International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | June 2026 www.irjet.net p-ISSN: 2395-0072

SEMANTIC SEGMENTATION ON PANORAMIC DENTAL X-RAY IMAGES USING U-NET ARCHITECTURES

Sangeetha S[1], Mr. R. Mohan kumar[2]

[1]PG Scholar, Bio-Medical Department Udaya School of engineering, Kanyakumari, Tamil Nadu, India.
[2]Assistant Professor, Bio-Medical Department, Udaya School of engineering, Kanyakumari, Tamil Nadu, India.

Abstract-*Accurate tooth segmentation from panoramic dental X-ray images is a critical task in computer-aided dental diagnosis, orthodontic treatment planning, and forensic dentistry. However, manual segmentation is time-consuming and prone to inter-observer variability, particularly in cases involving overlapping structures, crowding, and low-contrast tooth boundaries. This study proposes an automated tooth segmentation framework based on the Mask R-CNN deep learning architecture, which combines object detection and instance segmentation to identify and segment individual teeth with high precision. The proposed system incorporates image preprocessing, feature extraction using a pretrained convolutional neural network, region proposal generation, and pixel-level mask prediction. The model is trained and validated on annotated panoramic dental X-ray datasets and evaluated using Intersection over Union (IoU), Dice Coefficient, F1-score, and Accuracy metrics. Experimental results demonstrate that the proposed approach achieves robust segmentation performance and effectively distinguishes individual teeth even in complex dental conditions. The automated framework significantly reduces manual effort, enhances diagnostic consistency, and supports efficient clinical workflows. The findings highlight the effectiveness of Mask R-CNN as a reliable solution for AI-assisted dental image analysis and its potential integration into intelligent dental healthcare systems.*

Keywords: Panoramic Dental X-ray, Mask R-CNN, Tooth Segmentation, Instance Segmentation, Deep Learning, Computer-Aided Dental Diagnosis.

INTRODUCTION

Panoramic dental X-ray imaging is widely used in modern dentistry for comprehensive visualization of teeth, jawbones, and surrounding anatomical structures. It plays a crucial role in dental diagnosis, orthodontic treatment planning, implant placement, forensic identification, and oral disease assessment. However, manual delineation of teeth and related structures from panoramic radiographs is a labor-intensive and time-consuming process that is often affected by inter-observer variability. In addition, challenges such as overlapping teeth, low image contrast, varying tooth shapes, and imaging artifacts make accurate segmentation difficult. Recent advances in deep learning have demonstrated significant potential in automating dental image analysis, particularly through semantic segmentation techniques that can accurately identify tooth regions from panoramic X-ray images [1], [2], [4].

Traditional image processing approaches for dental segmentation rely on handcrafted features, thresholding, edge detection, and morphological operations. Although these methods can provide acceptable results in controlled scenarios, their performance often deteriorates when dealing with complex anatomical variations and noisy radiographic data. Deep convolutional neural networks (CNNs), especially encoder-decoder architectures such as U-Net, have emerged as powerful alternatives capable of learning robust hierarchical representations directly from image data. Recent studies have reported that U-Net-based frameworks achieve superior segmentation performance compared to conventional methods by effectively capturing both local structural details and global contextual information within dental radiographs [1], [3], [4].

To further improve segmentation accuracy, several researchers have incorporated transformer-based mechanisms into traditional U-Net architectures. Transformer-enhanced models utilize self-attention mechanisms to capture long-range dependencies and contextual relationships between distant anatomical structures, which are often difficult to model using convolutional operations alone. Recent frameworks such as STS-TransUNet and other hybrid CNN-transformer architectures have demonstrated improved performance in tooth segmentation tasks by combining the localization capability of CNNs with the global feature representation strength of transformers [2], [8], [10].

Another significant challenge in panoramic dental image analysis is the accurate separation of individual teeth in the presence of crowding, occlusions, and overlapping boundaries. Semantic segmentation models must distinguish subtle differences between adjacent teeth

while preserving fine structural details. Recent studies have proposed lightweight context-aware networks, hierarchical attention mechanisms, and instance-level segmentation strategies to enhance boundary detection and improve segmentation precision. These approaches have shown promising results in accurately identifying dental structures and reducing segmentation errors in complex clinical cases [3], [5], [7], [9].

Furthermore, the increasing adoption of artificial intelligence in dentistry requires models that are not only accurate but also computationally efficient and clinically reliable. Modern segmentation frameworks focus on reducing model complexity while maintaining high performance, enabling real-time deployment in computer-aided diagnosis systems. Advances in attention-guided learning, feature fusion strategies, and automated lesion segmentation have further contributed to improving the robustness and generalization capabilities of deep learning models across diverse dental datasets [6], [8], [9].

This study proposes a U-Net-based semantic segmentation framework for automatic tooth segmentation in panoramic dental X-ray images. By combining image preprocessing and deep feature extraction, the model effectively identifies tooth structures and improves segmentation accuracy in complex dental radiographs. The performance of the proposed system is evaluated using IoU, Dice Similarity Coefficient (DSC), F1-score, and Accuracy. The framework aims to support computer-aided dental diagnosis, orthodontic treatment planning, and intelligent dental healthcare by providing reliable and efficient tooth segmentation [1]–[10].

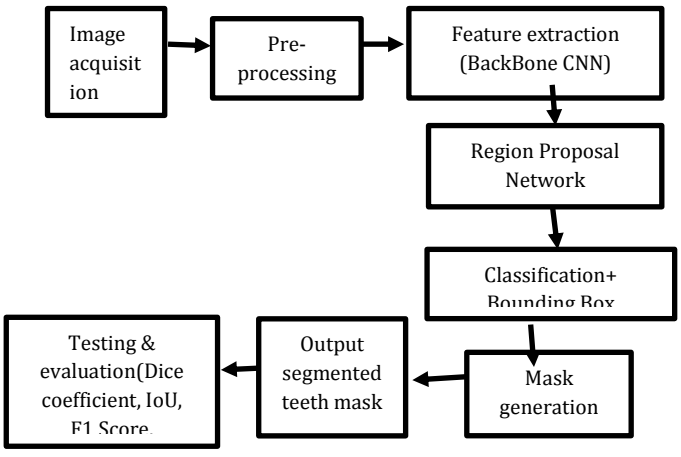

Fig.1 Architecture diagram

2. METHODOLOGIES

The proposed methodology employs Mask R-CNN for automated tooth segmentation in panoramic dental X-ray images. Unlike traditional semantic segmentation models such as U-Net, which classify pixels without distinguishing individual teeth, Mask R-CNN performs instance segmentation, enabling the detection and segmentation of each tooth as a separate object. This capability helps address challenges such as overlapping, crowded, and irregularly positioned teeth commonly found in panoramic radiographs.

The framework consists of image preprocessing, feature extraction using a deep convolutional backbone network, region proposal generation, object classification, bounding box regression, and mask prediction. By extending the Faster R-CNN architecture with a dedicated mask generation branch, the model produces precise pixel-level masks for each detected tooth while simultaneously identifying its location and class.

Through the integration of object detection and instance segmentation, the proposed system achieves accurate tooth-wise segmentation and improved interpretability. The generated instance masks facilitate detailed dental analysis and support clinical applications requiring individual tooth identification, making Mask R-CNN a robust and effective solution for AI-assisted dental image segmentation.

2.1 Dataset Collection

The dataset used in this study consists of panoramic dental X-ray images, also known as Orthopantomogram (OPG) images, collected from a publicly available Kaggle dataset. These radiographic images provide a comprehensive view of the upper and lower dental arches, making them suitable for automated tooth segmentation tasks. The dataset includes images captured under different imaging conditions, resulting in variations in contrast, brightness, and color tones. Visual inspection reveals that some images exhibit bluish tints while others appear grayish, reflecting differences in acquisition and processing techniques.

The collected dataset contains images from individuals of different age groups and genders, including children, adult men, and adult women. Such diversity is essential for developing a robust deep learning model capable of generalizing across varying dental structures and anatomical characteristics. The dataset encompasses a wide range of dental conditions, including fully developed

dentition, partially erupted teeth, missing teeth, and age-related dental variations. This diversity helps the model learn representative features from different patient populations.

Furthermore, the number and arrangement of teeth vary significantly across the dataset. While many adult radiographs contain the complete set of 32 teeth, others include missing teeth due to extraction, developmental stages, or dental conditions. Images of children aged between 6 and 10 years show mixed dentition patterns with both primary and permanent teeth, presenting additional segmentation challenges. Overall, the dataset provides a rich variety of dental structures and clinical scenarios, making it well-suited for training and evaluating deep learning models for accurate tooth segmentation in panoramic dental X-ray images.

2.2 Pre-Processing

The pre-processing stage plays a vital role in enhancing the quality of panoramic dental X-ray images before they are fed into the Mask R-CNN segmentation framework. Since raw radiographs often contain variations in illumination, contrast, and noise, preprocessing helps standardize the images and improve the visibility of important dental structures. Initially, the input images are converted into grayscale format, reducing computational complexity while preserving the essential anatomical information required for tooth segmentation.

To improve the distinction between teeth and surrounding tissues, histogram equalization is applied to enhance image contrast and highlight tooth boundaries. Subsequently, noise reduction techniques such as Gaussian filtering and median filtering are employed to remove unwanted artifacts and background disturbances. These filtering operations effectively suppress noise while retaining critical edge information, ensuring that important structural details are preserved for subsequent analysis.

After enhancement and denoising, the images are normalized to scale pixel intensity values within a uniform range, promoting stable and efficient model training. Finally, all images are resized to a fixed resolution compatible with the convolutional backbone network used in Mask R-CNN. This comprehensive preprocessing pipeline generates clean, consistent, and high-quality input images, thereby improving feature extraction and contributing to more accurate tooth detection and segmentation performance.

2.3 Feature Extraction (Backbone CNN)

The feature extraction stage is a fundamental component of the proposed Mask R-CNN framework, responsible for transforming pre-processed panoramic dental X-ray images into informative feature representations. In this study, a pretrained Convolutional Neural Network (CNN), such as ResNet50 or ResNet101, is employed as the backbone network. These deep architectures are designed to automatically learn and extract meaningful visual patterns from the input images, including edges, contours, textures, and structural details that characterize individual teeth and surrounding anatomical regions.

As the input image passes through multiple convolutional and pooling layers, the network generates hierarchical feature maps containing both low-level and high-level information. Lower layers capture basic image characteristics such as boundaries and shapes, while deeper layers learn more complex semantic features related to tooth morphology, alignment, spacing, and orientation. This multi-scale feature representation is particularly important for panoramic dental radiographs, where teeth may exhibit significant variations in size, position, and appearance.

To enhance learning efficiency and segmentation performance, the backbone network utilizes transfer learning from large-scale datasets such as ImageNet. This approach enables the model to leverage previously learned visual knowledge, resulting in faster convergence, improved generalization, and reduced overfitting when trained on limited dental datasets. The extracted feature maps serve as the foundation for the Region Proposal Network (RPN) and subsequent Mask R-CNN modules, enabling accurate detection and instance-level segmentation of individual teeth, even in challenging cases involving overlapping, crowded, or partially occluded dental structures.

2.4 Region Proposal Network (RPN)

The Region Proposal Network (RPN) is a crucial component of the Mask R-CNN architecture that identifies candidate regions likely to contain teeth within panoramic dental X-ray images. Operating on the feature maps generated by the backbone CNN, the RPN scans the image and generates multiple region proposals using anchor boxes of different sizes and aspect ratios. This enables the network to effectively detect teeth with varying shapes, sizes, and orientations.

For each proposed region, the RPN calculates an objectness score that indicates the probability of the region containing a tooth rather than background information. The generated proposals are then refined and filtered using Non-Maximum Suppression (NMS), which removes redundant and highly overlapping bounding boxes while retaining the most relevant candidate regions. This process ensures accurate localization of individual teeth and reduces unnecessary computational overhead.

By focusing the network's attention on tooth-specific regions, the RPN significantly improves the efficiency and accuracy of the segmentation framework. The selected region proposals are forwarded to the classification and mask prediction stages of Mask R-CNN, enabling precise detection and instance-level segmentation of teeth, even in challenging cases involving overlapping, crowded, or partially occluded dental structures.

2.5 Classification and Bounding Box Regression

After the Region Proposal Network (RPN) generates candidate regions, the Classification and Bounding Box Regression module is responsible for identifying and accurately localizing individual teeth within the panoramic dental X-ray image. Initially, the proposed regions are processed using ROI Align, which extracts fixed-size feature maps while preserving spatial information and ensuring precise feature representation.

The extracted features are then passed through fully connected layers that perform object classification. Each proposed region is assigned a class label indicating whether it represents a tooth or background. In multiclass settings, the model can further categorize specific tooth types such as incisors, canines, premolars, and molars. This classification process enables the network to distinguish relevant dental structures from surrounding anatomical regions and image artifacts.

Simultaneously, the bounding box regression branch refines the coordinates of each detected tooth to achieve more accurate localization. By predicting adjustments to the proposed bounding boxes, the model generates tighter boundaries around individual teeth. The network is optimized using regression loss functions such as Smooth L1 Loss, which improves localization accuracy while maintaining training stability. Together, classification and bounding box regression ensure precise tooth identification and localization, providing a strong foundation for the subsequent instance segmentation stage of the Mask R-CNN framework.

2.6 Mask Generation

The Mask Generation module is the final stage of the Mask R-CNN framework and is responsible for producing precise pixel-level segmentation masks for each detected tooth. After the classification and bounding box refinement stages, the features corresponding to each Region of Interest (RoI) are processed through a fully convolutional network (FCN) to generate high-resolution masks. This enables the model to accurately capture the shape, contours, and boundaries of individual teeth while preserving important spatial details.

Unlike conventional semantic segmentation approaches that assign a single class label to each pixel without distinguishing between separate objects, Mask R-CNN performs instance-level segmentation by generating an independent mask for every detected tooth. This capability is particularly valuable in panoramic dental X-ray images, where teeth may overlap, appear crowded, or have complex anatomical structures. Each predicted mask is aligned with its corresponding bounding box, ensuring accurate localization and segmentation of individual tooth instances.

The mask prediction branch is trained using a binary cross-entropy loss function, which compares the predicted masks with the ground-truth annotations to improve segmentation accuracy. The final output consists of detailed tooth-wise segmentation masks that clearly separate individual teeth from surrounding structures and the background. These instance-level masks provide enhanced visualization and support accurate dental analysis, diagnosis, and treatment planning.

2.7 Performance Evaluation

The performance evaluation stage is carried out to measure the effectiveness of the proposed Mask R-CNN model in accurately detecting and segmenting individual teeth from panoramic dental X-ray images. The predicted segmentation masks are compared with the corresponding ground-truth annotations using standard evaluation metrics, including Dice Coefficient, Intersection over Union (IoU), F1-Score, and Accuracy. These metrics provide a comprehensive assessment of the model's ability to identify tooth regions and generate precise instance-level segmentation masks.

Dice Coefficient

The Dice Coefficient is a similarity metric widely used in medical image segmentation to evaluate the overlap between the predicted mask and the ground-truth mask. It ranges from 0 to 1, where 1 indicates perfect overlap and 0 indicates no overlap.

$$\text{Dice Cofficient} = \frac{2\ X\ |X \cap Y|}{|X| + |Y|}$$

Where:

- X represents the predicted segmentation mask.
- Y represents the ground-truth mask.
- $|X \cap Y|$ denotes the overlapping pixels between the two masks.

A higher Dice value indicates better segmentation performance and stronger agreement between the predicted and actual tooth regions.

Intersection over Union (IoU)

Intersection over Union (IoU), also known as the Jaccard Index, measures the ratio between the overlapping area and the combined area of the predicted and ground-truth masks. The metric ranges from 0 to 1, with higher values indicating superior segmentation accuracy.

$$\text{IoU} = \frac{X \cap Y}{X \cup Y}$$

Where:

- $X \cap Y$ is the intersection of the predicted and ground-truth masks.
- $X \cup Y$ is the union of the predicted and ground-truth masks.

IoU provides a direct measure of how accurately the segmented tooth regions match the annotated dental structures.

F1-Score

The F1-Score combines Precision and Recall into a single metric and is particularly useful when evaluating imbalanced datasets. It balances false positives and false negatives, providing a reliable measure of classification performance.

$$\text{F1 score} = 2X\frac{\text{Precision X Recall}}{\text{Precision} + \text{Recall}}$$

Where:

$$\text{Precsion} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Postive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

A higher F1-Score indicates a better balance between precision and recall, reflecting robust tooth detection and segmentation performance.

Accuracy

Accuracy measures the proportion of correctly classified pixels or instances among all predictions made by the model.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total numer of predictions}}$$

Although accuracy provides a general measure of model performance, it can be misleading in segmentation tasks where background pixels significantly outnumber tooth pixels. Therefore, Dice Coefficient and IoU are often considered more reliable indicators of segmentation quality. Together, these evaluation metrics provide a comprehensive framework for validating the accuracy, robustness, and clinical applicability of the proposed Mask R-CNN-based tooth segmentation system.

RESULT & DISCUSSION

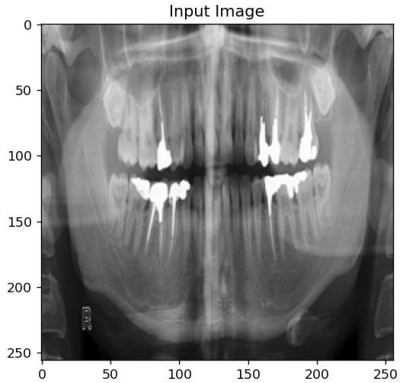

Fig.2 Input image

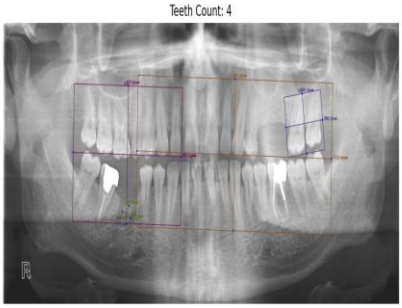

Fig.3 Final output

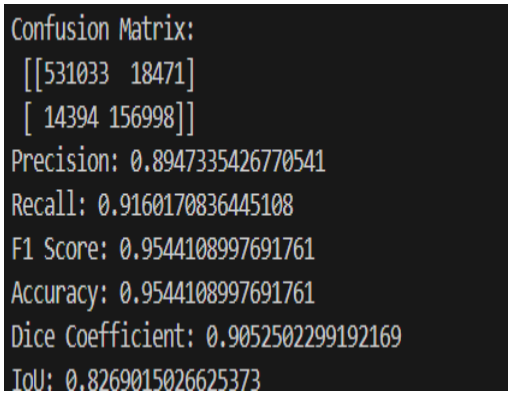

Fig.4 Performance metrices

The proposed Mask R-CNN-based tooth segmentation framework was evaluated using panoramic dental X-ray images to assess its effectiveness in detecting and segmenting individual teeth under varying dental conditions. As shown in Fig. 2, panoramic dental X-ray images were provided as the input for analysis, containing multiple teeth with overlapping structures and varying contrast levels. The segmentation output illustrated in Fig. 3 **(**demonstrates the capability of the Mask R-CNN model to accurately identify and segment individual teeth by generating precise instance-specific boundaries, even in crowded and complex regions. The model effectively distinguishes adjacent and overlapping teeth, improving segmentation clarity and interpretability for dental diagnosis. Furthermore, the quantitative evaluation presented in Fig. 4 confirms the robustness of the proposed approach, achieving high values for precision, recall, F1-score, Dice coefficient, IoU, and overall accuracy, indicating reliable tooth detection and segmentation performance. The results demonstrate that the integration of image preprocessing, feature extraction, region proposal generation, and mask prediction enables accurate and automated tooth segmentation while significantly reducing manual effort and inter-observer variability. Overall, the proposed framework provides a reliable and efficient solution for AI-assisted dental image analysis, supporting improved clinical diagnosis, orthodontic planning, and intelligent dental healthcare applications.

CONCLUSION

In conclusion, this study presented an effective deep learning-based tooth segmentation framework using Mask R-CNN for accurate instance-level segmentation of teeth in panoramic dental X-ray images. Unlike conventional semantic segmentation approaches, the proposed method enables precise identification and separation of individual teeth, including overlapping and adjacent structures, thereby improving segmentation accuracy and clinical interpretability. By integrating image pre-processing, feature extraction, region proposal, classification, and mask generation, the system successfully produced detailed tooth-specific segmentation masks suitable for dental diagnosis, orthodontic planning, and computer-assisted clinical applications. Despite its strong performance in complex dental imaging scenarios, the framework is constrained by the availability of limited annotated datasets, which may affect generalizability across diverse patient populations and imaging conditions. Future work may focus on expanding dataset diversity, optimizing computational efficiency, and improving inference speed to facilitate real-time implementation and deployment in practical dental healthcare systems.

REFERENCES

[1] R. Zannah, M. Bashar, R. B. Mushfiq, A. Chakrabarty, et al., “Semantic Segmentation on Panoramic X-ray Images Using U-Net Architectures,” IEEE Access, vol. 12, pp. 1–15, 2024.

[2] D. Sun, J. Wang, Z. Zuo, Y. Jia, and Y. Wang, “STS-TransUNet: Semi-supervised Tooth Segmentation Transformer U-Net for Dental Panoramic Image,” Mathematical Biosciences and Engineering, vol. 21, no. 2, pp. 2366–2384, 2024.

[3] A. Khaldi, B. Khaldi, and O. Aiadi, “LCAT-Net: Lightweight Context-Aware Deep Learning Approach for Teeth Segmentation in Panoramic X-rays,” International Journal of Computational Intelligence Systems, vol. 17, Art. no. 297, 2024.

[4] R. Pandey, A. Reddy, A. Waghmare, and A. Kamojjalwar, “Segmentation on Panoramic Dental X-Ray Images Using U-Net,” International Journal of Advanced Research in

Science, Communication and Technology, vol. 4, no. 5, pp. 1–8, 2024.

[5] D. Budagam, A. Kumar, S. Ghosh, A. Shrivastav, et al., "Instance Segmentation and Teeth Classification in Panoramic X-rays," arXiv preprint arXiv:2406.03747, 2024.

[6] M. Boztuna, M. Firincioglulari, N. Akkaya, and K. Orhan, "Segmentation of Periapical Lesions with Automatic Deep Learning on Panoramic Radiographs: An Artificial Intelligence Study," BMC Oral Health, vol. 24, Art. no. 1332, 2024.

[7] T. Ma, Z. Dang, Y. Yang, J. Yang, and J. Li, "Dental Panoramic X-ray Image Segmentation for Multi-feature Coordinate Position Learning," Digital Health, vol. 10, 2024.

[8] Y. Li and Co-authors, "A Dual-Stream Dental Panoramic X-Ray Image Segmentation Method Based on Transformer Heterogeneous Feature Complementation," Technologies, vol. 13, no. 7, Art. no. 293, 2025.

[9] M. Esmaeili, Z. Dalili, H. Sadr, A. Mousavie, and M. Nazari, "Hierarchical Attention Mechanism Combined with Deep Neural Networks for Accurate Semantic Segmentation of Dental Structures in Panoramic Radiographs," Scientific Reports, vol. 15, Art. no. 38725, 2025.

[10] Y. Wang, Z. Li, C. Wu, J. Liu, and H. Zhou, "MICCAI STS 2024 Challenge: Semi-Supervised Instance-Level Tooth Segmentation in Panoramic X-ray and CBCT Images," arXiv preprint arXiv:2511.22911, 2025.