Mask R-CNN (ResNet-50 vs ResNet-101): A Deep Learning Framework for Instance-Level Enamel Segmentation

Nandeesh Mahadevu
Department of Electronics and Communication Engineering, BGS Institute of Technology, Adichunchanagiri University, Karnataka, India
nandimalvi@gmail.com

Naveen Bettahalli
Department of Electronics and Communication Engineering, BGS Institute of Technology, Adichunchanagiri University, Karnataka, India
naveenb@bgsit.ac.in (corresponding author)

Srividya Chandagirikoppal Nagendra
Department of Electronics and Communication Engineering, BGS Institute of Technology, Adichunchanagiri University, Karnataka, India
srividyacn@bgsit.ac.in

Received: 21 November 2025 | Revised: 5 January 2026 and 20 January 2026 | Accepted: 23 January 2026

 DOI: https://doi.org/10.48084/etasr.16421

ABSTRACT

Dental cavities constitute a major global health issue and must be diagnosed reliably to enable timely and effective treatment. The identification of dental caries at an early stage is essential, as lesions typically begin at the enamel surface and, over time, progress into the deeper tooth structures, including dentin and pulp. Advancements in dental imaging, combined with artificial intelligence-based methodologies, offer promising solutions for improving diagnostic accuracy and efficiency. Therefore, the present study evaluates the performance of Faster Region-based Convolutional Neural Network (Faster R-CNN) and Mask Region-based Convolutional Neural Network (Mask R-CNN) with ResNet-50 and ResNet-101 backbones for automatic enamel detection and segmentation. All models exhibited excellent detection performance, obtaining perfect Average Precision (AP) scores at IoU thresholds of 0.50 (AP50) and 0.75 (AP75). Faster R-CNN has achieved an AP of 95.92%, while both Mask R-CNN variants, ResNet-50 and ResNet-100, achieved near-perfect bounding box detection with an AP of approximately 99%. For segmentation, Mask R-CNN with a ResNet-50 backbone achieved an AP of 86.30%, whereas the deeper ResNet-101 backbone significantly improved segmentation performance, achieving an AP of 98.44%. These results demonstrate that the Mask R-CNN architecture surpasses Faster R-CNN in detection accuracy and provides superior segmentation performance. Overall, Mask R-CNN with a ResNet-101 backbone can be considered the most effective model for enamel detection and segmentation. Nevertheless, the proposed model should be improved and externally validated. This work can be further carried out to detect carious lesions in the enamel portion for early detection and treatment.

Keywords-enamel; segmentation; deep learning; Mask R-CNN

I. INTRODUCTION

There are three parts to a tooth: the dentin, the pulp, and the enamel. There are different types of bacteria in the mouth, some of which may infect the dead teeth. These types of infections are known as dental caries. Cavities and damage resulting from poor dental care can persist for a long time. About 60–90% of school-aged children and almost all adults have Cavities and damage. Dentists usually diagnose cavities by looking at X-rays of the person's teeth and jaw. However, this hands-on method does not find all cavities, especially the ones that are still small or difficult to see on X-rays. It is also very hard to detect early cavities in X-ray images. There is a growing need for automation in dental diagnostics for improved detection, especially at early stages.

The present study proposes a novel methodology for detection and segmentation of enamel from teeth for dental

carries detection. While numerous studies have addressed automated caries detection, most existing methods analyze the tooth structure as a whole, without isolating enamel-specific regions. Since carious lesions typically originate at the enamel surface and progress inward, precise identification of enamel regions is essential for enabling earlier diagnosis. Therefore, this study focuses specifically on enamel localization and segmentation from teeth.

The proposed method uses the Faster R-CNN and Mask RCNN-based methods to find and separate teeth and enamel. In the enamel, the cavities can be detected earlier. Computer vision has advanced significantly due to CNNs. However, detection algorithms that use region proposals, such as RCNN in 2014, Fast-RCNN, and Faster R-CNN in 2015, have not shown similar rapid performance improvements [1, 2]. But these algorithms have significantly improved the analysis of medical images. The region-proposed method is 100-1000 times slower than a one-step network like the CNN, You Only Look Once (YOLO), and Single Shot Multi-box Sensor [3-6].

The performance of Mask R-CNN is evaluated in terms of prediction accuracy and time efficiency. Its regression branch refines the coordinates of proposals to enhance localization precision, and the bounding box classifier determines the probability of an object present in a proposal. Building on the foundation of two-stage detectors, Mask R-CNN extends the functionality of Faster R-CNN by adding a parallel mask prediction branch for each Region of Interest (RoI), enabling instance-level pixel segmentation. This is facilitated by a fully convolutional network, which allows precise segmentation of object boundaries. The key innovation in Mask R-CNN is the RoIAlign layer because it corrected spatial misalignment caused by quantization in earlier methods, thus improving segmentation accuracy significantly [7-9].

Mask R-CNN is the best deep learning framework for object detection and instance segmentation due to its ability to jointly handle classification, localization, and segmentation with better accuracy [10, 11]. Its flexible architecture and adaptability to different backbone networks and feature extractors, such as Feature Pyramid Networks (FPNs), have further improved its performance across different applications. This multi-task capability has proven remarkable in medical image processing. It has been demonstrated that models like Mask R-CNN and U-Net are well-suited for biomedical image segmentation and diagnostic support.

At its core, Mask R-CNN employs a two-stage process. In the first stage, a Region Proposal Network (RPN) identifies actual RoIs containing objects, and the second stage processes these RoIs for further refinement. A key feature of Mask R-CNN is parallel processing of multiple tasks. The bounding box regression branch fine-tunes the object boundaries to get better localization accuracy, and the classification branch estimates the probability of object presence within a given region. Additionally, a unique mask branch predicts a binary segmentation mask for each detected object, enabling pixel-level delineation of object shapes [12].

An important advancement in Mask R-CNN over Faster R-CNN is the integration of the RoIAlign layer, which resolves the issue of misaligned features due to quantization in the RoIPool layer. RoIAlign improves the spatial accuracy by maintaining precise mapping between the input and extracted features, ensuring better segmentation outcomes [13]. Mask R-CNN's architecture also has a mask prediction branch that operates independently of classification and localization tasks, thereby ensuring that segmentation accuracy is not compromised by the outputs of other branches [14]. By implementing pixel-to-pixel segmentation within each RoI, Mask R-CNN can achieve high performance across diverse datasets and complex visual environments.

The versatility and adaptability of Mask R-CNN make it a preferred choice for applications in computer vision, such as medical imaging, autonomous vehicles, and video analysis. It delivers high accuracy while efficiently handling both detection and segmentation. This has led to its deployment in real-world applications such as robotic perception and clinical diagnostics [15].

II. METHODOLOGY

Figure 1 outlines the complete process of the proposed model on dental bitewing X-rays. A dataset containing 600 dental X-ray images was acquired from a collaborating hospital under institutional ethical approval. Each image was subjected to manual annotation. These annotated images were then used to train multiple segmentation models, specifically Faster R-CNN and Mask R-CNN with two different backbone architectures—ResNet-50 and ResNet-101. By employing these networks, the system learns to distinguish and segment enamel structures within the dental X-rays based on the provided labeled data.

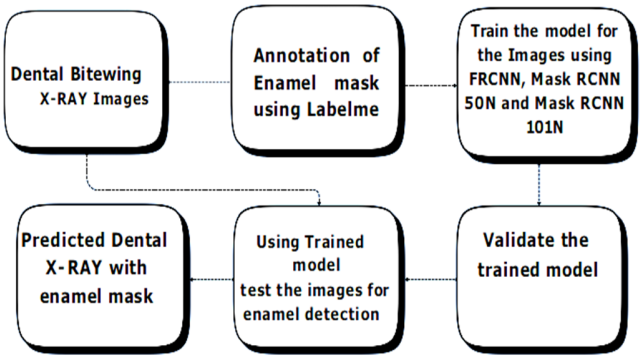


Fig. 1. Proposed Mask R-CNN model for enamel detection.

Once the training phase was completed, the trained models were used to validate and assess their segmentation accuracy and overall performance. The most accurate trained model was then used to test new X-ray images to automatically detect and segment the enamel region. This stage produces predicted X-rays showing enamel masks, allowing direct visual comparison with the manually annotated ground truth. The iterative validation and testing process ensures that the selected Mask R-CNN variant performs better for enamel detection and segmentation, providing a reliable framework for dental image analysis.

A. Data Collection

The proposed work compiled a collection of 600 intraoral and radiographic images sourced directly from the collaborating hospital. The study was conducted in accordance with the ethical standards of the institutional research committee. Dental images were collected after obtaining approval from the relevant institutional ethics committee. Due to ethical restrictions and patient privacy considerations, the data are not publicly available. The annotation process was carried out using LabelMe, an open-source graphical image annotation tool developed by the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. The importance of building domain-specific medical datasets has been emphasized in prior works on custom data acquisition in healthcare AI pipelines [16, 17].

B. Image Annotation and Labeling

Labelme allows annotators to manually draw polygons around target objects—here, tooth enamel structures—and save the annotation data as JSON files for later use in training. Each tooth region was assigned a class label of 1, while the background was labeled 0. While manual labeling can introduce minor inconsistencies, it remains one of the most reliable methods for generating high-quality training data for medical image segmentation tasks [18].

C. Data Preparation

The image dataset was split into training and validation, with 400 images (75%) allocated for training and 200 images (25%) reserved for testing and validation. All images were resized to a resolution of 640 × 480 pixels, with a resolution of 96 dpi and a bit depth of 24. The arrangement of images across the different lesion classes was kept consistent. Data augmentation techniques were not applied to the training set. The validation set was selected randomly while ensuring that the class proportions remained unchanged. Resizing and consistent preprocessing play an important role in the generalizability of deep learning models in medical imaging applications [19].

D. Deep Network Architecture

Mask R-CNN builds upon the Faster R-CNN framework by adding an additional segmentation branch that allows simultaneous object detection and instance-level segmentation. This improvement enabled Mask R-CNN to outperform existing single-model approaches across different visual tasks and was instrumental in exceeding the performance of many models in the COCO 2016 challenge [20]. The version used in the present study is implemented using the Detectron2 framework, developed by Facebook AI Research (FAIR), which is widely adopted for state-of-the-art object detection and segmentation tasks [21]. All training, validation, and testing operations were performed on a single workstation with an open-source GPU. The algorithms were trained and fine-tuned for 100 epochs, with a batch size of 6 images, a learning rate of 0.00025, and 10,000 iterations. The best algorithm in the study is Mask RCNN with ResNet-101, which utilizes a ResNet-101 backbone for feature extraction combined with an FPN for improved multi-scale object detection. FPN improves spatial hierarchy by transmitting high-level semantic information to lower-level feature maps, improving performance on small and medium-sized objects [22].

The overall structure has a two-stage process, similar to Faster R-CNN. In the first stage, an RPN generates candidate bounding boxes, referred to as anchors. Non-Maximum Suppression (NMS) is applied to eliminate anchor boxes with excessive overlap, streamlining the detection process [22]. To enhance precision, Mask R-CNN replaces traditional ROI pooling with RoIAlign. Unlike ROI pooling, RoIAlign avoids rounding coordinates during region alignment, which preserves spatial information and prevents inaccuracies in object positioning and shape. Typically, the RPN generates fractional box coordinates that require rounding, and that lead to minor errors in pooling. RoIAlign eliminates this rounding process, maintaining decimal precision and significantly improving detection and segmentation accuracy. For the second stage, Mask R-CNN applies separate network heads to each RoI for three tasks: classification, bounding box regression, and mask prediction. Each task is performed independently using a multi-task loss function, which combines the losses from classification, regression, and segmentation. The classification and regression losses are consistent with standard object detection networks, while the mask loss branch employs a CNN. This CNN takes the RoI as input and outputs the binary mask, ensuring accurate instance segmentation for each detected object [12, 14, 20].

The RoIAlign layer has better performance compared to Faster R-CNN because it addresses non-aligned features due to quantization in the RoIPool layer. RoIAlign improves spatial accuracy by ensuring correct mapping between the input and output features. Mask R-CNN has a separate, independent mask prediction branch and does not do the same as classification and localization. This ensures that the segmentation is always accurate. Mask R-CNN is effective under different datasets and complex visual settings, as it breaks each RoI down to single pixels.

III. RESULTS AND DISCUSSION

A. Faster R-CNN

1) Training Metrics

The training metrics for enamel detection demonstrate that the Faster R-CNN performs well at all IoU levels.

As presented in Table I, the F1-score stays at 0.9425 for IoU values between 0.5 and 0.9. The recall is perfect at 1.0, and the precision is about 0.97. This shows the effectiveness of the model to detect and label the enamel during training. This reduces the number of false negatives and provides balanced performance with high recall. The training performance surpasses that of previous studies. Authors in [16] assessed the performance of Faster R-CNN in detecting proximal caries, resulting in a detection accuracy of 92.3%. Additionally, in [17], the Faster R-CNN had AP values over 95% in both the upper and lower quadrants when used to count teeth in panoramic radiographs. In the present study, the training accuracy is 0.97 with a perfect recall. These results demonstrate the effectiveness of Faster R-CNN in enamel detection.

TABLE I. FASTER R-CNN TRAINING METRICS ACROSS DIFFERENT IoU

IoU threshold	Precision	Recall	F1-score
0.5	0.97	1	0.94
0.6	0.97	1	0.94
0.7	0.97	1	0.94
0.75	0.97	1	0.94
0.8	0.97	1	0.94
0.9	0.97	1	0.94

2) Validation

The validation metrics for Faster R-CNN's enamel detection are presented in Table II. The AP is 95.92%, and both AP scores at IoU thresholds of 0.50 (AP50) and 0.75 (AP75) achieved a score of 99%. In addition, the APm (medium-sized objects) achieved an accuracy of 95.92%, demonstrating the excellent performance of the method in detecting and locating enamel structures in dental X-rays. The model can also be used for any new dataset. The validation results are significantly better than those of previous studies. For example, Faster R-CNN used for detecting cavities in teeth had an accuracy of 92.3%, with a recall and F1-score of 100% and 96%, respectively [16]. Authors in [17] used panoramic radiographs to count teeth. The AP scores were 95.93% in the upper quadrants and 95.05% in the lower quadrants, which are comparable to the AP of 95.92% obtained in the present study. Authors in [18] used Mask R-CNN for segmentation, achieving an AP score of 66.73%-83.86%. In contrast, the proposed model has a higher AP score, demonstrating superior enamel detection.

TABLE II. VALIDATION METRICS FOR FASTER R-CNN'S ENAMEL DETECTION

Metric	Bounding box
AP	95.92%
AP50	99%
AP75	99%
APm	95.92%

The Fast R-CNN foreground classification accuracy (fast_rcnn/fg_cls_accuracy) score starts below 0.1, as shown in Figure 2, which indicates that the model has difficulty differentiating between the enamel and the background at first. After about 2,000 iterations, the accuracy significantly improved, indicating effective learning. The training data help the model perform better. The accuracy stays at about 1.0 after 8,000 iterations, indicating completion of the learning stage.

B. MASK RCNN ResNet-50

1) Training Metrics

The Mask R-CNN ResNet-50 backbone model obtained IoU values of 0.5 and 0.9 during training, with similar values across all levels, as depicted in Table III. The model had an F1-score of 0.988, a recall of 1.0, and a precision of 0.976. This shows that the detection accuracy for the model was 97.6% with 100% recall and no false positives. The F1-score remains the same across all IoUs, which shows that localization and segmentation are good even with significant overlap.

In [19], an accuracy of 96% was achieved, which further decreased with higher IoUs. Authors in [20] used a deep ensemble model, obtaining the highest F1-score of 0.92 on bitewing radiographs. These results indicate that further work is still needed to achieve the right balance between accuracy and recall.

In contrast, the present study achieves perfect recall and consistently high accuracy. In general, the model is better at detecting and locating enamel compared to other methods. Despite these promising training results, the potential for overfitting must be considered, which is often a challenge in medical imaging research. Models trained on homogenous datasets may not generalize effectively across varied clinical settings [20, 21]. To address this limitation, future work will focus on external validation using independent datasets from diverse imaging systems and patient populations, based on [22]. Such efforts are essential to ensure the model's robustness, generalizability, and practical applicability in real-world clinical environments.

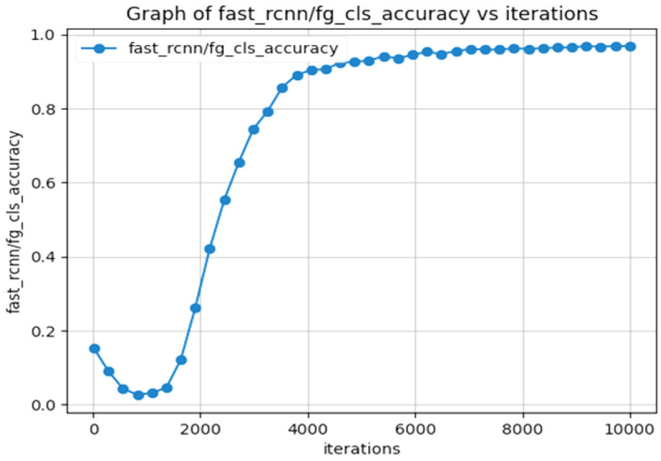


Fig. 2. Accuracy versus iterations for Faster R-CNN.

TABLE III. MASK R-CNN RESNET-50 TRAINING METRICS ACROSS DIFFERENT IoU

IoU threshold	Precision	Recall	F1-score
0.5	0.976	1	0.9879
0.6	0.976	1	0.9879
0.7	0.976	1	0.9879
0.75	0.976	1	0.9879
0.8	0.976	1	0.9879
0.9	0.976	1	0.9879

2) Validation

Table IV presents the performance of the Mask R-CNN model with a ResNet-50 backbone in detecting and segmenting enamel cavities in the validation set. The model achieved an AP score of 99.88% for bounding box detection and 86.298% for segmentation. This result demonstrates the accuracy of the model; however, the lower segmentation score indicates that there is still scope for improvement. Bounding box and segmentation achieved a score of 99.98% on the AP50 and AP75 tests, which is better than previous studies with Faster R-

CNN (94.8%). Additionally, authors in [19] obtained a U-Net AP score of 96%, while authors in [20] achieved an ensemble model AP score of 92%. In contrast, the present study achieved an APm score of 99.88% for bounding box detection and 86.298% for segmentation, suggesting that the model performs better on medium-sized enamel cavities. This is in line with [20], which emphasized the need for higher scaling for medical imaging. The internal validation results are excellent, which means that the detection is significantly reliable. However, in [20], it has been demonstrated that external validation is still necessary to ensure that the results can be used in real life and to avoid overfitting.

TABLE IV. VALIDATION METRICS FOR MASK R-CNN RESNET-50 ENAMEL DETECTION

Metric	Bounding box	Segmentation
AP	99.88%	86.298%
AP50	99.88%	99.98%
AP75	99.88%	99.98%
APm	99.88%	86.298%

The Mask R-CNN model with a ResNet-50 backbone achieved perfect scores of 99.88% for AP50, AP75, and APm for bounding box detection. It also had a segmentation score of 86.298% for AP/APm, which is strong but not as high as AP50 and AP75 scores. These results are a significant improvement over [19, 20], which achieved an AP50 of 94.8% and an AP score of 92%, respectively. As displayed in Figure 3, the accuracy for the Mask R-CNN model increased from less than 0.4 to 1.0 during training. The accuracy remains above 0.8 after 1,000 iterations and increases to 100% after 6,000 iterations. This trend shows the model's effectiveness in detecting and segmenting enamel cavities within a certain area.

C. MASK RCNN 101

1) Training Metrics

The training evaluation metrics for the Mask R-CNN model with a ResNet-101 backbone were tested with IoU thresholds ranging from 0.5 to 0.9. The results, as presented in Table V, show that all thresholds had a perfect precision, recall, and F1-score of 0.99 (99.98%). These findings confirm that all enamel cavities in the training set were detected with zero false positives. The model's consistently high and balanced performance indicates that it can effectively learn the shape and location of enamel cavities. These results surpass those reported in [19], which used a different version of U-Net and obtained an AP50 score of 96%; however, the AP50 declined as IoUs increased. Similarly, authors in [20] used a deep ensemble model and obtained the highest F1-score of 0.92.

Even with the training metrics used in the present study, there is still a possibility of overfitting. Authors in [20, 23] observed that models which are trained on certain datasets might not perform well in different clinical settings. Similar observations were made in [22], while emphasizing the importance of independent testing. The training results show that the model performs well with the dataset, but it needs to be tested on more clinical data to ensure that it can be used in real life.

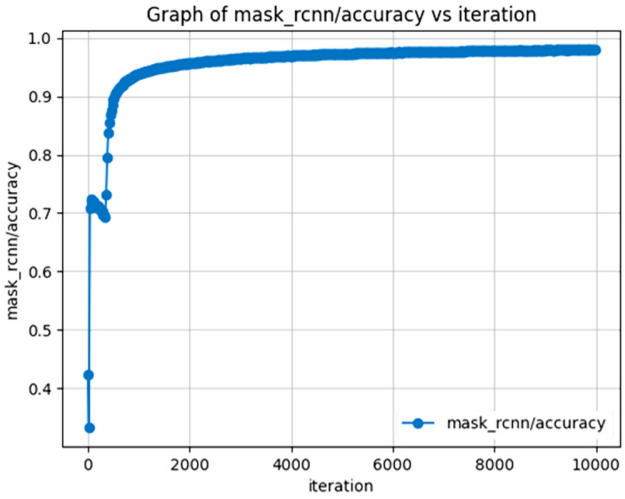


Fig. 3. Accuracy versus iterations for Mask R-CNN ResNet-50.

TABLE V. MASK R-CNN RESNET-101 TRAINING METRICS ACROSS DIFFERENT IoU

IoU threshold	Precision	Recall	F1-score
0.5	0.99	1	0.99
0.6	0.99	1	0.99
0.7	0.99	1	0.99
0.75	0.99	1	0.99
0.8	0.99	1	0.99
0.9	0.99	1	0.99

2) Validation

The validation metrics presented in Table VI show that the Mask R-CNN model with a ResNet-101 backbone is effective in detecting and segmenting enamel cavities. The model had a segmentation score of 98.44% and a bounding box detection score of 99.98%. This suggests a high degree of localization and segmentation accuracy, with few false positives and false negatives. The AP50 and AP75 scores for bounding box detection and segmentation are both 99.98%. This indicates high accuracy across all levels of IoU. These results surpass those of [19]. The APm segmentation score of 98.44% and the bounding box detection rate of 99.98% suggest that the models also perform well on enamel cavities of medium size. The proposed model is extremely good at detecting and segmenting cavities in enamel because these results are almost perfect. Nevertheless, authors in [24] demonstrated that internal validation is not enough by itself. Authors in [19] suggested that models should be able to handle images of all sizes and must be tested on a number of real-world datasets to avoid overfitting.

TABLE VI. VALIDATION METRICS FOR MASK R-CNN RESNET-101 ENAMEL DETECTION

Metric	Bounding box	Segmentation
AP	99.98%	98.44%
AP50	99.98%	99.98%
AP75	99.98%	99.98%
APm	99.98%	98.44%

The Mask R-CNN model with a ResNet-101 backbone is better for detecting and segmenting enamel cavities. It achieved

a perfect score on both the AP and the APm tests compared to other results. For example, authors in [20] used a U-Net-based method and achieved an AP value of about 96%. Figure 4 illustrates the accuracy of the Mask R-CNN model, which starts at about 0.7 and increases quickly after 1,000 iterations, when the accuracy increases to 0.99 and keeps increasing afterwards. The trend shows that the model gets better when the training iterations are increased. Figure 5 displays the output of testing using Mask RCNN 101, displaying how a Mask R-CNN model splits up enamel on a dental X-ray. As observed, each colored mask and bounding box shows a different part of the enamel.

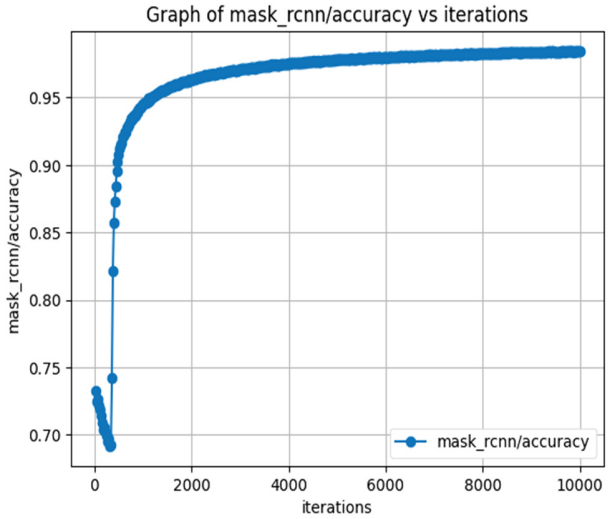
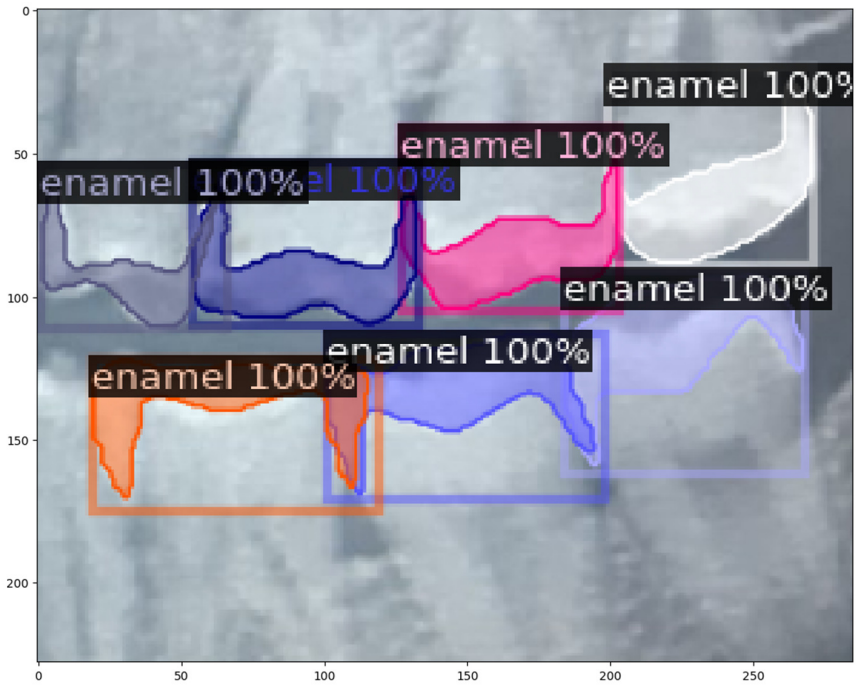


Fig. 4. Accuracy versus iterations for Mask R-CNN ResNet-101.

IV. COMPARATIVE ANALYSIS

In Table VII, the training data show that the Faster R-CNN model had an F1-score of about 0.94 and a precision of about 0.97 at all IoU levels, which were between 0.5 and 0.9. The Mask R-CNN with the ResNet-50 backbone always performs better than the Faster R-CNN. Its F1-score was about 0.9879, and its precision was about 0.976. At all IoU levels, Mask R-CNN with ResNet-101 as the backbone reached a Mask perfect precision of about 0.99, a perfect recall of 1, and an F1-score of 0.99.

The model had improved performance after switching from Faster R-CNN to Mask R-CNN with deeper backbones, such as ResNet-50 and then ResNet-101. The ResNet-101 backbone of Mask R-CNN is the best model in this test. It obtained perfect precision, recall, and F1-scores, resulting in excellent cavity detection. Similarly, as presented in Table VIII, all models had excellent performance at bounding box detection, with very high AP50 and AP75 scores. Faster R-CNN had an overall AP of 95.92%, while Mask R-CNN with ResNet-50 and Mask R-CNN with ResNet-101 achieved a better AP of 99.88% and 99.98%, respectively. This demonstrates that Mask R-CNN models are better at detecting enamel. In segmentation tasks, Mask R-CNN with ResNet-50 achieved an AP value of 86.3%, while Mask R-CNN with ResNet-101 achieved a significantly higher AP value of 98.44%. These results highlight the advantage of deeper residual networks in capturing subtle structural variations within enamel that are often missed by shallower architectures. Mask R-CNN with ResNet-101 had the best performance compared to all models discussed in the present study, making it the best choice for reliable enamel lesion detection. Its consistent superiority makes it particularly suitable for applications requiring highly accurate and reliable enamel mask predictions, especially in early dental lesion analysis, where precise boundary delineation is critical.



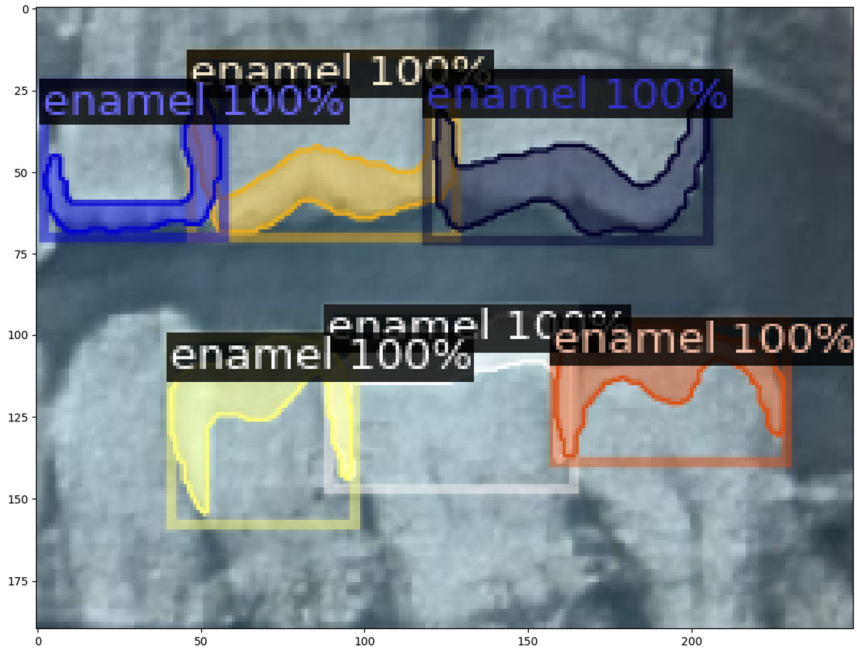


Fig. 5. Visualization of enamel detection on a radiographic image using Mask R-CNN ResNet-101.

TABLE VII. COMPARATIVE PERFORMANCE OF FASTER R-CNN AND MASK R-CNN MODELS UNDER TRAINING ANALYSIS ACROSS DIFFERENT IOU THRESHOLDS

IoU threshold	Faster R-CNN			Mask RCNN 50			Mask RCNN 101		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
0.5	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99
0.6	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99
0.7	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99
0.75	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99
0.8	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99
0.9	0.97	1	0.94	0.97	1	0.98	0.99	1	0.99

TABLE VIII. COMPARATIVE PERFORMANCE OF FASTER R-CNN AND MASK R-CNN MODELS UNDER VALIDATION ANALYSIS ACROSS DIFFERENT IOU THRESHOLDS

Metric	Faster R-CNN		Mask RCNN 50		Mask RCNN 101	
	Bounding box	Segmentation	Bounding box	Segmentation	Bounding box	Segmentation
AP	95.92%	--	99.88	86.298	99.98%	98.44%
AP50	99%	--	99.88	99.98	99.98%	99.98%
AP75	99%	--	99.88	99.98	99.98%	99.98%
APm	95.92%	--	99.88	86.298	99.98%	98.44%

V. CONCLUSION

This study proposed models for the detection and segmentation of enamel from teeth for the identification of dental caries. Previous studies have investigated automated caries detection; however, most existing approaches focus on the overall tooth structure rather than targeting enamel regions. Because carious lesions usually start on the surface of the enamel and move deeper, it is important to accurately identify enamel areas for an earlier diagnosis. This study concentrates on the localization and segmentation of the enamel from teeth.

The objective of this study is to evaluate the effectiveness of deep learning models in detecting and segmenting enamel by comparing the Faster Region-based Convolutional Neural Network (Faster R-CNN) and Mask Region-based Convolutional Neural Network (Mask R-CNN) architectures. The results demonstrated that all models achieving perfect AP50 and AP75 scores. However, the Mask R-CNN models were better at detecting enamel compared to Faster R-CNN models, with perfect AP and APm values. Mask R-CNN backbone with ResNet-101 showed the best performance with an AP of 98.44%, compared to ResNet-50, which achieved an AP score of 86.30%. These results show that making the backbone deeper improves segmentation considerably without reducing detection accuracy. The proposed system with Mask R-CNN ResNet-101 can help dentists to find cavities in enamel quickly and accurately. Mask R-CNN with ResNet-101 has an excellent detection capacity with negligible errors. Future research should focus on external datasets and test different frameworks to detect other dental issues, in addition to just cavities in the enamel.

REFERENCES

[1] Y. Liu *et al.*, "Richer Convolutional Features for Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1939–1946, Aug. 2019, https://doi.org/10.1109/TPAMI.2018.2878849.

[2] S. Kanagamalliga, R. Jayashree, and M. Chouksey, "Fast R-CNN Approaches for Transforming Dental Caries Detection: An In-Depth Investigation," in *2024 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Chennai, India, Mar. 2024, pp. 1–5, https://doi.org/10.1109/WiSPNET61464.2024.10532984.

[3] R. Hemalatha, G. Amulya, and Ch. S. N. S. Lalitha, "Computer Vision Techniques for Object Detection," in *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems*, Chennai, India, Dec. 2024, pp. 1–6, https://doi.org/10.1109/ICSES63760.2024.10910496.

[4] J. Sun and K. Jia, "Research on Semantic Segmentation of Ground-Based Cloud Image Based on Fully Convolutional Network," in *China Automation Congress*, Xiamen, China, Nov. 2022, pp. 5850–5854, https://doi.org/10.1109/CAC57257.2022.10055648.

[5] Y. E. Kang, W. Kang, T. Lee, and H. S. Chwa, "Paste-and-Cut: Collective Image Localization and Classification for Real-Time Multi-Camera Object Detection," in *2023 14th International Conference on Information and Communication Technology Convergence*, Jeju Island, Republic of Korea, Oct. 2023, pp. 740–742, https://doi.org/10.1109/ICTC58733.2023.10393851.

[6] S. Gobinath, M. Almusawi, G. Shalini, B. Pruthviraj Goud, and K. Sudheer Kumar, "Object Detection in High-Resolution Aerial Images Using Single Shot Multi-Box Detector," in *2024 International Conference on Data Science and Network Security*, Tiptur, India, Jul. 2024, pp. 1–4, https://doi.org/10.1109/ICDSNS62112.2024.10691194.

[7] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision*, Venice, Italy, Oct. 2017, pp. 2980–2988, https://doi.org/10.1109/ICCV.2017.322.

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, https://doi.org/10.1109/TPAMI.2017.2699184.

[9] Y. Gao, X. Wang, and Q. Li, "Improved Cervical Cell Segmentation Algorithm Based on Cascade Mask RCNN," in *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference*, Chongqing, China, Sept. 2023, pp. 1775–1780, https://doi.org/10.1109/ITOEC57671.2023.10291487.

[10] D. A. Agneya, M. S. Shekar, A. Bharadwaj, N. Vineeth, and M. L. Neelima, "Deep Learning in Medical Image Analysis: A Survey," in *2024 International Conference on Innovation and Novelty in Engineering and Technology*, Vijayapura, India, Dec. 2024, pp. 1–5, https://doi.org/10.1109/INNOVA63080.2024.10847040.

[11] R. Nambiar and R. Nanjundegowda, "Detection of Missing Tooth Regions Using Deep Learning in Panoramic Radiographs for Dental Implant Planning," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28071–28076, Oct. 2025, https://doi.org/10.48084/etasr.13101.

[12] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," in *2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Republic of Korea, Oct. 2019, pp. 9156–9165, https://doi.org/10.1109/ICCV.2019.00925.

[13] L. V. Comia and E. D. Festijo, "Performance Analysis of Original Implementation of ResNet50-Mask-RCNN using Transfer Learning: A Benchmark Data for Backbone-Improved Based Future Comparative Studies," in *2024 28th International Conference on Information Technology*, Zabljak, Montenegro, Feb. 2024, pp. 1–6, https://doi.org/10.1109/IT61232.2024.10475763.

[14] D. De Geus and G. Dubbelman, "Task-Aligned Part-Aware Panoptic Segmentation Through Joint Object-Part Representations," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, Jun. 2024, pp. 3174–3183, https://doi.org/10.1109/CVPR52733.2024.00306.

[15] V. Singh, K. Singh, R. Sehrawat, A. P. Singh, J. Augustine, and M. Verma, "Deep Learning Techniques for Dental Caries Detection and Prosthodontics: A Systematic Review," in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking*, Ghaziabad, India, Feb. 2025, pp. 651–655, https://doi.org/10.1109/CICTN64563.2025.10932642.

[16] Y. Zhu *et al.*, "Faster-RCNN Based Intelligent Detection and Localization of Dental Caries," *Displays*, vol. 74, Sept. 2022, Art. no. 102201, https://doi.org/10.1016/j.displa.2022.102201.

[17] H. Mohammad-Rahimi *et al.*, "Deep Learning for Caries Detection: A Systematic Review," *Journal of Dentistry*, vol. 122, July 2022, Art. no. 104115, https://doi.org/10.1016/j.jdent.2022.104115.

[18] T.-J. Lin *et al.*, "Evaluation of the Alveolar Crest and Cemento-Enamel Junction in Periodontitis Using Object Detection on Periapical Radiographs," *Diagnostics*, vol. 14, no. 15, Aug. 2024, Art. no. 1687, https://doi.org/10.3390/diagnostics14151687.

[19] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and Diagnosis of Dental Caries Using a Deep Learning-based Convolutional Neural Network Algorithm," *Journal of Dentistry*, vol. 77, pp. 106–111, Oct. 2018, https://doi.org/10.1016/j.jdent.2018.07.015.

[20] A. AlSayyed *et al.*, "Employing CNN Ensemble Models in Classifying Dental Caries Using Oral Photographs," *International Journal of Data and Network Science*, vol. 7, no. 4, pp. 1535–1550, 2023, https://doi.org/10.5267/j.ijdns.2023.8.009.

[21] O. Khattak *et al.*, "Deep Learning Applications in Dental Image-Based Diagnostics: A Systematic Review," *Healthcare*, vol. 13, no. 12, June 2025, Art. no. 1466, https://doi.org/10.3390/healthcare13121466.

[22] F. Schwendicke *et al.*, "Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection," *Journal of Dental Research*, vol. 100, no. 4, pp. 369–376, Apr. 2021, https://doi.org/10.1177/0022034520972335.

[23] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122-1131, Feb. 2018, https://doi.org/10.1016/j.cell.2018.02.010.

[24] M. He *et al.*, "Lightweight Mask R-CNN for Instance Segmentation and Particle Physical Property Analysis in Multiphase Flow," *Powder Technology*, vol. 449, Jan. 2025, Art. no. 120366, https://doi.org/10.1016/j.powtec.2024.120366.