Deep Learning Approach for Tooth Instance Segmentation on Panoramic Dental Radiographs Using U-NET

Dr. P Ammi Reddy[1], G Chiranjeevi[2], A Showri Joseph Kumar[3], A Siva Kumar[4] and B Jeevan Harsha[5]
[1-5] Department of ECE, Vasireddy Venkatadri Institute of Technology, Guntur

***Abstract—* Dental radiography plays a critical role in diagnosing and treating oral health conditions, with panoramic dental radiographs being commonly used for comprehensive assessments. Automating the segmentation of individual teeth within panoramic radiographs is a crucial step towards improving diagnostic accuracy and efficiency. In this study, we explore a deep learning approach tailored specifically for panoramic dental radiographs, aiming to automatically segment teeth using the U-Net architecture. We propose leveraging the U-Net network to achieve precise tooth instance segmentation in panoramic X-ray images. The proposed method achieves an impressive Dice overlap score of 95.4% in overall teeth segmentation. What sets this approach apart is the introduction of a novel post-processing stage that refines the segmentation maps by applying grayscale morphological and filtering operations to the output of the U-Net network before binarization.**

The obtained results concludes Deep Learning approach along with innovative post-processing techniques, holds great promise for advancing image analysis in dentistry and beyond, offering potential applications to similar challenges in various domains.

***Index Terms—* Dental, Panoramic, Segmentation, Counting, Deep learning, U-net.**

I. INTRODUCTION

Oral health plays a crucial role in overall well-being, encompassing various functions such as speaking, chewing, and conveying emotions. Dental diseases like caries, periodontal issues, and oral cancers affect billions worldwide, with caries being particularly prevalent [1,2]. Radiographs aid in diagnosis, with periapical and panoramic images being common tools in dental practice, although limited in revealing three-dimensional details [3]. Panoramic X-rays offer a comprehensive view of the oral region, aiding in diagnostics and treatment planning [4].

Segmentation of teeth in panoramic images is vital for various analyses, from lesion detection to age estimation [5,6]. Traditional methods face challenges due to image complexity, leading to the adoption of deep learning, notably convolutional neural networks (CNNs) [10]. Existing models like TSASNet [11], Faster R-CNN [12], and Mask R-CNN [18] target tooth segmentation or identification. TSASNet employs a two-stage approach, achieving high accuracy [11]. Faster R-CNN focuses on tooth detection and numbering [13,14,15]. Mask R-CNN excels in tooth instance segmentation [18].

The proposed model integrates U-Net with morphological processing for tooth instance segmentation on panoramic X-rays. By training a semantic segmentation CNN with manually annotated images, it detects and labels teeth and associated issues. Post-processing techniques refine the segmentation map for enhanced accuracy. Its goal is to provide detailed diagnostic information essential for effective dental management. Evaluations will compare its performance against existing CNNs in the field. This approach offers a promising solution for precise tooth segmentation and diagnosis in dental imaging.

II. LITERATURE SURVEY

Tooth instance segmentation in panoramic dental radiographs is pivotal for automated diagnosis and treatment planning in dental image analysis. Various models have been proposed.

TSASNet, a two-stage attention segmentation network proposed by Zhao et al. in 2020, achieved notable success in tooth segmentation on dental panoramic X-ray images. With an average dice overlap of 92.72% on a dataset comprising 1500 panoramic images, TSASNet demonstrates its effectiveness in accurately delineating tooth boundariess.

Mask R-CNN, an advanced model proposed by He et al. in 2017, has been employed for tooth instance segmentation. Studies have reported impressive F1 scores ranging from 88% to 87.5% on dental panoramic radiographs, showcasing its efficacy in segmenting individual teeth [9].

Faster R-CNN, a model introduced by Ren et al. in 2015, has been widely utilized for tooth detection and numbering tasks. Several studies report tooth numbering accuracy rates reaching up to 84.5%, highlighting its utility in dental image analysis [8].

III. METHODOLOGY

1. Gathering the Data and Data Preprocessing:

The dataset includes 116 panoramic dental x-ray images from Noor Medical Imaging Center, Iran. Images, taken by Soredex CranexD, are resized to 512x512 pixels and normalized. Edentulous cases are excluded. Manual labeling provides two tooth masks: one with teeth labeled together and another with separate teeth.

(a) (b) (c)

Fig 1: Examples of masks that have been used in this study. (a) The original radiograph. (b) Full mask obtained by manual labelling. (c) Split mask, in which each tooth is separated from the others by a narrow gap

2. Architecture Model

The U-net architecture, which is a fully convolutional neural network and popular in biomedical image segmentation, is used in this study. The network architecture is illustrated in Figure 2. The U net is a symmetric architecture consists of an encoder network, which maps the image into lower dimensional latent representation, followed by a decoder network, which reconstructs the output by up-sampling the latent vector back to the input size. At each level of the contracting path, two convolutional layers with 3x3 kernels and rectified linear unit (ReLU) activation functions, followed by a batch normalization, are applied. The feature maps are downsampled by a factor of 2, whereas the number of features are doubled, by 2x2 max pooling operations at each step. In the expanding path, the upsampling is performed by 4x4 transposed convolution. The rate of dropout operations applied at each level are 0.15, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, 0.1, in the order from the input to the output level. By the skip connections, the features are transferred from each level of the contracting path to the same level of the expanding path. Two 42 convolutional layers with 3x3 kernels and ReLU activation functions, followed by a batch normalization, are applied after each upscaling operation. In the final step, the output of the network is produced by applying a 1x1 convolution and a sigmoid activation function.

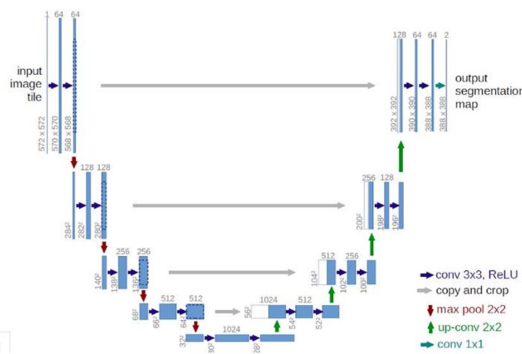


Fig 2: U-NET Architecture

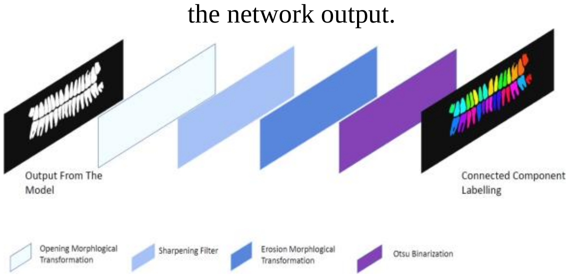
3. Training Details

The network is implemented in Python using the Keras library. Binary cross-entropy is used as the loss function, with weights initialized randomly from a truncated normal distribution. Optimization is done using the Adaptive Moment Estimation (ADAM) optimizer over 250 epochs with a batch size of 4 and a learning rate of 0.001. Data augmentation techniques include horizontal and vertical flipping, as well as the addition of random salt and pepper noise, with 5% of image pixels replaced with noise.

4. Post-processing

After obtaining the sigmoid output from the network, tooth instance separation is ensured through morphological operations using the OpenCV library. Firstly, the output map is resized to the original input size to maintain spatial alignment. Then, a morphological grayscale opening operation is applied to remove small details and separate teeth using a 5x5 square-shaped structural element. Subsequently, an image sharpening filter is employed to enhance details, followed by grayscale erosion morphological operation applied twice to detach connected teeth. Masks are then generated using Otsu's method for segmentation, and connected components are uniquely labeled with a cluster size threshold of 2000 pixels, ensuring accurate tooth separation and reducing noise in the final segmentation maps.

Fig 3: Steps of the post-processing that are applied to the network output.



IV.EXPERIMENT AND RESULTS

Five experiments are performed to determine the performance of the proposed method under various parameters including different output labels, data augmentation or post processing scenario. In these experiments, a 10-fold cross validation technique is applied. The comparison of the performed procedures is given in Table 1. Whenever the post-processing steps are not applied, segmentation masks are obtained by directly binarizing the sigmoid output of the network using a threshold level of 0.2.

	Sensitivity	*Specificity*	*PPV*	*NPV*
E1	96.6±0.9	98.4±0.4	92.6±1.9	99.3±0.2
E2	96.3±0.6	98.7±0.3	93.9±1.0	99.2±0.1
E3	96.1±0.6	98.8±0.3	94.4±0.8	99.2±0.1
E4	95.9±0.6	98.9±0.2	94.8±0.8	99.1±0.2
E5	99.1±0.3	98.6±0.2	91.9±1.2	99.9±0.0

	Jaccard	*Dice*
E1	89.7±1.2	94.5±0.7
E2	90.6±0.8	95.1±0.5
E3	90.9±0.6	95.2±0.3
E4	91.1±0.5	95.4±0.3
E5	91.1±1.2	95.3±0.6

Table 1: Evaluation metrics

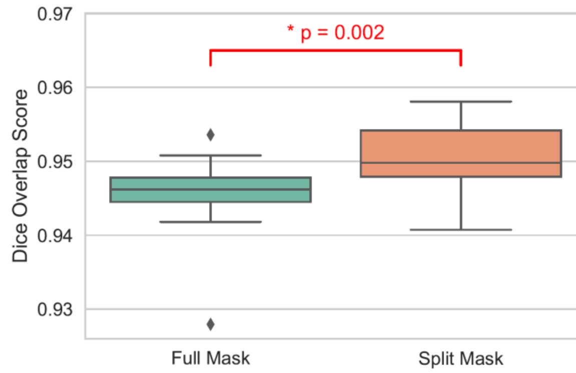


Fig 4: Comparison of the segmentation performance of the network on the full mask (E1) and on the split mask (E2), without any augmentation nor post processing

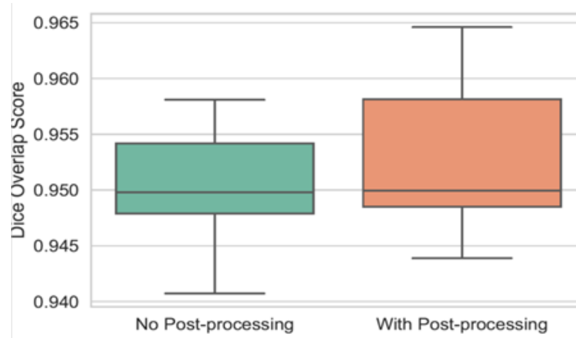


Fig 5: Comparison of the segmentation performance after post-processing with the proposed stages (E5) and without post-processing (E2).

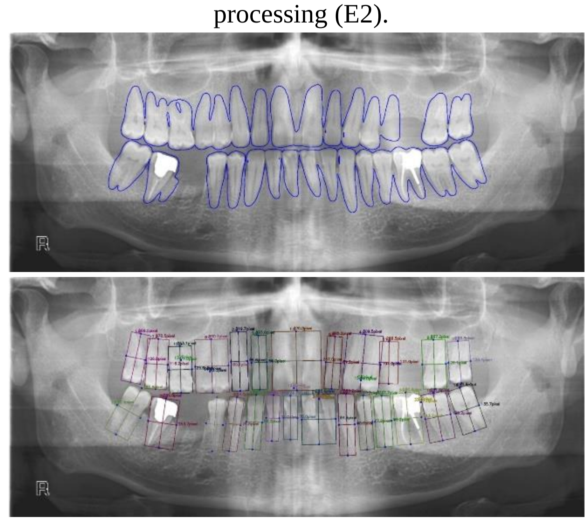


Fig 6: Segmentation results

VII. CONCLUSION

In this study, a method to segment and count the tooth instances on panoramic radiographs is proposed. The evaluations demonstrate the potential of the method to help clinical practice by serving as a prior step for further processing and analysis of dental images. The performances of both segmentation and tooth counting are the highest in the literature, to our knowledge. Moreover, this is achieved by using a relatively small training dataset, which consists of 105 images. The technique proposed in this study is based on image processing stages that are applied to the sigmoid output of the neural network before binarization. Although, the aim in this study is to segment tooth instances, the presented method is applicable to similar problems on other domains, such that separating the cell instances. A dice overlap score of 95.4±0.3% is obtained in overall teeth segmentation.

REFERENCE

[1] M. Glick, D. M. Williams, D. V. Kleinman, M. Vujicic, R. G. Watt, and R. J. Weyant, "A new definition for oral health developed by the fdi world dental federation opens the door to a universal definition of oral health," British dental journal, vol. 221, no. 12, pp. 792–793, 2016.

[2] M. A. Peres, L. M. Macpherson, R. J. Weyant, B. Daly, R. Venturelli, M. R. Mathur, S. Listl, R. K. Celeste, C. C. Guarnizo-Herreño, C. Kearns et al., "Oral diseases: a global public health challenge," The Lancet, vol. 394, no. 10194, pp. 249–260, 2019.

[3] N. Shah, N. Bansal, and A. Logani, "Recent advances in imaging technologies in dentistry," World journal of radiology, vol. 6, no. 10, p. 794, 2014.

[4] B. Vandenberghe, R. Jacobs, and H. Bosmans, "Modern dental imaging: a review of the current technology and clinical applications in dental practice," European radiology, vol. 20, no. 11, pp. 2637–2655, 2010.

[5] R. G. Birdal, E. Gumus, A. Sertbas, and I. S. Birdal, "Automated lesion detection in panoramic dental radiographs," Oral Radiology, vol. 32, no. 2, pp. 111–118, 2016.

[6] E. Avuclu and F. Basciftci, "Novel approaches to determine age and gender from dental x-ray images by using multiplayer perceptron neural networks and image processing techniques," Chaos, Solitons & Fractals, vol. 120, pp. 127 – 138, 2019.

[7] M. H. Bozkurt and S. Karagol, "Jaw and teeth segmentation on the panoramic x-ray images for dental human identification," Journal of Digital Imaging, pp. 1–18, 2020. 48

[8] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2018, pp. 400–407.

[9] G. Silva, L. Oliveira, and M. Pithon, "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.

[10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60–88, 2017.

[11] Y. Zhao, P. Li, C. Gao, Y. Liu, Q. Chen, F. Yang, and D. Meng, "Tsasnet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network," Knowledge-Based Systems, vol. 206, p. 106338, 2020.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[13] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," Scientific reports, vol. 9, no. 1, pp.1–11, 2019.

[14] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, and S. Youm, "Automatic tooth detection and numbering using a combination of a cnn and heuristic algorithm," Applied Sciences, vol. 10, no. 16, p. 5624, 2020.

[15] D. V. Tuzoff, L. N. Tuzova, M. M. Bornstein, A. S. Krasnov, M. A. Kharchenko, S. I. Nikolenko, M. M. Sveshnikov, and G. B. Bednenko, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," Dentomaxillofacial Radiology, vol. 48, no. 4, p. 20180051, 2019.
[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[17] 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.
[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.