Instance Segmentation and Labeling of Teeth from Dental X-Ray using Region Based Convolutional Neural Network

Sireesha Rodda[1,*] , Vaibhav Kovela , Sanjay Dokula
Department of CSE
GITAM Institute of Technology
GITAM (Deemed to be University), Visakhapatnam, India

* Corresponding Author: srodda@gitam.in

Abstract

Radiological Examination of teeth is a primary step that a dentist usually takes to diagnose the problem before further treatment. The diagnosis involves searching for diseases ranging from cavities to tumors, So, correct diagnosis is vital for timely and precise treatment. This paper attempts to solve one of the elementary steps in diagnosis i,e, Labeling of Teeth, using Region-Based Convolutional Neural Networks that help reduce monotonous work for a dentist and provide segments of each tooth for further diagnosis of diseases with the use of Mask R-CNN. We used 200 panoramic X-Ray images of 4 categories to train, test and validate the model. Mask R-CNN with pre-trained weights of COCO Dataset is employed. We further tuned the weights of the dental X-ray dataset considered in the paper for better performance. On testing the learned model, the performance measures were encouraging.

 Keywords: Panoramic X-Rays, Instance Segmentation, Mask R-CNN, Faster CNN, Dental Labeling.

1.INTRODUCTION

Dental Notation is vital in dentistry and one of the most frequently used norms for identifying teeth before further treatment. Tooth Notation is a means to locate the teeth and record the teeth-related information. From these notations, the dentists diagnose the disease and plan for treatment. This work uses **extraoral panoramic dental X-Rays or Orthopantomography Images** for labeling. Dentists most commonly use these kinds of x-rays for preoperative examinations and bone surgeries [1]. The Notation format for teeth is depicted in Figure 1. One of the significant challenges in the task is to handle the teeth gaps and irregularities, which is primarily addressed in this paper. A wide range of dental x-rays has been examined to address these irregularities and feed them into the training model. This paper aims to eliminate a significant disadvantage of manual notation, which is a time-consuming method and very arduous.

We use Mask R CNN Network [1], which improves the older Faster-R CNN[7] and provides much more spatial output to Segment the teeth with labels, masks and boundary boxes. The use of Mask-R CNN is not limited to segmenting and labeling the tooth but can also be used to detect many other dental conditions like caries, cavities, teeth reshaping, and many other issues.

The first process is to annotate the teeth with their corresponding numbering. The model can detect the image with greater accuracy with the quadrants considered and the teeth labeled according to them.
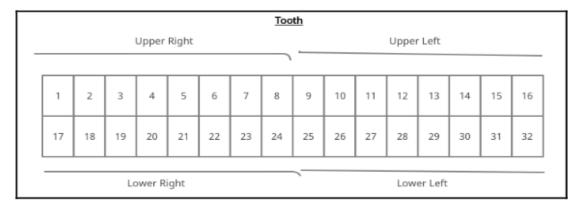

Figure 1: Dental Notation Format

2. RELATED WORK

Significant contributions have been made in this field of computational dentistry. Most contributions focus on the pixel-wise classification of images using unsupervised learning techniques or genetic algorithms. Moreover, various authors have tried various segmentation techniques to label or detect teeth edges in X-Rays. Our paper concentrates much on Teeth notation or Numbering the teeth and edge detection.

Jader et.al.[1] , proposed a way to instance segment teeth of various kinds of panoramic x-rays using the Mask-R CNN Network. This work proposes the detection of teeth and the production of masks—ten different kinds of x-rays where many attributes like dental implants and many others are considered. The Machine Learning model performs significantly better than the older approaches using unsupervised learning techniques.

Senthilkumaran[2] has given valuable insights as to why Genetic Algorithms are highly performant in a problem like teeth segmentation, which has a high search space and the ability of genetic algorithms always to find the global optimum, disregarding the noise or new input. This paper mainly focuses on edge detection by finding points of discontinuity in the image.

Lin et.al.[3] proposed a way to label and number the teeth using homomorphic filtering, homogeneity based contrast stretching and adaptive morphological transformation. Which isolate Regions of Interest (RoI) followed by contour extraction and then use Binary support vector machines to classify and thus to number them using simplified sequence alignment, which is famously used in the field of bioinformatics.

Ali et.al.[4] Introduced a faster way of instance segmenting dental images using parallel GPU instead of using the slower serial CPU's. This paper provides insight into the performance of GPUs in highly parallel processing compared to the multipurpose CPU. It uses a model that detects edges using curve evolution techniques based on the image's active contours[27].

Kaiming et.al.[5] proposed a much faster and scalable version of Faster-R CNN, for instance, segmentation of images. It is also better in detecting, resulting in higher accuracy. The additional overhead made over Faster-R CNN enables the Mask-R CNN to simultaneously detect masks with the boundary boxes detected by the former network. Replacement of ROIPool layer with ROIAlign layer helps in handling harsh quantisations made by ROIPool.

Yusra Y. et.al.[19] proposed a method to segment wisdom teeth accurately (the third molar) using various image morphological filter operators and dynamic length masks to diagnose wisdom teeth by classifying them as impacted, partially erupted, or completely erupted. A segmentation algorithm along with a preprocessing step is proposed in this paper.

Mahdi et.al.[11] worked on teeth recognition using Faster-R CNN, the predecessor of Mask-R CNN, which could take the instance of that class and produce boundary boxes for the instances—considering a Resnet50 as the backbone for this network. This paper is based on the principle of transfer learning.

Hui Gao et.al.[18] worked on a process to perform segmentation and 3D construction of individual teeth with complete crown and root parts. They proposed a single level set method for root segmentation and coupled level set method for crown segmentation to create virtual boundaries to separate touching teeth. A gradient direction is also introduced to the level set framework to avoid catching surrounding boundaries. The work is based on CT images.

Marques Lira et al. [20] have proposed a segmentation approach based on a supervised learning technique for texture recognition using some statistical methods for feature extraction and classification using Bayesian classifiers, which distinguish pixels into active (Teeth) and inactive (Other than teeth). Marques' is one of the earlier models of dental segmentation, which paved the way for newer segmentation approaches.

Poonsri et.al.[21] were the first to segment teeth using panoramic dental x-ray images automatically. They introduced a three-phase approach to automatic segmentation using otsu's thresholding and Mahalanobis distance technique; the second phase performs template matching with different image sizes. Finally, overlapping the matching templates is used to segment the teeth in the third phase.

3. METHODOLOGY

3.1 Mask R-CNN Architecture

This paper presents an instance-based segmentation method to classify dental images accurately.

Mask R CNN is used for this purpose. The segmentation process works in three phases. In the first phase, a backbone network, usually comprising a Convolutional Neural Network like ResNet50 or ResNet101, is employed to extract features from images. In the next phase of the architecture, an X-ray image is passed into a feature pyramid network which converts the image into various feature maps, which are then passed into regional proposal networks that generate multiple ROIs using binary classifiers. The ROI Align layer converts ROIs from the FPN into uniform-sized ROIs. The feature maps are processed for classification and regression in the final phase to generate Boundary boxes (bboxes) and Masks. Figure 1 depicts the architecture of the Mask R CNN network.
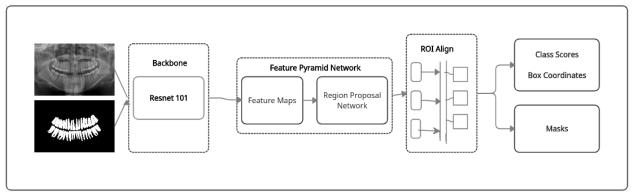
Figure 3: Mask-R CNN Architecture

3.2. Dataset Description

The dental X-ray dataset considered in this paper is sourced from ivision labs [26]. An expert manually performed annotations to associate the 200 dental x-ray images into a total of four categories. The dataset is then partitioned into a 100:30:70 ratio to perform training, validation, and testing, respectively. The various categories considered are depicted in Table 1. The notation for the numbering of teeth is followed, as shown in Figure 1.

Table 2 depicts the summary of all 32 teeth in the dataset being considered during training. It can be observed that some of the teeth along the edges do not have enough representation in the dental images.

Table 1: Types of X-Ray Images Considered

	Kind of X-Ray	**No. of Images Considered**
1	*X-Ray with all 32 Teeth Present*	70
2	*X-Ray with a nonuniform pattern of teeth*	65
3	*X-Ray with Gaps in the Molar and Premolar Regions*	40
4	*X-Ray with Gaps in the Canine and Incisor Regions*	25

Table 2:Cumulative Class Distribution of the Dataset

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Instances	114	118	121	126	144	147	156	172	162	153	134	115	132	124	120	105
Class	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Instances	105	113	120	124	136	148	164	185	179	164	157	139	116	110	103	95

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

All experiments were performed on Google Colab instance.with Intel ® Xeon Processor, 26 Gigabytes of memory and Nvidia Tesla P100-PCIe 16 Gigabytes of VRAM. Tensorflow version 1.5.2 and OpenCV 3.5 to avoid various dependency issues. The main Mask - R CNN headers are taken from matterplots' codebase. Some additional configuration changes were made to Mask-R CNN before training, and The changes were; Detection Confidence threshold was set to 95% to improve the chances of correct predictions; to efficiently use the GPU resources, the batch size was set to 4 images, which improves the GPU utilisation thus decreasing the time consumption. The steps per epoch are set to 500 to facilitate better training accuracy. Figure 4 depicts the full configuration of our custom Mask - R CNN model.

```
Configurations:
BACKBONE                        resnet101
BACKBONE_STRIDES                [4, 8, 16, 32, 64]
BATCH_SIZE                      4
BBOX_STD_DEV                    [0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE          None
DETECTION_MAX_INSTANCES         100
DETECTION_MIN_CONFIDENCE        0.9
DETECTION_NMS_THRESHOLD         0.3
FPN_CLASSIF_FC_LAYERS_SIZE      1024
GPU_COUNT                       1
GRADIENT_CLIP_NORM              5.0
IMAGES_PER_GPU                  4
IMAGE_CHANNEL_COUNT             3
IMAGE_MAX_DIM                   512
IMAGE_META_SIZE                 46
IMAGE_MIN_DIM                   512
IMAGE_MIN_SCALE                 0
IMAGE_RESIZE_MODE               square
IMAGE_SHAPE                     [512 512   3]
LEARNING_MOMENTUM               0.9
LEARNING_RATE                   0.001
LOSS_WEIGHTS                    {'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
MASK_POOL_SIZE                  14
MASK_SHAPE                      [28, 28]
MAX_GT_INSTANCES                100
MEAN_PIXEL                      [123.7 116.8 103.9]
MINI_MASK_SHAPE                 (56, 56)
NAME                            teeth_train
NUM_CLASSES                     34
POOL_SIZE                       7
POST_NMS_ROIS_INFERENCE         1000
POST_NMS_ROIS_TRAINING          2000
PRE_NMS_LIMIT                   6000
ROI_POSITIVE_RATIO              0.33
RPN_ANCHOR_RATIOS               [0.5, 1, 2]
RPN_ANCHOR_SCALES               (32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE               1
RPN_BBOX_STD_DEV                [0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD               0.7
RPN_TRAIN_ANCHORS_PER_IMAGE     256
STEPS_PER_EPOCH                 500
TOP_DOWN_PYRAMID_SIZE           256
TRAIN_BN                        False
TRAIN_ROIS_PER_IMAGE            200
USE_MINI_MASK                   True
USE_RPN_ROIS                    True
VALIDATION_STEPS                50
WEIGHT_DECAY                    0.0001
```

Figure 4: Mask-R CNN Custom Configuration

4.2 Evaluation Metrics.

The efficiency of the approach discussed in this paper is evaluated using various performance measures based on a confusion matrix, viz., comprising TP(True Positives), TN(True Negatives), FN(False Negatives), FP(False Positives), as shown in Table 3.

Table 3: Confusion Matrix

		Actual Result	
		Positive	Negative
Predicted	Positive	TP	FP
Result	Negative	FN	TN

To determine the model's performance and assess the model on various metrics, we used the following metrics :

***Precision*:** Ratio of Correctly Identified positive predictions and Total Positive predictions.

$$TP \div (TP + FP)$$

***Recall*:** Ratio of Correctly classified positive predictions and total correct predictions.

$$TP \div (TP + FN)$$

F-Score: Harmonic mean of precision and recall.

$$2 * \frac{Precision * Recall}{Precision + Recall}$$

***Average Precision (AP)*:** The average precision of all possible Intersection over Union (IoU) values for a class is known to be Average Precision. In this paper, the IoU threshold is considered to be 0.50. Fig. 4 shows the consideration of the IoU score where the ratio between the Intersection of Actual and Predicted Segment and Union of Actual and Predicted Segments are considered, which depicts the area of Actual Segment that our learning algorithm correctly predicted. So when the IoU is >0.5(threshold), the prediction is deemed favorable.

Mean Average Precision(mAP): Mean of Average Precision of all classes together is known as Mean Average Precision. Even though the AP values provide insight into the algorithm's performance over the range of classes, the mAP value determines the total performance of the whole problem set.
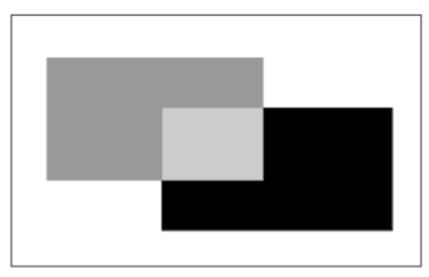
Figure 4: Intersection over Union Depiction

4.3 Results

A total of 148 images were annotated with the help of the makesense tool[28]. We have split the whole set into three parts i,e., 70 images for train, 30 for validation and 49 for the test. After Training the Mask R CNN for 15 epochs (500 steps per epoch), the learned model returned a decent performance. We used F-Score, Precision, Recall and Mean Average Precision(mAP) during the evaluation process. Fig. 2 corresponds to the graph of Precision per Class. Fig. 3 depicts the graph of recall and F-Score of model per class. Fig. 4 illustrates the graph of various Average Precision (APS) per class. Moreover, the Mean Average Precision(mAP) over all classes is 0.905.

Observing the results presented in Fig. 6 and Table 2, we could conclude that the model performs with low recall values and high precision values over all the classes. This is because the model performs very well in the edge tooth of the X-Ray i,e (Molars and Premolars) but not in the cases of Middle Tooth, i,e (Canines and Incisors). The model's performance is relatively low, resulting in lower Recall Values. The solution for this problem could be a grouping of teeth into different sectors. Segmenting might improve the performance of the model and the overdetection of incisors and canines.
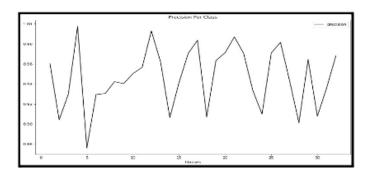

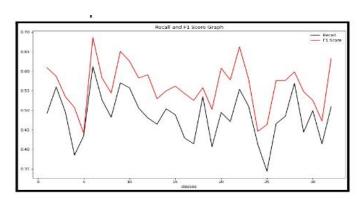**Figure 5: Precision of the model over the range of all classes**

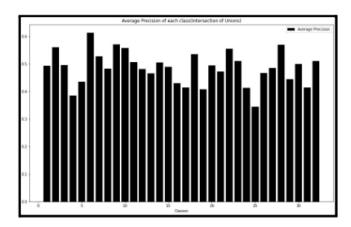**Figure 6: Recall and F-Score of the model over the range of all classes**

Figure 7: Average Precision(IoU) of various classes predicted by the model

Table 5: Cumulative Performance of model over all classes

Class	Precision	Recall	F1 Score	AP
1	0.97	0.50	0.68	0.49
2	0.90	0.49	0.56	0.59
3	0.92	0.49	0.53	0.49
4	0.99	0.43	0.50	0.38
5	0.88	0.61	0.44	0.44
6	0.92	0.52	0.68	0.62
7	0.93	0.48	0.58	0.53
8	0.94	0.56	0.54	0.47
9	0.94	0.55	0.65	0.58
10	0.95	0.50	0.62	0.55

11	0.95	0.48	0.58	0.50
12	0.99	0.46	0.59	0.47
13	0.96	0.58	0.52	0.48
14	0.90	0.48	0.54	0.42
15	0.96	0.42	0.56	0.44
16	0.97	0.41	0.54	0.56
17	0.98	0.43	0.52	0.41
18	0.90	0.53	0.55	0.65
19	0.94	0.44	0.50	0.38
20	0.97	0.49	0.60	0.53
21	0.98	0.47	0.57	0.48
22	0.97	0.55	0.66	0.59
23	0.93	0.50	0.57	0.51
24	0.90	0.41	0.66	0.42
25	0.98	0.34	0.57	0.34
26	0.94	0.46	0.44	0.49
27	0.90	0.48	0.46	0.51
28	0.96	0.56	0.57	0.59
29	0.90	0.44	0.59	0.54
30	0.90	0.49	0.52	0.52
31	0.93	0.41	0.47	0.43
32	0.96	0.50	0.63	0.51

We also observed that the variation of Average Precision (AP) over the classes is high. The molar and premolars have low AP, canines and incisors have high AP, which is due to the class imbalance problem where the former group have fewer number of instances. In contrast, the latter group has a higher number of cases, which is due to most of the tooth loss in patients belonging to the former group i,e. Molars and Premolars. This can also be solved using the grouping solution discussed earlier in this section.

Fig.7(a) and 8(a) illustrate the model's predictions over an X-Ray consisting of all teeth, and Fig.7(b) and 8(b) depict the model's predictions over an X-Ray consisting of teeth distributed sporadically.
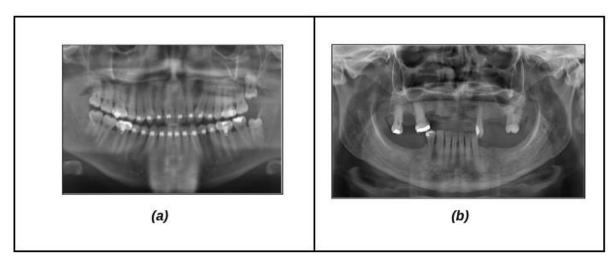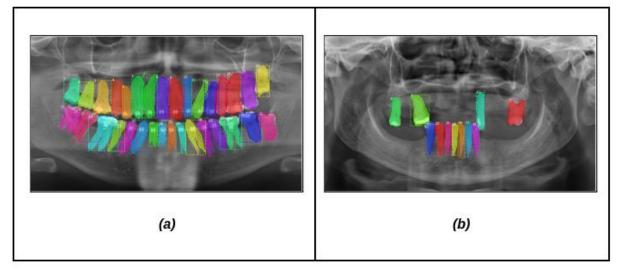
Figure 7: Actual X-Ray Images

Figure 8: Labelled and Numbered X-Ray Images

5. CONCLUSIONS AND FUTURE WORK

The quest for building a Deep Learning model for image Segmentation in Dental X-Rays has been ubiquitous these days. Even though much of the concentration was on developing machine learning models using unsupervised techniques, the perfection in the system is not achieved yet. Segmenting and labelling teeth reduces the tedious and unproductive work that has to be done by dentists before performing the necessary treatment. Our future work would include an additional grouping element of the tooth discussed in the result section and the detection of various minor diseases, like caries and cavities.

REFERENCES

[1] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon and L. Oliveira, "Deep Instance Segmentation of Teeth in Panoramic X-Ray Images," 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 400-407, doi:10.1109/SIBGRAPI.2018.00058.

[2] N. Senthilkumaran, "Genetic Algorithm Approach to Edge Detection for Dental X-ray Image Segmentation," International Journal of Advanced Research in Computer Science and Electronics Engineering, vol. 1, no. 7, pp. 5236–5238, 2012.

[3] P. L. Lin, Y. H. Lai, and P. W. Huang, "An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information," Pattern Recognition, vol. 43, no. 4, pp. 1380–1392, 2010.

[4] R. B. Ali, R. Ejbali, and M. Zaied, "GPU-based Segmentation of Dental X-ray Images using Active Contours Without Edges," in International Conference on Intelligent Systems Design and Applications, vol. 1, 2015, pp. 505–510.

[5] He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.

[6] G. Zhu, Z. Piao and S. C. Kim, "Tooth Detection and Segmentation with Mask R-CNN," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 070-072, doi: 10.1109/ICAIIC48513.2020.9065216.

[7] Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016): 1137-1149.

[8] Yuniarti, Anny, et al. "Classification and numbering of dental radiographs for an automated human identification system." *Telkomnika* 10.1 (2012): 137.

[9] Mahoor, Mohammad H., and Mohamed Abdel-Mottaleb. "Classification and numbering of teeth in dental bitewing images." *Pattern Recognition* 38.4 (2005): 577-586.

[10] Tangel, Martin Leonard, et al. "Dental numbering for periapical radiograph based on multiple fuzzy attribute approach." *Journal of Advanced Computational Intelligence and Intelligent Informatics* 18.3 (2014): 253-261.

[11] Mahdi, Fahad Parvez, Naomi Yagi, and Syoji Kobashi. "Automatic teeth recognition in dental X-ray images using transfer learning-based faster R-CNN." *2020 IEEE 50th International Symposium on Multiple-Valued Logic (ISMVL)*. IEEE, 2020.

[12] O. Nomir and M. Abdel-Mottaleb, "Hierarchical contour matching for dental X-ray radiographs," Pattern Recognition, vol. 41, no. 1, pp. 130–138, 2008.

[13] C. K. Modi and N. P. Desai, "A simple and novel algorithm for automatic selection of ROI for dental radiograph segmentation," in Canadian Conference on Electrical and Computer Engineering, 2011,pp. 000 504–000 507.

[14] R. B. Ali, R. Ejbali, and M. Zaied, "GPU-based Segmentation of Dental X-ray Images using Active Contours Without Edges," in International Conference on Intelligent Systems Design and Applications, vol. 1, 2015, pp. 505–510.

[15] H. Li, G. Sun, H. Sun, and W. Liu, "Watershed algorithm based on morphology for dental x-ray images segmentation," in International Conference on Signal Processing Proceedings, vol. 2, 2012, pp. 877–880.

[16] J. Kaur and J. Kaur, "Dental image disease analysis using pso and backpropagation neural network classifier," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, no. 4, pp. 158–160, 2016.

[17] Wang, C-W, Huang, C-T, Lee, J-H, Li, C-H, Chang, S-W, Siao, M-J, Lai, T-M, Ibragimov, B, Vrtovec, T, Ronneberger, O, Fischer, P, Cootes, TF & Lindner, C 2016, 'A benchmark for comparison of dental radiography analysis algorithms', Medical Image Analysis, vol. 31, pp. 63-76.

[18] Gao, Hui & Chae, Oksam. (2010). Individual tooth segmentation from CT images using level set method with shape and intensity prior. Pattern Recognition. 43. 2406-2417. 10.1016/j.patcog.2010.01.010.

[19] Yusra Y. Amer, Musbah J. Aqel, An Efficient Segmentation Algorithm for Panoramic Dental Images, Procedia Computer Science, Volume 65,2015, Pages 718-725, ISSN 1877-0509,

[20] Marques Lira, Pedro Henrique et al. "Dental R-Ray Image Segmentation Using Texture Recognition." IEEE Latin America Transactions 12 (2014): 694-698.

[21] Poonsri, A., Aim Jirakul, N., Charoenpong, T., & Sukjamsri, C. (2016). Teeth segmentation from dental x-ray image by template matching. *2016 9th Biomedical Engineering International Conference (BMEiCON)*, 1-4.

[22] Rad, A.E., Rahim, M.S., Kumoi, R., and Norouzi, A. (2013). Dental x-ray image segmentation and multiple feature extraction. *Global Journal on Technology, 2*.

[23] Indraswari, R., Arifin, A.Z., Navastara, D.A., & Jawas, N. (2015). Teeth segmentation on dental panoramic radiographs using decimation-free directional filter bank thresholding and multistage adaptive thresholding. *2015 International Conference on Information & Communication Technology and Systems (ICTS)*, 49-54.

[24] Razali et al. (2015) *Razali, M. R. M., Ahmad, N. S., Hassan, R., Zaki, Z. M., and Ismail, W. (2015).* Sobel and canny edges segmentations for the dental age assessment. In *Intl. Conference on Computer Assisted System in Health*, pages 62–66.

[25] Li et al. (2006) *Li, S., Fevens, T., Krzyzak, A., and Li, S. (2006).* An automatic variational level set segmentation framework for computer aided dental x-rays analysis in clinical environments. *Computerized Medical Imag. and Graph.*, 30(2):65–74.

[26] Jader G. (2019, February). Deep instance segmentation of teeth in panoramic X-ray images. https://github.com/IvisionLab/deep-dental-image .

[27] Chan, Tony F. and Luminita A. Vese. "Active contours without edges." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 10 2 (2001): 266-77 .

[28] Piotr Skalski (2019, February). Makesense . https://github.com/SkalskiP/make-sense.