An Attention U-Net for Semantic Segmentation of Dental Panoramic X-ray images

Dishant Padalia
Department of Electronics and Telecommunications
K.J. Somaiya College of Engineering
Mumbai, India
dishant.padalia@somaiya.edu

Kush Vora
Department of Computer Engineering
K.J. Somaiya College of Engineering
Mumbai, India
kush.v@somaiya.edu

Deepak Sharma
Department of Computer Engineering
K.J. Somaiya College of Engineering
Mumbai, India
deepaksharma@somaiya.edu

***Abstract*—Dental disorders can lead to serious implications such as heart attack or strokes if not diagnosed and treated early. The diagnosis of these disorders differs from dentist to dentist due to differences in perception, poor x-ray quality because of noise, and different types of patients. As a result, there is an urgent need to create automated, AI-driven diagnostic solutions for dental disorders. Deep learning solutions have shown outstanding results in automated medical image analysis tasks. Our work proposes a U-Net with attention blocks to segment teeth from dental panoramic X-rays. The proposed Attention U-Net consists of four encoding and decoding blocks and achieved Dice Coefficient, IOU score, Specificity, and F1 Score of 0.9318, 0.8724, 0.9910, and 0.9379, respectively. The Attention U-Net achieves a 1.7% better Dice Coefficient score and 2.9% better IOU (Intersection Over Union) score than one of the best segmentation models, the U-Net. The segmented output can be used in computer-aided diagnostic (CAD) systems to detect various mouth disorders, helping the dentist diagnose the problem efficiently and accurately.**

Keywords— Image Segmentation, Dental Panoramic X-Rays, U-Net, Image Processing

I. INTRODUCTION

There are different types of problems that can affect one's teeth. Some of the common issues are chipped or broken teeth, misaligned teeth (malocclusion), pus-causing tooth infection (abscess), and tooth decay leading to cavities. The cause of these disorders depends from person to person; for some, it is from birth, while for some, it is due to poor hygiene or accidents. These tooth disorders are diagnosed by dentists using different dental instruments. However, dental x-rays are the most widely used techniques to detect, diagnose, treat, and monitor a wide variety of oral conditions and diseases. Dental X-rays detect small concealed areas of decay infections, periodontal disease, abscesses or cysts, developmental anomalies, and malignancies that a visual dental examination cannot detect.

Dental x-rays are of two types: intraoral x-rays and extraoral x-rays. Intraoral x-rays are prevalent and are further categorised according to what dentists search for. Bitewing x-rays are used to examine a specific region of the mouth. These x-rays assist the dentist in detecting tooth decay and changes in the jawbone caused by gum infection. Periapical X-rays capture the whole tooth from the root to the crown, allowing the clinician to spot changes in bone structure or root structure. Full mouth survey x-rays are a mix of bitewing and periapical images that are used as a baseline to measure oral health. Occlusal X-rays assist the dentist in determining whether or not a child's teeth have broken through the gums. Extraoral x-rays are used to diagnose problems with the jaw and skull. In the proposed work, panoramic X-rays, a kind of extraoral x-ray, are utilised. These x-rays assist the dentist in detecting a variety of abnormalities allowing treatment before the situation worsens.

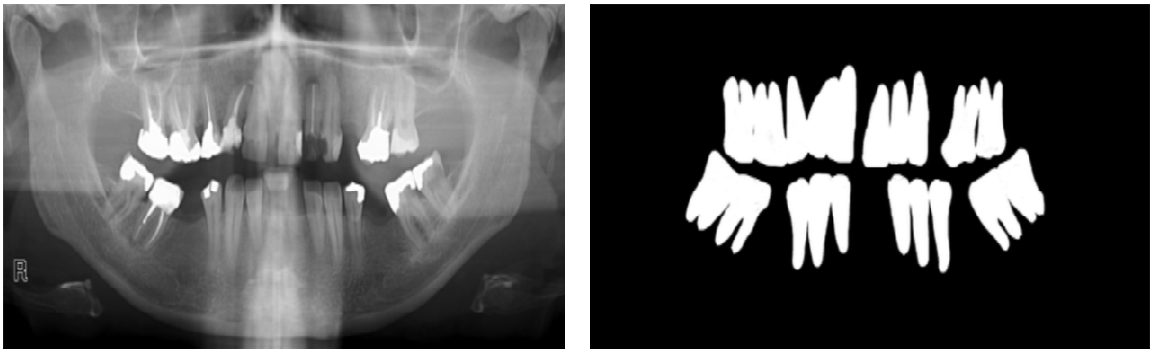
However, due to a lack of automated approaches, dental x-ray diagnosis differs from dentist to dentist based on their expertise and visual perception. Furthermore, a variety of circumstances make identifying disorders using dental x-rays challenging. Some of these issues are restoration artefacts (prostheses), different types of patients, and poor x-ray quality owing to noise or low contrast. To address this issue, tooth segmentation techniques have been developed, which serve as the foundation for a variety of computer-aided diagnostic (CAD) systems capable of diagnosing various dental problems. Dental segmentation not only aids in the diagnosis of tooth disorders but also assists the dentist in designing surgical plans. Moreover, the segmented output can be used for age and forensic identification. With the advent of deep learning solutions, the healthcare industry has revolutionised, and a number of solutions have been proposed which make the detection and diagnosis of diseases easy and accurate. Ronnenberg et al. [1] proposed the U-Net architecture, which is used extensively in a number of biomedical image segmentation tasks. This study implemented several image processing algorithms and deep learning-based image segmentation techniques. Of these, the Attention U-Net with four encoder-decoder pairs fared better than other techniques and achieved a 1.7% better Dice Coefficient score and 2.9% better IOU score than the five encoder-decoder pair U-Net model, and 27.23% better Dice Coefficient score and 38.04% better IOU score compared to the OTSU thresholding.

II. LITERATURE REVIEW

In the last few decades, numerous image processing techniques have been used in the medical domain. These methods generate high-quality, clear images for scientific and medical study, ultimately assisting clinicians in disease diagnosis. A number of researchers have also proposed different image processing-based techniques to segment teeth from dental X-rays. Eyad Haj Said et al. [2] proposed a mathematical morphology-based teeth segmentation approach. To segment the teeth, a number of morphological operations were carried out. The closing top-hat transformation was applied to suppress the teeth and accentuate internal noise. The teeth were then separated from the background using a thresholding process. This was followed by connected component labelling, in which pixels

2022 5th International Conference on Advances in Science and Technology (ICAST) | 978-1-6654-9263-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAST55766.2022.10039666

in the thresholded picture were grouped and labelled based on their connectivity. Finally, a refining procedure based on



(a) (b)

Fig. 1. Sample Images from the dataset. (a) panoramic dental x-ray, (b) ground truth annotated by experts

geometric features such as area and dimension was applied to the connected components. The researchers also employed grayscale contrast stretching to boost the efficacy of the segmentation approach in their study. This was primarily done for low-contrast dental imaging. YY. Amer et al. [3] developed a system for segmenting wisdom teeth, which they then utilised to identify and retrieve information including tooth shape and deviation angle. To segment the wisdom teeth, the researchers used a three-step pipeline that includes the pre-processing, area of interest (ROI) extraction, and post-processing. Contrast enhancement, OTSU thresholding, morphological dilation, connected component labelling, artefact removal, and image multiplication were performed on the X-Ray pictures in the first phase (pre-processing). Finally, the image was post-processed to eliminate the neighbouring teeth from the ROI that had been extracted. The steps in post-processing were similar to those in pre-processing.

The traditional image segmentation method employs edge detection filters and other mathematical techniques. However, in recent years, deep learning technology has been used to analyse medical images in a variety of domains, with exceptional results in applications such as segmentation and disease detection. G. Jader et al. [4] created a mask region-based convolutional neural network (Mask R-CNN) to conduct instance teeth segmentation from panoramic X-ray images. The R-CNN architecture was trained on 193 mouth images with an average of 32 teeth. ResNet101 was used to extract features, which were then combined to form a feature pyramid network from which the area of interest (ROI) was retrieved. Due to a paucity of training data, pre-trained weights from the MSCOCO dataset were employed in the Mask-R CNN architecture's backbone (ResNet 101). Using the transfer learning approach, the architecture achieved accuracy, precision, recall, specificity, and F1 scores of 98%, 94%, 84%, 99%, and 88% on 1224 unseen images, respectively. JH. Lee et al. [5] evaluated a Mask R-CNN by training it on 846 images with tooth annotations with only 20 images in the testing and validation set. An oral radiologist manually performed the teeth annotations. Different data augmentation techniques were employed to expand the dataset to 1024 training samples. The R-CNN approach achieved an F1-score and mean IOU of 0.875 and 0.877 respectively. S. Tian et al. [6] demonstrated using sparse voxel octree and 3D convolution neural networks to segment and classify tooth types in a 3D-Dental model (CNN). The researchers introduced a two-level hierarchical feature learning to tackle the issue of misclassification in similar teeth categories. They also devised a three-level hierarchical segmentation approach with an accuracy of 89.81% for segmenting teeth-gingiva and inter-teeth. Low contrast and uneven intensity distribution can cause problems segmenting the tooth boundary and root. Y. Zhao et al. [7] proposed a two-Stage Attention Segmentation Network (TSASNet) to overcome these problems. TSASNet consists of an attention model that has attention modules capable of localising the tooth region. In addition, a CNN architecture is used to segment the actual tooth area from the attention maps generated through the attention model. The proposed architecture achieved a dice score and accuracy of 92.72% and 96.64% on a dataset containing 1500 dental panoramic X-ray images.

The U-Net design, introduced by O Ronneberger et al. [1], is frequently used in medical picture segmentation. Given the U-Net architecture's success, several U-Net variations have been developed, each attempting to increase network performance while retaining the U-shaped structure. A comparative study was performed by E. da Silva Rocha et al. [8] where four different U-Net architectures, viz. a simple U-Net, Nano-Net (with a pre-trained MobileNet), Double U-Net, and Dual Channelled U-Net, were compared. All the U-Net architectures were trained on 1500 panoramic X-ray images with and without augmentation of the dataset. The double U-Net performed best with an accuracy and dice score of 96.59% and 92.886%, respectively, on the augmented data. The model also performed well without augmentations, with a dice score and accuracy of 92.695% and 96.552%, respectively. Compared to other models, double U-Net performed well on detailing the root of the teeth. The Nano-Net architecture also performed well on the images with only 235,000 trainable parameters. TL. Koch et al. [9] used a fully convolutional neural network based on a U-Net for semantic segmentation of dental panoramic radiographs. The images in the dataset were large enough to be sent at once to the

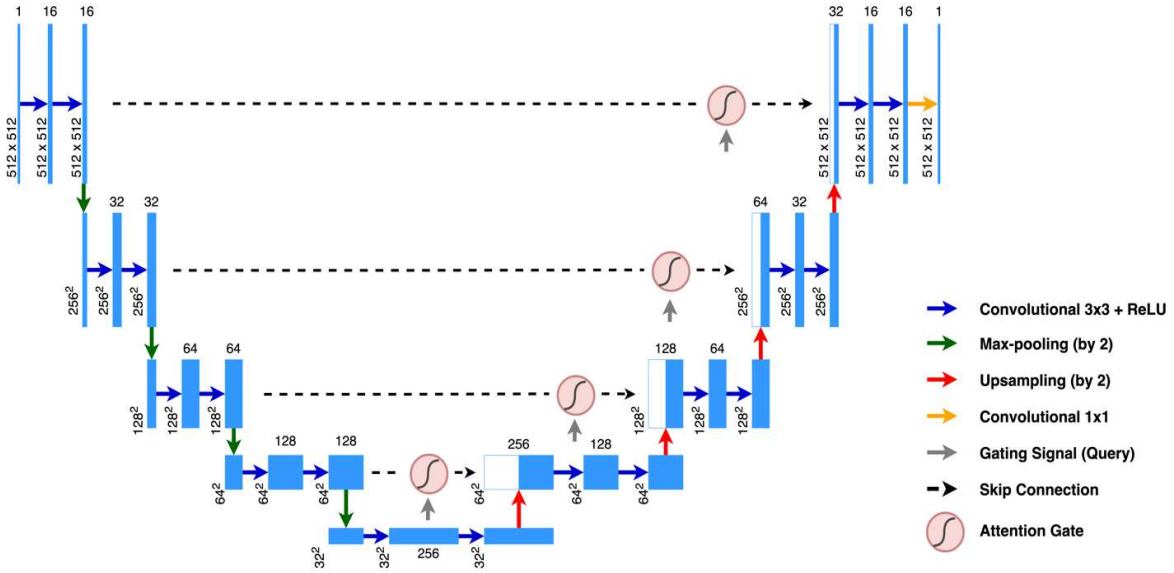


Fig. 2. Architecture diagram of the proposed architecture

architecture: hence patching was employed to feed random windows of a specified size to the model. The original radiographs were of the size 1127× 1991; after patching, over a billion radiographs were generated and used to train the network. The patches were generated randomly, leading to standardization variability, thus reducing overfitting. The researchers in the given work also experimented with augmentation during testing and exploited the symmetry. The system achieved a dice score of 0.936 by ensembling four different U-Net architectures. Most studies in the literature review have a key flaw in that they have not assessed their recommended architecture using IOU. IOU is a critical metric for evaluating segmentation model's performance since it quantifies the proportion of overlap between the predicted output and the ground truth.

III. Methods and Materials

A. Dataset

In the presented work, the deep learning algorithms were trained and evaluated on 116 panoramic dental x-rays with segmented mandibles by Amir Abdi et al. [10]. The images in the dataset ranged in size, with some measuring 2900 by 1250 pixels. The dental x-rays were captured using a digital panoramic x-ray machine called the Soredex CranexD. The dental x-ray scans ranged from healthy to completely and partially edentulous. Two expert dentists manually segmented these panoramic dental X-rays. Sample image from the dataset with its corresponding annotated ground truth can be seen in Fig 1.

B. Data Preprocessing

The dataset used in the proposed work consisted of 116 panoramic dental x-rays with manually annotated ground truths. Different augmentation techniques were used to deal with data scarcity to expand the dataset's size. The image and the corresponding ground truths were rotated to the left and right by 15 degrees. In addition to the rotation, the images were blurred with a maximum kernel size of 5. This was done to mimic real-world clinical settings. The augmentations expanded the dataset from 116 to 464 x-rays. These 464 photos were divided in an 80-20 ratio to create the train and test sets. The train set included 371 photos, whereas the testing set had 93 images.

C. Traditional Image Processing Techniques

Image segmentation techniques break apart an image into multiple groups (image segments). This is done to minimize the image complexity and to analyze the image. To put it another way, segmentation is the process of labelling pixels. The use of computer algorithms to enhance images or extract useful information is known as digital image processing. In this study, we have employed three prominent image processing techniques to segment dental X-ray images.

1) Global Thresholding

Image thresholding segmentation is the most basic type of image segmentation. It produces a binary or multi-color image by imposing a threshold value on the pixel intensity of the source image. Global thresholding uses a bimodal image to segment the entire image using a single thresholding value. It iterates through all the pixels in the image and assigns each pixel a value of 0 or 255 based on the grey level of that particular pixel value. On the other hand, local thresholding selects unique threshold values for all partitioned sub-images derived from the whole picture. In this research, the global thresholding value of 127 was selected after multiple trials, and the "Binary" method of thresholding was used, i.e., if the pixel intensity is higher than the specified threshold, the value is set to 255 (white), otherwise it is set to 0 (black). One downside of this sort of threshold is that it works badly on poorly lighted or degraded pictures and produces a significant number of false negatives and false positives.

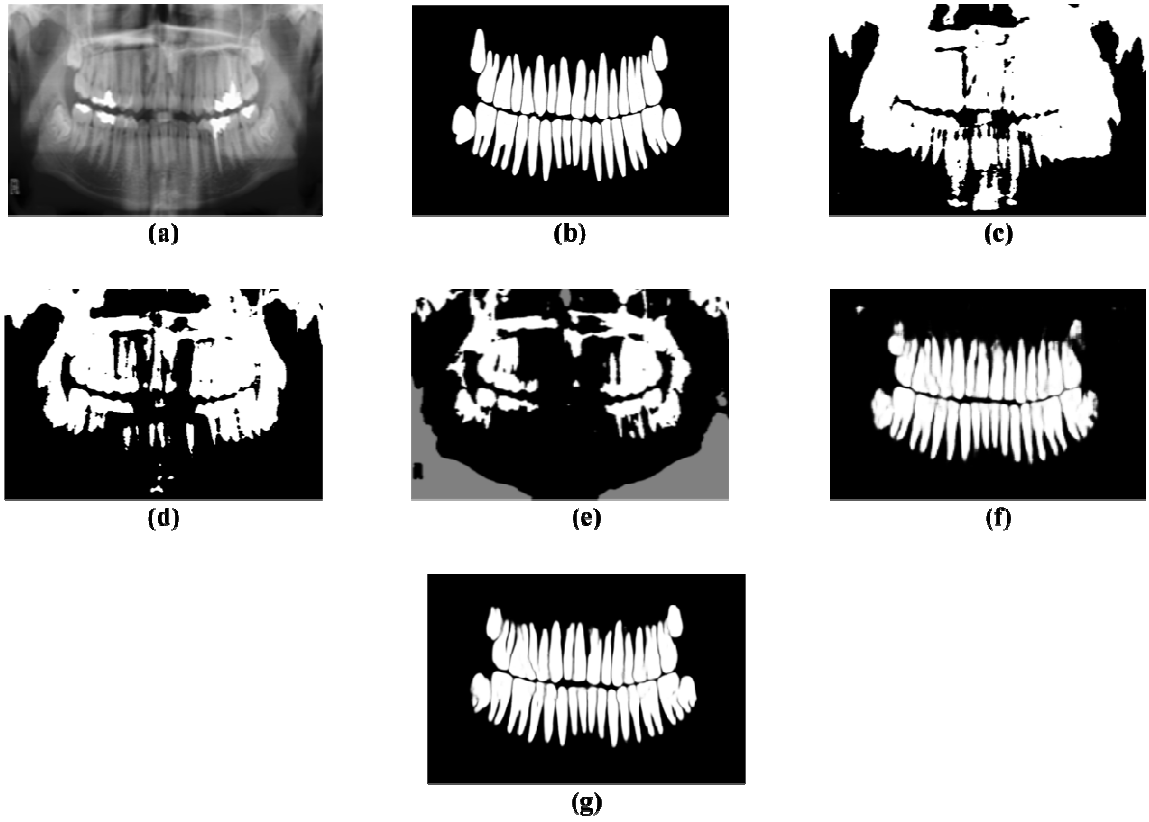


Fig. 3. Segmented output generated from different segmentation techniques. (a) panoramic dental x-ray, (b) original ground truth, (c) OTSU thresholding, (d) global thresholding, (e) region-based segmentation, (f) U-Net, (g) Attention U-Net

2) Ostu Thresholding

The main shortcoming of simple thresholding is that we must manually select the threshold value, and experimenting with different values until we discover the ideal one is time-consuming and tiresome. To address this problem, Nobuyuki Otsu [11] devised a thresholding approach in 1979 that automatically chooses the best value without any manual work. The Otsu thresholding technique employs an iterative approach, going through all potential threshold values and computing the measure of spread for pixels on each side of the threshold. It chooses the threshold value with the lowest dispersion.

3) Region Based Segmentation

Region-based segmentation [12] is a segmentation technique that grows patches in an image recursively and is also known as the region growing technique. An image area is a collection of connected pixels with similar properties. Pixels may be equivalent in intensity, colour, and so forth. There are two approaches to region-based segmentation: top-down and bottom-up. The top-down approach selects a seed pixel at random and then grows the region surrounding the pixel until all of the pixels are exhausted. In contrast, the seed in the bottom-up strategy is chosen exclusively from items of interest. In the study, we have employed the top-down approach, also called the region growing method; we start with some pixels as the seed pixel and then recursively check the subsequent pixels for any similarity until no similarity is left.

D. Deep Learning Techniques

Deep Learning algorithms have fared better than conventional computer vision and image processing algorithms for many applications [13]. The accuracy of many biological segmentation tasks has increased dramatically with the advent of the U-Net [1] architecture.

1) U-Net

O. Ronneberger et al. were the first to propose the U-Net architecture. The architecture originated from conventional CNN and was designed to localise abnormality areas. The design is structured like a U and has two paths: the contraction path (or encoder) and the expansion path (or decoder). The encoder consists of convolutional and max pooling layers and is responsible for capturing context and high-level spatial information in the image. In contrast, the decoder, consisting of transposed convolutional concatenation layers followed by regular convolutional layers, aids localization. The U-Net architecture can take images of any size as the input because it does not contain dense layers. The U-Net architecture is capable of image localization by predicting each pixel in an image and can produce good prediction results even if trained on a minimal number of image samples. In the presented work, a number of U-Net architectures were built and trained with varying numbers of encoder-decoder pairs (or depth).

2) Attention U-Net

The major shortcoming of regular U-Nets for medical imaging is that it focuses on all the image

regions and not specifically on the crucial portions. To overcome this, Ozan Oktay et al. [14] proposed the Attention U-Net architecture, which was initially presented for semantic segmentation on pancreatic images but can be extrapolated to various tasks. The primary distinction between conventional U-Net and Attention U-Net is that the latter includes an attention gate that focuses on target structures of various shapes and sizes while suppressing irrelevant areas. There are two main types of attention, hard and soft. Hard attention highlights the relevant areas by cropping the image. Hard attention chooses only one region at a time and hence is non-differentiable and requires reinforcement learning to train. Soft attention, on the other hand, considers several aspects of the image. High-importance areas are multiplied by larger weights, whereas low-importance areas are multiplied by smaller weights. As a consequence, regions with larger weights get more attention.

Soft attention was used in the presented work, implemented at the skip connections, and suppresses irrelevant regions, reducing the number of redundant features. The attention gate is fed two vectors, g and x. Vector g is drawn from the network's lowest layer and provides improved feature representation. Vector x is taken from a higher layer, contains good spatial information, and is twice the size of vector g. Vector x is reduced in size by a strided convolution, and the 2 vectors are added elementwise. The resulting vector is passed through a ReLU activation, followed by a sigmoid layer that scales the vector values between 0 and 1, with characteristics closer to 1 being more relevant. Finally, the attention coefficients are upsampled to their original dimensions, multiplied by the original vector x, and forwarded along the skip connection. We designed an Attention U-Net with four encoding-decoding layers in the presented work. Each encoding layer consists of eight layers in total, seven in the convolution block and one Max Pooling layer. The encoding block comprises two sets of convolutional, batch normalization, and ReLU activation layers. The convolutional layer has a filter size of 3x3 and padding as “same,” i.e., the output size was equal to the input size. In addition to these six layers, a dropout layer was added with a 20% rate, i.e., randomly drops 20 percent of the input neurons. The Max Pooling layer in the encoding block had a pool size of 2x2 to downsample the input image. After the four encoding blocks, two sets of convolutional, batch normalization, and ReLU activation layers and a dropout layer were added. Subsequently, there were four decoding blocks, each consisting of a gating signal, an attention block, and an Upsampling layer. The gating signal included a set of convolutional, batch normalization, and activation layers. The attention block creates a skip connection between the corresponding encoding layer and the gating signal to capture significant contextual information. The output from the attention block is forwarded to the Upsampling layer. Lastly, after the four decoding blocks, the output is passed to a 1x1 convolutional layer, and then a sigmoid activation is applied to obtain the same dimension output as the input image. Fig 2. showcases the architecture diagram of the proposed attention U-Net.

IV. Results and Discussion

In the presented work, a number of image processing techniques as defined in section 5, were employed to segment the dental panoramic x-rays. Table 1. shows a comparison of different techniques and their performance on the dataset. OTSU thresholding performed the best with an IOU score of 0.4920 and dice coefficient of 0.6595.

TABLE 1 PERFORMANCE OF THE IMAGE PROCESSING TECHNIQUES ON THE DATASET

Algorithm	IOU	Dice Coeff	Specificity	Precision	F1 Score
GLOBAL	0.4890	0.6568	0.8356	0.7078	0.7478
OTSU	0.4920	0.6595	0.8933	0.8579	0.6980
REGION BASED	0.3653	0.5351	0.7568	0.4787	0.7164

Since the image processing techniques did not perform well on the dataset, a number of deep learning architectures were trained. U-net with different numbers of encoder-decoder pairs were trained and tested on the dataset. The U-Net with a depth of 5 (5 encoder-decoder pairs) performed better than other U-Net architectures and achieved an IOU of 0.8430 and dice score of 0.9148. Table 2. shows the performance metrics of the different U-Net architectures trained in our research.

TABLE II PERFORMANCE OF THE U-NET ARCHITECTURES ON THE DATASET

Model	IOU	Dice Coeff	Specificity	Precision	F1 Score
UNET 3	0.7653	0.8670	0.9823	0.9683	0.9612
UNET 4	0.8368	0.9112	0.9850	0.9776	0.9735
UNET 5	0.8430	0.9148	0.9840	0.9761	0.9060
UNET 6	0.8144	0.8976	0.9779	0.9646	0.9016

Table 3 illustrates that Attention U-Net performed much better than regular U-Net architectures. This is because standard U-Nets focus on the entire image as a whole, capturing irrelevant portions as well as vital areas. The proposed Attention U-Net, on the other hand, has an attention gate that automatically suppresses extraneous information and concentrates on the relevant regions. In this study, several Attention U-Net models were trained and tested with varying numbers of encoder-decoder pairs, and eventually the Attention U-Net architecture with four encoder-decoder pairs performed the best.

TABLE III PERFORMANCE OF THE ATTENTION U-NET ARCHITECTURES ON THE DATASET

Model	IOU	Dice Coeff	Specificity	Precision	F1 Score
Attn-UNET 3	0.7758	0.8736	0.9896	0.9824	0.9634
Attn-UNET 4	0.8724	0.9318	0.9910	0.9862	0.9379
Attn-UNET 5	0.8559	0.8559	0.9772	0.9548	0.9524
Attn-UNET 6	0.8658	0.9280	0.9905	0.9867	0.9079

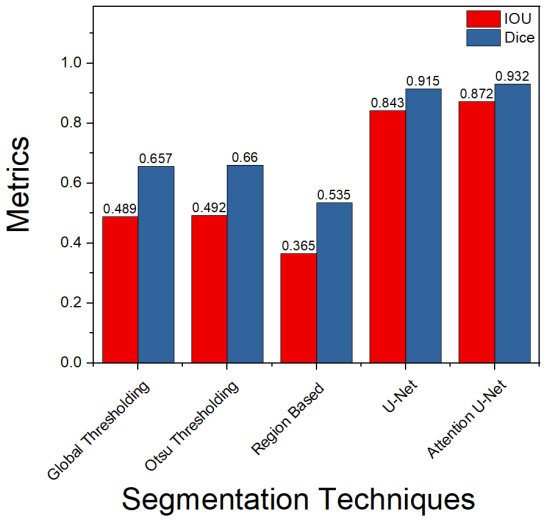


Fig. 4. A Quantitative Comparison of Different Segmentation Techniques with Respect to Dice Coefficient and IOU Score

The output generated from each of the segmentation techniques employed in this work can be seen in Fig 3 and their dice score and IOU score can be observed in Fig 4. The attention U-Net performed best on the given dataset with IOU and dice score of 0.8724 and 0.9318 respectively. Table 4 demonstrates the comparative study of the performance of the proposed model with the architectures discussed in the literature study. The dice coefficient is a popular metric for assessing the efficacy of an image segmentation model. Our model outperformed previously proposed architectures. Although, the system proposed by S. Sivagami et al. performs marginally better than our model in terms of dice coefficient, but has significantly lower specificity and precision compared to the architecture proposed in this study.

TABLE IV COMPARISON WITH OTHER ARCHITECTURES

Model	Dice Coeff	Specificity	Precision	F1 Score
Y Zhao [6]	0.9272	-	-	-
É da Silva Rocha [8]	0.9288	-	-	-
S Sivagami [15]	0.94	0.95	0.93	0.93
Proposed Architecture	0.9318	0.9910	0.9862	0.9379

V. Conclusion

Poor dental hygiene increases one's susceptibility to gum disease and tooth decay. This not only makes eating and speaking difficult, but it can also impact confidence and intensify social anxiety. Furthermore, studies have shown that poor oral hygiene can harm other body organs, perhaps leading to heart attacks or strokes. As a result, it is critical to diagnose specific oral problems early to begin a correct treatment regimen. However, manual diagnosis of dental problems is subjective and varies from dentist to dentist. As a result, there is an urgent need to create automated, AI-driven diagnostic solutions for these illnesses. We designed a U-Net with attention blocks for tooth segmentation in the proposed study. Our proposed Attention U-Net provided better segmentations than previous work. The incorporation of attention blocks to the standard U-Net design improves segmentation output by concentrating on vital dental features rather than unimportant background characteristics. The segmented output can be used in various computer-aided diagnosis systems that detect tooth diseases from dental x-rays. Furthermore, the segmented output can help the dentist plan surgery. The suggested architecture may be expanded and utilized for a variety of biomedical segmentation problems.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[2] Said, E. H., Nassar, D. E. M., Fahmy, G., & Ammar, H. H. (2006). Teeth segmentation in digitized dental X-ray films using mathematical morphology. IEEE transactions on information forensics and security, 1(2), 178-189.

[3] Amer, Y. Y., & Aqel, M. J. (2015). An efficient segmentation algorithm for panoramic dental images. Procedia Computer Science, 65, 718-725.

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

[5] Lee, J. H., Han, S. S., Kim, Y. H., Lee, C., & Kim, I. (2020). Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. Oral surgery, oral medicine, oral pathology and oral radiology, 129(6), 635-642.

[6] Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., & Cheng, X. (2019). Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access, 7, 84817-84828.

[7] Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., & Meng, D. (2020). TSASNet: Tooth segmentation on dental panoramic X-ray images by Two-Stage Attention Segmentation Network. Knowledge-Based Systems, 206, 106338.

[8] da Silva Rocha, É., & Endo, P. T. (2022). A Comparative Study of Deep Learning Models for Dental Segmentation in Panoramic Radiograph. Applied Sciences, 12(6), 3103.

[9] Koch, T. L., Perslev, M., Igel, C., & Brandt, S. S. (2019, April). Accurate segmentation of dental panoramic radiographs with U-NETS. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 15-19). IEEE.

[10] Abdi, A., and S. Kasaei. "Panoramic dental X-rays with segmented mandibles." Mendeley Data, v2 (2020).

[11] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

[12] Gould, S., Gao, T., & Koller, D. (2009). Region-based segmentation and object detection. Advances in neural information processing systems, 22.

[13] O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2019, April). Deep learning vs. traditional computer vision. In Science and information conference (pp. 128-144). Springer, Cham.

[14] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[15] Sivagami, S., Chitra, P., Kailash, G. S. R., & Muralidharan, S. R. (2020, August). Unet architecture based dental panoramic image segmentation. In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) (pp. 187-191). IEEE.