Tooth detection with small panoramic radiograph images datasets and Faster RCNN model

Milan Zdravković*, Zoran Pešić**, Pavle Pešić***
* Faculty of Mechanical Engineering, University of Niš, Serbia
** Medical Faculty, University of Niš, Department of Maxillofacial Surgery, Serbia
*** Clinical Center Niš, Department of Radiology, Serbia
milan.zdravkovic@gmail.com, pesic.z@gmail.com

***Abstract*— This paper deals with tooth detection in realistic situations, characterized with relatively small number (in this case, 114 panoramic radiographs) of images for training object detection models, with differing qualities (contrasts, lightness) and sizes. The objective of the work is to define the pipeline and framework of decisions in x-ray images object detection problems. Faster RCNN architecture is used for detection; the approach includes transfer learning, augmentation and normalization (Contrast Limited Adaptive Histogram Equalization) of x-ray images. Resulting model performance is comparable with or exceeding the state of the art in the field, with mean Average Precision for the test set of mAP=0.96-0.97.**

I. INTRODUCTION

Artificial Intelligence (AI) has become an integral part of radiology practice in the last two decades. Four main tasks of the AI-driven computer vision are classification, localization, object detection, and object segmentation.

In dentistry, panoramic radiography is the most common diagnosis technique because of its low cost, simplicity, informative content, and the reduced exposure of the patient. Object detection is used in dentistry for detection of teeth, caries, filled crown, prosthesis, dental implants, vertical root fractures, jaw fractures, unerupted teeth, endodontic treatment, periodontal bone loss, apical lesions, osteosarcoma, ameloblastoma, osteodystrophy and systemic bone diseases. Objects are detected from panoramic, periapical or bitewing radiograph images. Some other applications found in literature are bone healing analysis, diagnosis of osteoporosis, dental forensics, automatic segmentation of mandible in panoramic x-ray, dental implants placement, orthodontic assessment (mispositioned teeth and jaws), identification and filtering of x-ray artifacts.

Today, Region-based Convolutional Neural Network (R-CNN) is the first choice for image object detection problems [1]. Initially, R-CNN architecture consisted of three models:

- Region proposal - generation and extraction of region of interest, rectangle frame for possible object locations, independent of the class (Selective Search method for object recognition [2] was used for extraction).
- Proposed regions are then passed to a feature extractor - extraction of features from each of the individual regions (pre-trained AlexNet deep CNN was initially used),
- Classifier (Support Vector Machine - SVM).

Although successful, R-CNN model was slow because feature extraction is carried out from each of the regions of interest. On 2015, Girshick proposed Fast R-CNN architecture [3]. This model was integrated. For feature extraction, VGG-16 was used, after which RoI pooling layer is positioned to crop the parts of the feature map and resize those. Then, neural network follows with two outputs: class prediction and bounding box. A year later, Ren et al [4] proposed Faster R-CNN model. This model, during training forms region proposals (Regional Proposal Network - RPN). Thus, model consists of two modules: RPN and Fast R-CNN. The latest model was Mask R-CNN [5]. It builds upon Faster R-CNN by adding output model for image segmentation. The network backbone (feature extractor and classifier) is a Resnet101 architecture.

Before Region-based CNN introduced integration of localization of objects of interest in the image, different localization methods were combined with image classification algorithms in dentistry, with different success.

Oktay [6] used modified AlexNet CNN architecture to achieve "higher than 90%" accuracy. Nung-Hsiang et al [7] achieved up to 96% accuracy in teeth detection. Tuzoff et al [8] used Faster RCNN architecture with pre-trained ImageNet weights and VGG16 feature extraction blocks for teeth detection in panoramic radiographs. Training was carried out with 1352 and 222 images were used for testing (all images from a single machine). Precision was 0.9945. Up to date information on use of deep learning models in dentistry can be found in several survey papers which provided additional insight into the topic [9][10].

The general objective of the work behind this paper is to setup a pipeline for object detection in x-ray images and establish a framework of decisions and optional alternatives for achieving the best possible accuracies equal to or near the state of the art. The problem to be dealt with is a tooth detection from relatively small number of images of differing qualities and sizes.

II. METHODOLOGY

Open dataset consisting of anonymized and deidentified panoramic dental X-rays of 116 patients, taken at Noor Medical Imaging Center, Qom, Iran was used in this research [11]. 3314 teeth were annotated in images by the authors. Annotated image dataset (Pascal VOC format) consists of images and XML files generated for each of the image in the dataset. XML files store coordinates of

the bounding boxes/rectangles and annotations. 80% of all data will be used for training, where out of this number, 80% will be a training set, while remaining 20% will be used as validation set for backpropagation. 20% of data will be used only for evaluation, once the training is completed.

Some fixed decisions are made based on the existing experience in the object detection domain. First, Faster RCNN model is used (its Python Keras Mask RCNN implementation) as it is top of the shelf model today. Second, transfer learning based on the existing CoCo model will be used in convolutional layers for more efficient feature engineering, while only RPN, classifier and mask layers will be trained with data. Transfer learning helps to build efficient models with less training (less annotated data).

Following ranges of decisions will be tested during the experiments: different radiograph contrast and lightness; effect of re-training; different augmentation approaches; different learning rates. Normally, image data will be of variable size and quality; the latter one is especially important for grayscale x-ray images as they may have different contrasts and lightness and they often suffer from the occurrence of so-called artifacts (for example, abnormal shadows, degraded image quality, etc.). Second, retraining is sometimes carried out with goal to train convolutional layers to better highlight different geometric features occurring in the objects of interest; the effect of retraining will be investigated. Third, augmentation technique is often used to compensate for insufficient size of the training set, which is quite common for medical image recognition and object detection field; will augmentation improve the capability of the model to generalize and detection precision? Finally, different learning rates will be tested for the best accuracy.

The accuracy of the prediction can be evaluated based on how well the predicted and actual bounding boxes overlap. Accuracy is calculated by dividing the intersection area (overlap) by the total area of both bounding boxes. This is often referred to as "intersection over union" or IoU. A perfect bounding box prediction will have an IoU of 1. It is common to assume that a positive prediction of a bounding box is made if the IoU is greater than 0.5, e.g. they overlap by 50% or more.

Each prediction of the bounding box is associated with the confidence score - the probability that the box contains an object. In order to have a prediction, that probability needs to be greater than some threshold. All this is performed by the classifier.

Precision refers to the percentage of the correctly predicted bounding boxes (IoU > 0.5) out of all bounding boxes predicted in one image. Recall is the percentage of the correctly detected objects (predicted bounding boxes, IoU > 0.5) out of all objects in the image.

Average precision metric (AP) [12] is the most used metric for object detection problems. In short, AP is the area under the precision-recall curve, where this curve shows the combined precision and recall values for different IoU thresholds. The performance of a model for an object detection task is typically evaluated using the mean absolute precision, or mAP. The average or mean of the average precision (AP) across all the images in a dataset is called the mean average precision, or mAP.

III. Solution and Discussion

The project uses modified Keras implementation of Mask R-CNN model by Abdulla [13].

In the preparation phase, one issue to handle before training was variability in gray level and contrast, due to use of different scanners. This variability may make it difficult to a model to learn right features. Some sort of normalization is needed to obtain images with similar contrasts. CLAHE (Contrast Limited Adaptive Histogram Equalization) is a histogram equalization technique that allows to enhance contrast locally while limiting the amplification of noise.

The model is fit on the training dataset with passed training and validation datasets, specified learning rate parameter and number of epochs. Only head layers of Mask R-CNN model are trained. Feature engineering layers use existing weights while localizer and classifier are trained to detect objects of interests for this project (teeth).

Loss plots for the initially trained models are presented on Figure 1. They provide indications on speed of convergence over epochs, whether the model has converged at some epoch (plateau) or whether the model is over-fitting after some epoch (increasing validation loss after some epoch).

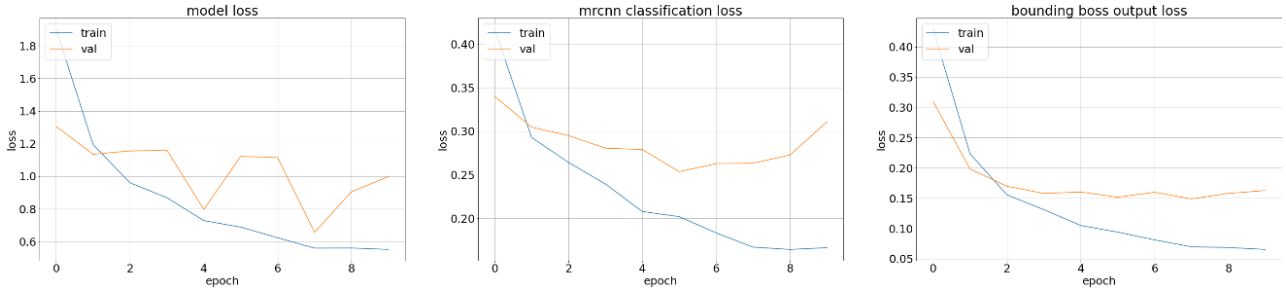

Figure 1. Loss over training epochs

The plots provide indications on speed of convergence over epochs, whether the model has converged at some epoch (plateau) or whether the model is over-fitting after some epoch (increasing validation loss after some epoch). Different train and test scores are reported during the training. When the problem addressed is object detection (and not segmentation), one should look at loss for the classification output on the train and validation datasets (mrcnn_class_loss and val_mrcnn_class_loss) and loss for

the bounding box output for the train and validation datasets (mrcnn_bbox_loss and val_mrcnn_bbox_loss).

Augmentation procedure was implemented by the imgaug library, used by Mask R-CNN model. Library converts a set of input images into a new, much larger set of slightly altered images. Alterations include Gaussian noise (contrast, sharpen), affine, cropping and padding, flipping, perspective and others. Only horizontal flipping was tested in augmentation step. Lightness is addressed by images normalization. Vertical flipping may confuse the model as lower and upper teeth looks different on x-ray due to different occurring artifacts in different regions of the image.

Several experiments were made with partial and full dataset. Some of the important conclusions were as follows:

- In general, conservative learning rate of LR=0.01 works better than the lower values, indicating possible overfitting in latter cases;
- Applying adaptive equalization (CLAHE) brings significant improvements. This is considered as a key for successful object detection model in this, realistic case (images from different machines, with different qualities, contrasts, lightness);
- Augmentation produces improvements in majority of experiments but not always. It is not possible to establish the rules to be followed. Further work is needed to find the effective configuration of the augmentation rules;
- Model fine tuning (retraining) does not improve the precision of object detection;
- Training with 5 epochs is sufficient for satisfactory results. After that, model starts to overfit (classification and bounding box losses increase).
- Model achieves the accuracies (mean Average Precisions) in the area of 0.95-0.97 which is considered high when comparing with the state of the art.

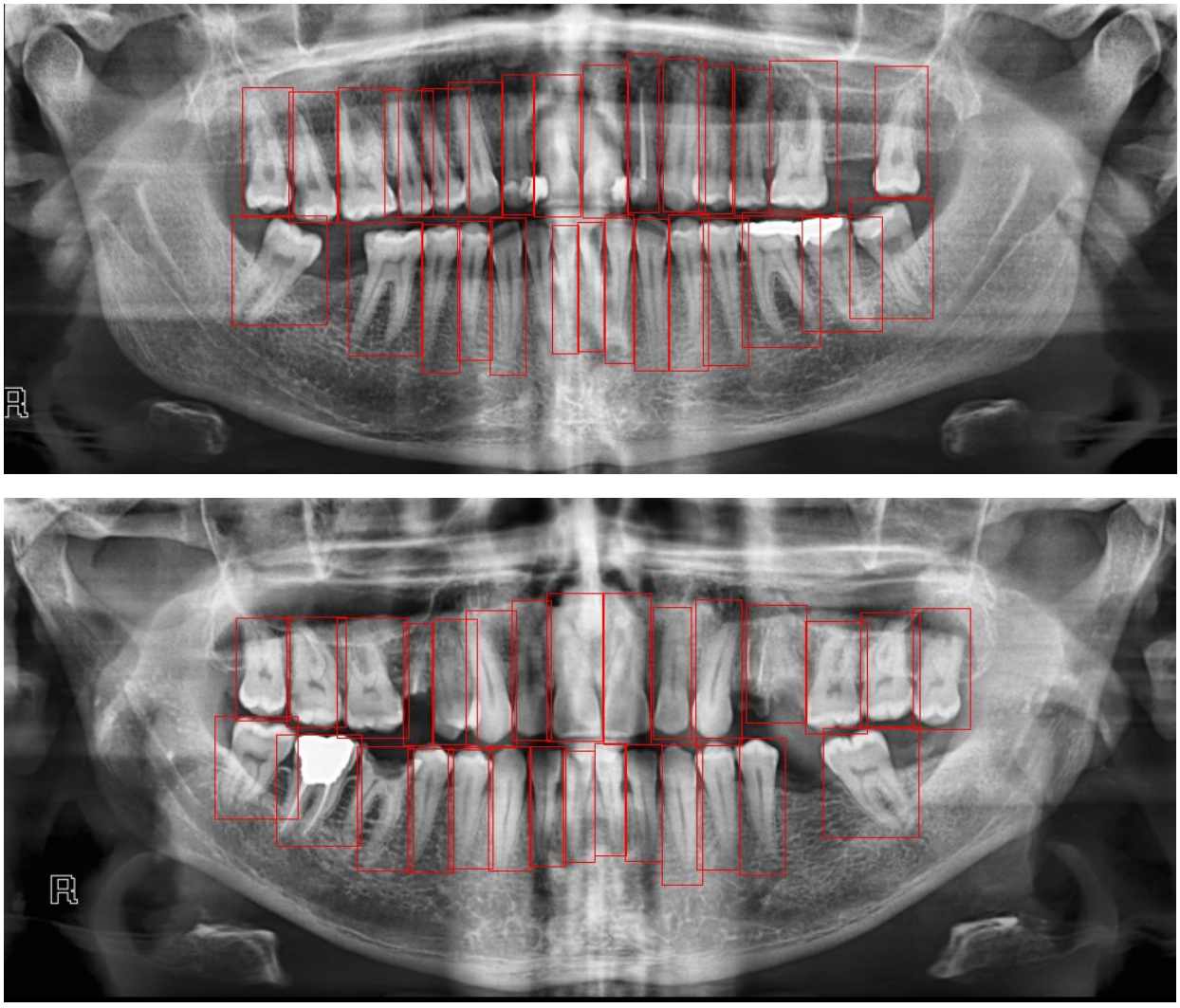

Figure 2. Examples of predicted bounding boxes for two radiograph images

Figure 2 above displays two examples of radiograph images with bounding boxes predicted with the implemented algorithm.

IV. CONCLUSIONS

With the most recent advances in computer vision algorithms and their implementations in popular programming languages, such as Python, processing of medical images is today becoming easier than ever, with excellent results, comparable to the performance of the humans. In this paper, we presented the process in which object detection problem over medical images is addressed. This process is characterized by the pipeline and a set of hypotheses to be tested in the set of experiments aiming at confirming the best performing model.

The research has confirmed the effectiveness of off-the-shelf models (such as Faster R-CNN for object detection), combined with several methods that introduced significant improvements over the vanilla model (such as image normalization and transfer learning). Even without the optimization, the resulting model has produced mean Average Precisions (with the used test set) in the area of 0.95-0.97 which is considered as excellent when comparing with the current state of the art.

In the future work, it is our objective to further investigate the effects of image augmentation, with goal to design the effective method and to develop the optimization procedure. Our interest is to implement the model for detecting other important features, specifically caries, filled crowns, vertical root fractures, unerupted teeth and periodontal bone loss. Another objective is identification and filtering of the medical image artifacts.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-9/2021-14/200109)

REFERENCES

[1] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. https://arxiv.org/abs/1311.2524

[2] Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M. 2013. Selective Search for Object Recognition. International Journal of Computer Vision volume 104, pages154–171(2013)

[3] Girshick, R. (2015) Fast R-CNN. https://arxiv.org/abs/1504.08083

[4] Ren, S., He, K., Girshick, R., Sun, J. (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. https://arxiv.org/abs/1506.01497

[5] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2018) Mask R-CNN. https://arxiv.org/abs/1703.06870

[6] Oktay, A.B. (2017) Tooth detection with Convolutional Neural Networks. 2017 Medical Technologies National Congress (TIPTEKNO). 10.1109/TIPTEKNO.2017.8238075

[7] Nung-Hsiang, L., Ting-Lan, L., Xiaoyue, W., Wan-Ting, K., Hua-Wei, T., Shih-Lun, C., Yih-Shyh, L., Jocelyn Flores, V., Yu-Fang, K. (2018) Teeth Detection Algorithm and Teeth Condition Classification Based on Convolutional Neural Networks for Dental Panoramic Radiographs. Journal of Medical Imaging and Health Informatics. 8(3) 507-515(9). 10.1166/jmihi.2018.2354

[8] Tuzzof, D.V., Tuzova, L.N., Bornstein, M.M., Krasnov, A.S., Kharchenko, M.A., Nikolenko, S.I., Sveshnikov, M.M., Bednenko, G.B. (2019) Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofac Radiol. 48(4). https://www.birpublications.org/doi/full/10.1259/dmfr.20180051

[9] Schwendicke, F., Golla, T., Dreher, M., Krois, J. (2019) Convolutional neural networks for dental image diagnostics: A scoping review. Journal of Dentistry. 91(2019) 103226

[10] Hwang, J.J., Jung, Y.H., Heo, M.S. (2019) An overview of deep learning in the field of dentistry. Imaging Sci Dent. 49(1):1-7. DOI:10.5624/isd.2019.49.1.1

[11] Abdi, A.; Kasaei, S. (2020) Panoramic Dental X-rays With Segmented Mandibles, Mendeley Data, v2 http://dx.doi.org/10.17632/hxt48yk462.2

[12] Salton, G., McGill, M. J. (1983). Introduction to modern information retrieval. McGraw-Hill. Retrieved from https://dl.acm.org/citation.cfm?id=576628

[13] Abdulla, W. (2017) Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub. https://github.com/matterport/Mask_RCNN