



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

**February 2024**

**Master's Degree Thesis**

**A Study on Multi-Class Teeth  
Segmentation in Dental 2D Panoramic  
X-ray Images**

Graduate School of Chosun University  
Department of Information and Communication  
Engineering

**Ghafoor Muhammad Afnan**

# **A Study on Multi-Class Teeth Segmentation in Dental 2D Panoramic X-ray Images**

치과 2D 파노라마 X-ray 영상에서 멀티클래스  
치아 분할에 관한 연구

February 23, 2024

Graduate School of Chosun University  
Department of Information and Communication  
Engineering

Ghafoor Muhammad Afnan

# **A Study on Multi-Class Teeth Segmentation in Dental 2D Panoramic X-ray Images**

Advisor: Prof. Bumshik Lee

This thesis is submitted to Chosun University in partial fulfillment of  
the requirements for a Master of Engineering degree.

October 2023

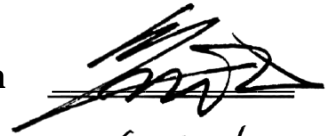
Graduate School of Chosun University  
Department of Information and Communication  
Engineering

Ghafoor Muhammad Afnan

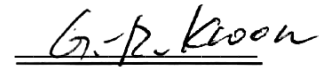
This is to certify that the master's thesis of  
**Ghafoor Muhammad Afnan**

has been approved by the examining committee for the  
thesis requirement for the master's degree in engineering.

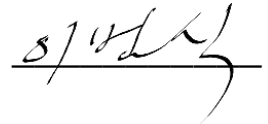
**Committee Chairperson: Prof. Jae-Young Pyun**



**Committee Member: Prof. Goo-Rak Kwon**



**Committee Member: Prof. Bumshik Lee**



December 2023

Graduate School of Chosun University

## Table of Contents

<b>List of Figures</b> .....	<b>iii</b>
<b>List of Tables</b> .....	<b>iv</b>
<b>Abstract</b> .....	<b>v</b>
<b>한 글 요 약</b> .....	<b>vii</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 Overview .....	1
1.2 Research Objective.....	6
1.3 Thesis Layout .....	6
<b>2. Related Works</b> .....	<b>7</b>
<b>3. Proposed Network</b> .....	<b>14</b>
3.1 Overall Architecture .....	14
3.2 Pre-Processing .....	15
3.3 Encoder and Decoder .....	17
3.4 Swin Transformer Blocks.....	18

3.5 Teeth Attention Block .....	19
3.6 Loss Functions and Supervision.....	23
<b>4. Experimental Results.....</b>	<b>26</b>
4.1 Implementation Details .....	26
4.2 Experimental Setup .....	26
4.3 Results and Discussion.....	31
4.4 Ablation Study.....	37
<b>5. Conclusion .....</b>	<b>42</b>
<b>Acknowledgment.....</b>	<b>43</b>
<b>References .....</b>	<b>44</b>
<b>Publications .....</b>	<b>49</b>

## List of Figures

Figure 1-1.	Panoramic image obtained from the supervisory.....	1
Figure 3-1.	Block diagram of pre-processing.....	15
Figure 3-2.	Overall architecture of the proposed teeth segmentation network (TAB: Teeth Attention Blocks).....	16
Figure 3-3.	Teeth Attention Block in the proposed teeth segmentation network.....	20
Figure 3-4.	Visual Comparison between the proposed model and the conventional method.....	33
Figure 3-5.	Visual Comparison between the proposed model and the conventional method.....	34



## List of Tables

Table 1.	Comparisons between the proposed method and conventional ones.....	29
Table 2.	Comparisons of running times under the stopping criteria in experimental setup.....	30
Table 3.	Components of the Variations in the Ablation Study.....	38
Table 4.	Performance Metrics for Variations in the Ablation Study.....	40

## **Abstract**

### **A Study on Multi-Class Teeth Segmentation in Dental 2D**

#### **Panoramic X-ray Images**

Ghafoor Muhammad Afnan

Advisor: Prof. Bumshik Lee

Department of Information and Communication Engineering

Graduate School

Chosun University

This thesis proposed a cutting-edge multiclass teeth segmentation architecture that integrates an M-Net-like structure with Swin Transformers and a novel component named Teeth Attention Block (TAB). Existing teeth image segmentation methods have issues with less accurate and unreliable segmentation outcomes due to the complex and varying morphology of teeth, although teeth segmentation in dental panoramic images is essential for dental disease diagnosis. An M-Net-like structure with Swin Transformers and TAB is incorporated into the proposed novel teeth segmentation model. The proposed TAB utilizes a unique attention mechanism that focuses specifically on the complex structures of teeth. The attention mechanism in TAB precisely highlights key elements of teeth features in panoramic images, resulting in more accurate segmentation outcomes. The proposed architecture effectively captures local and global contextual information, accurately defining each

tooth and its surrounding structures. Furthermore, a multiscale supervision strategy is employed, which leverages the left and right legs of the U-Net structure, boosting the performance of the segmentation with enhanced feature representation. The squared Dice loss is utilized to tackle the class imbalance issue, ensuring accurate segmentation across all classes. The proposed method was validated on a panoramic teeth X-ray dataset, which was taken in a real-world dental diagnosis. The experimental results demonstrate the efficacy of the proposed architecture for tooth segmentation on multiple benchmark dental image datasets, outperforming existing state-of-the-art methods in objective metrics and visual examinations.

## 한 글 요약

### 치과 2D 파노라마 X-ray 영상에서 멀티클래스 치아 분할에 관한 연구

가푸어 무함마드 아프난  
지도교수: 이범식  
정보통신공학과  
조선대학교 대학원

본 논문에서는 M-Net 기반 Swin Transformers(Swin Transformers) 및 TAB (Teeth Attention Block)라는 새로운 구성 요소를 갖는 다중 클래스 치아 분할 아키텍처를 제안하였다. 기존 치아 영상 분할 방법은 복잡하고 다양한 치아 형태로 인해 정확하고 신뢰할 수 없는 분할 결과가 발생하는 문제가 존재한다. 제안하는 치아 분할 방법에서 TAB는 치아의 복잡한 구조에 초점을 맞추는 새로운 어텐션 메커니즘을 이용한다. TAB의 어텐션 메커니즘은 치과 파노라마 X-ray 영상에서 치아 특징의 주요 요소를 정확하게 강조하여 보다 정확한 분할 결과를 도출할 수 있도록 도움을 준다. 또한 제안하는 치아 분할 아키텍처는 지역 및 전역의 치아 정보를 효과적으로 캡처하여 각 치아와 그 주변 구조를 정확하게 분할할 수 있도록 한다. 또한 U-Net의 왼쪽 다리(left-leg)와 오른쪽 다리(right-leg)를 활용하는 다중 스케일 감독 방법 채택하여 향상된 영상의

특징 표현으로 분할 성능을 향상시킨다. 제안 방법에서는 클래스 불균형 문제를 해결하기 위해 자승 Dice 손실 함수를 적용하여 보다 정확한 치아 분할할 수 있도록 하였다. 제안된 방법은 실제 치과 진단에서 촬영한 파노라마 치아 X-ray 데이터 세트에서 검증하였다. 실험 결과는 여러 치과 이미지 데이터 세트에서 제안된 치아 분할 아키텍처의 성능을 검증하였고 객관적인 분할 성능 수치 및 시각적 검사에서 기존 치아 분할 방법을 크게 능가하는 것을 실험적으로 검증하였다.

# 1. Introduction

## 1.1 Overview

Dental imaging is essential for oral healthcare because it helps in the diagnosis and treatment of various dental conditions [1]. For example, dentists can recognize jaw-related conditions and identify anatomical characteristics such as teeth, maxillary sinus, and alveolar bone using panoramic dental X-ray images [2]. Furthermore, the precise measurements offered by this technique provide technical assistance in the preoperative diagnosis, surgical planning, and postoperative evaluation [3].



Figure 1-1. An example of panoramic dental X-ray image

Teeth image segmentation is a vital process in computer-assisted dentistry diagnostics and serves as an initial step in analyzing the tooth status. Dentists

can use panoramic radiographs to assess a range of dental conditions, including missing teeth, dental development, impacted teeth, and adjacent relationships [2]. This is achieved through image segmentation. Current technology employs a ground-truth identification mechanism for panoramic X-ray images as shown in Fig. 1-1 and utilizes a segmentation architecture to generate precise segmentation outcomes that can potentially facilitate clinical diagnosis. Panoramic dental X-ray scans and tooth image segmentation technology are essential components of computer-assisted dentistry diagnostics because they enable accurate measurements and provide a comprehensive view of the jaw and teeth. For example, accurate teeth segmentation from panoramic images is essential for diagnosing serious dental conditions like periodontitis, which is a severe gum infection leading to potential tooth loss. Through detailed segmentation, dentists can identify anomalies in the tooth and surrounding structures, such as enlarged periodontal ligament spaces or bone loss. However, while segmentation is crucial for initial evaluations, more specific imaging techniques, like bitewing X-rays or CBCT, are often required for a thorough diagnosis and treatment plan. As a result, teeth segmentation remains a cornerstone in dental diagnostic tools.

The precise categorization of teeth into distinct groups poses a significant challenge in dental image analysis despite its critical importance in various applications such as orthodontic treatment planning, dental implant surgery, and forensic odontology. Manual, semiautomatic, and automatic approaches

have been devised to segment teeth in dental images [3]. Despite the progress made in this field, dental restorations, malocclusions, and pathological conditions can affect the segmentation performance.

Artificial intelligence applications in dentistry are growing, as they help practitioners increase patient safety while simplifying complicated procedures and offering predictable outcomes [5]. Medical image analysis uses deep learning techniques that provide several benefits, including anomaly detection, image segmentation, and classification [3]. Hence, AI systems can potentially improve health data outcomes, lower healthcare costs, and advance medical research [6]. For example, in dental image analysis, deep learning techniques have shown promising results in segmenting and classifying teeth and dental structures, resulting in enhanced diagnosis and treatment planning for various dental conditions [3], [6].

Convolutional neural networks (CNNs) have emerged as the primary technique for image segmentation in dental imaging because of their ability to collect local spatial data and learn feature representation [7]. However, recent advancements in deep learning architectures, such as transformers, have demonstrated their potential to outperform CNNs in several computer vision tasks by effectively modelling long-range dependencies and global context [9]. The progress of CNN-based segmentation models in accurately segmenting teeth and other dental structures from background noise enables the precise analysis and diagnosis of dental conditions. With the increasing availability of



large-scale dental image datasets and the ongoing advancements in deep learning techniques, CNN-based models are expected to play an even more significant role in dental imaging by facilitating rapid, accurate, and automated image analysis.

Boundary box filters, also known as region proposal methods, have been extensively used in medical image segmentation tasks to improve the performance of deep learning models by focusing on specific areas of interest. Boundary box filters were used to identify nodules on CT scans in [11, 14] when segmenting lung nodules accurately. Similarly, Oktay et al. [12] used boundary box filters to increase the pancreatic segmentation accuracy. Fan et al. [13] successfully segmented lesions in colonoscopy images using boundary box filters. By focusing on specific areas of interest in medical images, Nader et al. [10] demonstrated that boundary box techniques can enhance segmentation precision. In dental imaging, boundary box techniques have been used to segment teeth and other dental structures by identifying the regions of interest on panoramic radiography and dental cone-beam computed tomography (CBCT) [2]. These studies suggest that boundary box filters have the potential to considerably enhance the accuracy and efficiency of dental and medical image segmentation, thereby improving diagnosis and treatment planning for a variety of conditions.

Transformers have recently emerged as solid deep-learning architectures with excellent results in various computer vision applications, including image

segmentation [9]. Transformers successfully represent long-term dependencies and the global environment using self-attention processes, enabling them to record connections between pixels or features in an input image regardless of their geographical distance. This characteristic makes transformers a potentially suitable choice for dental image segmentation tasks where accurately capturing the context and relationships between teeth and their surrounding structures is crucial.

A cutting-edge deep neural network model is designed to segment teeth into 32 distinct categories based on panoramic dental radiography images. The proposed model achieved an accuracy rate of 97.26%, a Dice Similarity Coefficient of 0.9102, and a Jaccard Index of 0.8501, all of which represent significant improvements over previous methodologies, showing significant advancements in dental 2D X-ray image segmentation. These enhancements are the result of a novel methodology integrating CNNs and transformers, combined with TAB, as a novel addition. Utilizing these blocks facilitates the model's ability to focus on regions of interest, thereby effectively capturing both local and global contexts. TABs address significant challenges associated with dental image segmentation such as overlapping structures and varied tooth shapes, thereby enhancing the overall performance and accuracy of the model.

## **1.2 Research Objective**

The proposed tooth segmentation model offers significant contributions to dental diagnostics and treatment planning. By providing precise tooth segmentation, the proposed architecture enables early detection and diagnosis of various dental diseases. For instance, it can facilitate the detection of periodontal diseases or dental caries by identifying changes in tooth shape or the appearance of lesions. Moreover, the accuracy achieved in tooth segmentation can greatly assist in creating detailed treatment plans. Orthodontists, for example, can use the segmentation results to plan braces placement or determine the necessity of tooth extraction in overcrowded mouths. These applications underscore the clinical relevance and potential impact of the proposed segmentation model in the field of dental medicine.

## **1.3 Thesis Layout**

This paper is organized as follows. An overview of related works on dental picture segmentation is presented in Section 2, emphasizing deep learning-based methods. Section 3 describes the proposed deep learning architecture. The experimental setup, including the dataset, assessment criteria, and implementation information, is presented in Section 4. Section 5 presents experimental results and comparisons with other state-of-the-art models. Finally, Section 6 concludes the paper by addressing possible future research directions in this field.

## 2. Related Works

Medical image segmentation has witnessed significant advancements with the advent of deep learning-based approaches. Yamanakkanavar and Lee [14] presented a novel M-SegNet architecture that used a global attention CNN model for automated brain MRI segmentation. The base architecture, M-Net [41], is also used for brain image segmentation. Badrinarayanan et al. [15] developed SegNet, a deep convolutional encoder–decoder architecture, with significant success in medical imaging applications. Gu et al. [16] proposed CE-Net, a situation-encoding network for 2D medical-picture segmentation. Lin et al. [17] presented an efficient piecewise training approach for deep-structure models in semantic segmentation. Slic-Seg, a minimally interactive segmentation approach for the placenta using fetal MRI, was proposed by Wang et al. [18]. Lee et al. [19] proposed a patchwise U-Net structure for automated brain MRI segmentation. Deep learning-based techniques have demonstrated exceptional efficacy in various medical image-related assignments, underscoring their potential.

Dental image segmentation has recently emerged as a popular research area. In this regard, deep learning-based algorithms have exhibited encouraging results. This section provides a comprehensive review of the literature on dental image segmentation, focusing on deep learning-based techniques. The intricate nature and proximity to adjacent anatomical structures render dental image segmentation a formidable task. Nevertheless, deep learning

methodologies have been demonstrated to overcome these obstacles and exhibit encouraging outcomes. Various deep-learning architectures and techniques have been utilized in numerous studies on dental image segmentation. These include U-Net [20], Mask R-CNN [21], and ResNet [22]. The results of these studies suggest that deep learning-based methodologies can yield positive results in terms of precision and expediency in the dental segmentation of image tasks.

Researchers have employed several methods to enhance dental segmentation. Tekin et al. [23] segmented and numbered teeth in dental imaging panoramic images using a Mask R-CNN, yielding high-quality segmentation masks. Similarly, Yang et al. [24] developed an automated system for dental image analysis that included dental image diagnostic knowledge, drastically reducing the amount of human labor necessary for data preparation. Xia et al. [25] presented a method that successfully separated individual teeth from CT images of the upper and lower natural contact-scanned teeth.

CNNs have been extensively used in various medical image segmentation applications because of their ability to gather local spatial inputs and generate hierarchical representations [7]. Several CNN-based algorithms for tooth segmentation have been introduced for dental image analysis. Hou et al. [2] introduced Teeth U-Net, a segmentation approach for tooth panoramic X-ray images that uses a U-Net structure to capture contextual semantics and improve image contrast. Similarly, Tekin et al. [6] developed an improved

tooth segmentation and numbering technique for bitewing radiographs using a machine-learning algorithm based on the U-Net architecture. These studies showed that CNNs efficiently segment teeth and dental structures using different dental images.

Recent advancements in deep learning architectures, such as transformers, have shown their potential to outperform CNNs in several computer vision tasks by effectively modelling long-range dependencies and global context [8]. Although Transformers have mainly been used for natural language processing applications, their use in medical-picture analysis is gaining popularity. Transformers have been used for image segmentation, classification, and anomaly detection tasks and have shown promising results in various medical domains. The utilization of bounding box techniques to concentrate on regions of interest has been observed in medical imaging. This approach serves to decrease the intricacy of segmentation tasks. Nader et al. [12] proposed an automatic tooth segmentation method for panoramic X-rays using deep neural networks with bounding boxes to enhance the accuracy of the segmentation process. El Jurdi et al. [11] presented BB-Unet, a U-Net design that includes bounding box priors to improve segmentation results for medical imaging tasks. The aforementioned studies demonstrated the capacity of bounding box methodologies to enhance the segmentation outcomes and augment the overall efficacy of deep-learning models.

U-Net [20] has been widely used for dental segmentation tasks. Koch et al. [26] employed a U-Net architecture to segment panoramic images of teeth, resulting in enhanced sample segmentation using a more compact and less complex network design. Similarly, Kong et al. [27] proposed an efficient encoder–decoder network (EED-Net) for the fast and accurate segmentation of maxillofacial images. Zhao et al. [28] developed a two-stage attention segmentation network (TSASNet) to locate and segment teeth in dental panoramic X-ray images that can combine pixel-level contextual information and identify fuzzy tooth areas. Cui et al. [29] proposed a tooth segmentation network (TsegNet) for 3D scanning of dental structures. Some researchers have also improved the U-Net architecture by enhancing the encoder and decoder, modifying the convolutional layers, and improving skip connections.

Attention techniques play a crucial role in boosting the performance of CNNs in medical image analysis tasks. These techniques allow the rescaling of extracted features through skip connections, thereby enhancing high-level representation learning. For example, Jin et al. [30] proposed residual attention U-Net (RAUNet) for liver tumor segmentation, which includes a backbone branch for learning original features and a soft mask branch to reduce noise and enhance positive features. Similarly, Liu et al. [31] introduced the deep residual attention network (DRANet), which improved the feature processing between the encoder and decoder, leading to more accurate lesion-type classification. Moreover, establishing extensive connections between encoders

and decoders can enhance the links between different modules. To address this, Jose et al. [32] proposed the intervertebral disc network (IVD-Net), which utilized a dense technique to link encoders layer by layer, with each encoder processing a distinct image pattern. In addition, Zhang et al. [33] proposed a multiscale densely connected U-Net (MDU-Net) that fuses neighboring feature maps of multiple sizes at high and low levels to improve the encoder, decoder, skip connection performance, and segmentation accuracy. The related works explained above are mentioned in Table I for easy comparison.

By merging these related work sections, it can be observed the evolution of dental image segmentation using deep learning techniques, from the early use of the Mask R-CNN to the more recent incorporation of transformers, bounding box techniques, and improvements to the U-Net structure. These advancements provide a strong foundation for the proposed method and exciting possibilities for future research in this field.

This diverse range of methods demonstrates the ongoing advancements and potential for further improvements in dental image segmentation using deep learning techniques. A novel deep-learning architecture is introduced for tooth segmentation based on panoramic images to address the weaknesses of existing approaches. The advantages of CNNs and transformers are combined with a unique tooth-bounding box technique that improves the accuracy of tooth segmentation while resolving the challenges that currently exist in dental image analysis, going beyond the integration of existing tools.



The critical contribution of the proposed methodology is the strategic integration of CNNs, Transformers, and the novel TAB. CNNs are used to extract features from dental images and to capture specific local characteristics. When encapsulating the global context, a domain in which CNNs fall short, transformers are utilized. Additionally, the proposed model incorporates novel TAB, allowing for a focused understanding of the overall dental arch structure, a factor in previous approaches.

The significant contribution, TAB, is intended to improve segmentation outcomes. This technique addresses the sensitivity issues encountered in previous models by sharpening the focus on the teeth, thereby reducing the impact of noise and irrelevant regions.

The proposed approach makes several significant contributions to the field of dental image analysis.

1. This study proposes a cutting-edge multiclass teeth segmentation architecture that combines an M-Net-like structure with Swin Transformers. This architecture integrates various components to efficiently capture local and global contextual information, enabling the accurate delineation of teeth and their adjacent structures.
2. This study introduces an innovative component called TAB, which plays a crucial role in the proposed architecture. TAB enhances the segmentation

performance by selectively attending to teeth-related features, further improving the accuracy of tooth segmentation.

3. This study incorporates a multiscale supervision strategy by utilizing the left and right legs of a U-Net structure. This strategy aids in precise feature representation and boosts segmentation performance by providing supervision at different scales.
4. A thorough examination was conducted, and it was demonstrated that proposed model outperforms state-of-the-art techniques in several important metrics.

### 3. Proposed Network

#### 3.1 Overall Architecture

A tooth segmentation architecture that integrates an M-Net-like structure with an encoder and decoder, Swin Transformer [34] and TAB is proposed to accurately segment dental images. The proposed architecture aims to capture both local and global contextual information effectively, resulting in the precise delineation of teeth and surrounding structures. The U-Net-like structure consists of an encoder that extracts feature representations through downsampling and a decoder that reconstructs the segmentation mask through upsampling. The skip connections between the encoder and decoder layers preserve spatial information. Additionally, left- and right-leg supervision is employed for the encoder and decoder to facilitate accurate feature representation learning and improve segmentation performance.

Swin Transformer blocks, placed at the bottleneck, effectively capture long-range dependencies using the self-attention mechanism, which models nonlocal information and relationships between distant regions in dental images. This enhanced the model's understanding of complex structures and relationships. Furthermore, TAB in skip connections refine segmentation by focusing on object boundaries, leading to more precise delineations between different teeth and structures. This architecture effectively captures both local and global contextual information, resulting in accurate tooth segmentation.

## 3.2 Pre-Processing

In the proposed approach, pre-processing steps are performed to enhance the overall quality of panoramic images before training and testing the model.

Figure 3-1 shows the block diagram of the pre-processing steps.

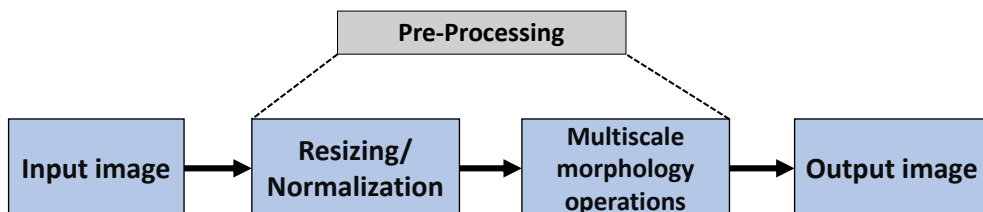


Figure 3-1. Block diagram of pre-processing

As shown in Figure 3-1, the pre-processing steps perform image resizing, normalization of pixel intensities to a range of  $[0, 255]$ , and multiscale morphology in sequence. Multiscale morphology [38] employs a range of morphological operations at different image scales to reduce noise, improve contrast, and highlight the salient features of dental imagery. Such pre-processing is essentially required in refining the input data for the model, ensuring enhanced performance during the training and testing stages.

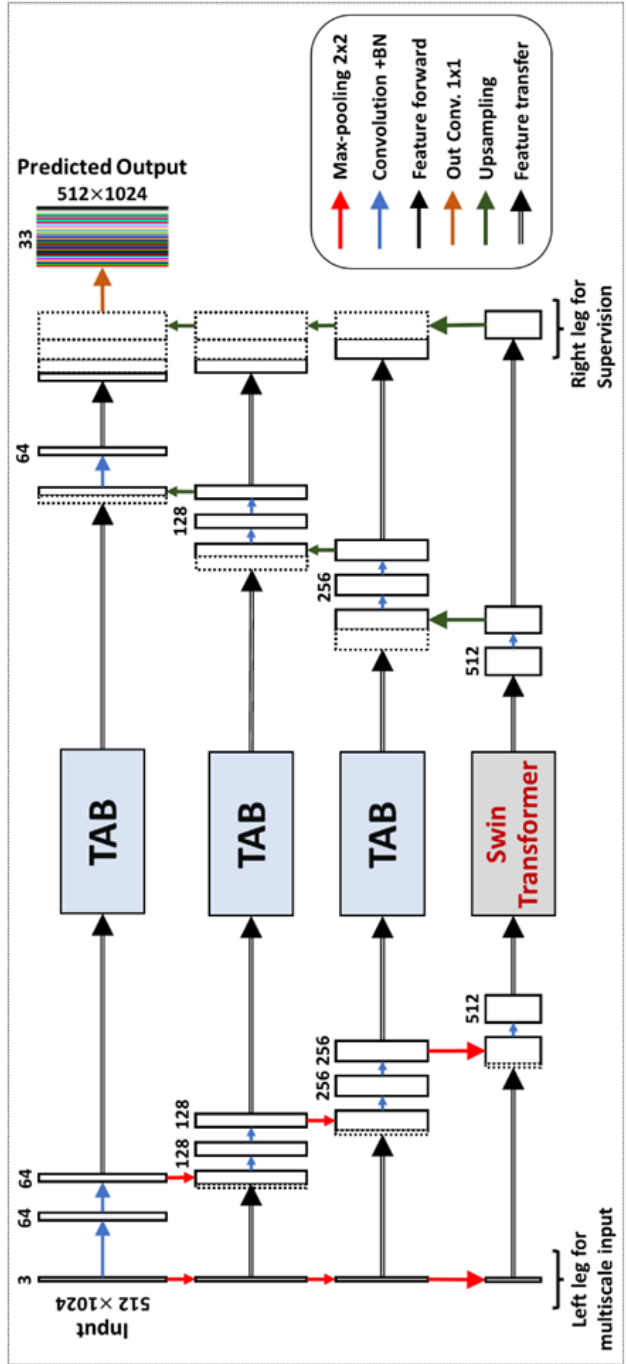


Figure 3-2. Overall architecture of the proposed teeth segmentation network.

(TAB: Teeth Attention Blocks)

### 3.3 Encoder and Decoder

Hierarchical dental image features are extracted in the encoder, where multiple convolutional layers are utilized with a  $3 \times 3$  kernel size in each layer, which is the same size used in U-Net-like architectures. The encoder utilizes these layers to learn varying levels of features from the input dental images, from basic to complex. Batch normalization is performed after each convolution to improve the stability of the model and speed up learning. The activation function known as the Rectified Linear Unit (ReLU) is responsible for introducing non-linearity into the model, enhancing its ability to learn intricate patterns. The inclusion of max-pooling layers is implemented to decrease the spatial dimensions of the feature maps and increase the receptive field. The features retrieved by the encoder substantially impact the overall performance of tooth segmentation. They are responsible for identifying different forms and patterns, hence facilitating accurate tooth segmentation.

The decoder recovers the spatial resolution of the feature maps and reconstructs the segmented teeth. This upsampling is achieved through transposed convolutions, effectively increasing the height and width of the feature maps while preserving their depth. Segmentation masks are reconstructed from these up-sampled feature maps simultaneously, resulting in pixel-wise class predictions for the input dental image.

Moreover, skip connections play a pivotal role in the decoder by bridging the gap between the encoder and decoder layers. These connections send the

feature maps from the encoder to their corresponding decoder layers through the intermediate layers. This process allows for the incorporation of high-resolution details from the encoder's earlier layers with the abstract, lower-resolution features from the deeper layers. This fusion of features aids in the more accurate reconstruction of segmentation masks, as it captures both local details and global context, thereby enhancing the precision of the tooth.

### **3.4 Swin Transformer Blocks**

Swin Transformer [34] is employed in deep learning architectures to effectively capture local and global contextual information using a self-attention mechanism. They divided the input feature maps into non-overlapping local windows, enabling efficient processing and utilizing multihead self-attention layers to learn multiple relationships simultaneously. Swin Transformers merge and shift windows after each self-attention layer to capture long-range dependencies, whereas position-wise feed-forward layers help learn complex nonlinear relationships. The multihead self-attention layers in the Swin Transformer are followed by position-wise feed-forward layers and layer normalization, which allow the Swin Transformer to successfully manage the multiclass tooth segmentation task. Specifically, the multihead self-attention mechanism helps to capture intricate spatial relationships across different parts of the dental image, while the position-wise feed-forward layers enhance the local representations with non-overlapping local windows within input feature maps.

Following each self-attention phase, the Swin Transformer highlights its adaptability by merging and shifting windows, which is for capturing long-range dependencies. Additionally, the inclusion of position-wise feed-forward layers enhances the model's ability to identify complex nonlinear relationships. In aggregate, these methods enhance the effectiveness of the Swin Transformer in addressing dental image segmentation challenges. Since teeth exhibit diverse morphologies, it is essential to recognize subtle patterns and distant relationships in dental images. The Swin Transformers are specifically positioned to address the issues of tooth segmentation in the proposed method. The Swin Transformer blocks are strategically placed in the bottleneck of the proposed design to effectively capture long-range dependencies. The utilization of nonlocal information management is of utmost importance, as it allows the model to effectively analyze complex connections between teeth and different dental diseases.

### **3.5 Teeth Attention Block**

TABs, a key contribution of this study, act as boundary-aware or boundary refinement filters, playing a critical role in segmentation tasks to improve the recognition of boundaries between different objects or structures in an image. TABs enhance the focus of the tooth segmentation model on a dental image by observing a local receptive field around each pixel. The implementation of a filtering operation in this receptive field emphasizes the boundaries of the objects, specifically the edges of single teeth and their adjacent structures.



The attention mechanism employed by TAB plays a crucial role in enhancing the accuracy and precision of the segmentation masks. Specifically, this is achieved by effectively refining the differentiation between the individual and surrounding teeth. In simple terms, TAB reduces the effect of noise and meaningless information in dental images by selectively optimizing the features of teeth and their boundaries while minimizing irrelevant features or noise that are not beneficial to the teeth segmentation task. This improved attention helps achieve more precise and consistent segmentation results, providing a cleaner and clearer illustration of the individual teeth and their

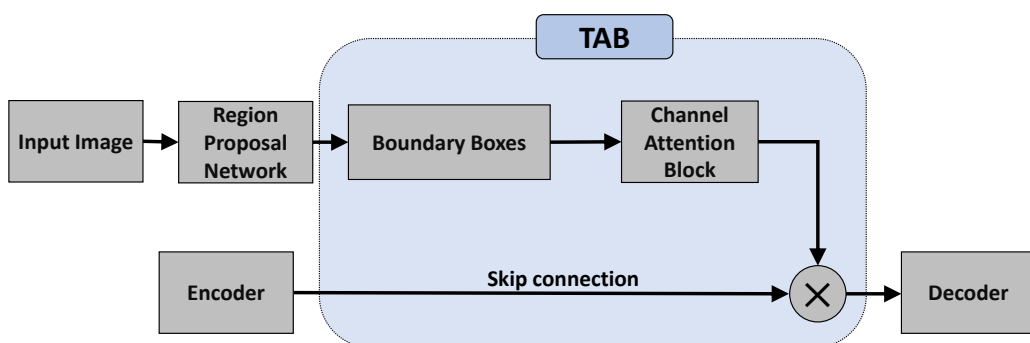


Figure 3-3. Teeth Attention Block in the proposed teeth segmentation network

boundaries in dental panoramic images. The proposed TAB using a self-attention mechanism enables the model to assess the significance of various tooth parts within the image by leveraging acquired contextual knowledge. In the training stage, the TAB assigns increased attention scores to boundary pixels around separate teeth. This indicates that greater attention is given to

the boundary regions while generating feature maps, hence improving the capability of the model to distinguish between individual teeth.

The improvement of the overall segmentation performance is facilitated by the attention capability of the TAB on these boundary regions. Specifically, it aids in enhancing the demarcation of tooth boundaries inside the segmentation masks, minimizing the occurrence of overlapping between neighboring teeth and improving the precision and accuracy of the segmentation process. The integration of the TAB method into proposed model yields a notable advantage, particularly in complex dental images characterized by densely arranged or slightly overlapping teeth, therefore mitigating the limitations of conventional segmentation techniques.

The TABs were incorporated into the skip connections in the tooth segmentation architecture between the encoder and decoder layers which can be seen in Figure 3-3. RPN detects boundary boxes that indicate potential regions containing dental structures. Then, the boundary boxes pass through the Channel Attention Block (CAB), fine-tuning the feature's attention. The utilization of specific filters in TABs significantly improves the segmentation accuracy of the model in accurately delineating complex object boundaries, particularly those pertaining to individual teeth and neighboring structures. The upgraded feature maps are subsequently multiplied with the features from the encoder via the skip connections. After the combination, the merged entities undergo processing by the decoder to achieve final segmentation.

TABs operate by considering a local neighboring receptive field around each pixel and implementing a filtering operation to highlight the boundaries of the teeth. The filtering operation of the TAB, which is one of the novel contributions, can utilize techniques such as convolutional layers and attention mechanisms to learn and target object boundaries. The incorporation of TAB within skip connections confers multiple benefits to the architecture of tooth segmentation, as follows:

- Improved segmentation accuracy: The model can distinguish between each tooth and its surrounding structures by concentrating on object boundaries, thereby producing more precise segmentation masks.
- Smoother and sharper object boundaries: The utilization of TAB has the potential to reduce the presence of unusual or uneven edges within the segmentation masks, thereby resulting in more refined and distinct object boundaries.
- Better handling of overlapping or adjacent objects: Teeth are frequently shown in proximity as well as overlapping in dental images. The implementation of TAB can enhance the model's ability to differentiate between teeth that are adjacent or overlapping with enhanced performance.

In summary, the proposed TAB plays a crucial role in enhancing the precision of tooth segmentation outcomes in the dental image segmentation framework.

This is achieved by highlighting object boundaries and enhancing the differentiation between various teeth and structures.

### **3.6 Loss Functions and Supervision**

The left and right legs of the U-Net structure were employed for supervision during the multiclass tooth segmentation task. Incorporating the multiscale supervision within both the downsampling (encoding) and upsampling (decoding) components of the U-Net, the model is supervised at multiple scales. Such a technique helps to capture and recognize the objects well and accurately segment the different classes of teeth found in dental images. This approach not only facilitates feature representation across various scales but also boosts the differentiation capacity of the model to distinguish between distinct teeth categories.

The proposed model is trained using a custom square Dice loss function. Dice loss, which is commonly employed in medical image segmentation tasks, calculates the overlap between the predicted and ground-truth images, making it ideal for addressing the class imbalance frequently observed in such tasks.

In the conventional Dice loss function, a score is calculated as twice the region of intersection between the predicted and true segmentation maps divided by the total number of pixels in both maps. The Dice loss was

calculated as one minus the Dice score to obtain the best overlap between the predicted and actual segmentations.

Before formulating the loss, the Dice loss function was changed by squaring the pixel values. The following modification, described as square Dice loss, gives greater emphasis to every pixel, which makes the model more sensitive to segmentation boundary changes. The significance of precise segmentation boundaries in dental imaging cannot be overstated, because even minor deviations can significantly affect the quality of the resulting output.

The square Dice loss function includes a normalization factor that avoids division by a zero. The computation involves determining the intersection between the ground-truth segmentation map and the predicted segmentation map while considering the combined sum of the squares of both maps. This causes the neural network to prioritize accurate predictions for each pixel, which improves its precision and recall.

$$L = 1 - \left( 2 \sum (y_t \cdot y_p)^2 + \epsilon \right) / \left( 2 \sum (y_t^2 + y_p^2) + \epsilon \right) \quad (1)$$

where  $L$  denotes the loss function.  $y_t$  and  $y_p$  represent the ground truth and predicted segmentation maps, respectively,  $\epsilon$  is a smoothing factor for avoiding division by zero. The squared Dice loss is chosen for the proposed method due to its efficiency, leading to superior segmentation results. Several loss functions, such as soft Dice loss [44], Tversky loss [43], and Log-Cosh Dice loss [43] functions, which are popularly used in medical

segmentation tasks, were tried. It was observed that the squared Dice loss function achieved significantly better segmentation performance in DSC and JI, etc.

## 4. Experimental Results

The experimental results for the proposed tooth segmentation architecture are discussed in this section. The present study commenced by providing a detailed account of the dataset and the pre-processing procedures employed in the training and testing of the model. Subsequently, the evaluation metrics, experimental setup, and comparison with established methods are discussed. The results are evaluated while the performance of the proposed model is addressed.

### 4.1 Implementation Details

A dataset consisting of dental panoramic images has been compiled through a collaborative effort between a dental college and its students. These panoramic images were annotated meticulously using a supervisory platform, resulting in a detailed categorization of separate teeth across multiple classes. In total, the dataset comprises 540 annotated images. These images were resized to dimensions of  $1024 \times 512$  pixels to ensure computational efficiency and reduce memory demands, with care taken to preserve critical anatomical landmarks.

### 4.2 Experimental Setup

Uniform settings were utilized to train and evaluate the proposed tooth segmentation network, ensuring a fair and consistent comparison with existing methodologies. The Keras framework and an NVIDIA GeForce RTX 3090 graphics processing unit (GPU) were utilized for model training and

evaluation. The dataset was divided into training (70%), validation (15%), and testing (15%) datasets. The Swin Transformer model with  $2 \times 2$  regions was applied and trained over 50 epochs. The initial learning rate was set to  $10^{-4}$  and subsequently adaptively decreased to  $10^{-7}$  to address the potential overfitting issue. A batch size of two was maintained throughout the training process. Dropout layers are added after the encoder convolution layers to overcome overfitting. Furthermore, a strategy to decrease the learning rate by 10% was employed if the validation loss did not decrease over five consecutive epochs.

The effectiveness of the proposed teeth segmentation model was quantitatively evaluated using a total of five standard evaluation metrics, which are Accuracy (ACC), Jaccard Index (JI) [35], Precision [36], Recall [36], and Specificity [37]. These metrics are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$\text{JI} = \frac{|P \cap G|}{|P \cup G|} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$



$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

where True Positive (TP) and True Negative (TN) represent the number of pixels accurately classified as teeth and non-teeth, respectively. Conversely, FP (False Positive) and False Negative (FN) refer to the number of pixels incorrectly categorized as teeth and non-teeth, respectively.

These metrics offer a comparable scale from zero to one, with one indicating an exact match between the predicted and actual values. Higher scores across these parameters denoted better segmentation performance, indicating that the model was efficient in accurately segmenting teeth and distinguishing between them and other structures in the dental images. These evaluation metrics are applied to provide a comprehensive performance assessment of the proposed tooth segmentation model, facilitating its comparative analysis with other established methods in the field.

Table I

Comparisons between the proposed method and conventional ones

Models	ACC	DSC	JI	Precision	Recall	Specificity
U-Net [20]	0.9720	0.7602	0.6871	0.7458	0.8366	0.9725
Attention_U-Net [12]	0.9720	0.7846	0.7132	0.7557	0.8391	0.9725
ResNet-50 Attention U-Net [39]	0.9721	0.7875	0.7172	0.7487	0.8544	0.9726
Swin U-Net [40]	0.9712	0.6348	0.5296	0.6107	0.7192	0.9721
Modified-U-Net [10]	<b>0.9726</b>	0.9004	0.8489	0.7898	0.9366	0.9728
Proposed	<b>0.9726</b>	<b>0.9102</b>	<b>0.8501</b>	<b>0.8046</b>	<b>0.9389</b>	<b>0.9730</b>

Table II

Comparisons of running times under the stopping criteria in experimental setup

Models	Run-times (in minutes)	DSC	JI
U-Net [20]	61	0.7556	0.6815
Attention_U-Net [12]	59	0.7796	0.7071
ResNet-50 Attention U-Net [39]	61	0.7804	0.7096
Swin U-Net [40]	60	0.5432	0.4302
Modified-U-Net [10]	61	0.9003	0.8490
<b>Proposed</b>	<b>60</b>	<b>0.9102</b>	<b>0.8501</b>

### 4.3 Results and Discussion

This section presents a comprehensive analysis and discussion of the performance of the proposed model for dental segmentation. Several well-established segmentation models, including the traditional U-Net [20], Attention U-Net [12], ResNet-50 Attention U-Net [39], Swin U-Net [40], and a Modified U-Net [10], which is identical to BB-Unet [11], were compared to proposed segmentation model. The performance of the proposed model was evaluated using multiple critical metrics, including ACC, Dice Similarity Coefficient (DSC), JI, precision, recall, and specificity. Each tooth in an X-ray image is aimed to be segmented into 32 categories based on the World Dental Federation (FDI) notation [42], where each tooth is categorized into #11 to #18, #21 to #28, #31 to #38, and #41 to #48. Each pixel is classified into a specific number with multiclass 32-categorized pixels, and the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are also measured to obtain objective evaluation metrics for segmentation. TP, TN, FP, and FN are computed for each tooth type, considering tooth number as one class and all other teeth as the other, and this process was repeated for each tooth type in each class.

A computational cost analysis was performed, where the training times for each method were measured under identical experimental conditions. The proposed method required approximately 60 minutes for training to achieve the segmentation performance in Table II. Other methods achieve DSC values

of 0.7602, 0.7846, 0.7875, 0.6348, and 0.9004 and run-times of 45, 48-, 48-, 68-, and 64-minute training times for [20], [12], [39], [40] and [10], respectively. These values are obtained from the above-mentioned experimental setup. Since the stopping criteria and learning rates are variable for each method during training, it is not difficult to judge the superiority of the complexity-performance trade-off. The change in segmentation performance for each method was investigated by setting a similar run-time by adjusting the stopping criteria and learning rate values. Table III shows the comparison of segmentation performances, such as DSC and JI, under almost identical run times. As shown in Table III, the proposed method can achieve significantly higher segmentation performance under similar complexity. It also indicates that the proposed method requires much lower run times to achieve the same segmentation performance.

The accuracy score of the proposed (0.9726) is comparable to that of the other models. However, based on the Dice Coefficient (0.9102) and JI (0.8501), the proposed model significantly outperformed the other models. These scores demonstrate that the proposed model distinguishes true positives while minimizing false positives and false negatives. In addition, the precision and recall scores of 0.8046 and 0.9389, respectively, provided further evidence. This table demonstrates the superior performance of the proposed model compared to the others.

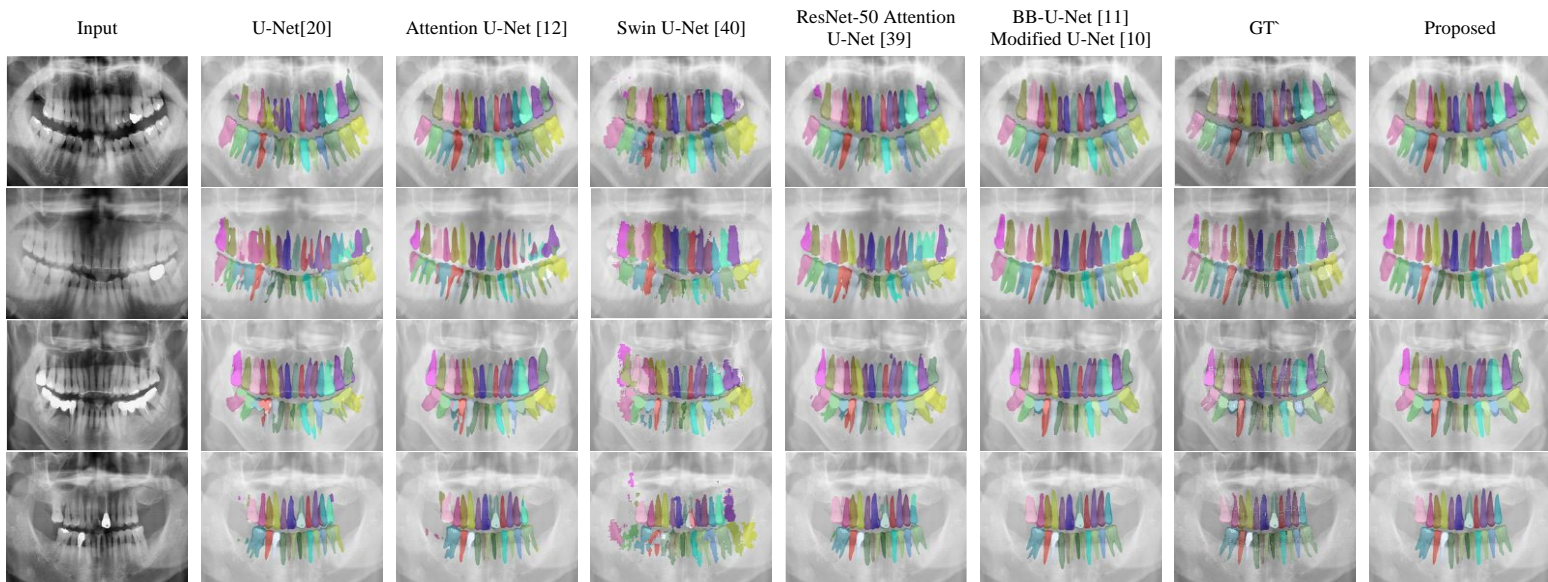


Figure 3-4. Visual Comparison between the proposed model and the conventional method

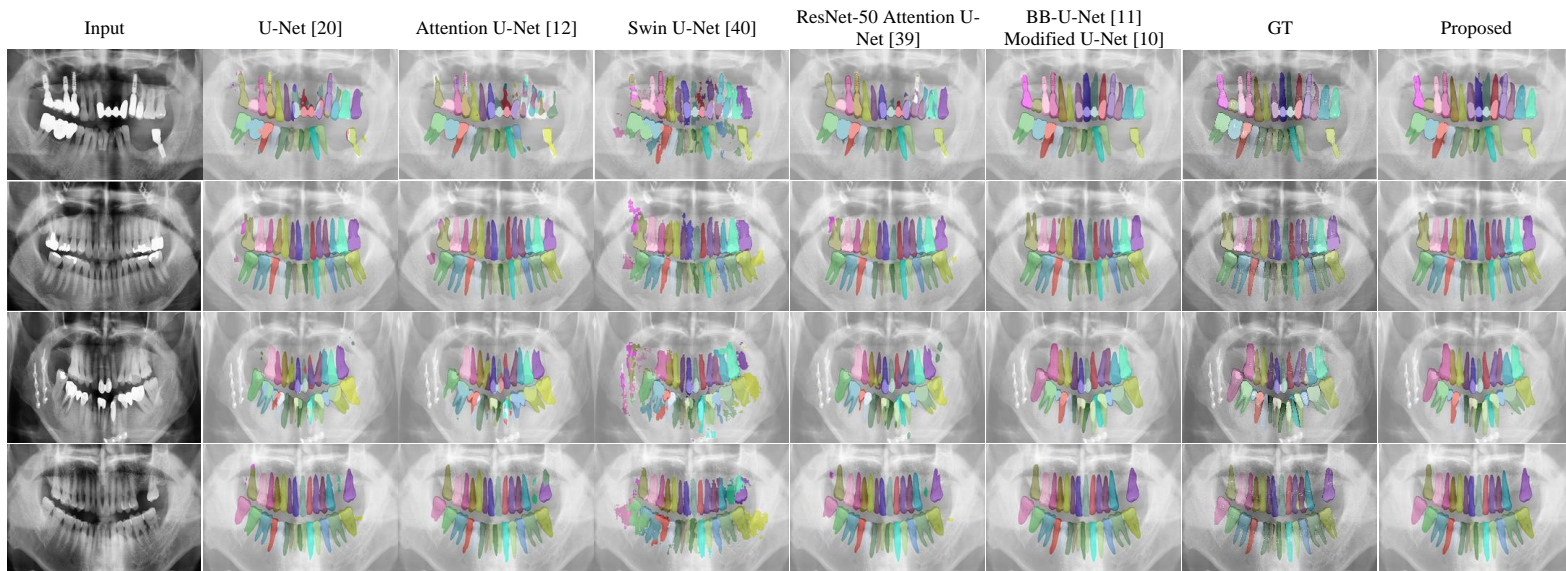


Figure 3-5. Visual Comparison between the proposed model and the conventional method

The performance of each model was critically analyzed. In this study, the traditional U-Net model exhibited commendable performance in terms of ACC, Dice score, and other metrics. However, it does not surpass the performance of the proposed model. Similarly, although the Attention U-Net and ResNet-50 Attention U-Net demonstrated improvements over the traditional U-Net model, they still did not match the performance of the proposed model. The performance of the Swin U-Net model did not align with those of the other models, whereas the Modified U-Net showed a performance close to that of the proposed model.

The superior performance of the proposed model can be attributed mainly to the incorporation of the Swin Transformer and boundary boxes and the application of a modified loss function utilizing the squared Dice loss. These design decisions enabled the proposed model to learn and segment dental structures in the input images, thereby improving the performance across all evaluation metrics.

The segmentation results of all models were visually compared in addition to a quantitative comparison, which can be seen in Figure 4 and Figure 5. This visual comparison provides evidence that the proposed model accurately segments dental structures. Furthermore, it consistently produced more accurate and consistent segmentation outcomes, highlighting the benefits of



the Swin Transformer, boundary boxes, and modified loss function in the proposed dental segmentation model.

Furthermore, the improved generalizability of the approach used for different dental images can be primarily attributed to the addition of TAB and the squared Dice loss function. The use of TABs in dental imaging improves the accuracy of boundary delineation and increases the level of detail in individual tooth analyses. This innovative approach has demonstrated effectiveness in decreasing the impact of noise and artifacts that are commonly found in dental images. A closer analysis of row 1 of Fig. 3, illustrates the effectiveness of the proposed approach. For example, in the case of tooth #48 (FDI notation), most existing models struggle to precisely define the boundaries. However, the approach exhibits exceptional precision, providing segmentation outcomes that closely match the ground truth. This is largely due to the ability of TAB to focus on the local receptive fields surrounding each pixel, thereby highlighting the boundaries of objects and significantly improving the differentiation between individual teeth and their neighboring teeth.

This analysis highlights the potential advantages of the proposed approach over other approaches, particularly when dealing with complex dental structures and obtaining accurate and uniform segmentation results. The incorporation of TABs into the model demonstrated a marked improvement in the overall performance, indicating its potential as an asset in the progression of dental imaging analysis and diagnostics.

The results presented in the table and figures demonstrate that the proposed custom dental segmentation model outperformed several state-of-the-art models in terms of evaluation metrics and visual comparisons. This significant boost in performance can be primarily attributed to the unique TABs. The use of TAB as a boundary refinement filter significantly enhances the identification of each tooth and its adjacent structures. Although the Swin Transformer and the specific loss function play crucial roles, the use of TAB further propels the precision and efficiency of dental image analysis. This study provides a robust platform for future advances in dental image analysis and enhances the potential impact of dental procedures, including diagnosis, treatment planning, and patient monitoring.

#### **4.4 Ablation Study**

An ablation study was performed to assess the contribution of each component to the proposed tooth segmentation network. Each variation in the network under similar training conditions successively included essential components of the basic U-Net model. The components investigated in this study include Deep Supervision, Swin Transformers, and TAB. The effectiveness of each model, named Variations A to D and the complete proposed network was examined using several important metrics, such as ACC, DSC, JI, precision, recall, and specificity. The performance results for each variable are presented in Tables III and IV. The effectiveness of each model, named Variations A to

Table III

Components of the Variations in the Ablation Study

Variations	U-Net	Deep Supervision	Swin Transformer	TAB
Variation A	✓	✗	✗	✗
Variation B	✓	✓	✗	✗
Variation C	✓	✗	✓	✗
Variation D	✓	✗	✗	✓
Proposed	✓	✓	✓	✓

D and the complete proposed network were examined using several important metrics, such as ACC, DSC, JI, precision, recall, and specificity.

Table IV provides the details of the components of each variation. Variation A is the basic U-Net structure, and each variation includes an additional component, namely Deep Supervision (Variation B), Swin Transformers (Variation C), and TAB (Variation D). Table V shows the segmentation performances for the ablation study. As shown in Table V, the accuracy of the segmentation results gets higher over the variation number. Small gains were observed in DSC and JI in Variation B, where Deep Supervision is used. This improvement is due to better feature propagation throughout the network, enhancing the model's distinction between teeth classes. Swin Transformers yields a marginal enhancement in the DSC, as shown in the result of Variance C. It is because Swin Transformers, with the self-attention mechanism, enables

capturing local and global contextual information, which is a crucial factor for segmenting the complex structure of dental images where each tooth can influence the context of neighboring teeth. The improvement in the performance was notably observed in Variation D, where the proposed TAB is solely performed. TAB enhances model performance by selectively focusing on teeth boundaries, enhancing the accuracy and precision of segmentation masks. TAB refines differentiation between teeth and surrounding structures by assigning higher attention scores to boundary pixels, resulting in more distinct edges of individual teeth. This enhances the model's overall performance, resulting in more accurate and detailed tooth segmentation results which can be observed from the performance metrics.

Table IV lists the performance metrics associated with each variation in the ablation study. This demonstrates that each successive variant, with an additional component, results in a gradual increase in the performance metrics.

The basic U-Net structure was set as a base. Features like Deep Supervision and Swin Transformers were added, and improved results are shown in the performance of models. Among the added featured tools, the most significant boost in performance is shown for the proposed TAB.

Although this thesis work proposes a novel tooth segmentation approach, it has certain limitations that guide future works. The dental images used in this

Table IV

Performance Metrics for Variations in the Ablation Study

Variations	Accuracy	DSC	Jl	Precision	Recall	Specificity
Variation A	0.9720	0.7602	0.6871	0.7458	0.8366	0.9725
Variation B	0.9722	0.7846	0.7132	0.7557	0.8391	0.9725
Variation C	0.9721	0.7644	0.6905	0.7569	0.8477	0.9726
Variation D	0.9725	0.9001	0.8476	0.7908	0.9312	0.9728
<b>Proposed</b>	<b>0.9726</b>	<b>0.9102</b>	<b>0.8501</b>	<b>0.8046</b>	<b>0.9389</b>	<b>0.9730</b>

study contains complete teeth sets with relatively fewer images with dental diseases, which may restrict the learning capability of the proposed model. Although the proposed model achieves promising results in segmenting teeth into many classes, further studies can be feasible with a more extensive set of dental health issues.

## 5. Conclusion

An innovative tooth segmentation model for dental panoramic images is introduced, aiming to enhance the accuracy and efficiency of the segmentation process. It incorporates an M-Net-like structure with Deep Supervision, Swin Transformers, and TAB. The proposed model efficiently leverages local and global contextual information, resulting in significantly more accurate tooth segmentation. In particular, the proposed TABs show remarkable proficiency in highlighting complex dental anatomy and finely delineating tooth borders. The novel attention mechanism embedded in the TAB precisely highlights complex tooth structures, resulting in highly accurate segmentation outcomes. Using multiscale supervision and the squared Dice loss, the proposed architecture effectively tackles class imbalances and enhances feature representation, ultimately achieving precise tooth delineation and surrounding structure definition. The proposed method demonstrated its effectiveness and reliability in dental diagnosis applications on a real-world panoramic teeth X-ray dataset. Furthermore, the proposed method shows the feasibility of automated disease diagnosis and treatment planning owing to the precise segmentation performance. For example, it enables the early detection of periodontal diseases or dental caries by identifying changes in tooth shape or the appearance of lesions. However, although the proposed model achieves significantly better results over the state-of-the-art, the investigation of a more extensive set of dental health issues remains as further studies.

## Acknowledgment

I am profoundly thankful to Allah Almighty, the Most Merciful and Compassionate, for bestowing upon me strength, guidance, and blessings throughout the fulfilling journey of completing this thesis.

I extend my sincere and deepest appreciation to Prof. Bumshik Lee, a mentor for his invaluable guidance, unwavering support, and scholarly insights that have significantly enriched the quality of this research. I am truly grateful for the knowledge and skills I have acquired under his mentorship.

Heartfelt thanks are due to my family for their enduring love, encouragement, and unwavering understanding. Their constant support has been my steadfast pillar of strength.

Finally, I extend my gratitude to my friends and fellow lab mates for their support and encouragement during challenging times. Their camaraderie has been a constant source of motivation and joy.

May Allah's grace continue to illuminate our paths and bless us all abundantly.



## References

- [1] P. Deshpande, P. Jain, and S. Patil, “Review on dental image segmentation and analysis techniques,” *J. Dent. Orofac. Res.*, vol. 16, no. 1, pp. 27–36, 2020.
- [2] S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, and H. Shi, “Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement,” *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106296.
- [3] M. Abadi, M. Ahmadi, and S. Salehi, “A review of dental image analysis methods for tooth segmentation,” *Comput. Methods Programs Biomed.*, vol. 208, Dec. 2021, Art. no. 106207.
- [4] J. Choi, S. M. Kang, S. J. Park, and G. C. Kim, “Dental panoramic image analysis using a deep learning algorithm for the early diagnosis of dental diseases,” *Diagnostics*, vol. 11, no. 1, p. 84, 2021.
- [5] Y. W. Chen, K. Stanley, and W. Att, “Artificial intelligence in dentistry: Current applications and future perspectives,” *Quintessence Int.*, vol. 51, no. 3, pp. 248–257, 2020.
- [6] B. Y. Tekin, C. Ozcan, A. Pekince, and Y. Yasa, “An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs,” *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105547.
- [7] A. M. Alsharif, H. M. Alharbi, and N. M. Alsharif, “Deep learning for medical image segmentation: A review,” *Front. Artif. Intell. Appl.*, vol. 8, Mar. 2021, Art. no. 572366.
- [8] M. H. Alsharif, S. El-Sappagh, M. Elmogy, and A. Riad, “Deep learning in dental image analysis: A review,” *J. Imag.*, vol. 7, no. 2, p. 29, 2021.
- [9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16 16 words: Transformers for image recognition at scale,” 2020, arXiv:2010.11929.
- [10] R. Nader, A. Smorodin, N. D. La Fourniere, Y. Amouriq, and F. Autrusseau. (May 28, 2022). Automatic Teeth Segmentation on Panoramic X-Rays Using Deep Neural Networks. [Online]. Available: <https://hal.science/hal-03671003>

- [11] R. El Jurdi, C. Petitjean, P. Honeine, and F. Abdallah, “BB-UNet: U-Net with bounding box prior,” *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1189–1198, Oct. 2020, doi: 10.1109/JSTSP.2020.3001502.
- [12] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, and B. Glocker, “Attention U-Net: Learning where to look for the pancreas,” 2018, arXiv:1804.03999.
- [13] L. Fan, J. Z. Cheng, Y. H. D. Fang, X. Li, X. Cai, and D. Shen, “Automatic polyp segmentation using a boundary-aware U-Net network in colonoscopy,” *Med. Phys.*, vol. 46, no. 4, pp. 1744–1754, 2019.
- [14] N. Yamanakkanavar and B. Lee, “A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI,” *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104761, doi: 10.1016/j.combiomed.2021.104761.
- [15] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder–decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [16] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “CE-Net: Context encoder network for 2D medical image segmentation,” *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019, doi: 10.1109/TMI.2019.2903562.
- [17] G. Lin, C. Shen, A. van den Hengel, and I. Reid, “Efficient piecewise training of deep structured models for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3194–3203.
- [18] G. Wang, M. A. Zuluaga, R. Pratt, M. Aertsen, T. Doel, M. Klusmann, A. L. David, J. Deprest, T. Vercauteren, and S. Ourselin, “Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views,” *Med. Image Anal.*, vol. 34, pp. 137–147, Dec. 2016.
- [19] B. Lee, N. Yamanakkanavar, and J. Y. Choi, “Automatic segmentation of brain MRI using a novel patch-wise U-Net deep architecture,” *PLoS ONE*, vol. 15, no. 8, Aug. 2020, Art. no. e0236493.
- [20] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015, arXiv:1505.04597.

- [21] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., Venice, Italy, Jun. 2017, pp. 2961–2969.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 770–778.
- [23] Y. Tekin and L. Tekin, “Automatic tooth detection and segmentation in occlusal radiographs using mask R-CNN,” *J. Digit. Imag.*, vol. 31, no. 6, pp. 679–689, 2018.
- [24] J. Yang, Y. Xie, L. Liu, B. Xia, Z. Cao, and C. Guo, “Automated dental image analysis by deep learning on small dataset,” in Proc. IEEE 42nd Annu. Comput. Softw. Appl. Conf. (COMPSAC), vol. 1, Jul. 2018, pp. 492–497.
- [25] Z. Xia, Y. Gan, L. Chang, J. Xiong, and Q. Zhao, “Individual tooth segmentation from CT images scanned with contacts of maxillary and mandible teeth,” *Comput. Methods Programs Biomed.*, vol. 138, pp. 1–12, Jan. 2017.
- [26] T. L. Koch, M. Perslev, C. Igel, and S. S. Brandt, “Accurate segmentation of dental panoramic radiographs with U-NETS,” in Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI), Apr. 2019, pp. 15–19, doi: 10.1109/ISBI.2019.8759563.
- [27] S. Bozóki, Z. Szádóczi, and H. A. Tekile, “Filling in pattern designs for incomplete pairwise comparison matrices: (quasi-)regular graphs with minimal diameter,” 2020, arXiv:2006.01127.
- [28] 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,” *Knowl.-Based Syst.*, vol. 206, Oct. 2020, Art. no. 106338.
- [29] Z. Cui, C. Li, N. Chen, G. Wei, R. Chen, Y. Zhou, D. Shen, and W. Wang, “TSegNet: An efficient and accurate tooth segmentation network on 3D dental model,” *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101949.
- [30] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, “RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans,” *Frontiers Bioeng. Biotechnol.*, vol. 8, Dec. 2020, Art. no. 605132.

- [31] J. Wu, W. Hu, Y. Wen, W. Tu, and X. Liu, “Skin lesion classification using densely connected convolutional networks with attention residual learning,” *Sensors*, vol. 20, no. 24, p. 7080, Dec. 2020.
- [32] J. Dolz, C. Desrosiers, and I. Ben Ayed, “IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet,” in *Proc. Int. Workshop Challenge Comput. Methods Clinical Appl. Spine Imag.*, 2018, pp. 130–143.
- [33] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, “MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation,” 2018, arXiv:1812.00352.
- [34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin Transformer: Hierarchical vision transformer using shifted windows,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 10012–10022.
- [35] L. da F. Costa, “Further generalizations of the jaccard index,” 2021. [Online]. Available: <http://arxiv.org/abs/2110.09619>
- [36] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [37] R. Koch, M. P. Heinrich, and M. Koller, “U-Net for semantic segmentation of teeth panoramic images,” *J. Digit. Imag.*, vol. 32, no. 3, pp. 411–421, 2019.
- [38] J. C. M. Román, V. R. Fretes, C. G. Adorno, R. G. Silva, J. L. V. Noguera, H. Legal-Ayala, J. D. Mello-Román, R. D. E. Torres, and J. Facon, “Panoramic dental radiography image enhancement using multiscale mathematical morphology,” *Sensors*, vol. 21, no. 9, p. 3110, Apr. 2021, doi: 10.3390/s21093110.
- [39] O. Oktay, J. Schlemper, L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention U-Net: Learning where to look for the pancreas,” 2018. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [40] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-UNet: UNet-like pure transformer for medical image segmentation,” 2021, arXiv:2105.05537.
- [41] R. Mehta and J. Sivaswamy, “M-net: A convolutional neural network for deep brain structure segmentation,” in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 437–440, doi: 10.1109/ISBI.2017.7950555.

- [42] R. Hickel, A. Peschke, M. Tyas, I. Mjor, S. Bayne, M. Peters, K. A. Hiller, R. Randall, G. Vanherle, and S. D. Heintze, “FDI world dental federation: Clinical criteria for the evaluation of direct and indirect restorations-update and clinical examples,” *J. Adhes. Dent.*, vol. 12, no. 4, pp. 259–272, Aug. 2010, doi: 10.3290/j.jad.a19262.
- [43] S. Jadon, “A survey of loss functions for semantic segmentation,” in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Via del Mar, Chile, Oct. 2020, pp. 1–7, doi: 10.1109/CIBCB48159.2020.9277638.
- [44] Z. Wang, T. Popordanoska, J. Bertels, R. Lemmens, and M. B. Blaschko, “Dice semimetric losses: Optimizing the dice score with soft labels,” 2023, arXiv:2303.16296.
- [45] A. Ghafoor, S. -Y. Moon and B. Lee, “Multiclass Segmentation Using Teeth Attention Modules for Dental X-Ray Images,” in *IEEE Access*, vol. 11, pp. 123891-123903, 2023, doi: 10.1109/ACCESS.2023.3329364.

## Publications

### International Journal Paper:

1. A. Ghafoor, S. -Y. Moon and B. Lee, "Multiclass Segmentation Using Teeth Attention Modules for Dental X-Ray Images," in IEEE Access, vol. 11, pp. 123891-123903, 2023, doi: 10.1109/ACCESS.2023.3329364.

### International Conference Paper:

1. A. Ghafoor and B. Lee, "Multi-class Segmentation of Panoramic X-ray Dental Images Using a Hybrid ResNet50-U-Net Model," in ISIS, Gwangju, 6-9 Dec, 2023

### Domestic Conference Paper:

1. Afnan Ghafoor and Bumshik Lee, "A Multi-class Teeth Segmentation Method using Attention Modules", 인공지능신호처리 학술대회, Daegu, 2023.
2. Afnan Ghafoor and Bumshik Lee, "A patch-wise M-Net with attention for skin lesion segmentation", in Proceedings of KIIS Autumn Conference, Mokpo, 2022.