Automatic Tooth Segmentation and Classification in Dental Panoramic X-ray Images

Shuxu Zhao

Lanzhou Jiaotong University

QING LUO (✉ 1242578958@qq.com)

LZJTU: Lanzhou Jiaotong University <https://orcid.org/0000-0003-2711-6145>

Changrong Liu

Lanzhou Jiaotong University

Technical advance

Keywords: Dental Panoramic X-ray Images, Mask R-CNN, Tooth Segmentation, Tooth Classification, Deep Learning

Posted Date: October 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-89894/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

TECHNICAL ADVANCE ARTICLES

Automatic Tooth Segmentation and Classification in Dental Panoramic X-ray Images

Shuxu Zhao, Qing Luo* and Changrong Liu

*Correspondence:

0618562@stu.lzjtu.edu.cn

School of Electronics and
Information Engineering, Lanzhou
Jiaotong University, Anning West
Road, Lanzhou, China
Full list of author information is
available at the end of the article

Abstract

Background: The information of tooth shape, type and tooth position plays an important role in the understanding of pathological features in dental X-ray films. It is of great significance to realize the accurate tooth segmentation and tooth classification of dental panoramic X-ray images for the construction of an intelligent dental diagnosis system. At present, the segmentation results of teeth are relatively rough, and most methods realize tooth recognition and segmentation as independent tasks, ignoring the parameter sharing between the two tasks. Therefore, an instance segmentation method which can realize tooth recognition and tooth segmentation at the same time is proposed.

Methods: In model designing, the Mask R-CNN, an instance segmentation model, is adopted, which includes classification branches and segmentation branches. The classification branch can be used to complete the tooth recognition task and the segmentation branch to complete the tooth segmentation task. On this basis, the U-Net architecture is integrated to modify the segmentation branch to improve the segmentation effect. In data engineering, two classification schemes are designed, one according to the function of teeth, the other according to the position of teeth.

Results: Based on the data of 400 panoramic X-ray films of teeth, we combined migration learning to conduct experiments on the TensorFlow deep learning framework. The experimental results show that compared with other methods, the classification and segmentation of teeth can be realized simultaneously in this paper, with an accuracy of more than 90%. Compared with the original model, the improved Mask R-CNN proposed in this paper improves the segmentation recall rate by 10%. In the proposed classification scheme, the accuracy of classification based on tooth function is 3% higher than that based on tooth position.

Conclusions: The model proposed in this paper combines the two tasks of classification and segmentation, avoids the repetitive training of the model, and improves the segmentation precision with the improved segmentation branch. Compared with the recall rate traditional methods of tooth function classification, the proposed method based on tooth function has better classification effect.

Keywords: Dental Panoramic X-ray Images; Mask R-CNN; Tooth Segmentation; Tooth Classification; Deep Learning

Background

The development of economy has promoted the improvement of people's living standard and the change of people's concept, which makes them pay more attention to the problem of oral cavity and tooth disease. In China, there are only 167,300

dental surgeons in the whole country as of 2018, with only over 100 dentists per million people. However, the number of dentists per million people are 500 to 1000 in the developed countries, which indicates that there is actually a shortage of dentists in China[1, 2]. In addition, in developed countries such as Europe and America, although there are more dentists than China and private dental clinics are common, treatment costs are quite high. The outbreak of COVID-19 in 2019 has exposed the shortage of global medical resources. Faced with these problems and the needs of actual diagnosis and treatment, considering the rapid development of current computer image processing technology, it is gradually possible to build an intelligent diagnostic system using computer graphics image processing technology[3].

In the process of diagnosis of oral diseases, dental X-ray film is an important auxiliary diagnostic tool. Experts rely on it to display information such as the structure and shape of tooth bones to screen embedded teeth, tooth loss, bone abnormalities, cyst, tumor, infection, fracture and other problems[4]. But sometimes just relying on the naked eye of an expert can lead to a difference in diagnosis, which can lead to treatment errors. Realizing automated analysis of dental X-ray images can improve the accuracy and convenience of medical diagnosis and treatment, which can quickly process large-scale X-ray image data and reduce mechanical repetitive work, thereby improving the efficiency of medical staff and patient satisfaction, reducing medical costs, and easing the shortage of dentists and medical resources.

In the current research on the automatic analysis of dental X-ray images, its processing is mainly reflected in the segmentation[5, 6] and classification[7, 8, 9] of teeth, lesions and other targets. For example, Hasan et al.[10] used GVF Snake to automatically segment jaw bones from panoramic dental X-ray images. Rana et al.[11] used a convolutional neural network to segment teeth. Choi et al.[12] used the method of combining variational method and convolutional neural network to detect periodontal damaged teeth. Patil et al.[13] proposed a decayed tooth detection method using PCA and neural network for dental X-ray image analysis. With the development of deep learning, more and more scholars apply it to image processing. In the study of tooth classification and segmentation, Koch et al.[14] proposed the use of U-Net network for edge segmentation of tooth pictures. However, each tooth is independent and the edges are connected but discontinuous in the segmentation results, which is impossible to extract a single tooth in the picture. Jader et al.[15] proposed the use of Mask R-CNN for segmentation of teeth. It discussed in detail the detection and segmentation effect of missing teeth in dental X-rays, but it classified all teeth into one category and ignored different teeth positions. The semantic difference between teeth (such as incisors and molars) cannot be distinguished.

In the actual diagnosis process, numbering teeth is a necessary step in the diagnosis process of dental diseases. The stomatologist tend to quickly understand the dental conditions according to the records of the dental parts in the medical history, which is also convenient for the doctor to make medical records for the subsequent work. Chen et al.[16] first proposed using Faster R-CNN to detect and count teeth in dental images, providing a new direction for dental panoramic X-ray analysis. However, compared with the Mask R-CNN[15], which can complete both target detection and semantic segmentation tasks at the same time, the detection precision of Faster R-CNN is not high, and it can only complete the task of target detection, which consumes a lot of computing resources.

Considering these deep learning methods for tooth segmentation and tooth classification are completed as two separate tasks, they require a lot of time and computing power in the practical application process, which greatly reduces their practicability. In the learning process of neural network, there are a large number of characteristic parameters for tooth segmentation and recognition that can be shared. The integration of these two tasks and the reuse of resources can reduce the computation and workload of the overall task. Therefore, this paper aims at the field of image segmentation and target detection, using Mask R-CNN to complete the tooth segmentation and tooth classification of panoramic dental X-ray film at the same time. By comparing two experimental schemes and modifying the mask branch, it fully proves the effectiveness of Mask R-CNN in both tooth segmentation and tooth classification.

Methods

In this work, the improved Mask R-CNN was used to segment and classify the teeth in X-ray images, which combines the two classic computer vision tasks of semantic segmentation and target detection, and each detected object is classified, positioned and segmented. Each pixel of the target is divided into known categories by semantic segmentation, which can accomplish the task of tooth segmentation. Using target detection to locate and classify a single target can complete the task of tooth classification. In addition, two kinds of tooth classification schemes were designed and experimented respectively.

Mask R-CNN

As an instance segmentation framework, the Mask R-CNN is mainly composed of backbone network, RPN, ROIAlign module and three task branches. The network structure is shown in Figure 1. The three task branches are target classification, target bounding box regression and target object segmentation[17]. Multi-task learning method is adopted to realize the learning of these three branches. The classification and regression parts were completed by the parallel processing method of Faster R-CNN[18], which was used to realize the classification of different tooth positions and generate the bounding box of each tooth instance. For the segmentation of the target object, FCN[19] independently generates a binary mask for each class, that is, generates a mask for the teeth in the panorama. In the feature extraction part of the backbone network, we use the ResNet-101-FPN with the best performance in the Mask R-CNN[17], and the resulting image feature map will be input into the RPN[20] to obtain the ROI. Before running the ROI classification and bounding box regressor, the ROIAlign[21] algorithm is used to properly align the input feature map with a fixed size (such as 7×7), operations and uses the pixels-level sigmoid activation function to obtain the output of $K \times m \times m$ dimension, where K is the number of categories of the detection target and m is the size of the feature map. Since the mask branch selects the output mask according to the category label predicted by the classification branch, it allows the network to generate a mask for each category, and there is no competition among different categories, which makes the classification and mask generation split apart, improving the effect of instance segmentation. The network structure of the Mask branch is shown in Figure 2.

Proposed Mask branch

In this part, U-Net architecture[22] is added to the Mask branch to improve the effect of fine-grained segmentation on the target. Its network structure is illustrated in Figure 3. The contracting path consists of the repeated application of a 3×3 convolutions, each followed by a BN layer, a rectified linear unit (ReLU) and a 3×3 convolution with stride 1 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature followed by a 3×3 deconvolution that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and a 3×3 convolution, each followed by a ReLU. Then input into a deconvolution layer with stride 2 to expand the size of the feature map to 28×28 . At the final layer a sigmoid function is used to generate masks.

Materials and Data

The following experiment uses python and the TensorFlow to implement code in Pycharm. The experimental environment of this paper is shown in Table 1.

Table 1 The experimental environment.

Item	Content
OS	Windows 10
CPU	Intel i7-9700K
GPU	NVIDIA GTX 1080Ti
Python	Python3.6.5
TensorFlow	Tensorflow1.8.0

The experimental data used a dataset of 400 dental panoramic X-rays, in which each image size is 1024×2161 . The original image is shown in Figure 4.

Data on the label marked by dentists use VIA tools in the CSV file, after data cleaning we convert the format to Mask R-CNN model requires the json data format label, the json file contains the tooth profile coordinate, FDI tooth number information, which is shown in the Table 2.

Table 2 Label information.

Field	Description
Filename	Image name
Region_id	Region ID, where the region refers to a single tooth instance
Region_shape_attributes	Contour coordinates of each area
Region_attributes	FDI tooth number of each area

Design

In the current tooth segmentation method[3], all teeth are classified into one category, which means that different teeth learn from each other, ignoring their different features. For example, incisors and fangs have wedge-shaped crowns, while premolars and molars have cube crowns. Maxillary molars have three teeth, mandibular molars have two, and other teeth are mostly single. In order to distinguish different teeth, we proposed two experimental schemes according to the function of teeth and the method of dental position recording while the experimental framework remained unchanged.

Classification by the tooth function: According to the morphological and functional characteristic, teeth are divided into incisors, fangs, premolars and molars[23].

Incisors are located in the front of the mouth. There are 8 incisors left, right, upper and lower. Fangs, commonly known as canine teeth, are located in the corner of the mouth, left, right, up and down a total of 4. Premolars are located behind the fangs and before the molars. There are 8 teeth in total, including left, right, top and bottom. Molars are located behind the premolars. There are 12 molars on the left, right, top and bottom. The Figure 5 shows the distribution of teeth.

Classification by dental recording: Dental recording is a method used in dentistry to number each human tooth. The tooth number has its unified standard, and the text uses FDI tooth notation[24], where each tooth is represented by two Arabic numerals, which was proposed by the International Dental Union in 1970 and is universal. The first represents the quadrant in which the teeth are located: the patient's upper right, upper left, lower left, and lower right are 1, 2, 3 and 4, respectively. The second tooth position is 1 to 8 from incisor to molars. The Figure 6 shows the representation of FDI tooth position.

In experiment 1, we divided the teeth into four categories according to their functions, and the teeth with different tooth positions in the same category were only input into the network as different instances. Therefore, five categories should be defined for the network, namely incisors, fangs, premolars, molars and backgrounds. The label diagram is as Figure 7.

In experiment 2, teeth with different dental positions were input into the network as separate instances for training according to the representation of tooth positions. Considering that adults normally have 32 teeth, we defined 33 classes for the network, in which the teeth of each tooth position are treated as a separate class, that is, 32 tooth position classes plus a background class, with only one instance of each class. The label diagram is as Figure 8.

Setting

The two schemes were experimented respectively, and the super parameters were set as follows: learning rate was 0.001, batch size was 100, and epochs were 35. Eighty percent of the data in the experiment was used for the training set, and the remaining 20 percent was used for the test. Considering that tooth segmentation is a pixel-level classification task, it requires more data than tooth position recognition. Due to the small amount of data, we cannot train the whole deep learning network from scratch. Therefore, we consider to adopt the transfer learning method to improve the training effect of the network. Transfer learning[25] refers to a new task that improves learning by transferring knowledge from related tasks already learned, that is, the model developed for task A is taken as the initial point and reused in the process of developing the model for task B. Because the training of neural network is more and more time-consuming, and the data set size required by neural network can not be satisfied in all cases. Therefore, for small sample data, adopting the method of transfer learning and using the trained neural network for training can save time and effort. In the model we used, the trunk network feature extraction part imported the pre-training weight on the MSCOCO data set with the migration learning method, and then used the existing data to fine-tune the network head.

Results

To evaluate the experimental results, we adopted precision and recall at an IOU threshold of 0.50, where the precision value represents how many of the positively predicted samples are true positive samples, and the recall rate represents how many positive examples in the sample are correct prediction, the calculation formula is as :

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

Where, TP, FN and FP represent true positive, false negative and false positive, respectively. True positive means positive in prediction, positive in fact; false negative means that the prediction is negative and the fact is positive. A false positive is when the prediction is positive and the reality is negative. In addition, there are true negative indicators, is to predict negative, actually negative. For true and false, the prediction is true if it is consistent with reality, and false if it is inconsistent.

From the segmentation results, the results of the two sets of experiments and other tooth segmentation methods in this paper are shown in Table 3. Among them, Mask R-CNN-1 and Mask R-CNN-2 are the segmentation results of experiment 1 and experiment 2 respectively.

Table 3 Segmentation results.

Methods	Precision	Recall
U-Net	0.92	0.93
Mask R-CNN	0.98	0.84
Mask R-CNN-1	0.967	0.948
Mask R-CNN-2	0.958	0.806

It can be seen that the original Mask R-CNN[15] has the best performance in the precision and Mask R-CNN-1 has the best performance in the recall. Compared with U-Net[14], the precision of Mask R-CNN-1 and Mask R-CNN-2 were improved by more than 3%, and the recall rate of Mask R-CNN-1 was also improved by 1%. Compared with the original model, the improved Mask R-CNN-1 proposed in this paper improves the segmentation recall rate by 10%. However, in the comparison between Mask R-CNN-1 and Mask R-CNN-2, the precision rate and recall rate of Mask R-CNN-2 are much lower than that of Mask R-CNN-1, and their segmentation renderings are shown in Figure 9.

In experiment 1, each tooth has a mask, but the mask coverage is incomplete, such as 16 and 17 teeth shown in Figure 9. (a). In experiment 2, there were more teeth without forming masks, such as 31, 32, 33 and 34 teeth. The reason for this is that the teeth are divided into four categories by experiment 1, and different instances of the same category contribute losses to each other, while the teeth are divided into 32 categories by experiment 2, and there is only one instance under each category, so they cannot contribute losses to each other. Therefore, the segmentation effect of experiment 2 is not as good as that of experiment 1.

From the classification results, the results of the two groups of experiments and the results of the Faster R-CNN[10] method are shown in Table 4, where the classification results of experiment 1 and experiment 2 are shown in Mask R-CNN-1 and Mask R-CNN-2 respectively.

Table 4 Classification results.

Methods	Precision	Recall
Faster R-CNN	0.92	0.91
Mask R-CNN-1	0.968	0.939
Mask R-CNN-2	0.943	0.844

As shown in the Table 4, Mask R-CNN-1 has the best performance and can accurately recognize the teeth and classify them according to their functions. However, as it only does four classifications and cannot achieve the effect of tooth position recognition. Faster R-CNN and Mask R-CNN-2 both do 32 classifications, which can effectively identify different tooth positions. But the precision of Mask R-CNN-2 is improved 2% compared with Faster R-CNN[10]. In addition, in the comparison between Mask R-CNN-1 and Mask R-CNN-2, the precision and recall rate of Mask R-CNN-2 are much lower than that of Mask R-CNN-1. The Figure 10 shows Mask R-CNN -1 and Mask R-CNN-2 target detection results, the text above the target detection box is the target classification label and its score value.

It can be seen that the recall rate in experiment 1 is high, each tooth is detected and correctly classified, and the recognition score can reach more than 0.98. There are undetected targets in experiment 2. As shown in the Figure 10. (b), the teeth at 14, 24, 27, and 44 have no bounding box; and the detected target score is lower than experiment 1, the average around 0.95.

In general, the functions of each method are compared in Table 5. Compared with other methods, which can only complete one task, our method can realize segmentation and classification simultaneously.

Table 5 Function comparison.

Methods	Precision	Recall
U-Net	✓	-
Mask R-CNN	✓	-
Faster R-CNN	-	✓
Mask R-CNN-1	✓	✓
Mask R-CNN-2	✓	✓

Figure 11 is the visual results of the two experiments in this paper. It can be seen from Figure 11. (e) that the segmentation of teeth is relatively complete and each instance does not affect each other, but the fangs are often not detected. In Figure 11. (f), the result of tooth position recognition has a good identification accuracy, but there are many instances that have not been identified.

Combining the two experimental schemes, experiment 1 performs better in tooth segmentation task than experiment 2, but its degree of tooth classification is not as fine as experiment 2, so it cannot achieve the effect of tooth position recognition. Experiment 2 can better complete the tooth position recognition task, but it also results in poor segmentation results because the segmentation loss values between each tooth position cannot be shared. However, the two sets of experiments fully

proved that Mask R-CNN has the ability to achieve tooth segmentation and classification at the same time guaranteeing the classification effect of 90% precision and 95% precision of tooth segmentation.

Discussion

At present, there are many researches on tooth apical slices, but it is obviously insufficient for analyzing all patients' dental conditions and doing data integration. Dental X-ray panoramic image is widely used and has a large amount of information, which can provide reference for a variety of oral diseases such as orthodontics. The method in this paper can effectively detect, segment and number all teeth on the tooth panorama, and can classify different kinds of teeth. In the past research[26], we carried out the texture analysis of the gray value statistical method for a single tooth, which can reflect the normality of the tooth to a certain extent. Combined with the method in this article, it can provide a basis for the pathological analysis of a single tooth, and it has an important role in the diagnosis and analysis of the dental caries, periodontal disease, and apical periodontitis of each tooth. It also helps to systematically manage basic dental information and disease information.

In experiment 2, the teeth were divided into 32 categories, resulting in insufficient training data and unsatisfactory segmentation effect. Moreover, since the left and right relative positions of teeth are in a mirror relationship, such as 14 and 24, 37 and 47, and the convolutional neural network has the characteristic of translation invariance, which increases the difficulty of classification. In experiment 1, the teeth were divided into 4 categories according to their morphological and functional characteristics, which to some extent improved the problem of insufficient training data in experiment 2. However, compared with Mask R-CNN[15], there was still much room for improvement. For example: use its mirror relationship to change the category to 16 or 8 categories, and then obtain the final classification effect according to the quadrant; or to improve the network structure of the Mask branch, so that different types of teeth can learn from each other, which can help improve segmentation accuracy.

From the perspective of practicality, as the images used in the experiment are all from clinical medicine and the label data are marked by professional dentists, the annotation cost is high. Therefore the data set is only 400, which is a small sample data set for the deep learning task, which is also the main reason for the rough segmentation results. The experiment in this paper combines the idea of transfer learning, which can improve the interpretability of the algorithm for small sample data set. However, the pre-training model used in the experiment is based on the MSCOCO data set, which has a low similarity with the dental X-ray data set in the paper. In future studies, other X-ray data sets can be considered, which can reduce the difficulty of fine-tuning and obtain better experimental results.

Although our experimental results prove the universality of this method on complete teeth, dental X-rays in the actual diagnosis and treatment often have abnormal conditions such as missing teeth, residual roots and implants, which is a challenge for our model. Other post-treatment methods, such as applying the rules of tooth alignment, can be considered for targeted analysis, which is also the work we need to do in the future.

Conclusion

Based on the X-ray panorama of teeth, the Mask R-CNN model was used to segment and classify teeth, so as to describe the shape and position of each tooth in the X-ray panorama more accurately. To solve the problem that different teeth can not be classified in previous studies, we conducted an experimental analysis on the two classification methods of teeth from functional and FDI tooth position representation. We used the modified Mask R-CNN to perform both tooth segmentation and tooth classification tasks. The results show that the method in this paper can achieve more than 90% accuracy in both tasks, but there is still some room for improvement in segmentation. This problem may be caused by the number of images in the data set and the X-ray image irradiation equipment. In the future studies, we will consider further improving the accuracy of the experiment and conduct intelligent analysis system research based on the experimental results.

Abbreviations

PCA: Principal Components Analysis
 RPN: Region Proposal Network
 ROI: Region of Interest Alignment
 FCN: Fully Convolutional Networks
 BN: Batch Normalization
 VIA: VGG Image Annotator
 CSV: Comma-Separated Values
 FDI: Federation Dentaire Internationale
 IOU: Intersection over Union

Declarations

Ethics approval and consent to participate
 Not applicable.

Consent for publish

All the authors agree to the terms of the BioMed Central Copyright and License Agreement.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

ZSX analyzed and explained the related literature, and proposed the improvement direction of Mask RCNN. LQ proposed a classification scheme using two different classification methods, carried out experimental verification of theoretical research, and was the main contributor to the manuscript writing. Liu Changrong analyzed the references. All the authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Wang, X.: Current situation, opportunities and challenges of stomatological medical industry in china. *Technology and Finance* **05**, 5–8 (2018)
2. Feng, X.: Residents of oral health, china - the fourth chinese oral health epidemiological survey report. In: *Compilation of Papers for the 18th Annual Meeting of The Chinese Dental Association on Oral Preventive Medicine in 2018*, p. 2 (2018). Chinese Stomatological Association
3. Geert, L., Kooi, Thijs, Beijndorf: A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 66–88 (2017)
4. Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: Overview, challenges and future (2017)

5. Gil, S., Luciano, O., Matheus, P.: Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, 0957417418302252 (2018)
6. Abdi, A., Kasaei, S., Mehdizadeh, M.: Automatic segmentation of mandible in panoramic x-ray. *Journal of Medical* **2**, 044003 (2015)
7. Vidar, L.J., Morten, G.: Automated dental identification with lowest cost path-based teeth and jaw separation. *Scandinavian Journal of Forensic Science* **22**, 44–56 (2016)
8. Keerthana, K.M., Rajeshwari, B., Keerthi, S., Menon, H.P.: Classification of tooth type from dental x-ray image using projection profile analysis. In: 2017 International Conference on Signal Processing and Communication (ICSPC), pp. 394–398 (2017)
9. Ainas, A., Hazem, E., Sameh, A.: Detection of caries in panoramic dental x-ray images using back-propagation neural network. *International Journal of Electronics Communication and Computer Engineering* **7**, 250–256 (2016)
10. Mosaddik, H.M., Waidah, I., Rozita, H., Atsuo, Y.: Automatic segmentation of jaw from panoramic dental x-ray images using gvf snakes. In: 2016 World Automation Congress (WAC), pp. 1–6 (2016)
11. Anantharaman, R., Anantharaman, V., Lee, Y.: Oro vision: Deep learning for classifying orofacial diseases. In: 2017 IEEE International Conference on Healthcare Informatics (ICHI), pp. 39–45 (2017)
12. Ainas, A., Hazem, E., Sameh, A.: Boosting proximal dental caries detection via combination of variational methods and convolutional neural network. *Journal of Signal Processing Systems for Signal Image & Video Technology* **90**, 87–97 (2018)
13. Shashikant, P., Vaishali, K., Archana, B.: Caries detection using multidimensional projection and neural network. *International Journal of Knowledge-Based and Intelligent Engineering Systems* **22**, 155–166 (2018)
14. Loring, K.T., Mathias, P., Christian, I., Sebastian, B.S.: Accurate segmentation of dental panoramic radiographs with u-nets. In: IEEE 16th International Symposium on Biomedical Imaging (ISBI), pp. 15–19 (2019)
15. Jader, G., Fontinele, J., Ruiz, M., Abdalla, K., Pithon, M.: Deep instance segmentation of teeth in panoramic x-ray images. In: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 400–407 (2018)
16. Chen, Zhang, K., Peijun, Lyu, Hong, Ludan, Lee, C.: A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Scientific reports* **9**, 1–11 (2019)
17. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **PP**, 1–1 (2017)
18. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **39**, 1137–1149 (2017)
19. Jonathan, L., Evan, S., Trevor, D.: Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 640–651 (2015)
20. Meshan, K., Attique, K.M., Fawad, A., Mamta, M., Chandra, S.S.: Gastrointestinal diseases segmentation and classification based on duo-deep architectures. *Pattern Recognition Letters* **131**, 193–204 (2019)
21. Wangzhe, D., Hongyao, S., Jianzhong, F., Ge, Z., Quan, H.: Approaches for improvement of the x-ray image defect detection of automobile casting aluminum parts based on deep learning. *NDT&E international* **107**, 102144–110214412 (2019)
22. Olaf, R., Philipp, F., Thomas, B.: U-net: Convolutional networks for biomedical image segmentation, 234–241 (2015)
23. Chen, S.: Introduction to Dentistry. Peking University Press, Peking (1992)
24. YongZhan, H., JuXiang, R.: Suggested to record the location and name of teeth uniformly. *Journal of Dental pulp periodontology* **22**, 488–488 (2012)
25. Zhuang, F., Luo, P., He, Q., Shi, Z.: Survey on transfer learning research. *Journal of software* **26**, 26–39 (2015)
26. Zhao, S., Liu, C., Luo, Q.: Dental data analysis based on dental x-ray panorama. In: ISICDM 2019: Proceedings of the Third International Symposium on Image Computing and Digital Medicine, pp. 133–137 (2019)

Figures

Figure 1 Network Structure of the Mask R-CNN.

Figure 2 The original Mask branch.

Figure 3 Improved Mask branch.

Figure 4 The original image.

Figure 5 Classification by the tooth function.

Figure 6 Classification by dental recording.

Figure 7 Labels of experiment 1.

Figure 8 Labels of experiment 2.

Figure 9 Segmentation results. (a) is the result of experiment 1, and (b) is the result of experiment 2.

Figure 10 Classification results. (a) is the result of experiment 1 that is to divide the teeth into four categories, and (b) is the result of experiment 2 that is to divide the teeth into 32 categories.

Figure 11 Results. Among them, (a) (b) is the original image, (c) (d) is the ground truth, (e) is the result of experiment 1, and (f) is the result of experiment 2.

Figures

Figure 1

Network Structure of the Mask R-CNN.

Figure 2

The original Mask branch.

Figure 3

Improved Mask branch.

Figure 4

The original image.

Figure 5

Classification by the tooth function.

Figure 6

Classification by dental recording.

Figure 7

Labels of experiment 1.

Figure 8

Labels of experiment 2.

Figure 9

Segmentation results. (a) is the result of experiment 1, and (b) is the result of experiment 2.

Figure 10

Classification results. (a) is the result of experiment 1 that is to divide the teeth into four categories, and (b) is the result of experiment 2 that is to divide the teeth into 32 categories.

Figure 11

Results. Among them, (a) (b) is the original image, (c) (d) is the ground truth, (e) is the result of experiment 1, and (f) is the result of experiment 2.