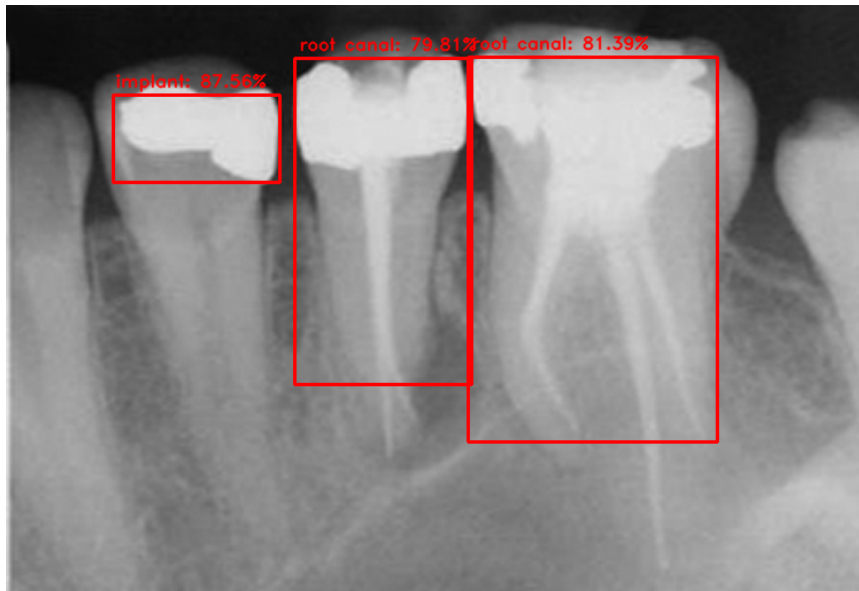


Fig. 5: Detected object on dental x-ray image

4 Result

In order to detect objects, there are a number of metrics that are variations of intersection over union (IoU). The overlap between two borders is measured by the IoU. The overlap between our anticipated border and the ground reality is then calculated (the real object boundary).

Let $n_{i,j}$ be the number of pixels of class i predicted to belong to class j , where there are n_{cl} different classes, and let $t_i = \sum_j n_{ij}$ be the total number of pixels of class i . Mean IoU is defined as:

$$Mean_{IoU} = \left(\frac{1}{n_{cl}}\right) \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (6)$$

To determine if an item is present in the image, the mean average precision (mAP) is utilized as a measure of accuracy. It is important to know the accuracy and recall of a classifier in order to comprehend the AP. To train RPNs, we assign each anchor a binary class label (whether it is an object or not). Two types of anchors are given a good designation. Intersection over Union (IoU) anchors having the highest overlap with a ground truth box, or any anchor with an IoU overlap greater than 0.7 with a ground truth box, are highlighted. Assigning positive labels to numerous anchors is possible with a single ground-truth box. All ground truth boxes with an IoU ratio less than 0.3 will be labeled as negative. Following the multi-task loss in Fast R-CNN, we minimize an objective function with these definitions in place.

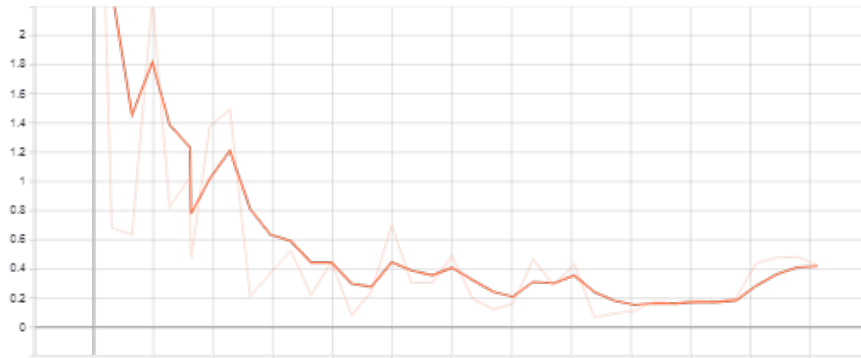


Fig. 6: Total Loss function per epoch (in thousand)

The result shows that the proposed methodology can find out 83.45% accurate dental object detection with an inaccuracy of 16.55%. The results comparing the proposed method with three others are shown in Table 1. It shows that the mAP of RCNN is better than the earlier results. If we increase the size of the training data, more mAP can be obtained. The existence of a large volume of data is always useful for the RCNN model.

Table 1: Comparison with existing models

Models	mAP(Digital Dental Periapical X-Ray Database)
RFCN Resnet 101	68.3%
GoogleNet Inception V3	55.01%
SSD Inception V1	73.56%
Our study	83.45%

5 Conclusion

As well as reducing the expense of oral health care, accurate identification of dental decay and root canals improves the chance of natural tooth preservation in the long run. In this research, deep learning techniques are used to recognize and classify dental X-ray items. In order to recognize generic objects, we have utilized the Faster RCNN framework which features feature concatenation, multi-scale training, hard negative mining, and correct tuning of anchor sizes for RPN, amongst other things. Because this framework combines a variety of approaches, it is able to overcome many of the limitations of single methods. Both dental cavities and root canals have been detected. As a result of the classification accuracy, our technique has been proven to be reliable and efficient. When training and testing were done on the dataset, classification accuracy increased. 83.45% of caries and root canals were detected using our technique, which is better than any other method. A better local minimum is achieved with this strategy compared to a conventional random initialization of network weight. Our approach may be utilized to detect dental caries and root canals from dental X-rays, therefore in conclusion, we can conclude that it works. In future we will try to use a better deep learning method with multiple dataset.

References

1. M. M. Lakshmi and P. Chitra, "Classification of Dental Cavities from X-ray images using Deep CNN algorithm," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), 2020.
2. J.H. Lee, D.H. Kim, S.N. Jeong and S.H. Choi, "Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm", Journal of periodontal implant science, vol. 48, no. 2, pp. 114-123, 2018.
3. A.A. Al Kheraif, A.A. Wahba and H. Fouad, "Detection of dental diseases from radiographic 2d dental image using hybrid graph-cut technique and convolutional neural network", Measurement, vol. 146, pp. 333-342, 2019.
4. Oral health, <https://www.who.int/news-room/fact-sheets/detail/oral-health>. Last accessed 19 Aug 2021.
5. J. Hatvani, A. Horváth, J. Michetti, A. Basarab, D. Kouamé and M. Gyöngy, "Deep learning-based super-resolution applied to dental computed tomography", IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 2, pp. 120-128, 2018.

6. G. Silva, L. Oliveira and M. Python, "Automatic segmenting teeth in X-ray images: Trends a novel data set benchmarking and future perspectives", *Expert Systems with Applications*, vol. 107, pp. 15-31, 2018.
7. A. E. Rad, I. B. M. Amin, M. S. M. Rahim, and H. Kolivand, "Computer-aided dental caries detection system from x-ray images," *Computational Intelligence in Information Systems*. Springer, 2015, pp. 233–243.
8. M. Ali, M. Khan and N.T. Tung, "Segmentation of dental Xray images in medical imaging using neutrosophic orthogonal matrices", *Expert Systems with Applications*, vol. 91, pp. 434-441, 2018.
9. J. Choi, H. Eun, and C. Kim, "Boosting proximal dental caries detection via combination of variational methods and convolutional neural network," *Journal of Signal Processing Systems*, vol. 90, no. 1, pp. 87–97, 2018.
10. A. E. Rad, M. S. M. Rahim, H. Kolivand, and A. Norouzi, "Automatic computer-aided caries detection from dental x-ray images using intelligent level set," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28 843–28 862, 2018.
11. Lee, Jae-Hong, et al. "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm." *Journal of dentistry* 77 (2018): 106-111.
12. Yang, Y. Xie, L. Liu, B. Xia, Z. Cao, and C. Guo, "Automated dental image analysis by deep learning on small dataset," *IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. IEEE, 2018, pp. 492–497.