Detection and classification of 32 tooth types in dental panoramic radiographs using single CNN model and post-processing

Morishita, Takumi, Muramatsu, Chisako, Seino, Yuta, Takahashi, Ryo, Hayashi, Tatsuro, et al.

Takumi Morishita, Chisako Muramatsu, Yuta Seino, Ryo Takahashi, Tatsuro Hayashi, Wataru Nishiyama, Xiangrong Zhou, Takeshi Hara, Akitoshi Katsumata, Hiroshi Fujita, "Detection and classification of 32 tooth types in dental panoramic radiographs using single CNN model and post-processing," Proc. SPIE 12177, International Workshop on Advanced Imaging Technology (IWAIT) 2022, 121771T (30 April 2022); doi: 10.1117/12.2625961

Event: International Workshop on Advanced Imaging Technology 2022 (IWAIT 2022), 2022, Hong Kong, China

Detection and classification of 32 tooth types in dental panoramic radiographs using single CNN model and post-processing

Takumi Morishita[a], Chisako Muramatsu*[b], Yuta Seino[c], Ryo Takahashi[d], Tatsuro Hayashi[d], Wataru Nishiyama[e], Xiangrong Zhou[c], Takeshi Hara[c], Akitoshi Katsumata[e], Hiroshi Fujita[c]

[a]Department of Intelligence Science and Engineering, Graduate School of Natural Science and Technology, Gifu University, Gifu 501-1194, Japan

[b]Faculty of Data Science, Shiga University, Hikone, Shiga 522-8522, Japan

[c]Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, Gifu 501-1193, Japan

[d]Media Co., Ltd, Tokyo 113-0033, Japan

[e]Department of Oral Radiology, School of Dentistry, Asahi University, Mizuho, Gifu 501-0296, Japan

ABSTRACT

The purpose of this study is to analyze dental panoramic radiographs for completing dental files to contribute to the diagnosis by dentists. In this study, we recognized 32 tooth types and classified four tooth attributes (tooth, remaining root, pontic, and implant) using 925 dental panoramic radiographs. YOLOv4 and post-processing were used for the recognition of 32 tooth types. As a result, the tooth detection recall was 99.65%, the number of false positives was 0.10 per image, and the 32-type recognition recall was 98.55%. For the classification of the four tooth attributes, two methods were compared. In Method 1, image classification was performed using a clipped image based on the tooth detection result. In Method 2, the labels of tooth attributes were added to the labels of tooth types in object detection. By providing two labels for the same bounding box, we performed multi-label object detection. The accuracy of Method 1 was 0.995 and that of Method 2 was 0.990. Method 2 uses a simple and robust model yet has comparable accuracy as Method 1. In addition, Method 2 did not require additional CNN models. This suggested the usefulness of multi-label detection.

Keywords: Dental panoramic radiographs, Deep learning, Tooth recognition, Computer-assisted diagnosis, YOLOv4

1. INTRODUCTION

Medical images used in the field of dentistry include dental panoramic radiographs and dental periapical films, as well as dental CT images. In particular, dental panoramic radiographs are the most common images in the dental field, and they are routinely taken and interpreted. However, it is difficult for dentists to make an accurate diagnosis of all the teeth because the panoramic radiographs contain various information including treatment status, caries, and apex lesions. In addition to the teeth, there are many lesions such as periodontal disease and jawbone lesions that can be visualized in the images. Therefore, computer-aided diagnosis is expected to assist dentists in reading the images [1]. The purpose of this study is to extract information necessary for reading dental panoramic radiographs to contribute to dentists' diagnosis.

In this study, two problems were addressed: the first one is the recognition of 32 tooth types (including dental prosthesis), and the second is to classify the detected teeth into four tooth attributes (tooth, remaining root, pontic, and implant). In this study, the word "recognition" is used as a series of detection and classification.

Previous studies have been reported for the recognition of 32 tooth types in dental panoramic radiographs [2, 3]. We have proposed a method to improve the accuracy of tooth classification by extending the horizontal range of the clipped images [4]. These methods can detect and classify tooth types with high performance, but often exclude dental prostheses such as implants. In this study, the recognition targets are teeth and dental prostheses. In addition, we propose a method to perform recognition of 32 tooth types and classification of four tooth attributes simultaneously.

*chisako-muramatsu@biwako.shiga-u.ac.jp

International Workshop on Advanced Imaging Technology (IWAIT) 2022, edited by Masayuki Nakajima, Shogo Muramatsu, Jae-Gon Kim, Jing-Ming Guo, Qian Kemao, Proc. of SPIE Vol. 12177, 121771T

doi: 10.1117/12.2625961

2. METHODS

2.1 Databases

In this study, 925 dental panoramic radiographs obtained at ten facilities, including a university hospital and dental clinics, were used. The images were captured using various imaging machines, and the matrix size ranged from pixel resolutions of 783 × 380 to 3369 × 1350. Approximate outlines of the tooth and dental prosthesis were annotated by experts. Square bounding boxes that confined the outlines were used as labels of object detection in this study. The detection target was teeth and dental prosthesis.

For input to the detection model, the dentition area was automatically extracted from a panoramic radiograph and resized so that the longest side has 512 pixels. It was then zero-padded to form a 512×512 square image (Figure 1). Single shot multibox detector (SSD) [5] was used to extract the dentition area.

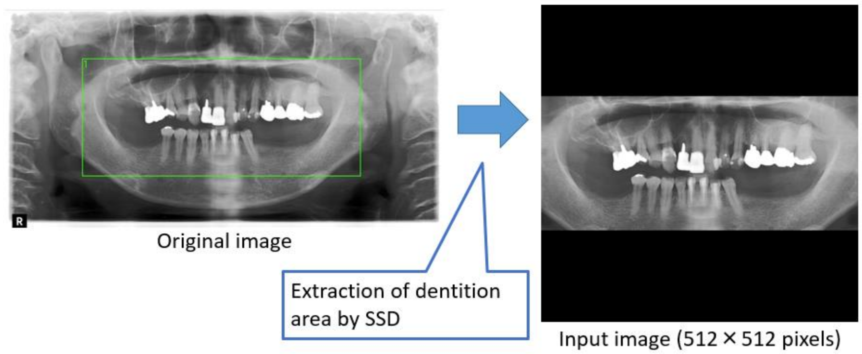

Figure 1. Input image

2.2 Recognition of 32 tooth types

The process flow of the recognition of 32 tooth types is shown in Figure 2. Padded regions were omitted in Figure 2 for presentation.

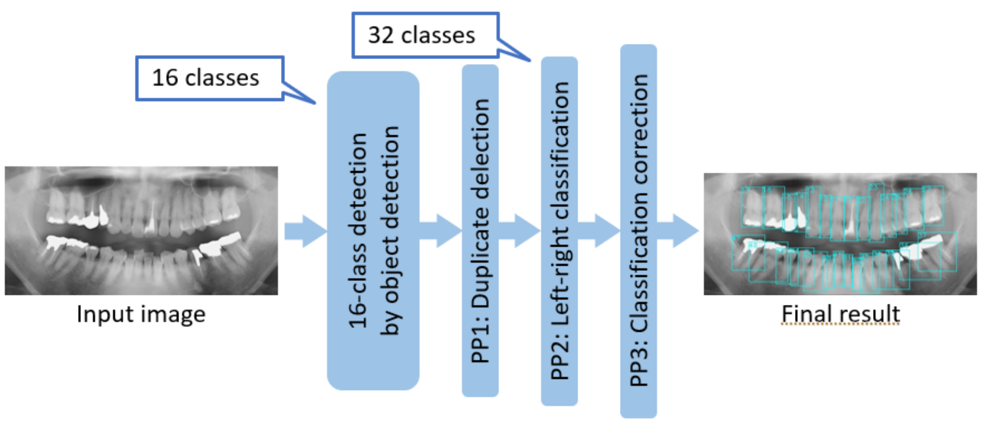

Figure 2. Flow of recognition of 32 tooth types

First, object detection was performed with 16 classes (i.e., the central incisor, lateral incisor, canine, first premolar, second premolar, first molar, second molar, and third molar, distinguished by the upper and lower jaws). We labeled 16 classes instead of 32 because there was no significant difference in shape between the right and left teeth of the same type, and we judged that it would be difficult to distinguish them using deep learning. In addition, it can increase the number of samples by approximately two-fold. In this study, we used SSD (the feature extraction part was changed from VGG16 [6] to ResNet-101 [7]) and YOLOv4 [8] as object detectors and compared their results.

For the 2D-bounding boxes detected by SSD and YOLOv4, we applied three post-processing steps in order: PP1 duplicate deletion, PP2 left-right classification, and PP3 classification correction. The output of PP3 was used as the final output for the recognition of 32 tooth types in this study.

In PP1, if the intersection over union (IoU) of the two detected boxes exceeds the threshold, the one having lower output confidence was deleted. The threshold was set to 0.3 when the classes of the two detected boxes were the same, and 0.5 when the classes were different. In PP2, for each jaw, the detection boxes were classified into left and right based on the central incisors according to the results of the 16-class detection. Using this process, the 16 classes were converted into 32 classes (32 tooth types). In PP3, when two detection boxes had the same class, the class of the box with the lower confidence was modified or the box was deleted. When modifying the class, we focused on the classes of the neighboring boxes to the left and right of the target box and modified the class so that it would be sequentially numbered with them. If the neighboring boxes are sequentially numbered, the target detected box is deleted.

2.3 Classification of four tooth attributes

In this study, we defined four tooth attributes: (1) tooth, (2) remaining root, (3) pontic, and (4) implant (Figure 3). (2) remaining root is a diseased tooth in which only the root remains without the crown, and teeth other than (2) were defined as (1) tooth. (3) pontic is a dental prosthesis without a root and is attached to the right and left teeth, and (4) implant is that embedded in the jawbone.

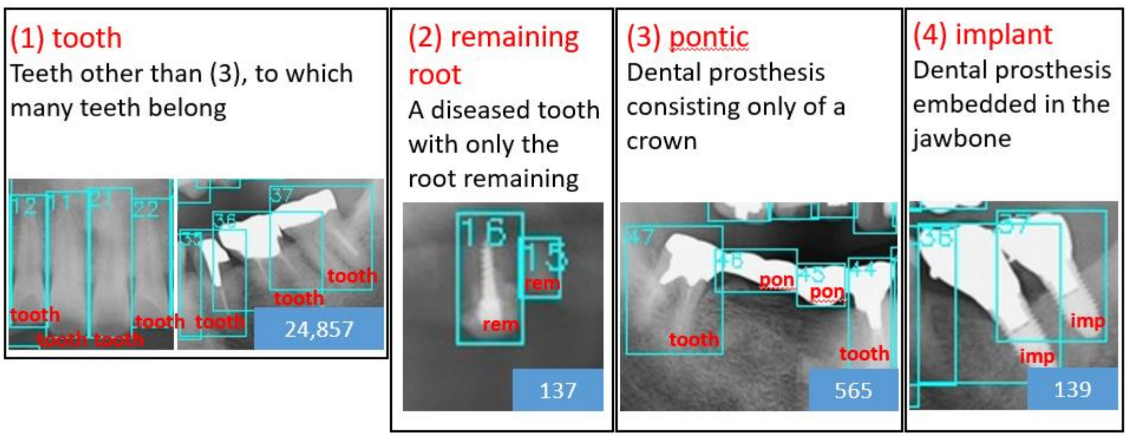

Figure 3. The four tooth attributes are shown. The red text is the attribute, and the blue text is the tooth type in the box. The number at the bottom right of each image indicates the number of samples for each attribute in 925 cases.

In this study, we compared two methods for classifying these tooth attributes (Figure 4). In Method 1, after recognizing 32 tooth types, image classification was performed using the clipped images. ResNet-101 was used as the CNN for image classification. In Method 2, we added the labels of tooth attributes to the labels of 16 tooth types in object detection. By providing two labels for the same bounding box, we performed multi-label object detection and simultaneously handled two types of problems: tooth type recognition and tooth attribute classification. For the (1) tooth, we did not provide attribute labels because the number is so large that it accounts for more than 96% of the total. Instead, we considered boxes with no attributes as the tooth attributes. Therefore, 19-class (16 tooth types + three tooth attributes) detection was performed in Method 2.

2.4 Evaluation methods

Of the 925 cases, 740 cases were used as training data and 185 as test data in a 5-fold cross-validation. For the detected box, if the IoU with the labeled box exceeded 0.5, it was considered a successful detection; otherwise, it was considered a false positive (FP). The following were used as indices for tooth recognition performance of the 32 tooth types.

- Perfect Detection (PD): 100% detection recall with no false positives in a case
- Perfect Recognition (PR): 100% recognition recall with no false positives in a case (perfect detection and classification)

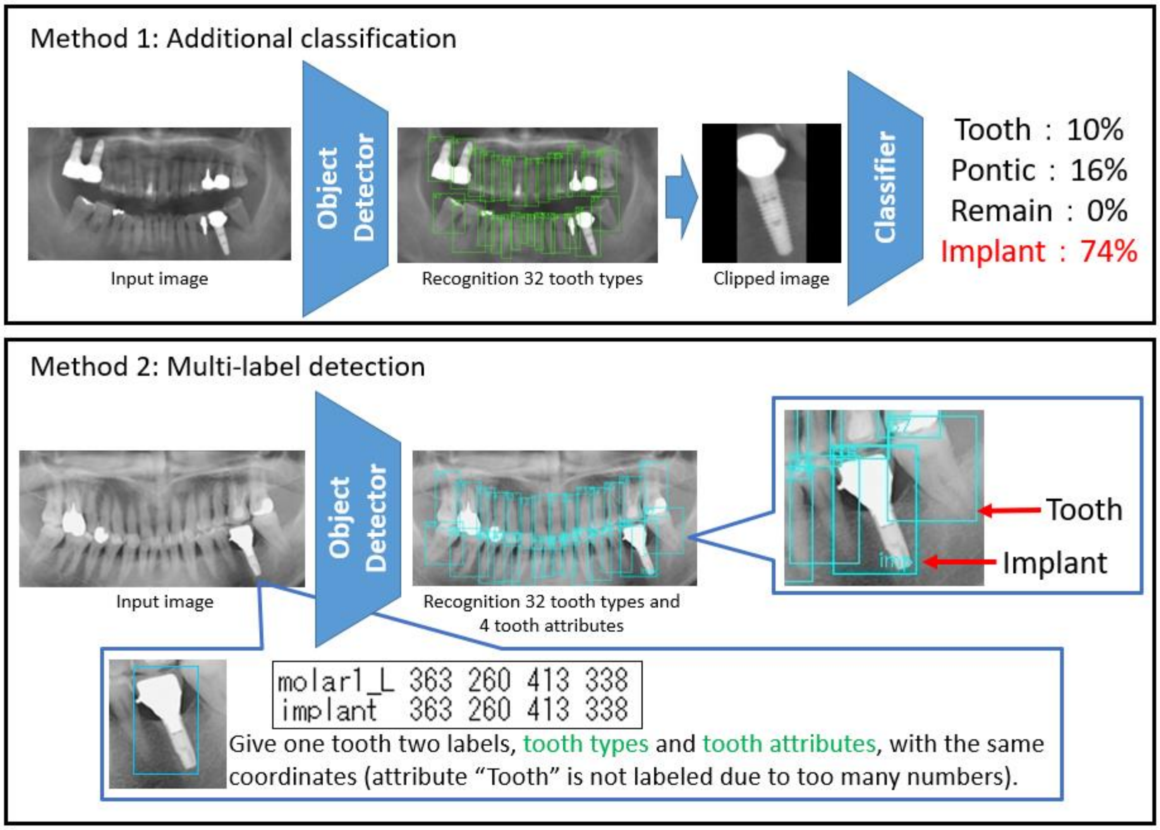

Figure 4. Two methods for classification of tooth attributes

3. RESULTS

3.1 Recognition of 32 tooth types

Table 1 shows the tooth recognition performance of 16-class detection and the effect of post-processing for SSD and YOLOv4. PP3 by YOLOv4 is the final output for the 32 tooth-type recognition of this study. All the numbers reported here are the average results of the 5-fold cross-validation.

Table 1. Recognition of tooth types

	Classes	Detection recall	Number of FP per image	Recognition recall	Classification rate	Number of cases of PD	Number of cases of PR
SSD	16	99.50	2.82	97.89	98.38		
SSD (after PP3)	32	99.33	0.12	97.79	98.45	757	629
YOLOv4	16	99.71	0.70	98.63	98.91		
YOLOv4 (after PP1)	16	99.65	0.11	98.58	98.93		
YOLOv4 (after PP2)	32	99.65	0.11	98.51	98.86	808	695
YOLOv4 (after PP3)	32	99.65	0.10	98.55	98.89	815	705
Tuzoff [3]	32	99.41	0.13	97.41	98.00		

3.2 Classification of four tooth attributes

The average recognition recall of the three attributes added by Method 2 was 64.45% for SSD and 93.58% for YOLOv4. Table 2 shows the confusion matrices of Method 1 and Method 2 by YOLOv4. The accuracy and other indicators are shown in Table 3.

Table 2. Confusion matrix ((1) tooth = T, (2) remaining root = R, (3) pontic = P, (4) implant = I)

Method 1	Prediction					Method 2 (YOLOv4)	Prediction				
Correct		T	R	P	I	Correct		T	R	P	I
	T	24827	12	8	10		T	24823	70	55	12
	R	62	75	0	0		R	33	104	0	0
	P	27	1	545	2		P	12	0	551	2
	I	20	0	0	119		I	7	0	0	132

Table 3. Classification of four tooth attributes

Method	Accuracy	Recall	Precision	F-measure
1	0.995	0.842	0.936	0.886
2 (YOLOv4)	0.990	0.920	0.856	0.887

To investigate the influence of adding attribute labels in Method 2, Table 4 shows the comparison of 32 tooth type recognition performance between 16-class detection and 19-class detection. The effect on the tooth type recognition is considered minimal.

Table 4. 32 tooth type recognition of 16-class and 19-class detection

	Classes for training	Detection recall	Number of FP per image	Recognition recall	Classification rate	Number of cases of PD	Number of cases of PR
SSD	16	99.33	0.12	97.79	98.45	757	629
	19	99.35	0.11	97.73	98.37	755	627
YOLOv4	16	99.65	0.10	98.55	98.89	815	705
	19	99.67	0.09	98.53	98.85	816	699

An example of PR and successful classification of all attributes is shown in Figure 5.

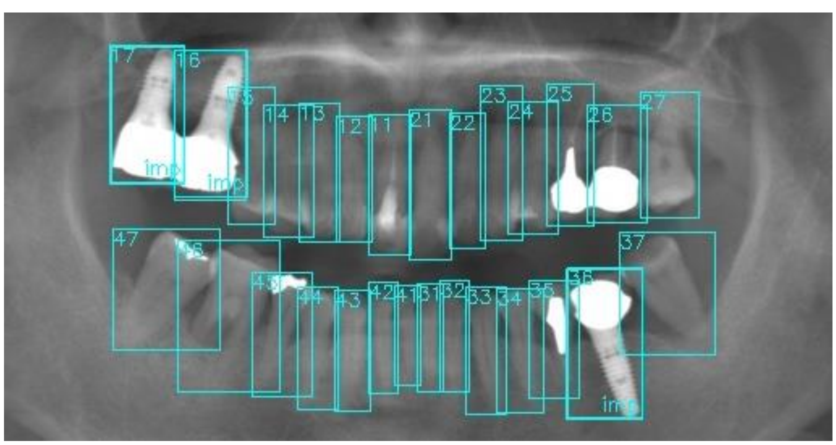

Figure 5. A case of PR and successful classification of all attributes

4. DISCUSSION

The accuracy and F-measure of Methods 1 and 2 were almost the same, but recall was higher in Method 2 than in Method 1. In Method 1, there were many misclassification of the (2) remaining root as the (1) tooth. The result may be due to imbalanced sample size that the number of (1) tooth samples is much larger than others. On the other hand, the precision of Method 1 was higher than that of Method 2. However, Method 2 does not require additional CNN models, but merely requires additional labels for object detection, which is considered useful for its ease and generality.

In Method 2, there was a large difference in the attribute recognition performance between SSD and YOLOv4, which is thought to be due to the difference in class prediction methods. While SSD uses a softmax function for class prediction, YOLOv4 uses logistic regression, which is presumed to have been effective in preventing conflicts between classes. Table 4 shows that there is almost no difference in tooth type recognition performance between 16-class detection and 19-class detection, which is one of the advantages of Method 2.

5. CONCLUSIONS

In this study, 32 tooth types with four attributes were recognized with high accuracy in dental panoramic radiographs. We proposed an easy and robust scheme of assigning two labels to a single object to handle two different tasks simultaneously.

ACKNOWLEDGEMENTS

This study was partially supported by JSPS Grants-in-Aid for Scientific Research JP19K10347 and by a joint study grant from Eye Tech Co. The authors would like to express their gratitude to Y. Ariji, DDS, PhD, (Aichi-Gakuin University, currently with Osaka Dental University) and E. Ariji, DDS, PhD, (Aichi-Gakuin University), and all who provided useful advice in conducting this study.

REFERENCES

[1] Fujita, H., "Ai-based computer-aided diagnosis (ai-cad): the latest review to read first," Radiological Physics and Technology 13(1), 6-19 (2020).

[2] Chung, M., Lee, J., Park, S., Lee, M., Lee, C. E., Lee, J., and Shin, Y. G., "Individual tooth detection and identification from dental panoramic x-ray images via point-wise localization and distance regularization," Artificial Intelligence in Medicine 111, 101996 (2021)

[3] Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., Sveshnikov, M. M., and Bednenko, G.B., "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," Dentomaxillofacial Radiology 48(4), 20180051 (2019)

[4] Muramatsu, C., Morishita, T., Takahashi, R., Hayashi, T., Nishiyama, W., Ariji, Y., Zhou, X., Hara, T., Katsumata, A., Ariji, E., and Fujita, H., "Tooth detection and classification on panoramic radiographs for automatic dental chart filtering: improved classification by multi-sized input data," Oral Radiology 37(1), 13-19 (2021).

[5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C., "Ssd: single shot multibox detector," In European conference on computer vision Springer Cham, 21-37 (2016)

[6] Simonyan, K., Andrew Z., "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).

[7] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778 (2016).

[8] Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M., "Yolov4: optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934 (2020).