Detection of Missing Tooth Regions Using Deep Learning in Panoramic Radiographs for Dental Implant Planning

Rajashree Nambiar
Department of Robotics & AI, NMAM Institute of Technology (NMAMIT), Nitte (Deemed to be University), Nitte, Karnataka, India | Faculty of Engineering and Technology, JAIN (Deemed to be University), Bengaluru, India
raji24oct@gmail.com (corresponding author)

Raghu Nanjundegowda
Faculty of Engineering and Technology, Department of Electrical and Electronics Engineering, JAIN (Deemed to be University), Bengaluru, India
n.raghu@jainuniversity.ac.in

Received: 1 July 2025 | Revised: 22 July 2025, 3 August 2025, and 17 August 2025 | Accepted: 20 August 2025

 DOI: https://doi.org/10.48084/etasr.13101

ABSTRACT

The advances in dental radiology, particularly the utilization of panoramic radiographs, have significantly enhanced the precision of dental implant planning. This study introduces a novel deep learning-based approach for detecting missing tooth regions in panoramic radiographs, leveraging a region-based Convolution Neural Network (Mask R-CNN) with a Residual Neural Network (ResNet-101) to enhance the extraction of features from input data, such as the backbone for tooth segmentation and numbering. By integrating a heuristic algorithm, the proposed method improves detection accuracy and addresses common challenges such as multiple numbering errors and misalignments. The model was evaluated using a robust dataset, demonstrating superior performance metrics, including a precision of 0.9566, a recall of 0.9635, and a mean Average Precision (mAP) of 0.9241, compared to conventional methods. The results affirm the potential of this automated system to streamline dental implant planning, reduce clinician workload, and support advanced diagnostic and educational tools.

Keywords-artificial intelligence; deep learning; dental implant; dentistry; image segmentation; Mask R-CNN; pragmatic algorithm

I. INTRODUCTION

Dental implant planning is an essential component of restorative dentistry that requires a precise evaluation of the anatomical components to ensure appropriate implant positioning. This often involves locating missing teeth and determining appropriate implant sites using panoramic radiography or Cone Beam Computed Tomography (CBCT) scans. Although CBCT provides three-dimensional data, panoramic radiographs continue to be prevalent because of their lower cost, fast collection, and lower radiation exposure. Historically, implant planning has relied on manual interpretation, which is labor-intensive and prone to fluctuation. In clinical situations necessitating numerous implants, the ability to identify several missing teeth is crucial. This process is confounded by anatomical heterogeneity and inconsistent patterns of tooth loss, which can range from single missing teeth to many edentulous areas. Artificial Intelligence (AI), especially Deep Learning (DL), has demonstrated significant promise in automating diagnostic procedures in dentistry, including caries diagnosis, molar extraction, and anatomical segmentation [1]. In [2], a method was presented to detect wisdom teeth from panoramic images, but this study was restricted to classification and did not include tooth numbering. In [3], it was observed that segmentation accuracy can be improved with the use of larger datasets. In [4], Faster R-CNN and VGG-16 were combined to identify and enumerate teeth, but this multi-stage DL pipeline proved to be computationally demanding. In [5], a comparable method was proposed for periapical films, which relied on complete panoramic images. The study in [6] focused on enumerating only the teeth associated with lesions in the identification of periodontal bone loss. In [7], R-CNN, Single-Shot Detector (SSD), and heuristics were combined for tooth counting, but certain metrics were still inadequate. In [8], spatial correlation was used to improve detection accuracy. In [9], YOLOv4 was used as a real-time object detection model for pediatric panoramic images, achieving an mAP of 92.22%. As can be observed, most of these investigations prioritize detection and enumeration.

The key contributions of this study are threefold.

1. Integrates a pragmatic algorithm with a DL-based segmentation model to enhance tooth numbering accuracy.
2. Applies a fully automated method, using Mask R-CNN with ResNet-101 specifically trained on panoramic radiographs, achieving high mAP and precision.
3. Performs a structured comparison against state-of-the-art methods under identical dataset conditions.

Unlike existing approaches that often separate detection and enumeration processes, the proposed one simultaneously performs both with improved consistency, making it highly practical for real-world dental implant planning.

II. PROPOSED METHOD

The proposed approach seeks to improve the results reported in the literature and achieve high accuracy by utilizing the outputs generated from a traditional DL algorithm combined with a pragmatic one. This study also presents a novel approach to reduce dentists' effort and facilitate diagnostic procedures. The first phase comprises the automated identification, segmentation, and enumeration of teeth based on the Federation Dentaire Internationale (FDI) globally used numbering system. The subsequent step focuses on identifying missing teeth in the images using a missing tooth number detection model.

A. Dataset Details

The datasets used are publicly available and accessed under an open-source license [10, 11]. Combining these datasets offers 1856 panoramic radiographs after augmentation, annotated with the help of dental radiologists using a web-based labeling tool (VGG Annotator) to define tooth boundaries.

Covering a broad spectrum of dental disorders, including partial and total edentulism, the dataset comprises individuals aged 18–70 with a near-equal gender distribution (52% female, 48% male). To evaluate the model's accuracy, it is crucial to avoid using images that were part of the training process. The test dataset consisted of 279 images, distinct from the training set. Out of the total dataset, 1577 images were randomly selected for training the model. K-fold cross-validation ($k = 5$) was used to validate the model, with each fold including 315 validation images and 1261 training images. Figure 1 shows a sample image from the dataset.

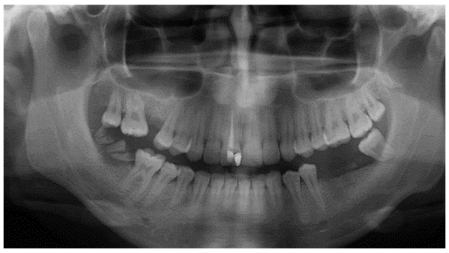


Fig. 1. A panoramic view from the dataset.

B. Tooth Detection Model

The proposed model is designed to segment 32 teeth in panoramic radiographic images. For this purpose, Mask R-CNN was utilized, as it has demonstrated excellent performance in image segmentation tasks. ResNet-101, a deep neural network architecture frequently used for feature extraction in applications like segmentation and object recognition, is the foundation of the segmentation model. With its 101 layers and residual blocks, ResNet-101 tackles the degradation issue by keeping the input-output disparity to a minimum. Figure 2 shows a simplified diagram of the Mask R-CNN network structure applied to the dental dataset.

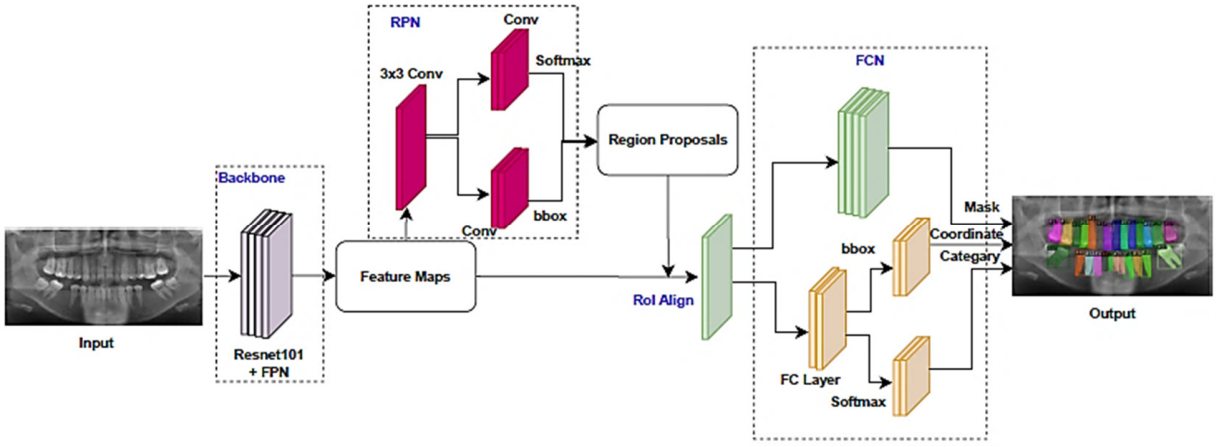


Fig. 2. System architecture for tooth numbering.

The most important feature of Mask R-CNN is its capacity to carry out instance segmentation, which entails locating and identifying every object of interest within an image. Mask R-CNN uses a CNN as a feature extractor. The ResNet-101 network is commonly used as the backbone for extracting image features [12]. The first part of Mask R-CNN is the Region Proposal Network (RPN), which finds potential image object Regions of Interest (RoIs). The second part, in

conjunction with classification and bounding box predictions, creates binary masks for each of these RoIs. Panoramic images undergo processing via a two-step RPN mesh. Within the residual backbone network, the features of the dental image are identified and passed on to the pooling layer.

Non-Maximum Suppression (NMS) is applied to get rid of boxes that overlap each other. The model uses a threshold value to segment regions that include objects using a connection-based method. The CNN uses the pooling layer to choose the highest values. For every detected area in the image, the model creates a final output that includes bounding boxes and masks. The class-specific segmentation masks stand in for the target entities that need to be identified.

$$L = L_{cls} + L_{bbox} + L_{mask} \quad (1)$$

where L_{cls} and L_{bbox} are the class and bounding box losses, respectively, while L_{mask} represents the segmentation mask loss, expressed as follows:

$$L_{mask} = -\frac{1}{m^2}\sum_{i \le j \le m} \left[y_{ij} log(\hat{y}_{ij}^{k} + 1 - y_{ij}) log\ (1 - \hat{y}_{ij}^{k}) \right] \quad (2)$$

where k is the aggregate quantity of classes. In recognition of each class and RoI, a mask of specified dimensions $m \times m$ is generated, with the output size set to $k \times m^2$. $\hat{y}_{ij}^{k}$ represents the predicted value for class k, while y_{ij} corresponds to the tag at point i, j in the generated mask.

The model was trained with 32+1 classes, processing 2 images per GPU with a total batch size of 2. ResNet-101 serves as the backbone architecture, and the feature pyramid network strides are set at 4, 8, 16, 32, and 64. To avoid overfitting, the learning rate is set at 0.0001, while each training epoch includes 630 steps based on the batch size of 2, and validation is performed over 137 steps. ResNet-101 was employed for its better feature extraction, as preliminary comparisons showed that ResNet-50 and MobileNet were less accurate.

C. Pragmatic Algorithm

The output of the Mask R-CNN is a set of dental masks that can be used to identify teeth in panoramic images. The masks are subsequently run through an algorithm that was designed to improve the model's accuracy. This technique gets rid of duplicate number assignments and fixes errors in numbering for the same teeth. A two-step pragmatic approach was proposed to enhance the model's prediction accuracy. Figure 3 shows the structure of the proposed pragmatic algorithm.

The model sometimes generates multiple FDI numbers for a tooth that exhibits high similarity with others. The initial phase of the pragmatic method tackles this concern. Initially, the predicted teeth are sorted based on their prediction confidence scores. Through this sorting process, teeth assigned multiple FDI numbers are identified by examining the prediction results. In these instances, the mask, prediction score, and class data linked to the lowest confidence score are eliminated. The second step resolves errors in numbering caused by misalignment. The anticipated tooth mask coordinates are ordered ascendingly along the x-axis to ensure that the predicted teeth keep this sequence. The FDI values of neighboring teeth are compared, beginning with the first tooth. If a discrepancy in the sequence is detected, the misaligned tooth is identified and assigned the correct FDI number.

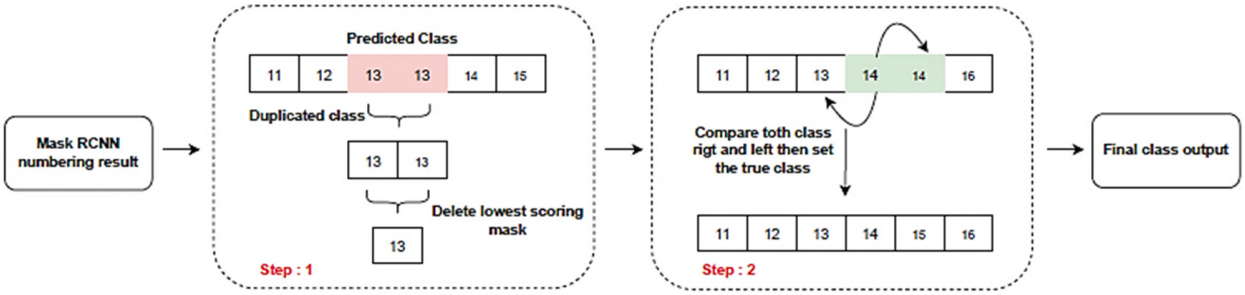


Fig. 3. Pragmatic algorithm steps.

D. Missing Tooth Number Detection

After obtaining the predicted tooth regions from Mask R-CNN, the missing teeth and their locations are identified by analyzing the predicted adjacent teeth. The detected tooth numbers are stored in an array, called *Predicted_Teeth_Array*, which is then compared with the complete FDI numbering array (*FDI_Array*), containing all tooth numbers (11–48). Any tooth present in the FDI array but absent in the predicted array is considered missing and added to the *Missing_Teeth* list. Once missing teeth are identified, their regions are classified based on their FDI numbers. Teeth numbered between 11 and 28 are classified as Maxillary, while those numbered between 31 and 48 are classified as Mandibular. This categorization is useful to determine if the lost tooth is from the top or bottom jaw. Finally, the model outputs the missing tooth numbers along with their respective regions, providing a comprehensive prediction of missing teeth and their positions for further dental analysis or implant planning.

III. RESULTS AND DISCUSSIONS

The validation loss values from each of the 5-fold cross-validation were used to assess the model's performance after training. Figure 4 highlights the best epoch value corresponding to the lowest validation loss achieved in each fold. Among the cross-validations, the 275th epoch in the fifth fold demonstrated the lowest validation loss. This model was used on the test dataset, which consisted of 279 images. To enhance the results, the previously described pragmatic

algorithm was applied. Table I summarizes the results of the baseline DL model and the proposed pragmatic approach. The pragmatic algorithm outperformed the baseline model by approximately 4% across all evaluation measures, showing that the suggested method is effective.

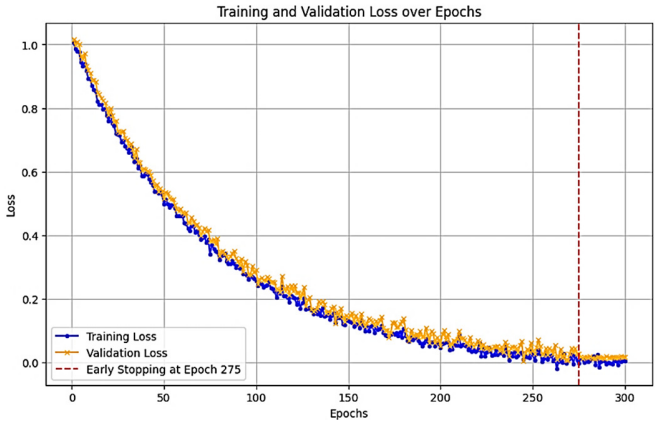


Fig. 4. Evaluation of training and validation loss for the fifth-fold cross cross-validation. The best epoch is 275.

The following precision and recall metrics were used to gauge how accurate tooth detection is:

$$Detection\ Precision = \frac{N_{mbp}}{N_{db}} \quad (3)$$

$$Detection\ Recall = \frac{N_{mbp}}{N_{gtb}} \quad (4)$$

where N_{mbp} the number of matched box-pairs, N_{db} is the number of detected boxes, and N_{gtb} is the number of ground truth boxes. The average Intersection over Union (IoU) value of the matching boxes was examined to determine how well the identified boxes relate to the ground truth ones.

$$Mean\ IoU = \frac{\sum IoU_{mbp}}{N_{mbp}} \quad (5)$$

When a detected box and its corresponding ground truth box share the same tooth number label, it is considered correctly numbered, which is referred to as a true positive numbering (tpn). The precision and recall for teeth numbering can be computed as follows:

$$Numbering\ Precision = \frac{N_{tpn}}{N_{db}} \quad (6)$$

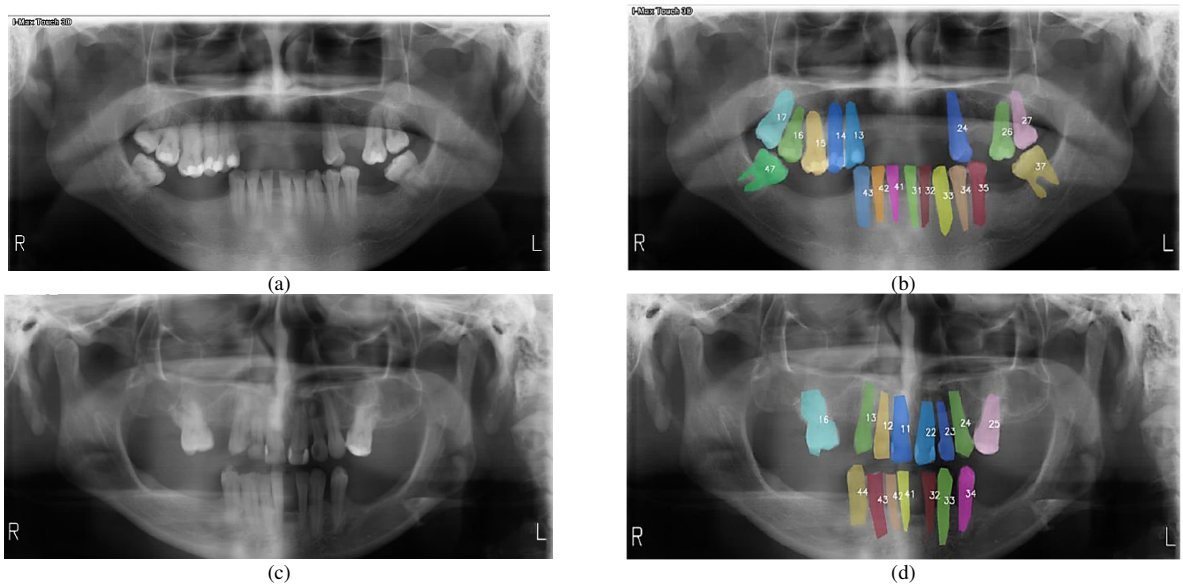
$$Numbering\ Recall = \frac{N_{tpn}}{N_{gtb}} \quad (7)$$

Table I shows the sequential order of the steps involved in the automated tooth-detecting and counting system. Stage 1 uses a trained Mask R-CNN to find the tooth's bounding boxes, Stage 2 eliminates duplicate classes, Stage 3 matches with nearby classes, and Stage 4 detects whether a tooth is missing and ensures that everything lines up with the template.

Figure 5(a) and (c) show test input images, and Figure 5(b) and (d) illustrate the masked outputs and numbering tasks.

TABLE I. ACCURACY, RECALL, AND IOU FOR BOXES DISCOVERED IN THE TEST DATASET

Object	Stage1	Stage 2	Stage 3	Stage 4
Detection precision	0.895	0.985	0.985	0.985
Detection recall	0.982	0.982	0.982	0.982
Mean IoU	0.90±0.05	0.91±0.05	0.91±0.05	0.91± 0.05
Numbering precision	0.720	0.800	0.890	0.902
Numbering recall	0.780	0.790	0.880	0.904



(a) (b) (c) (d)

Fig. 5. (a, c): Input images from the dataset, (b, d): segmented results of the input image.

The confusion matrices in Figure 6 classify predictions as True Positives (TP), where a tooth is correctly detected ($IoU \geq threshold$) and correctly numbered, False Positives (FP), where a tooth is detected but has either an IoU below the threshold or incorrect numbering, and False Negatives (FN), where a tooth present in the ground truth is not detected by the model. From these categories, precision is calculated to show the proportion of correctly detected and numbered teeth among all predictions. Recall is the percentage of projected teeth in the ground truth that really exist. The F1-score strikes a balance between recall and precision by taking their harmonic mean. IoU evaluates the overlap between predicted tooth masks and ground truth, with predictions requiring an IoU above a defined threshold (e.g., 0.5) to be considered TP.

Finally, mAP provides a comprehensive assessment of detection and numbering performance by computing precision-recall curves for each tooth class and calculating the area under these curves. The mean of average precision across all tooth classes yields the mAP, providing an overall performance metric for models handling up to 32 classes in adult dentition. The results indicate low false-positive rates, which are important for implant planning. The mAP of 0.9241 of the proposed method is higher than PANet's (0.88), indicating that the pragmatic improvement improves numbering accuracy. Table II shows the performance metrics for each tooth. Compared to Mask R-CNN, the suggested Mask R-CNN with the pragmatic algorithm clearly achieves better results in terms of numbering.

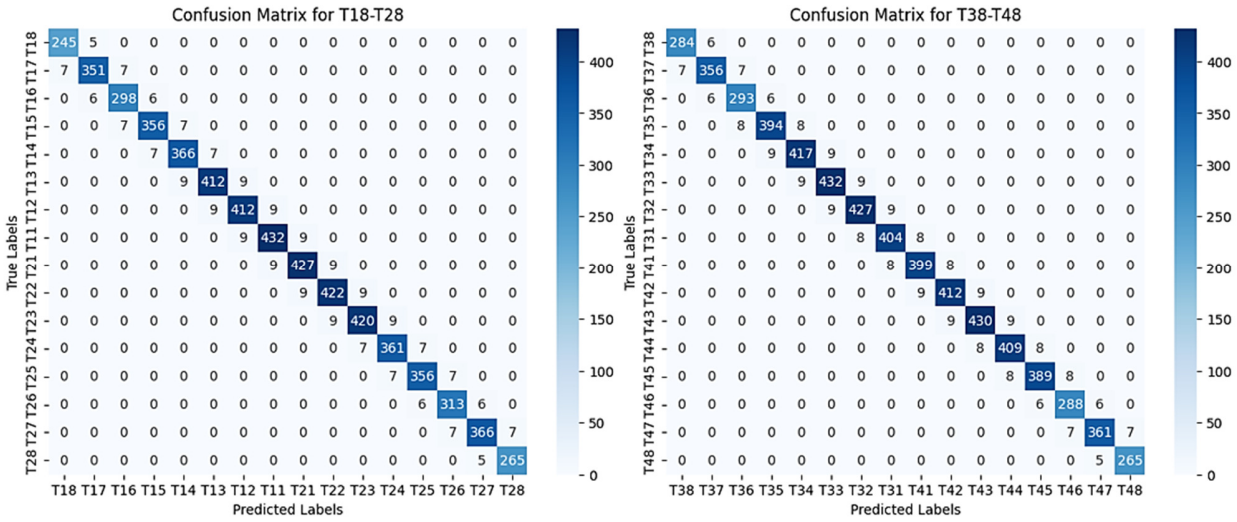


Fig. 6. Confusion matrices for maxillary and mandibular jaw teeth.

TABLE II. PERFORMANCE METRICS FOR VARIOUS TEETH (T#)

T#	Tp	Fp	Fn	Pre	Rec	F1	mAP	IoU	T#	Tp	Fp	Fn	Pre	Rec	F1	mAP	IoU
T18	250	28	25	0.90	0.91	0.91	0.85	0.88	T38	290	14	13	0.95	0.95	0.95	0.91	0.93
T17	365	36	30	0.91	0.92	0.91	0.87	0.89	T37	370	15	10	0.96	0.97	0.96	0.93	0.95
T16	310	15	13	0.95	0.96	0.96	0.91	0.93	T36	305	18	15	0.94	0.96	0.95	0.91	0.93
T15	370	25	20	0.94	0.95	0.94	0.90	0.92	T35	410	10	7	0.98	0.98	0.98	0.95	0.97
T14	380	13	12	0.97	0.96	0.96	0.94	0.96	T34	435	8	6	0.99	0.99	0.99	0.96	0.98
T13	430	8	7	0.98	0.98	0.98	0.95	0.97	T33	450	6	5	0.99	0.99	0.99	0.97	0.99
T12	430	7	6	0.99	0.99	0.99	0.96	0.98	T32	445	9	8	0.98	0.98	0.98	0.95	0.97
T11	450	4	3	0.99	0.99	0.99	0.97	0.99	T31	420	14	12	0.96	0.97	0.96	0.93	0.95
T21	445	5	6	0.98	0.99	0.99	0.96	0.98	T41	415	12	10	0.97	0.98	0.98	0.94	0.96
T22	440	6	5	0.99	0.99	0.99	0.96	0.98	T42	430	10	9	0.98	0.98	0.98	0.95	0.97
T23	438	7	5	0.98	0.99	0.98	0.95	0.97	T43	448	8	7	0.98	0.98	0.98	0.95	0.97
T24	375	13	11	0.96	0.97	0.96	0.92	0.94	T44	425	12	9	0.97	0.98	0.97	0.94	0.96
T25	370	15	14	0.94	0.96	0.95	0.91	0.93	T45	405	14	13	0.96	0.96	0.96	0.92	0.94
T26	325	30	25	0.91	0.93	0.92	0.88	0.90	T46	300	30	28	0.91	0.92	0.91	0.88	0.90
T27	380	28	26	0.92	0.93	0.93	0.89	0.91	T47	375	24	22	0.94	0.94	0.94	0.90	0.92
T28	270	25	20	0.91	0.93	0.92	0.87	0.89	T48	270	16	14	0.93	0.95	0.94	0.89	0.91

Table III presents a comparative evaluation of the proposed method against traditional Mask R-CNN and PANet architectures using the Kaggle panoramic radiograph dataset, following the implementation guidelines specified in [13]. These models were trained and tested under consistent conditions to ensure a fair comparison. Notably, while PANet achieved a respectable mAP of 0.88, the proposed method outperformed it with 0.92, primarily due to the integration of the pragmatic algorithm, which improved numbering accuracy and reduced misclassification in anatomically complex regions.

TABLE III. COMPARISON OF THE PROPOSED METHOD WITH CONVENTIONAL METHODS

Architecture	Recall	Precision	F1-score	mAP	Dice coeff
Mask R-CNN	0.91	0.92	0.92	0.89	0.88
PANet	0.89	0.94	0.91	0.88	0.90
Proposed	0.96	0.95	0.95	0.92	0.92

Using a pragmatic method with Mask R-CNN, this study achieved an mAP of 0.9241 when it came to counting. The use of a post-processing algorithm specifically designed for dental images is the main reason why the suggested approach performed better than PANet. The proposed method uses heuristic adjustments and anatomical ordering to preserve tooth placement consistency.

IV. CONCLUSION

This work fills a crucial gap in implant design, where manual interpretation is often time-consuming and prone to errors, by presenting an automated approach to identify missing tooth areas on panoramic radiographs. The proposed method outperformed traditional techniques, such as PANet, by combining Mask R-CNN with ResNet-101 and a pragmatic algorithm, resulting in high accuracy (precision 95.66%, recall 96.35%, mAP 0.9241). The main innovation in this study is the combination of a pragmatic correction technique with DL-based segmentation, which reduces computational cost and increases numbering consistency. The suggested approach provides a more comprehensive and effective solution for automated analysis when compared to previous research that just focuses on detection or counting. However, there are still difficulties, especially when it comes to treating complicated cases with significant anatomical variances or radiographic abnormalities and diverse datasets. To improve applicability, future studies should investigate implementation in real-time clinical settings, integration with CBCT, and validation on multi-center datasets. All things considered, this study is a solid step toward clinically deployable tools for precise and effective dental implant design, and it enhances automated dental image analysis.

REFERENCES

[1] W. Fan, J. Zhang, N. Wang, J. Li, and L. Hu, "The Application of Deep Learning on CBCT in Dentistry," *Diagnostics*, vol. 13, no. 12, Jun. 2023, Art. no. 2056, https://doi.org/10.3390/diagnostics13122056.

[2] E. Gardiyanoğlu, G. Ünsal, N. Akkaya, S. Aksoy, and K. Orhan, "Automatic Segmentation of Teeth, Crown–Bridge Restorations, Dental Implants, Restorative Fillings, Dental Caries, Residual Roots, and Root Canal Fillings on Orthopantomographs: Convenience and Pitfalls," *Diagnostics*, vol. 13, no. 8, Apr. 2023, Art. no. 1487, https://doi.org/10.3390/diagnostics13081487.

[3] Y. Y. Amer and M. J. Aqel, "An Efficient Segmentation Algorithm for Panoramic Dental Images," *Procedia Computer Science*, vol. 65, pp. 718–725, 2015, https://doi.org/10.1016/j.procs.2015.09.016.

[4] D. V. Tuzoff *et al.*, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiology*, vol. 48, no. 4, May 2019, Art. no. 20180051, https://doi.org/10.1259/dmfr.20180051.

[5] H. Chen *et al.*, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, no. 1, Mar. 2019, Art. no. 3840, https://doi.org/10.1038/s41598-019-40414-y.

[6] J. Kim, H. S. Lee, I. S. Song, and K. H. Jung, "DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs," *Scientific Reports*, vol. 9, no. 1, Nov. 2019, Art. no. 17615, https://doi.org/10.1038/s41598-019-53758-2.

[7] C. Kim, D. Kim, H. Jeong, S. J. Yoon, and S. Youm, "Automatic Tooth Detection and Numbering Using a Combination of a CNN and Heuristic Algorithm," *Applied Sciences*, vol. 10, no. 16, Aug. 2020, Art. no. 5624, https://doi.org/10.3390/app10165624.

[8] F. P. Mahdi, K. Motoki, and S. Kobashi, "Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs," *Scientific Reports*, vol. 10, no. 1, Nov. 2020, Art. no. 19261, https://doi.org/10.1038/s41598-020-75887-9.

[9] N. Akkaya, Ö. Kansu, H. Kansu, L. Çağirankaya, and U. Arslan, "Comparing the accuracy of panoramic and intraoral radiography in the diagnosis of proximal caries," *Dentomaxillofacial Radiology*, vol. 35, no. 3, pp. 170–174, May 2006, https://doi.org/10.1259/dmfr/26750940.

[10] Humans In The Loop, "Teeth Segmentation on dental X-ray images." Kaggle, https://doi.org/10.34740/KAGGLE/DSV/5884500.

[11] A. Abdi and S. Kasaei, "Panoramic Dental X-rays With Segmented Mandibles," Jul. 2020, https://doi.org/10.17632/hxt48yk462.2.

[12] S. Aparna, H. Gottumukkala, N. Shivampet, K. Muppavaram, and C. C. V. Ramayanam, "Advancements in Dental Filling Detection Technologies and Strategies for Comprehensive Oral Health Care," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14470–14474, Jun. 2024, https://doi.org/10.48084/etasr.7400.

[13] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 8759–8768, https://doi.org/10.1109/CVPR.2018.00913.