Deep Instance Segmentation of Teeth on Dental X-rays using U-Nets

**Aishwarya Virigineni1, Maturi Tanuj1, Apoorva Mani1, Dr.Suja P2**

*1Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India*

*2Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India* [*aishwaryavirigineni@gmail.com***1**](mailto:aishwaryavirigineni@gmail.com1)*,*[*charantanujma63@gmail.com*](mailto:charantanujma63@gmail.com)**1***,* [*apoorvaa.mani@gmail.com*](mailto:apoorvaa.mani@gmail.com)**1** ,*p\_suja@blr.amrita.edu* **2**

***Abstract* - Automatic tooth detection and segmentation has been an important research topic of image analysis in dentistry for at least the last two decades. In this paper, a post-processing stage is proposed to obtain a segmentation map that differentiates the objects in an image, and we use the U-Net network to apply this approach to tooth instance segmentation. Grayscale morphological and filtering procedures are done to the network's sigmoid output before binarization as part of the post-processing. To our awareness, the segmentation and tooth counting results are the finest in the literature. Furthermore, this is accomplished with a relatively modest training dataset of 105 photos. Even though the goal of this research is to segment tooth instances, the approach given here can be used to solve various problems in other domains, such as separating cell instances.**

**Keywords—Dental, Panoramic, Segmentation, Instance Segmentation, U-net architecture.**

# I. Introduction

Oral health is now considered a key predictor of general health and quality of life. Oral health encompasses the ability to confidently talk, smile, smell, taste, touch, chew, swallow, and express a variety of emotions through facial expressions while avoiding pain, discomfort, and craniofacial disease. Dental caries (tooth decay), periodontal disease (gum disease), and oral malignancies are the most common oral diseases. Oral diseases are extremely common, affecting more than 3.5 billion people worldwide, despite the fact that they are generally preventable. Dental caries is the most common disease in the world, and its incidence is rising. Radiographs are useful instruments for diagnosing oral disorders and complementing the clinical examination. Periapical and panoramic radiographs are commonly utilized and give information for regular procedures. However, projecting the complete or a portion of the mouth onto a two-dimensional picture plane has limitations, and improved three-dimensional imaging techniques that disclose extra information are required in the diagnosis or treatment planning of particular patients. Cone beam computed tomography (CBCT), computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonography are three-dimensional imaging techniques used in dentistry. Despite the aforementioned drawback, panoramic dental x-ray is a regularly performed medical examination by dentists and oral surgeons in everyday practice and is an important diagnostic tool due to its wide availability, lower dosage of ionizing radiation (relative to CBCT), and patient comfort. Panoramic radiographs show the orofacial region in its entirety, including the jaws, teeth, sinuses, and temporomandibular joint (TMJ). They're particularly useful for demonstrating dental development stages or anomalies, as well as serving as an initial evaluation for broad disease or many issues. Teeth segmentation is frequently required in the processing of dental pictures for purposes such as lesion detection, age or gender determination, and human identification. Automatic teeth segmentation in panoramic x-ray pictures is an important image analysis research topic in dental medicine. It's difficult to isolate teeth on panoramic radiographs because other elements of the patient's body are visible (e.g., chin, spine and jaws).With the advent of computation \of Deep learning algorithms, increase in the amount of data, and superior success with simpler UNet architecture, when compared to the performance of more sophisticated models, tested on the same dataset, deep learning algorithms, in particular convolutional neural networks (CNNs), have rapidly become a methodology of choice for analyzing medical images. Instead of separate teeth segmentations, the network's output is an overall segmentation map. In this paper, we suggest using a post-processing stage to obtain a segmentation map in which the objects in the image are segregated, and then using the U-Net network technique for tooth instance segmentation. Grayscale morphological and filtering processes are applied to the network's sigmoid output before binarization as part of the post-processing activities. Their method for removing small light details was mostly based on gray-scale morphology opening procedures, and they observed a high success rate in separation while maintaining the shape of the objects. Segmentation of each cell instance independently for a comparable challenge in another area. The goal of this project is to design a method for performing tooth instance segmentation on panoramic dental radiographs using a U-Net network and morphological processing in order to give diagnostic information for the management of dental problems, diseases, and conditions.

# II. Dataset

The dataset consists of unknown and de-identified x-ray teeth images of 116 patients, taken at Noor Medical Imaging Center, Qom, Iran. The manual segmentations of mandibles are also available in the dataset, but those segmented images are irrelevant to the subject of our study and only original images are used. The images have been taken by the Soredex CranexD digital panoramic x-ray unit. The widths of all images vary between 2600-3138 pixels, their heights are between 1050-1380 pixels. Complete edentulous cases are excluded from this study. All panoramic dental x-ray pictures are scaled to 512x512 pixels and standardized in the range of 0 to 1 during data preprocessing.



**Figure 1**. Panoramic radiograph image.



**Figure 2**. Mask image.



**Figure 3**. Split mask image.

Figure 1 shows an example of an input image (a). Manual labelling yields two separate teeth masks for each panoramic dental x-ray image in the dataset. All teeth are labelled in the first mask, as seen in Figure 1. (b). Although the teeth are positioned in contact in the second one, each tooth is separated from the others by a thin gap, as seen in Figure 1. (c).

# III. The network architecture

For classification, the convolution neural network (CNN) model needs a little coding and gives a class probability as the result. On the contrary to classification, encoder and decoder for the convolutional units is essential for segmentation. An encoder is required to encode the input image onto maps with lower dimensional representation. The decoder is used to carry out de-convolution by up-sampling to generate a segmentation map which dimensionally similar to the actual input image. “U-Net” architecture is one of the earliest and most famous technique for bio-medical image segmentation. The basic architecture diagram can be seen in Figure 1. The U-Net network is a fully convolutional symmetric neural network which comprises of two networks: the encoder part and the decoder part.

In both the networks, convolutional networks go along with ReLU activation. In the encoder network, a 2x2 max-pooling operation is performed for

Figure 4. The U-Net architecture
down-sampling. And to sample the feature map in the decoder network, a de-conv\solution operation to perform up-sampling is applied. The initial version of the U-Net was utilized to cut and copy the feature map from the encoder and paste it on the decoder. The U-Net architecture has various benefits in terms of segmentation. One among them is it permits the usage of global contexts and locations simultaneously. And the other one is this architecture can perform better segmentation even with a smaller training dataset sample and also the full-on processing of the complete image in the forward pass and producing the segmentation map directly. This guarantees us a big advantage of the U-Net keeping up the whole context of our input image.

Figure 4. The U-Net architecture

The U-Net model comprises of three main parts: contraction path, bottleneck, and expansion path. The contraction part is made up of numerous contraction box. Every box comprises of two convolution layers with 3 × 3 kernels and rectified linear unit (ReLU) function followed by a dropout layer, batch normalization and a maximum pooling layer with 2 × 2 kernel. The bottom-most layer intercedes between the contraction path and the expansion path. Now, coming to the expansion path, it also contains several expansion blocks similar to contraction blocks. Every block comprises of two convolution layers with 3 × 3 kernels and rectified linear unit (ReLU) function followed by a dropout layer, batch normalization and a transpose convolutional layer with 4 × 4 kernel along with a ReLU activation function. Now with the help of skip

connections, features from each contracting path level are transferred to the corresponding expanding path level. In the end, 1 × 1 convolution along with a sigmoid activation function is applied to produce the segmented teeth image.

# IV. Training details

Kera’s Library is used to implement U-Net network in python. Loss function used in implementation is binary cross-entropy. Adaptive Moment Estimation (ADAM) optimizer is used to optimize weights in the network. The model is trained with a batch size of 8 for 200 epochs. The training dataset is augmented with varied combinations of horizontal flipping, vertical flipping, and adding random salt and pepper noise in this study.

# V. Post-processing

Rather than binarizing the network's sigmoid output right away, morphological processes ensure that tooth instances in the final map are separated. To do this, each network output map is subjected to a number of fundamental image processing operations from the "Open Computer Vision (OpenCV)" package. The goal of these processes is to reduce noise in the final segmentation maps while simultaneously increasing the separation of the teeth from one another.

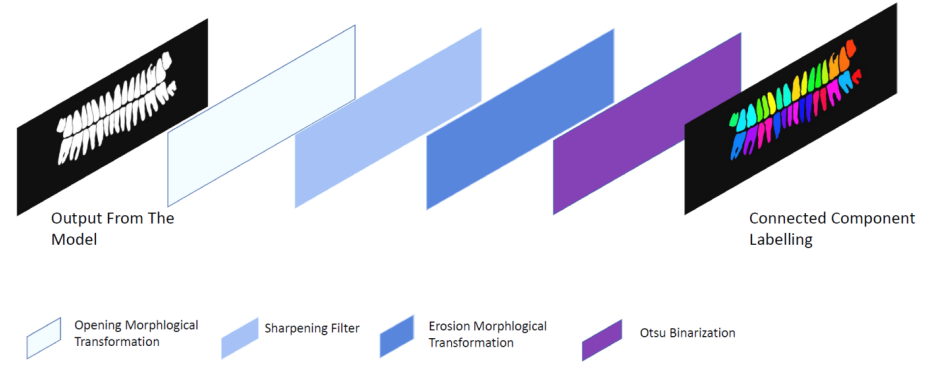
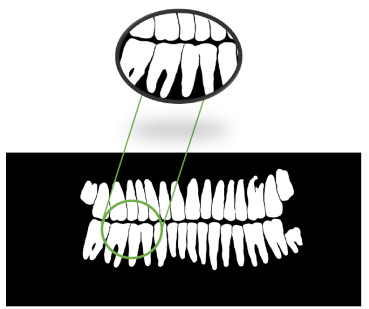


Figure 5. Post-processing steps applied to the network output

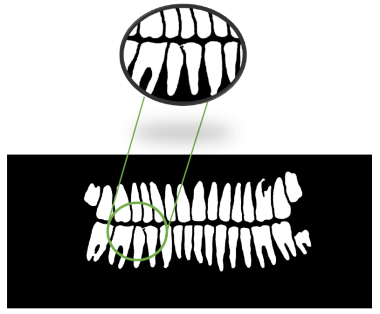
To begin, the network's output was shrunk to match the original input by applying a Lanczos filter to all pixels that could add to the output value. Second, on the scaled output map, a morphological grayscale opening procedure is used to eliminate small bright details in the image and separate teeth from one another. A 5x5 matrix consisting of ones is the square-shaped structural element used in all morphological procedures. Grayscale erosion morphological method is employed twice to the output map of the sharpening filter before segmentation to remove the joined teeth. A structural element with the same square shape as the opening is used. The masks are created after the erosion operation, using the Otsu's approach to establish the best threshold. Ultimately, using the connected technique and a cluster size cutoff of 2000 pixels, subsets of connected components are uniquely labelled on the masks.

# VI. Performance evaluation

When the proposed post-processing stages are used, the mean error of tooth count is 6.15 percent, but the error without post-processing is 26.81 percent.



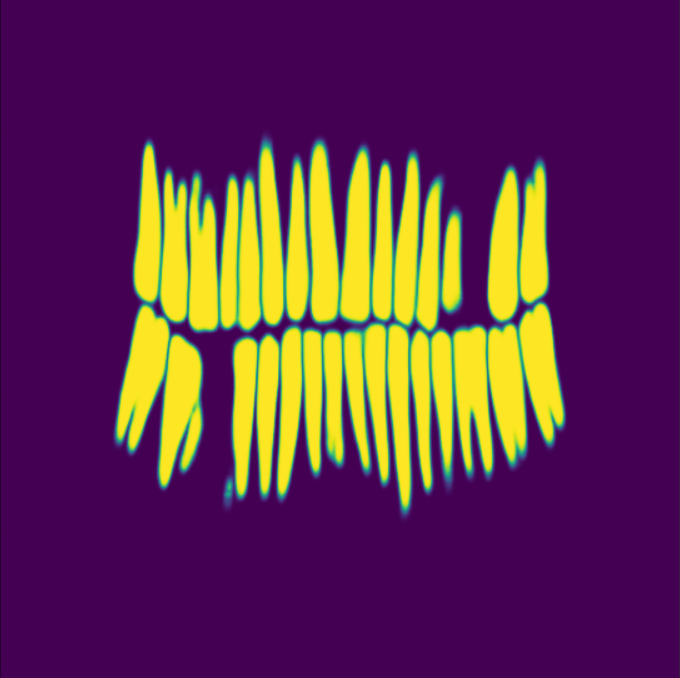
**Figure 6**. The network's output without any post-processing actions.



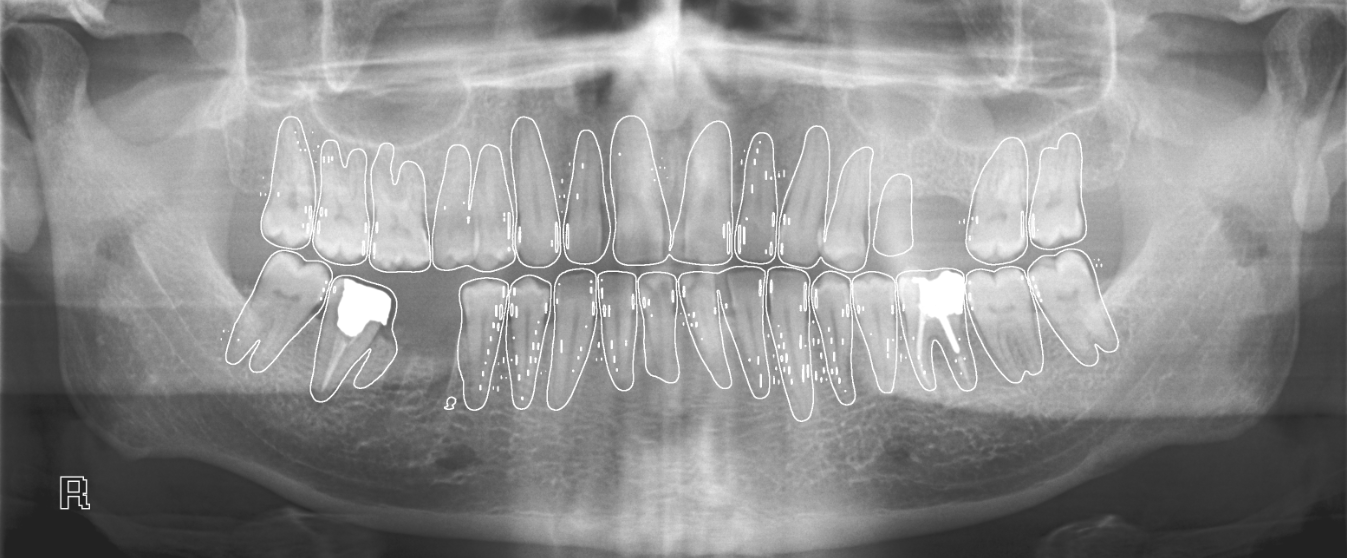
**Figure 7**. The network's output with proposed post-processing actions.

Figure 6. and Figure 7. shows the influence of post-processing methods on the segmentation result on a sample output.

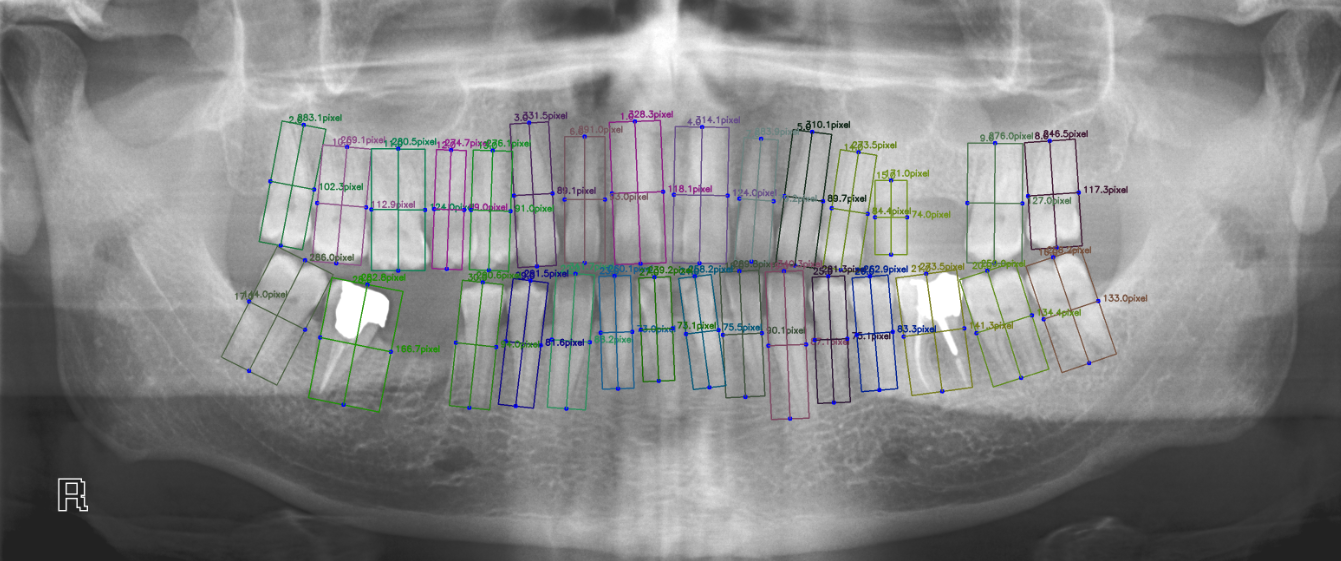
# VII. Results



**Figure 8**. Full mask obtained by manual labeling.



**Figure 9**.Image after post processing



**Figure 10**.Counting and labeling tooth after post processing

# VIII. Discussion

A method for segmenting tooth instances on panoramic pictures is proposed in this work. Teeth separation is achieved by applying post-processing procedures to the output of the neural network's sigmoid classification layer. We used the entire segmentation maps without separation of tooth instances to compare the proposed approach to the relevant literature. Using the U-Net neural networks, one of the highest achievements in segmenting the teeth in panoramic radiographs was reported in the literature. We artificially constructed gaps between the teeth on the ground truth label maps to differentiate the occurrences, despite the teeth being continuous on panoramic radiographs. Nonetheless, the split mask's segmentation performance is much better than the entire masks. In comparison to other studies in the literature, the number of radiographs used to train the network is fewer. We trained the network from scratch in this research by randomly initializing the weights. The key reason permitting this is because the separation of instances is achieved by morphological processes, instead of a trainable network. Data augmentation procedures were used to compensate for the dataset restriction. The performance was slightly improved by adding salt and pepper noise and horizontal flipping.

The use of post-processing steps considerably improves the performance of counting the number of teeth. The mean error of tooth count in our study using the proposed technique is 6.15 percent, which is the lowest in the literature to our knowledge. The network's sigmoid output was subjected to post-processing techniques, which not only reduced the inaccuracy in counting the teeth but also greatly improved the performance of the final segmentation map.

# IX. Conclusion

A method for segmenting and counting tooth occurrences on panoramic radiographs is proposed in this work. The results show that the method has the capacity to aid medical practice by serving as a pre-processing and analysis step for dental pictures. To our knowledge, the segmentation and tooth counting results are the best in the literature. Furthermore, this is accomplished with a relatively modest training dataset of 105 photos. The method suggested in this paper is based on image processing stages applied to the neural network's sigmoid output before binarization. Although the goal of this research is to segment teeth occurrences, the technology given here can be used to solve similar challenges in other domains, like separating cell instances.

# X. References

1. Fariza, Arna & Arifin, Agus Zainal & Astuti, Eha. (2020). Automatic Tooth and Background Segmentation in Dental X-ray Using U-Net Convolution Network. 144-149. 10.1109/ICSITech49800.2020.9392039.
2. Li, Yunxiang & Wang, Shuai & Wang, Jun & Zeng, Guodong & Liu, Wenjun & Zhang, Qianni & Jin, Qun & Wang, Yaqi. (2021). GT U-Net: A U-Net Like Group Transformer Network for Tooth Root Segmentation.
3. Sivagami, S. & Chitra, P. & Kailash, G. & Muralidharan, S.R.. (2020). UNet Architecture Based Dental Panoramic Image Segmentation.187-191. 10.1109/WiSPNET48689.2020.9198370.
4. Jader, Gil & Fontineli, Jefferson & Ruiz, Marco & Lima, Kalyf & Pithon, Matheus & Oliveira, Luciano. (2018). Deep Instance Segmentation of Teeth in Panoramic X-Ray Images. 400-407. 10.1109/SIBGRAPI.2018.00058.
5. Yin, Xiao-Xia & Sun, Le & Fu, Yuhan & Lu, Ruiliang & Zhang, Yanchun. (2022). U-Net-Based Medical Image Segmentation. Journal of Healthcare Engineering. 2022. 1-16. 10.1155/2022/4189781.
6. Dhar, Mrinal & Yu, Zeyun. (2022). S-R2F2U-Net: A single-stage model for teeth segmentation.
7. Zheng, Youyi & Shen, Yuefan & Shen, Kaidi. (2022). TeethGNN: Semantic 3D Teeth Segmentation with Graph Neural Networks. IEEE Transactions on Visualization and Computer Graphics. PP. 1-1. 10.1109/TVCG.2022.3153501.
8. Hu, Zhanli & Wu, P.Z. & Gui, Jianbao & Chen, Yao & Zheng, Hairong. (2015). Teeth segmentation using dental CT data. Proceedings - 2014 7th International Conference on BioMedical Engineering and Informatics, BMEI 2014. 81-84. 10.1109/BMEI.2014.7002747.
9. Vila Blanco, Nicolás & Cootes, Timothy & Lindner, Claudia & Tomás Carmona, Inmaculada & Carreira, María. (2019). Fully Automatic Teeth Segmentation in Adult OPG Images. 11-21. 10.1007/978-3-030-11166-3\_2.
10. Gao, Yonghui & Li, Xiaoxiao. (2013). Teeth segmentation via semi-supervised learning. Proceedings of the 2013 6th International Conference on Biomedical Engineering and Informatics, BMEI 2013. 558-563. 10.1109/BMEI.2013.6747003.
11. Mairaj, Danish & Wolthusen, Stephen & Busch, Christoph. (2010). Teeth Segmentation and Feature Extraction for Odontological Biometrics. 323-328. 10.1109/IIHMSP.2010.86.
12. Thakur, Niharika & Juneja, Mamta & Jindal, Prashant. (2017). Tooth/Teeth Segmentation and modeling from X-ray/CT images: A Survey. International Journal of Control Theory and Applications. 10.
13. Grzegorzek, Marcin & Trierscheid, Marina & Papoutsis, Dimitri & Paulus, Dietrich. (2010). A Multi-stage Approach for 3D Teeth Segmentation from Dentition Surfaces. 521-530. 10.1007/978-3-642-13681-8\_61.
14. Bozkurt, Hakan & Karagol, Serap. (2020). Jaw and Teeth Segmentation on the Panoramic X-Ray Images for Dental Human Identification. Journal of Digital Imaging. 33. 10.1007/s10278-020-00380-8.
15. Hao, Jin & Liao, Wen & Zhang, Yueling & Li, Peilin & Yi, Jianru & Peng, Jerry & Zhao, Zeu & Chen, Zhang & Shi, Wenxuan & Chen, Tingyu & Zhou, Bowen & Feng, Yang & Fang, Bing & Hu, Haoji & Yang, Howard & Li, Erping & Liu, Zuozhu & Zhao, Zhihe. (2020). Clinically Applicable System For 3D Teeth Segmentation in Intraoral Scans using Deep Learning. 10.21203/rs.3.rs-103285/v1.
16. Le, Binh & Xia, James & Chang, Yu-Bing & Zhou, Xiaobo. (2009). An Interactive Geometric Technique for Upper and Lower Teeth Segmentation. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention. 12. 968-75. 10.1007/978-3-642-04271-3\_117.