ISSN:2790-1688 Volume-14-(2025)

Teeth segmentation and recognition on dental panoramic radiographs using improved Mask RCNN

Shuying Liu [1, a], Wu Wang [1, b], Yunhao Wu [1, c], Qinqin Chai [1, d, *]

[1] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108

[a] 17716593820@163.com, [b] wangwu@fzu.edu.cn, [c] w352275906@gmail.com,

[d, *] qq.chai@fzu.edu.cn

Abstract. Accurately identify the type of tooth and its morphology is essential for the planning of dental implant surgery. However, manual recognition relies on the experience of dentists and usually has low recognition accuracy. And the existing instance segmentation models are difficult to realize accurately segmentation and identification at the same time due to the characteristics of highly imbalanced in teeth scale and low contrast of the real dental panoramic radiograph datasets. For this, this study proposes an automatic tooth recognition method based on improved Mask RCNN. In which, area detection module of Mask RCNN is improved through path enhancement and balancing mechanism is designed to solve the problem of low recognition accuracy caused by insufficient feature extraction of Mask RCNN. Experiments on real panoramic radiograph dataset show that the proposed method achieves fusion extraction of tooth type and morphology information. While the average accuracy of instance segmentation mAP(0.5) and mAP(0.5:0.95) for the testing set were 96.69% and 74.2%, respectively, and the recognition rate of the missing tooth reaches 93.81%. Ablation experimental results verify the effectiveness of the proposed path enhancement and balancing mechanism in increasing the accuracy of tooth classification and segmentation. The research results can promote the application of artificial intelligence-assisted diagnosis and treatment in the field of oral implants.

Keywords: Recognition and Segmentation; Panoramic Radiographs; Mask RCNN; Path Enhancement.

1. Introduction

Before placing a dental implant, locate the area of the tooth on the patient's dental panoramic radiographs is important for a dentist to determine the appropriate implant products[1].However, at present the recognition of teeth in panoramic radiographs is mainly dependent on the doctor's experience. Misrecognition and omission recognition phenomenon usually occur, especially in the areas where medical resources are scarce, which greatly affects the therapeutic effect of dental diseases[2]. To overcome this problem, the application of artificial intelligence technique to automatically realize teeth recognition and segmentation in the oral cavity become a new way to improve the misrecognition rate, which has also become a hot research topic.

There are quite a few papers in the literature which are devoted to the study of segmentation and recognition models for key targets in panoramic radiographs. For segmentation models, a segmentation model based on contours is proposed in to segment the cyst area in the dental film and promoted dentists' research on cyst diseases[3]. The apical lesion area on oral panoramic dental radiographs is segmented by U-Net model in to improve the diagnostic success rate of apical lesions[4]. And a U-Net model is implemented in to segment tooth edges using a weighted loss function[5]. However, these works only consider the segmentation of teeth, fail to localize the tooth numbers in the panoramic radiographs. For the tooth recognition models, a YOLOv3 model is used in to classify and detect dental diseases(cavities, root canals, crowns, and fractured root canals) and developed an automated tool for diagnosing dental abnormalities[6]. And Various deep pre-trained models are discussed in to design an automatic diagnosis system to detect caries in panoramic dental radiographs[7]. A faster R-CNN is proposed to extract rectangular regions of all teeth and achieved automatic detection of tooth types in [8]. However, the above classification or segmentation methods can only extract tooth type information or morphology information. They

cannot fully meet the needs of dentists for type and shape information of the teeth in assisted diagnosis and treatment.

To meet the real need of realizing tooth recognition and segmentation at the same time, a Mask-RCNN is utilized in [9]. However, the datasets constructed in only contain the real dental panoramic radiographs of young patient. In reality, there are significant differences in dental status among different age groups. Usually, the dental problems of the elderly groups are more serious and complex. The incompleteness of the dataset limits the application of Mask-RCNN proposed in [9]. Moreover, the imbalance of the data scale makes the Mask-RCNN model susceptible to the influence of artifacts and implants in the panoramic radiographs, resulting in a low accuracy of the model recognition. To overcome this drawbacks, this paper proposes an improved Mask R-CNN model to solve the problem of low accuracy of tooth segmentation and tooth type recognition. The effectiveness of the model is also verified on real data set with more complex dental morphology. The results of this study can provide diagnostic help for dentists and planning for subsequent surgery.

The remainder of this paper is organized as follows. Section 2 describes the datasets. Section 3 introduces the improved Mask R-CNN model. In Section 4, experimental results are conducted and discussed. Finally, the conclusions are summarized in Section 5.

2. Data Acquisition

Dental panoramic radiographs used in this paper are obtained from a dental hospital in Fujian province, China. Totally, 130 samples are collected by specialized doctors in the hospital containing four types of dental morphology: full teeth, missing teeth, intra-mandibular implants, and dental restorations. The original radiographs are cropped from 1194×821 pixels to 1057×690 pixels and converted to PNG images to remove redundant information and reduce the image size, so as to improve the accuracy and speed of training. The dataset is divided into training, validation, and testing sets at a ratio of 10:1:2. Part of the PNG images are shown in Figure 1.

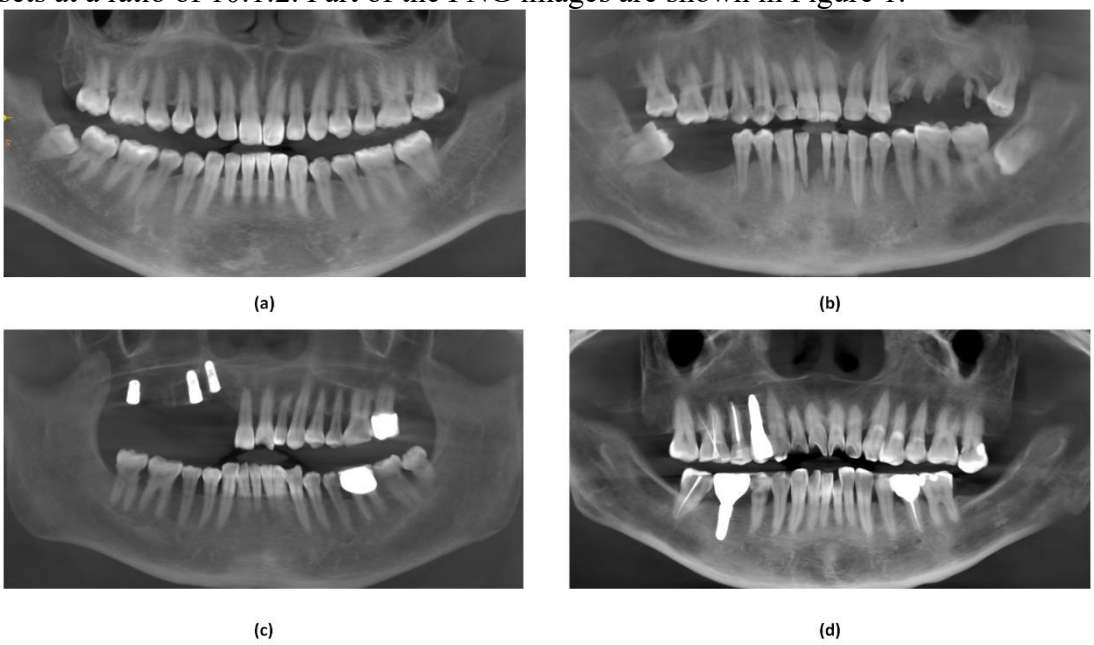

Figure 1 Part of the original PNG images in the training set
(a) Full teeth (b) Missing teeth (c) Intramandibular implants (d) Dental restorations

Then, the tooth positions are numbered according to the World Dental Federation tooth numbering system(ISO-3950).Considering that the third molars (also known as wisdom teeth) are located in the deepest part of the mouth and are often removed due to oral problems ,and the third molars are not included in the dental implant surgery, so the third molars are not considered in this study. The numbered tooth positions are shown in Table 1.

Table 1 FDI/ISO-3950 tooth position indication for permanent teeth

upper right	upper left
17 16 15 14 13 12 11	21 22 23 24 25 26 27
47 46 45 44 43 42 41	31 32 33 34 35 36 37
lower right	lower left

The teeth are then labeled using the image annotation software Labelme by polygons. The annotated original image and its mask form are shown in Figure 2. The annotated sample images are transformed into Common Objects in Context format for subsequent recognition modeling research.

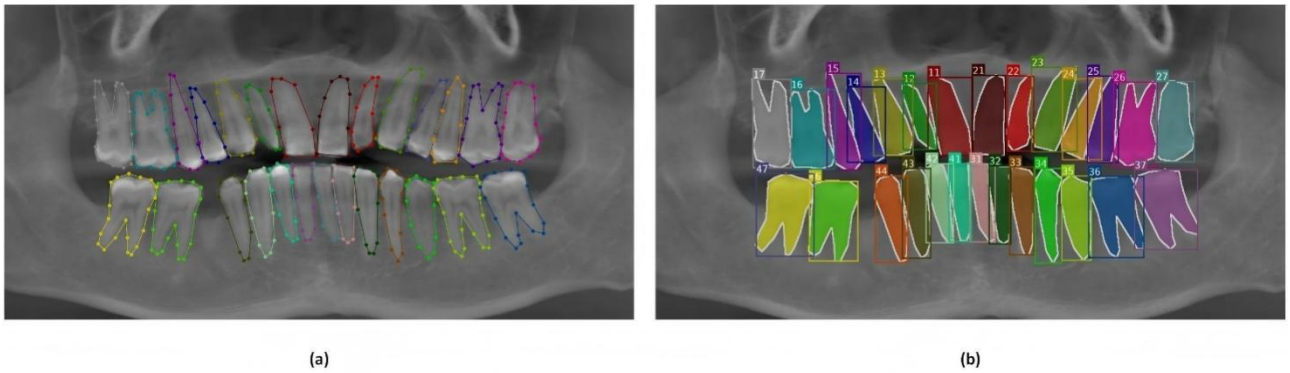

Figure 2 Part of the panoramic image annotation data. (a) Annotated original image (b) Masked image

As seen in Figure 1, there are various types of teeth in the dataset, and there are great difference in the patient's tooth positions and tooth sizes, Using the Mask-RCNN model directly achieves poor segmentation and recognition performances. Therefore, the performance of the Mask RCNN model needs to be improved.

3. Instance Segmentation Model Based on Improved Mask RCNN

3.1 Improved Mask RCNN model

Mask RCNN[10] is an instance segmentation model proposed by He et al. It can perform pixel-level segmentation for each detected target by adding a branch to Faster RCNN. The Mask RCNN network first uses the residual network (ResNet50) and the feature pyramid network (FPN) to extract effective features of the input images. The extracted feature map is then used to extract region proposals through the Region Proposal Network(RPN).Then, these region proposals are input into ROI Align and mapped into fixed-dimensional feature vectors through bilinear interpolation. Finally, the feature vectors are input into the fully connected layer for classification and bounding box regression.

Since Mask RCNN extracts the target region features before segmentation, its instance segmentation is affected by the feature levels in the target region feature map. Usually, higher-level features have stronger semantic information that facilitates object classification; low-level features have stronger location information that facilitates object localization, which is evident in instance segmentation[11]. Although FPN adopts a top-to-bottom approach for feature fusion to contain more layers of semantic information, there are too many feature extraction layers from the original map to the feature map, causing loss of location information containing in the low layer. Thus, it cannot solve the tooth position recognition problem well. In order to enable the model to focus on the location information more effectively, this paper introduces the Bottom-Up Path Enhancement Module (BUP) again after the FPN. The structure of the improved mask RCNN is shown in Fig. 3. The path enhancement structure of BUP shown in Fig.3 (c) firstly performs a 2*2 convolution operation on each feature map Ni to reduce the size, and then adds it to each element of the feature map Pi+1 via lateral connection to generate the fused feature map Ni+1 until N5 is generated. By

using BUP the lower location information reaches the feature map faster, making the output feature map have both high semantic and location information.

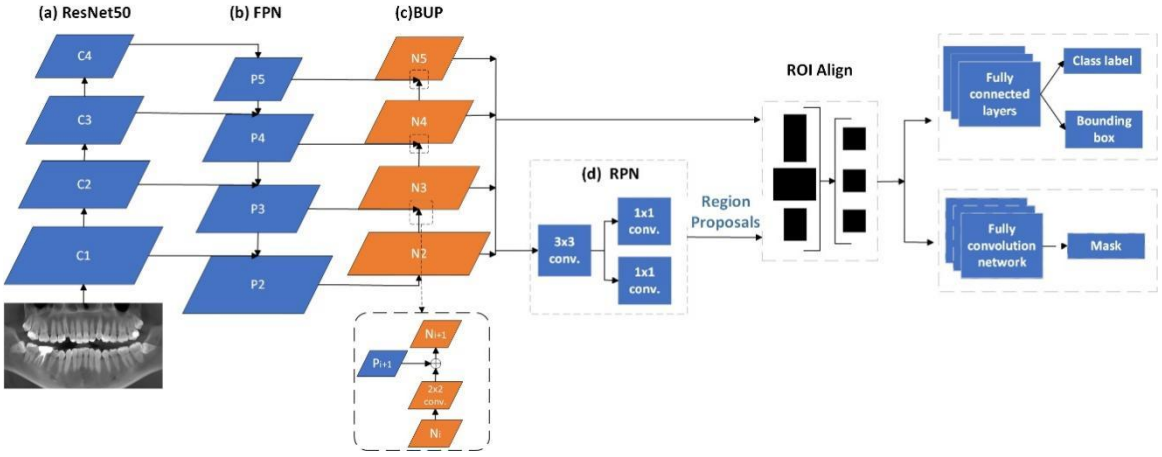

Figure 3 The structure of the Improved Mask RCNN

3.2 Loss Function with gradient balancing

The total loss function of the Mask RCNN consists of the RPN classification loss, the RPN target frame offset loss, the head classification loss, the head target frame offset loss, and the head pixel segmentation mask loss. Commonly, the head pixel segmentation mask loss is the average binary cross entropy loss[10], the head classification loss and RPN classification loss are the cross entropy loss function[12], and the RPN target box offset loss and head target box offset loss are the Huber function[12]. The Huber function introduces the hyperparameter to reduce the impact of outliers on the gradient, making the model convergence more stable. However, it does not effectively utilize the gradient information of the low-difference samples, resulting in a low convergence rate in training. In view of this, this study balances the gradient of the Huber loss function, and designs an improved Huber loss function as below.

$$loss(x,y)=\begin{cases}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{0.5}{b}(b|y_i-x_i|+1)\ln(b|y_i-x_i|+1)-0.5|y_i-x_i|\right], & if\ |y_i-x_i|<1\\ \frac{1}{n}\sum_{i=1}^{n}\left[1.5|y_i-x_i|+\frac{1.5}{b}-0.5\right], & otherwise\end{cases} \quad (1)$$

where $b=e^3-1$ .x_i and y_i are the predicted value and the true value respectively. The gradient of equation (1) is:

$$\frac{\partial loss}{\partial|y_i-x_i|}=\begin{cases}0.5\ln(b|y_i-x_i|+1), & if\ |y_i-x_i|<1\\ 1.5, & otherwise\end{cases} \quad (2)$$

From formula (2), we can see that when the difference between the true value and the predicted value is less than 1, 0.5 controls the improvement of the gradient. A suitable value will improve the gradient of the key sample without affecting other values; On the contrary, 1.5 is used to adjust the upper bound of the regression error, which can make different tasks more balanced. Through the two parameters of 0.5 and 1.5, the model can be trained more balanced to achieve better learning results.

4. Experimental results and Analysis

4.1 Training settings and evaluation indicators

The entire process of model training and testing is implemented on a server running Linux, with a memory size of 24G, a CPU of Intel(R) Xeon(R) Gold 6253CL CPU @ 3.10GHz, and a GPU of NVIDIA GeForce RTX 4090. The parameters for model training are: base learning rate is e-2, batch size is 8, number of classes is 28, epoch is 100, and the maximum iteration is 1200.

This study uses the mean average precision (mAP) of each type of tooth to evaluate the segmentation and recognition effect. The calculation formula is:

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \quad (3)$$

where k represents the total number of categories, and AP_i is the detection accuracy of the ith category, and

$$AP_i = \int_0^1 P(\mathrm{R})\,dR \quad (4)$$

$$\text{where } P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \quad (5)$$

TP indicates the number of samples predicted as positive and actually positive; FP indicates the number of samples predicted as negative and actually positive; TN indicates the number of negative samples predicted as negative; and FN indicates the number of samples predicted as positive but actually negative. This paper uses mAP with an Intersection Over Union(IOU) threshold of 0.5 and 0.5:0.95 as the test indicator. The higher the mAP value, the better the performance of the model.

At the same time, in view of the demand for dental implants, the accuracy of missing tooth detection is used as another evaluation indicator, it is:

$$Am = \frac{\mathrm{Im}}{Rm} \quad (6)$$

Among them, Im is the number of missing teeth detected, Rm is the actual number of missing teeth, and Am is the missing teeth detection accuracy.

4.2 Segmentation and recognition results using the improved model

After 100 epochs of training, the average location accuracy of 28 teeth (IOU=0.5) is 96.69%. The recognition accuracy mAP(0.5) of teeth in different tooth positions are listed in Table 2. It see that except for the teeth numbered 37, 41, and 47, the detection accuracies of the remaining teeth are above 90%. The main reason for the low recognition accuracy of the teeth numbered 37, 41, and 47 is that each patient has large individual differences and the scale imbalance of the data set is significant. For example, there is a significant difference in the morphology of teeth 41 and 42, and the position deviation is large. The second molar and the third molar have very similar tooth morphology and are located inside the dental mouth. There will be a relative deviation in their actual positions, which makes these two teeth be easily confused, resulting in the lowest prediction result.

Due to space limitations, only the segmentation results of the most complex panoramic radiographs containing intact teeth, implants, missing areas, and restorations are presented in Figure 4. It can be seen that the numbers of the teeth in the radiograph are all recognized with the corresponding confidence level (upper left of the box) and contour edges. And the intact teeth, implants, missing areas, and restorations (24, 25, and 46) are well avoided by the improved model.

Fig. 4 Example segmentation results for panoramic dental radiographs

Table 2 Results of tooth recognition at different tooth positions

Tooth position number	mAP%(0.5)	Tooth position number	mAP%(0.5)	Tooth position number	mAP%(0.5)	Tooth position number	mAP%(0.5)
11	97.5%	21	100%	31	90.47%	41	88.88%
12	98.5%	22	100%	32	95%	42	90.73%
13	100%	23	97.56%	33	100%	43	97.47%
14	97.73%	24	100%	34	100%	44	100%
15	94.44%	25	95.24%	35	97.79%	45	94.87%
16	97.94%	26	95.24%	36	97.33%	46	97.16%
17	90.47%	27	90.33%	37	85.71%	47	83.33%

Finally, the overall accuracy of missing teeth detection in dental panoramic radiographs is calculated, and the total number of missing teeth in the selected panoramic radiographs is 97. The total number of missing teeth identified in the test is 91, and the total recognition rate reaches 93.81%, which shows that the model has good recognition ability for missing teeth.

4.3 Model comparison analysis

In order to verify the effectiveness of the improved path aggregation network and balanced loss function (LF), ablation experiments are carried out based on the original Mask RCNN as the basic architecture. The experimental results are shown in Table 3. In Table 4, ‘O’ means that the modified module is not added to the original Mask RCNN model, and ‘P’ means that the improvement is used.

Table 3 Comparison of ablation test results

Number	Network structure		Average recognition accuracy	
	PA	Balanced LF	mAP(0.5)	mAP(0.5:0.95)
1	O	O	95.8%	73.5%
2	O	P	96.8%	73.8%
3	P	O	96%	74.1%
4	P	P	96.9%	74.2%

In Table 3, the results of Experiment 4 are the best, with 1.1% improvement in recognition accuracy compared to the original model Experiment 1. It can more accurately segment teeth and classify them according to tooth position numbers. The mAP(0.5) of Experiment 2 is 1% higher than that of Experiment 1, indicating that the loss function balance processing helps to improve the gradient information of the training process and achieves better training results. The mAP(0.5:0.95) of Experiment 3 is 0.6% higher than that of Experiment 1, indicating that the added feature level enhancement module is conducive to enriching feature information and improving position accuracy. The experiment fully proves that Mask RCNN has the ability to real tooth segmentation and teeth classification at once, achieving a tooth detection accuracy of more than 96%, and can well complete the tooth recognition task.

5. Conclusion

In this study, the instance segmentation model Mask RCNN is used to recognize and segment teeth. The model achieved good results during both the training process and the testing process. Regardless of the presence of teeth, missing teeth, implants and restorations in the panoramic dental radiographs, the improved algorithm in this paper achieves a recognition accuracy of 96.69%. Meanwhile, this model can be used to predict missing teeth in panoramic dental radiographs with an accuracy of 93.81%, which provides an objective method for predicting missing teeth. In this paper, the recognition and segmentation of teeth in panoramic dental radiographs are realized, and further research can be carried out subsequently in the recognition of missing teeth morphology.

References

[1] Turkyilmaz I ,Feldman Z D ,Suer T B .Imaging techniques in dental implant planning: Understanding the paradigm shift from periapical radiograph to cone beam computed tomography (CBCT)[J].Primary Dental Journal,2024,13(4):50-52.

[2] Can Z ,Isik S ,Anagun Y .CVApool: using null-space of CNN weights for the tooth disease classification[J].Neural Computing and Applications,2024,36(26):16567-16579.

[3] Divya, V. K., et al. (2016). Appending Active Contour Model on Digital Panoramic Dental Radiographs Images for Segmentation of Maxillofacial Region. IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, MALAYSIA.

[4] Bayrakdar, I. S., et al. (2022). "A U-Net Approach to Apical Lesion Segmentation on Panoramic Radiographs." Biomed Research International 2022.

[5] Nishitani, Y., et al. (2021)."Segmentation of teeth in panoramic dental Radiographs images using U-Net with a loss function weighted on the tooth edge." Radiological Physics and Technology 14(1): 64-69.

[6] Almalki, Y. E., et al. (2022). "Deep Learning Models for Classification of Dental Diseases Using Orthopantomography Radiographs OPG Images." Sensors 22(19).

[7] Haghanifar, A., et al. (2023). "PaXNet: Tooth segmentation and dental caries detection in panoramic Radiographs using ensemble transfer learning and capsule classifier." Multimedia Tools and Applications 82(18): 27659-27679.

[8] Mima, Y., et al. (2022). "Tooth detection for each tooth type by application of faster R-CNNs to divided analysis areas of dental panoramic Radiographs images." Radiological Physics and Technology 15(2): 170-176.

[9] Brahmi, W. and I. Jdey (2023). "Automatic tooth instance segmentation and identification from panoramic Radiographs images using deep CNN." Multimedia Tools and Application..

[10] He, K., et al. (2017). Mask RCNN. 16th IEEE International Conference on Computer Vision (ICCV), Venice, ITALY.

[11] Liu, S., et al. (2018). Path Aggregation Network for Instance Segmentation. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT.。

[12] Girshick, R. and Ieee (2015). Fast R-CNN. IEEE International Conference on Computer Vision, Santiago, CHI.