Teeth Detection and Dental Problem Classification in Panoramic X-Ray Images using Deep Learning and Image Processing Techniques

Mircea Paul Muresan, Andrei Răzvan Barbura, Sergiu Nedevschi
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
mircea.muresan@cs.utcluj.ro, razvan.barbura@yahoo.ro, sergiu.nedevschi@cs.utcluj.ro

***Abstract*— Deep convolutional neural networks, have gained a lot popularity in medical research due to their impressive results in detection, prediction and classification. Analysis of panoramic dental radiographies help specialists observe problems in poor visibility areas, inside the buccal cavity or in hard to reach areas. However, poor image quality or fatigue can cause the diagnosis to vary, which can ultimately hinder the treatment. In this paper we propose a novel approach of automatic teeth detection and dental problem classification using panoramic X-Ray images which can aid the medical staff in making decisions regarding the correct diagnosis. For this endeavor panoramic radiographies were collected from three dental clinics and annotated, highlighting 14 different dental issues that can appear. A CNN was trained using the annotated data for obtaining semantic segmentation information. Next, multiple image processing operations were performed for segmenting and refining the bounding boxes corresponding to the teeth detections. Finally, each tooth instance was labeled and the problem affecting it was identified using a histogram-based majority voting within the detected region of interest. The implemented solution was evaluated with respect to several metrics like intersection over union for the semantic segmentation and accuracy, precision, recall and F1-score for the generated bounding box detections. The results were compared qualitatively with the data obtained from other approaches illustrating the superiority of the proposed solution.**

Keywords—CNN; Medical Imaging; Image Processing; Machine Learning; Dental Informatics; X-Ray Images; Semantic Segmentation;

I. Introduction

In recent years, medical imaging technologies such as computed tomography (CT) or X-rays have aided the treatment and diagnosis of different diseases [1]. In the field of dentistry, dental informatics is an emergent field, which, in addition to helping improve the treatment and diagnosis process, saves time and reduces stress during the daily routine [2]. The use of high-resolution imaging sensors and biosensors has led to the generation of massive amounts of data, which can be interpreted by computer programs that can help dental professionals in making decisions related to prevention, diagnosis or treatment planning, among others [3]. Radiographies are obtained by the passage of X-rays, produced by an X-Ray generator, through the oral cavity. Radiation can be absorbed by some tissues, or it can pass through the patient being absorbed by a detector. This process is called projective radiography and it generates two-dimensional images which represents internal structures of the human body [4]. Dental radiographies can be classified in two categories: intraoral where the film is positioned inside the buccal cavity, and extraoral where the patient is positioned between the source that emanates X-rays and the radiographic film. Most common types of dental X-rays are bitewing, periapical, which are intraoral, and panoramic which is extraoral.

A panoramic dental radiography shows the entire mouth area where all the teeth can be seen. It also shows the jaws and the skull thus giving the dentist an overview about the patient's problems. The panoramic dental radiography is used by dentists to observe problems in hard-to-reach areas or with a poor visibility inside the buccal cavity. The interpretation of the radiography is done manually by the dentist, who identifies each tooth and the existing problem where appropriate. However, if the X-ray radiography is not clear it can cause problems when analyzed and thus lead to misinterpretation.

Convolutional Neural Networks (CNN) are the preferred solutions for medical imaging analysis, and have been employed in many clinical fields [5][6].

Some successful applications in which CNNs have been applied include the assessment of breast cancer in mammograms [7], skin cancer in clinical skin screenings [8] or recognition of hepatocellular carcinoma areas from ultrasound images [9]. In the field of dentistry CNNs have been applied to detect periodontal bone loss [10], carries [11] apical lesions [12]. CNNs can also be used to detect different structures, classify them and segment them [13]. When using supervised learning, Neural Networks need to be trained and optimized on an image database in order to obtain an accurate result.

In this paper we propose a deep learning solution that helps dentists make the correct diagnosis using panoramic dental X-rays images. The main contributions of the paper are illustrated bellow:

- We manually annotate panoramic radiographies in order to train the semantic segmentation CNN
- We segment semantically the panoramic X-Ray image, for 15 semantic classes depicting different dental problems, using a CNN

978-1-7281-9080-8/20/$31.00 ©2020 IEEE

- We detect and label each tooth or group of teeth (depending on the scenario) and the dental problem affecting it using multiple image processing techniques
- We implement a refinement method, in order to eliminate small inconsistencies
- We evaluate and compare the proposed solution, with other CNNs created for the same task.

The rest of the paper is, organized as follows: in section II we discuss about Related work. Section III contains details about system pipeline and the implementation. The experimental results and evaluation are discussed in section IV. Section V concludes the paper.

II. Related work

In the field of dental informatics there are many approaches developed for teeth segmentation using different types of radiographic images such as bitewing, periapical and panoramic images. In [14] the authors provide a comparative analysis of 10 segmentation methods applied in dental imaging. The presented solutions are grouped in five categories and they were evaluated and classified according to the following metrics: Accuracy, Specificity, Precision, Recall and F1-score. Unfortunately, none of these 10 segmentation methods was able to completely isolate the teeth due to bone parts present inside the buccal cavity. The authors in [15] proposed a method for teeth instance segmentation in panoramic images using a mask region-based convolution neural network in order to accomplish instance segmentation. After using Resnet-101 to extract features, a feature pyramid network (FPN) is built where anchors are defined and region of interest are extracted. The FPN and the anchors form the region proposal network (RPN). After this step the regions of interest are aligned in order to have the same size. Furthermore, each feature is classified as a tooth or background and then it is localized by the bounding box coordinates. Finally, in the last step the tooth is segmented and a bounding box is drawn around it. The drawback of this method is that it focuses only on the detection of teeth leaving aside other types of problems such as dentures and regions where teeth are missing.

The work presented in [16] presents a deep neural transfer network which detects periodontal bone loss (PBL) on panoramic dental radiographs, called DeNTNet. The detection process involves the training of several convolutional neural networks. Firstly, a segmentation network is trained in order to extract the teeth from the desired region of interest, then secondly a segmentation network was trained to predict the periodontal bone loss lesions. These two networks are constructed on an encoder-decoder architecture and using the encoder part of the lesion segmentation network as a pre-trained model, a classification network is built in order to predict the existence of PBL in each tooth. For a performance improvement a classification network is also trained for detecting PBL for premolar and molar teeth. In the end, these two classification networks are connected to make the final prediction. This solution achieves good values in terms of performance measure indicators compared with human competitors. An advantage of DeNTNet is that it provides the corresponding teeth numbers which are affected by periodontal bone loss according to dental federation notation, but the disadvantage of this solution is that it detects only this dental problem.

The authors in [17] proposed a solution to identify a person after death by comparing a postmortem dental radiograph with a database of antemortem dental radiographs according to some specific features. The extracted feature which is proposed in this solution is the teeth contour because it remains invariant over time compared with other features. This step is achieved by radiograph segmentation and contour extraction. The segmentation process consists of two steps. First step, the gap valley detection is done by using projections for each jaw on the x axis and y axis. Due to the fact that teeth have a higher gray level intensity than jaws, a gap between upper and lower teeth will form a valley in the y-axis projection histogram which is called gap valley. The second step is done by determining a curve (separates upper and lower teeth) that defines the boundary of each row teeth and perpendicular lines are drawn on the curve determining a ROI (region of interest) for teeth. Based on the segmentation output an enclosing rectangle is constructed for each tooth. Contour extraction is done by crown contour extraction and root contour extraction, process achieved using image processing algorithms and Bayes probability rule. The results of this solution are good for a relatively small image database, however human intervention is required to initialize the algorithm parameters and correct the errors that occur.

Zhang et al. [18] applied a deep learning approach in order to detect and classify teeth of dental periapical radiographs. The combination of faster R-CNN and region based fully convolutional neural networks (R-FCN) were used to identify problems such as tooth loss, decayed tooth and filled tooth, which frequently appear on patients.

An approach that is engineering features and feeding them in a multi-layer perceptron neural network with the purpose of identifying dental caries is presented in [19]. In the paper presented in [20] the authors use two multi-sized CNN models for the task of detecting and classifying teeth in dental panoramic radiographs for the automatic structured filing of the dental charts. The object detection network using a four-fold cross-validation method achieves a high accuracy on the testing dataset. The method however, is only focused on detecting the teeth not classifying the problems from which each of them suffers. In the paper [21] Fukuda et. al. uses CNNs for detecting the vertical root fracture (VRF) on panoramic radiographies. The used CNN was built using DetectNet with DIGITS version 5.0, and fivefold cross-validation was performed in order to increase the model reliability.

Our paper builds upon the state of the art by proposing a novel segmentation and classification method using a semantic segmentation CNN and multiple image processing techniques, which are applied on panoramic X-Ray images. Each tooth is labeled and segmented and the main problem that is affecting it is determined.

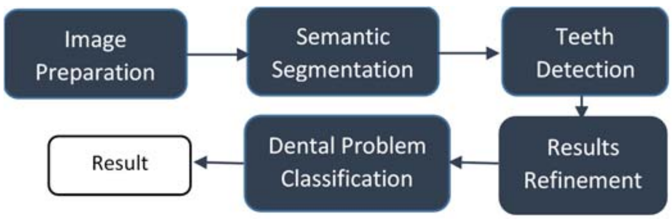


Figure 1. Main components of the processing pipeline.

III. PROPOSED SOLUTION

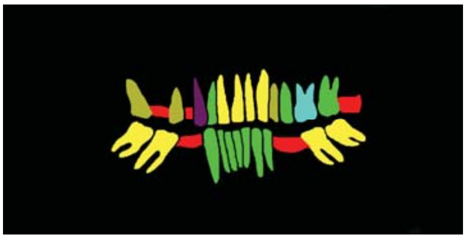
In this section the main contributions of this paper will be explained in more detail. The panoramic X-ray radiographies, used in this study, were obtained from three different dental clinics. The images contained various tooth problems such as restorations, dental implants, dentures and others. The main modules of the processing pipeline are presented in Figure 1.

A. Image Preparation

In this stage all of the images were cropped in order to remove any name that was present on the radiography and then they were renamed thus anonymizing the identity of the people. Afterwards, the images were resized to the dimensions of 2048x1024 pixels. The next step was annotating at pixel level all these images with 14 different classes each class corresponding to a specific tooth problem and another one for background, summing 15 classes in total. From the original dataset that contains approximatively 2000 images, 1000 images were selected and annotated for semantic segmentation. The selected semantic classes are: healthy tooth, missing tooth, dental restoration, implant, fixed prosthetics work, mobile prosthetics work (dentures), root canal device, fixed prosthetic work and root canal device, fixed prosthetic work and implant, fixed prosthetic work and devitalized tooth, devitalized tooth and restoration, dental inclusion, polished tooth, another problem and background. The last step in this stage was generating the corresponding labels for these classes.

a.

b.



c.

Figure 2. The data preparation process. In figure a, the cropped radiography is shown, figure b is the annotated ground truth image, each color representing a certain specific semantic class and c represents the label image.

Labels have values between 0 and 13 for the dental issues and 255 for background. The label images are represented as grayscale images, to conserve storage space. The graphical depiction of the preparation process is displayed in Figure 2.

B. Semantic Segmentation

The semantic segmentation of the X-Ray radiographies was performed using the ERFNet [22] neural network model. ERFNet stands for Efficient Residual Factorized Convolutional Network and it represents a convolutional neural network which performs real-time semantic segmentation. The neural network is built on encoder-decoder architecture, having a total of 23 layers. The first sixteen layers form the encoder and the last seven the decoder. The network uses residual functions in order to significantly reduce the degradation problem and 1D convolutions (non-bottleneck-1D design) to increase the computational efficiency thus obtaining a good trade-off between efficiency and speed. Others mechanism used to achieve this trade-off are the use of Dropout technique for avoiding the phenomenon of overfitting and the Batch Normalization technique to increase the speed, performance and the stability of the convolutional neural network.

The non-bottleneck design for residual functions was originally proposed in [23] and it uses two 3x3 convolutions with a ReLU activation function between the convolutions. As demonstrated in [24], any 2D filter can be represented as a combination of 1D filters so in the ERFNet the non-bottleneck design was rewritten using four 1D convolutions as follows: 3x1, 1x3, 3x1 and 1x3, using ReLU as activation function. This new design for the residual was named non-bottleneck-1D and it improves the neural network efficiency.

To extract features from the image the first step is downscaling the image. The down sampler block performs down sampling by uniting the parallel outputs of a single 3x3 convolution and a Max Pooling module. The activation function used is ReLU. After the feature extraction process is finished the image must be reconstructed from feature maps. The up-sampler block uses simple deconvolution layers instead Max un-pooling operation for the up sampling, the main advantage of this method being the fact that it is not necessary to share the pooling indexes from the encoder. Also, the deconvolution simplifies memory and computation requirements.

Before training the neural network, the initial classes weights must be computed. The equation responsible for computing the weights was initially proposed in [25] and adapted by the authors in [26] with a slightly difference for the c hyper-parameter value.

$$w_{class} = \frac{1}{\ln(c+p_{class})} \quad (1)$$

In equation (1) c represents an additional hyper-parameter which, for ERFNet, is set to 1.10, p_{class} represents the frequency of the pixels of the dataset.

The weights must be computed for both encoder and decoder. The difference in the calculation of the weights is made by the fact that for the encoder the calculations are made on images with a resolution of 128 x 64 pixels and for decoder the images resolution is 1024 x 512 pixels. The difference in images resolution is a downscaling factor of 8. Due to the downscaling process some shapes are changing their forms causing differences in pixel frequencies and so we will have different weights values for the encoder and decoder. The results of the image segmentation using this CNN model are depicted in Figure 3.

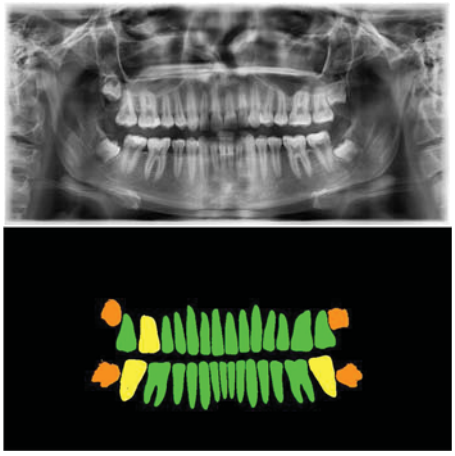


Figure 3. The top image illustrates the original radiography. The bottom image shows the semantic segmentation result.

C. Tooth Detection

A common problem which can appear in panoramic dental radiographies is the overlapping of teeth. This issue may occur because the dental X-Ray was not done properly or because the persons teeth are very close to each other, a phenomenon which is caused by the natural change in a person's dentition (the transition of primary teeth to permanent teeth). This situation can cause inaccurate tooth segmentation leading to the inconvenient case of seeing two or more teeth grouped as one instance. In this section we present a set steps that are applied on the semantic segmentation image, which are necessary for separating the teeth, especially in the overlapping scenario.

The first post-processing step consists in splitting the original segmented image in multiple images, equal to the number of classes, such that all teeth having the same semantic class are placed in the same image. Then, each image is binarized using a threshold value equal to the semantic class color for that specific image, such the background will be white and the foreground containing the teeth, dentures, implants and so on will have a black color. This process is graphically depicted in Figure 4.

The next step consists in using morphological on the obtained binary images, operations for some semantic classes, in order to better separate the teeth. To achieve this, the erosion operation was used in order to separate objects that are connected through unwanted edges. Two types of structural elements were applied, either combined or alone, a 5x5 element and one having a 15x1 dimension. Different structural elements were used in different scenarios, since there are situations in which a certain problem can contain multiple united teeth, like in the case of fixed prosthetics work, that causes a bridge to be formed between the teeth which does not need to be eliminated. The morphological operations were applied to different classes as follows: for the classes healthy tooth and dental restoration three consecutive erosion operations are performed, two with a 5x5 kernel and one with 15x1 kernel; for the semantic classes missing tooth, implant, fixed prosthetic work, mobile prosthetic work and polished tooth we have not applied any morphological operations; and finally for the remaining classes two consecutive erosion operations are performed using a 5x5 structural element.

In the final part of this processing stage, a two-step labeling algorithm [27] was used to identify each object from each binary image, assigning to it a unique numeric label. Finally, the coordinates of the bounding box of each tooth, or dental problem (depending on the scenario) is computed for each instance from label images. The coordinates of the bounding boxes are computed as the minimum and maximum pixel coordinates for each label

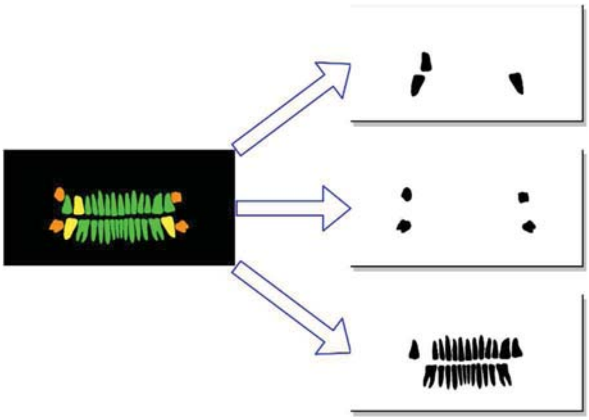


Figure 4. Segmentation of the teeth based on the semantic class and the binarization process of each segmentation image

D. Refinement and Classification

In the refinement step, all the previously detected bounding boxes are re-evaluated, and the boxes having an area smaller than a predefined threshold T, equal to 600 pixels in our application, are removed. These regions, usually correspond to small inconsistencies caused by semantic segmentation errors.

The classification procedure is done using a majority voting algorithm. The previously detected bounding boxes are projected onto the semantic segmentation image. A histogram having the length equal to the number of semantic classes is initialized for each identified tooth. The region of interest, belonging to a detected tooth, is traversed and for each semantic class identified, different from the background, a vote is cast in the histogram on the position corresponding to that semantic class. Afterwards, we iterate through the histogram and we select the position of the semantic class having the maximum number of votes for the region of interest.

This process is repeated for each tooth and a number is associated with a dental problem and drawn on the radiography. The teeth from the upper jaw are numbered first and then the teeth from the lower jaw. The numbering is done from left to right. For splitting the two jaws, a horizontal projection is made for all the pixels in the semantic segmentation image different from the background. The projection has a bimodal distribution, where one mode corresponds to the upper jaw and the other to the lower jaw. The region having the lowest value between the two formed picks from the horizontal projection corresponds to the splitting region between the two jaws.

The dental problem is not written onto the panoramic radiography in order not to overload the image. The processed panoramic radiography as well as the identified dental problems are written in a pdf report file, allowing the medical staff to easily review the patient's data anytime they want. In Figure 5 the results depicting the segmented teeth of a panoramic radiography are displayed. The different colors of the bounding boxes depict the different classes identified for each tooth.

Figure 5. Dental segmentation result of the proposed solution.

IV. EXPERIMENTAL RESULTS

In this section we evaluate the results of the proposed solution with respect to two metrics accuracy and intersection over union. Furthermore, the proposed solution is compared with other alternative methods that solve the same problems. The system on which the method was tested contains an Intel i7-4720HQ CPU with 2.60 GHz and a NVIDIA GeForce 950M GPU. The solution was developed in python, and no hardware acceleration methods were used.

A. *Different comparison methods*

In addition to the proposed solution, two alternatives have been developed. The first method is similar to the proposed solution, the differences being represented by the fact that the segmented image is binarized using a global threshold and only one binary image is generated. In this solution an opening operation using a 17 X 3 structural element, followed by a dilation with 13 X 3 and another 17x3 erosion. The area based bounding box refinement step used a threshold T=175. The following steps containing the labeling process, the voting phase and the bounding box generation remain as in the proposed solution. We will refer to this solution as Method 1.

The second method uses YOLO9000 [28] for generating the tooth bounding boxes, ERFNet for dental problem classification and multiple post processing operations for refining the results. The YOLO algorithm uses a 0.5 confidence factor for detections and was modified such that it can detect 2 classes, a tooth class and non-tooth class. The bounding boxes generated by the YOLO solutions are filtered, removing the ones that have an area smaller than 175 pixels. The semantic segmentation image is binarized using a global thresholding algorithm and one binary image is obtained. An erosion using a structural element of 17x3 is applied, followed by a 13x3 dilation and another 17x3 erosion. Next, every bounding box generated by YOLO is overlapped on the binary image. An assumption is made that our object of interest should be the largest object within the bounding box generated by YOLO. The region covered by the bounding box is cropped and a two-step labeling algorithm is applied on that region. The object having the largest area from the region of interest is extracted. In case the bounding box does not cover all the tooth, a region growing algorithm is used, starting from the already detected positions, and the whole tooth is associated to the bounding box. The voting phase and bounding box numbering remain the same as in the proposed solution. In the qualitative evaluation part this method will be referred to as Method 2.

B. *Qalitative Evaluation*

The original data set consists of 1000 annotated images images. We split these images in three according to the following proportions 70% training, 10% cross-validation and 20% test. In order to cover unexpected scenarios that may arise and to make the proposed segmentation solution more robust, we augment the initial dataset. The augmentation methods used contain the following operations: Translations, Flip, Gaussian Noise, Gaussian Blur and Gamma Contrast.

The CNN responsible for semantic segmentation was trained for 200 epochs, 100 epochs for encoder and 100 for decoder, using a learning rate of $5e^{-4}$ and batch size equal to 1. A small batch size was used because of the reduced capability of the GPU on which the model was trained. The intersection over union (2) obtained for the ERFNet on the test set is 60.03%. In equation 2, TP, FP and FN represent the number of true positives, false positives and false negative values.

$$IoU = \frac{TP}{TP+FP+FN} \quad (2)$$

The quality of the teeth detection was evaluated using the Accuracy, Precision, Recall and F1-score measures. The proposed solution together with other two methods were compared and evaluated on more than 200 ground truth images. The results are presented in Table I.

TABLE I. Teeth segmentation results

Method	**Metric**			
	Accuracy	**Precision**	**Recall**	**F1-score**
Proposed solution	0.89	0.98	0.91	0.93
Method 1	0.87	0.98	0.89	0.93
Method 2	0.68	0.98	0.68	0.80

The average running time of the proposed solution, as well as the running time of Method 1 and 2 are presented in Table II. Even though Method 1 has a better running time, the proposed solution outperforms it qualitatively.

TABLE II. Running Time

Method	Metric
	Average Running Time(s)
Proposed solution	60
Method 1	50
Method 2	800

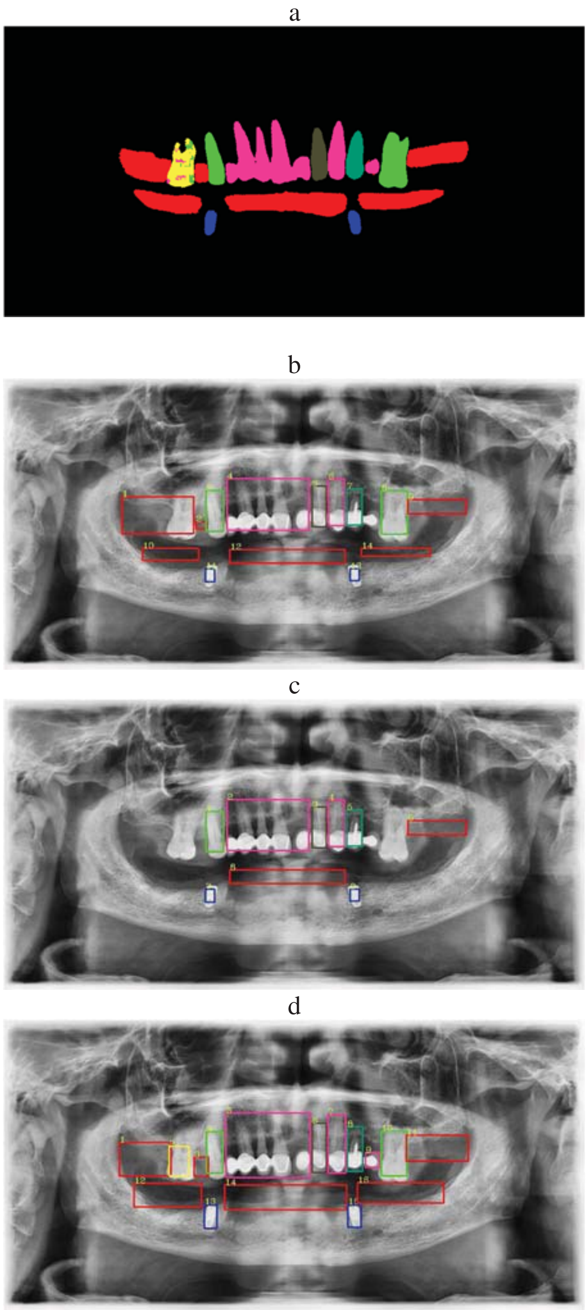


Figure 6. a) the resulted semantic segmentation image; b) result of Method 1; c) result of Method 2; d) result of the proposed solution.

In Figure 6 we illustrate the comparative results of the three methods on a complex scenario, where the patient has a missing tooth, a restoration, 2 healthy teeth, a fixed prosthetic work, fixed prosthetic work on a devitalized tooth, fixed prosthetic work and root crown device and 2 implants. As it can be seen the proposed solution is able segment accurately the teeth and identify the problems correctly. The other methods are not able to identify all the problems correctly.

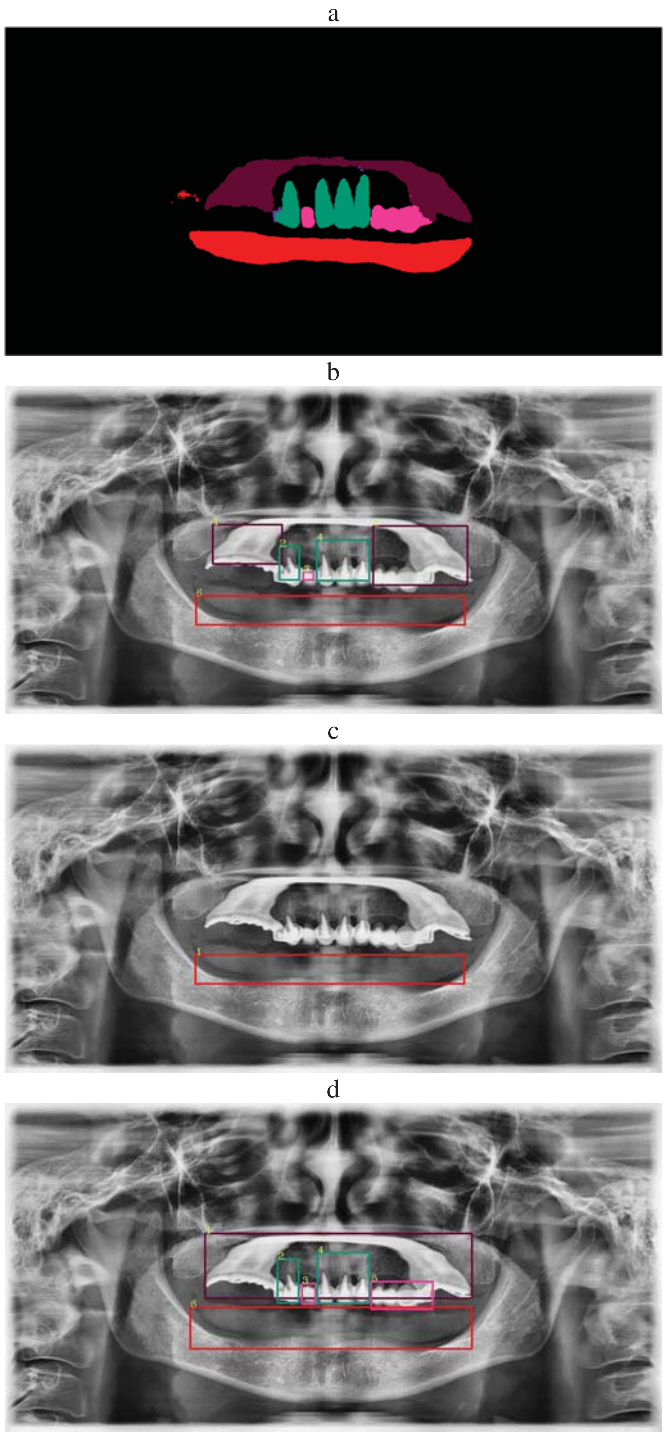


Figure 7. a) the resulted semantic segmentation image; b) result of Method 1; c) result of Method 2; d) result of the proposed solution.

In Figure 7 another scenario is shown where the patient has a denture, 4 teeth with fixed prosthetic work and root crown device, fixed prosthetic work on the other teeth and on the bottom jaw all the teeth are missing. Method 1 is splitting the denture, incorrectly identifying two dentures, and is unable to detect the right fixed prosthetic work. Method two is only identifying the bottom missing teeth. The proposed solution is able to identify all dental problems from the presented scenario.

V. CONCLUSION AND FUTURE WORK

This paper presented a novel teeth detection and dental problem classification approach using panoramic dental radiographies. For achieving the desired results images were collected from three different dental clinics and were annotated at pixel level, highlighting 14 different problems that can affect teeth. The annotated data was augmented using multiple operations, and a semantic segmentation CNN was trained using these images. Afterwards, the semantic segmentation image was binarized using multiple thresholds and a two-step labeling algorithm was used to detect each tooth instance. The bounding boxes corresponding to each instance are determined, and a refinement algorithm is applied in order to remove the regions that resulted from inconsistences in the semantic segmentation image. The bounding boxes of each instance is projected onto the semantic segmentation image and a histogram-based majority voting operation is performed in order to find the main semantic class of each tooth, which corresponds to the dental problem affecting the tooth. Each tooth or group of teeth, depending on the scenario, are numbered and a report containing the dental problems for each instance is generated aiding the medical staff in the diagnosis process. The implemented solution is evaluated using multiple metrics and compared to similar algorithms.

For future work, we aim to improve the running time of the proposed solution using hardware acceleration methods. Furthermore, we would like to increase the accuracy of the proposed solution and include more semantic classes.

REFERENCES

[1] D. Shen, G. Wu, H.-I. Suk, “Deep Learning in Medical Image Analysis”, Annual review of biomedical engineering, 19, pp. 221-248, 2017.

[2] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S.I. Nikolenko, M.M. Sveshnikov, G.B. Bednenko, “Tooth detection and numbering in panoramic radiographs using convolutional neural networks”, Dentomaxillofacial Radiol, 48, 20180051, 2019.

[3] E.A. Mendonça, “Clinical decision support systems: Perspectives in dentistry”, Journal of dental education, 68, pp. 589–597, 2004.

[4] J.I. Haring, L. Jansen, Dental radiography: principles and techniques, Second edition, W.B. Saunders Company, 2000.

[5] M. Prados-Privado, J. G. Villalón, C. H. Martínez-Martínez, C. Ivorra, Dental Images Recognition Technology and Applications: A Literature Review, Applied Sciences, 10(8) :2856, 2020.

[6] T. Zhou, S. Ruan, S. Canu, “A review: Deep learning for medical image segmentation using multi-modality fusion”, Array, 100004, 2019.

[7] D. Abdelhafiz, C. Yang, R. Ammar, S. Nabavi, “Deep convolutional neural networks for mammography: advances, challenges and applications”, BMC Bioinformatics, 20(S11), 2019.

[8] M.A. Kadampur, S. Al Riyaee, “Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images”, Informatics in Medicine Unlocked, 18, 100282, 2020.

[9] R. Brehar, D.-A. Mitrea, F. Vancea, T. Marita, S. Nedevschi, M. Lupsor-Platon, M. Rotaru, R.I. Badea, “Comparison of Deep-Learning and entional Machine-Learning Methods for the Automatic Recognition of the Hepatocellular Carcinoma Areas from Ultrasound Images”, Sensors 20, 3085, 2020.

[10] J. Krois, T. Ekert, L. Meinhold, T. Golla, B. Kharbot, A. Wittemeier, C. Dörfer, F. Schwendicke, “Deep Learning for the Radiographic Detection of Periodontal Bone Loss”, Scientific Reports (Nature Publisher Group), 9, 8495, 2019.

[11] J.-H. Lee, D.-H. Kim, S.-N. Jeong, S.-H. Choi, “Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm”, Journal of dentistry, 77, pp. 106-111, 2018.

[12] T. Ekert, J. Krois, L. Meinhold, K. Elhennawy, R. Emara, T. Golla, F. Schwendicke, “Deep Learning for the Radiographic Detection of Apical Lesions”, Journal of endodontics, 45(7), pp. 917-922, 2019.

[13] F. Schwendicke, T. Golla, M. Dreher, J. Krois, “Convolutional neural networks for dental image diagnostics: A scoping review”, Journal of dentistry, 91, 103226, 2019.

[14] G. Silva, L. Oliveira, M. Pithon, “Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives”, Expert Systems with Applications, 107, pp. 15-31, 2018.

[15] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira, “Deep instance segmentation of teeth in panoramic X-ray images” 31 st SIBGRAPI Conference on Graphics, Patterns and Images, October 2018.

[16] J. Kim, H. S. Lee, I. S. Song & K. H. Jung, “DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs”, in Scientific reports, vol.9(1), pp. 1-9, Nov. 2019.

[17] A. K. Jain, H. Chen, “Matching of dental X-ray images for human identification”, in Pattern Recognition, vol. 37(7), pp. 1519-1532, July 2004.

[18] K. Zhang, J. Wu, H. Chen, and P. Lyu, "An effective teeth recognition method using label tree with cascade network structure”, Computerized Medical Imaging and Graphics, vol. 68, pp. 61-70, 2018.

[19] V. Geetha, K. S. Aprameya, and D. M. Hinduja, ”Dental caries diagnosis in digital radiographs using back-propagation neural network”, Health Information Science and Systems, vol. 8(1), pp. 1-14, 2020.

[20] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, and H. Fujita, "Classification of teeth in cone-beam CT using deep convolutional neural network", Computers in biology and medicine, vol. 80, pp. 24-29, 2017.

[21] M. Fukuda, K. Inamoto, N. Shibata, Y. Ariji, Y. Yanashita, S. Kutsuna, and E. Ariji, ”Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography”, Oral Radiology, pp. 1-7, 2019.

[22] E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, "ERFNet: Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation" in IEEE Transactions on Intelligent Transportation Systems (T-ITS), vol. 19, pp. 263-272, Dec. 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2015.

[24] J. Alvarez, L. Petersson, “Decomposeme: Simplifying convnets for end-to-end learning”, arXiv preprint arXiv:1606.05426, 2016.

[25] Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “Enet: A deep neural network architecture for real-time semantic segmentation”, pp. 1-10, 2016.

[26] E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, "Efficient ConvNet for Realtime Semantic Segmentation", in IEEE Intelligent Vehicles Symposium (IV), Redondo Beach, California, USA, pp. 1789-1794, Jun. 2017.

[27] R. M. Haralick, L. G. Shapiro, Computer and Robot Vision, Addison-Wesley Longman Publishing Co., Inc., 1993.

[28] J. Redmon, A. Farhadi, ”YOLO9000: Better, Faster, Stronger, University of Washington, Allen Institute for AI”, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271, 2017.