Transformer-Based Tooth Segmentation, Identification and Pulp Calcification Recognition in CBCT

Shangxuan Li[1,2], Chichi Li[2], Yu Du[3,4,5], Li Ye[3,4,5], Yanshu Fang[6], Cheng Wang[2(✉)], and Wu Zhou[1(✉)]

[1] School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou, China
zhouwu@gzucm.edu.cn
[2] Hanglok-Tech Co., Ltd., Zhuhai, China
cheng.wang@hanglok-tech.cn
[3] Department of Operative Dentistry and Endodontics, Sun Yat-sen University, Guangzhou, China
[4] Affiliated Stomatological Hospital, Guangzhou, China
[5] Guangdong Provincial Key Laboratory of Stomatology, Guangzhou, China
[6] First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China

Abstract. The recognition of dental pulp calcification has important value for oral clinic, which determines the subsequent treatment decision. However, the recognition of dental pulp calcification is remarkably difficult in clinical practice due to its atypical morphological characteristics. In addition, pulp calcification is also difficult to be visualized in high-resolution CBCT due to its small area and weak contrast. In this work, we proposed a new method of tooth segmentation, identification and pulp calcification recognition based on Transformer to achieve accurate recognition of pulp calcification in high-resolution CBCT images. First, in order to realize that the network can handle extremely high-resolution CBCT, we proposed a coarse-to-fine method to segment the tooth instance in the down-scaled low-resolution CBCT image, and then back to the high-resolution CBCT image to intercept the region of the tooth as the input for the fine segmentation, identification and pulp calcification recognition. Then, in order to enhance the weak distinction between normal teeth and calcified teeth, we proposed tooth instance correlation and triple loss to improve the recognition performance of calcification. Finally, we built a multi-task learning architecture based on Transformer to realize the tooth segmentation, identification and calcification recognition for mutual promotion between tasks. The clinical data verified the effectiveness of the proposed method for the recognition of pulp calcification in high-resolution CBCT for digital dentistry.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_68.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 706–714, 2023.
https://doi.org/10.1007/978-3-031-43904-9_68

Keywords: Pulp calcification recognition · Transformer · Instance correlation · Tooth Segmentation · CBCT

1 Introduction

Pulp calcification is a type of pulp degeneration, characterized by the deposition of calcified tissue in the root canal. Clinically, pulp calcification usually occurs with pulp periapical diseases, which brings great challenges to root canal therapy. Finding pulp calcification before root canal treatment is very important for dentists to decide treatment strategies [1]. However, teeth with pulp calcification usually show few clinical symptoms, which are mainly found and diagnosed by radiographic examination. Compared with X-ray, cone beam computed tomography (CBCT) performs better in displaying root canal structure and pulp disease, so it is widely used for pulp calcification detection [2]. On CBCT images, pulp calcification showed partial or complete high attenuation root canal occlusion as shown in Fig. 1. Currently, the diagnosis of pulp calcification mainly depends on the image recognition of root canal occlusion by dentists. On the one hand, although high-resolution CBCT is used, the image contrast of calcified area is rather low, and human cannot easily find all calcified tubes. On the other hand, human recognition is time-consuming and laborious, and the agreement between observers is far from satisfactory. Therefore, an intelligent recognition method of pulp calcification is urgently needed for digital dentistry in clinical practice.

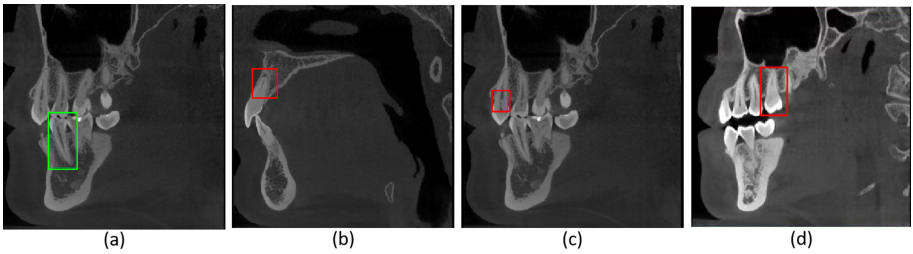

Fig. 1. Representative tooth images in CBCT. (a) Normal teeth; (b) Diffuse calcified teeth with a few root canal residues; (c) Calcified teeth with pulpal stones; (d) Calcified teeth with difficulty in recognition. The red box corresponds to the pulp area. (Color figure online)

With the introduction of artificial intelligence into various fields of medical images, research has proposed tooth and root canal segmentation models based on deep learning networks and CBCT images [3–7], which have achieved good performance in segmentation. However, how to intelligently recognize pulp calcification in small root canals is still very difficult and has not yet been explored. First, the calcified area has no clear morphological characteristics and is difficult for feature representation. Then, the resolution of CBCT image has a high

resolution ($672 \times 688 \times 688$ in this work), while the tooth calcification areas are relatively small and in low contrast. It is very difficult to directly process such large volume data based on deep learning networks. Reducing the high-resolution of CBCT image will weaken the local tooth calcification area information, which brings certain challenges to the intelligent recognition of pulp calcification in CBCT. In addition, the current digital dentistry needs the whole process from 3D volume input, tooth segmentation, identification, and lesion recognition. However, the current relevant research [3–7] separately focuses on functions such as tooth segmentation, root canal segmentation or lesion detection, and has not yet built an integrated intelligent diagnosis process.

To this end, we propose a new method of tooth segmentation, identification and pulp calcification recognition based on Transformer to achieve accurate recognition of pulp calcification in high-resolution CBCT images. Specifically, we propose a coarse-to-fine method to segment the tooth instance in the low-resolution CBCT image, and back to the high-resolution CBCT image to intercept the region of the tooth as the input for the fine segmentation, identification and calcification recognition of the tooth instance. In order to enhance the weak distinction between normal teeth and calcified teeth, we put forward tooth instance correlation and triple loss to further improve the recognition performance of calcification. Finally, we introduce transformer to realize the above three tasks in an integrated way, and achieve mutual promotion of task performance. The clinical oral CBCT image data is used to verify the effectiveness of the proposed method.

2 The Proposed Method

2.1 The Proposed Framework

The network structure designed in this work is shown in Fig. 2. It mainly includes two modules: tooth segmentation and identification module, and pulp calcification recognition module. First, we stack the swin-transformer as the backbone of the network, and save the computation of processing high-resolution CBCT images through down-sampling. Then, we introduce shallower features through skip connection. Those features with higher resolution and shallower layers will contain relatively rich low-level information, which is more conducive to accurate tooth segmentation and identification recognition. In addition, through the multi-task learning mechanism based on Transformer, the performance of tooth segmentation, identification and calcification recognition can be mutually improved.

In the pulp calcification recognition module, we extract the features of each tooth from the deep feature through the results of tooth segmentation, and input them into the pulp calcification recognition module. Specifically, we design an instance correlation transformer (ICT) block. This block allows teeth to learn information from other teeth, so that different teeth can interact, which enables the network itself to explore the relationship between instances, thus improving the recognition performance of calcified teeth. In addition, we introduce a

discriminator in the ICT block, which uses triple loss to learn the spatial distribution between categories, so as to learn better classification embedding.

2.2 Tooth Segmentation and Identification Module

In order to obtain the features of each tooth from the high-resolution CBCT image, we first recognize the segmented and identificated teeth,and combine the result of them for tooth instance segmentation. We use swin-transformer as the network backbone. Swin-transformer [8] is based on the idea of ViT model, which innovatively introduces the sliding window mechanism, so that the model can learn cross-window information. Meanwhile, through the down-sampling layer, the model can process super-resolution images, save computation and focus on global and local information. In addition, it has been proved to have advantages in the process of modeling an end-to-end multi-task system [9]. Unlike the typical swin-transformer, which only uses the Patch Merging layer in the deep layer, we first reduce the resolution of the input feature map through the Patch Merging layer in each stage, and then conduct the next sampling operation. This is conducive to our processing of high-resolution CBCT images. In addition, we have adopted the reverse skip connection (RSC). The information from the deep layer can reversely guide the learning of the shallow details to achieve better segmentation effect, which has been proved effective in segmentation task [10]. In this way, we can distinguish the edges of adjacent teeth more clearly, thus promoting better segmentation recognition.

The segmentation loss L_{seg} of the model can be defined as $L_{seg} = L_{cs} + \gamma_1 L_{dice} + \gamma_2 L_{bs}$, where γ_1,γ_2 are the balance parameters, where L_{cs} is the 33 categories of pixels (32 teeth classes + background) cross entropy loss. L_{dice} is the dice loss of each tooth instance segmentation. L_{bs} is binary segmentation. We calculate the cross entropy loss between teeth and background. The purpose of this step is to assist the model to distinguish foreground and background

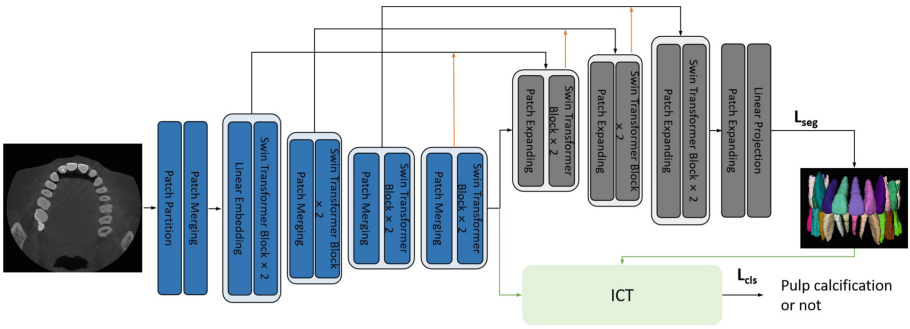

Fig. 2. The proposed framework. The orange arrow indicates the RSC process. (Color figure online)

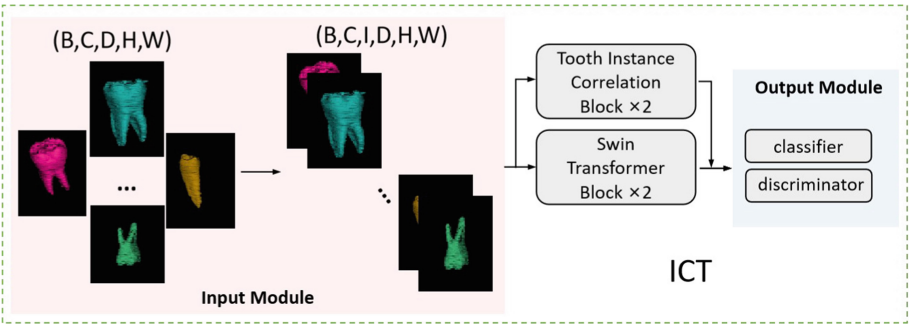

Fig. 3. The instance relevance transformer module

2.3 Calcification Recognition Module and Tooth Instance Correlation Block

Calcified root canals, especially small root canal calcification, are shown on CBCT images as the shadow in the center of the cross section faded and disappeared, and the density of the shadow is close to or the same as that of the surrounding dentin, which is significantly different from the root canal images of other root canals of the same affected tooth and normal adjacent teeth. Based on the above clinical observation, we design the tooth instance correlation block for better calcification recognition. The multi-head attention layer is widely used to build image relationship models in channel and spatial dimensions [11,12], so we believe it can be extended to explore the relationship between tooth instances. Specifically, we propose an ICT related to tooth instances to better identify calcified teeth. The ICT module is shown in Fig. 3.

First, in the input module of ICT, we extract the tooth deep feature from the output feature of the encoder through the prediction of instance segmentation. The specific method is to extract the tooth deep feature one by one according to the tooth id after the prediction label is de-sampled. We splice the extracted tooth instance features in a new dimension I. Then, we reduce the channel dimension to C/16 by convolution operation for this high-dimensional feature (dimensions: (B, C, I, D, H, W)). The purpose of this step is to reduce the heavy calculation caused by the excessive size of the channel dimension. Then, we divide the reshape into (B, C, I, N), where B is the batch size, I is the number of tooth samples, C and N are the number of channels and pixels, respectively. Through the tooth instance correlation block, we learn the cross attention in the I dimension. Given a CBCT image X_i contains multiple tooth instances $x_{i,1}$,$x_{i,2}$,...,$x_{i,n}$. The tooth instance correlation block is constructed as follows:

$$X_i^0 = (x_{i,class}; f(x_{i,1}); f(x_{i,2}); ...; f(x_{i,n})), X_i^0 \in R(n+1) \times d \tag{1}$$

$$X_i^l = MSA(LN(X_i^{l-1})) + X_i^{l-1}, l = 1....L \tag{2}$$

$$X_i^l = LN(MLP(X_i^l)) + X_i^l \tag{3}$$

where MSA is multi-head self attention, L is the layer of MSA, and LN is the standardization layer.

2.4 Triple Loss and Total Loss Function

In order to make the model more discriminative for the recognition of calcified teeth, a discriminator is designed in this work, which uses triplet loss to make the embedding of the input classifier more discriminative. This process is to make the features of the same category as close as possible, and the features of different categories as far away as possible. Meanwhile, in order to prevent the features of the instance from converging into a very small space, it is required that for two positive cases and one negative case, the negative case should be at least margin away from the positive case [13].

Specifically, we randomly selected an instance in the I dimension: Anchor (a), and randomly selected an instance that belongs to the same class as Anchor: Positive (p), and randomly selected an instance that belongs to a different class from Anchor: Negative (n). The goal of model learning is to make the distance $D(a, p)$ between Positive and Anchor as small as possible, and the distance $D(a, n)$ between Negative and Anchor as large as possible L_t. It is defined as follows:

$$L_t = max(D(a, p) - D(a, n) + \alpha), 0) \tag{4}$$

where D is the European distance, α is a margin between positive and negative pairs.

The classification loss is defined as $L_{cls} = L_{pc} + \gamma_3 L_t$, where γ_3 is the balance parameter, L_{pc} is the cross entropy loss as the loss of calcified tooth classification. Finally, the total loss function of the network is defined as $L_{total} = L_{seg} + L_{cls}$.

2.5 Implementation

The initialization setting of the learning rate is 1e-3, with 60000 iterations. The Adam algorithm is used to minimize the objective function. Two RTX3090 GPUs are used, each with 24G memory. The attenuation setting of the learning rate is 0.99 for every 500 iterations. All parameters, including weights and deviations, are initialized using a truncated normal distribution with a standard deviation of 0.1. In the tooth instance segmentation task, we use connected component analysis to extract the maximum area of predicted voxels and remove some small false-positive voxels. Code is available at: https://github.com/Lsx0802/ToothICT.

3 Experimental Results

3.1 Clinical Data, Experimental Setup and Evaluation Metric

This study was performed in line with the principles of the Declaration of Helsinki. In this work, 151 CBCT imaging data from the Imaging department of

the local institute were acquired. The image resolution of the CBCT equipment used was $0.2 \sim 1.0\,\text{mm}$, and the size of the CBCT volume is $672 \times 688 \times 688$. The bulb voltage was $60 \sim 90\,\text{kV}$, and the bulb current was $1 \sim 10\,\text{mA}$. 151 cases of dental symptoms were identified as CBCT oral indications by two dentists with 10 years of clinical experience. Among them, 60 patients had dental pulp calcification. In addition, each tooth was also marked for calcification. One dentist is responsible for the data label, and two doctors review it. When they disagree, they will reach an agreement through negotiation.

The CBCT data is preprocessed as follows. First, considering the balance between computational efficiency and instance segmentation accuracy, all CBCT images are normalized to $0.4 \times 0.4 \times 0.4\,\text{mm}^3$. Then, in order to reduce the impact of extreme values, especially in the metal artifact area, we cut the voxel-level intensity value of each CBCT scan to [0, 2500], and finally normalized the pixel value to the interval [0,1]. For Pulp calcification recognitionin, the adopted evaluation metrics include: Accuracy, Precision, Recall, F1, Dice. The measurement results were conducted with 10 times of four-fold cross validation. In addition, we have compared the performance of the proposed method with the relevant tooth segmentation methods, we used typical segmentation metrics for performance evaluation: Dice, Jaccard similarity coefficient (Jaccard), 95% Hausdorff distance (HD95), Average surface distance (ASD) and have conducted the ablation study of the proposed method.

3.2 Performance Evaluation

As shown in Table 1, Dice is based on the instance segmentation performance of each tooth. Backbone1 uses the swin transformer with skip connection to segment the teeth. w/o RSC is a model for eliminating reverse skip connection design. The segmentation results of tooth instance show that the proposed method is superior to other relevant segmentation methods. The main reason is that the proposed method uses the transformer structure. Its multi-head attention mechanism can capture global information, which is superior to the U-net structure based on CNN local features in the relevant methods. In addition, the ablation study for instance segmentation also shows the effectiveness of the proposed module.

Our model adopts a network based on swin transformer. Through its powerful global and local modeling ability, while retaining the jumping connection in UNet to retain the shallow features, the performance of Backbone1 is better than the previous segmentation model. In particular, we use reverse skip connection and use deep features to guide shallow feature learning, which has achieved obvious improvement in segmentation performance. After combining the task of calcification classification, the segmentation network has been improved a little, which benefits from the ICT module we adopted, because it not only learns the correlation characteristics between calcified teeth and normal teeth, but also learns the morphological correlation between teeth, which is beneficial to tooth segmentation.

Table 2 shows the ablation experimental results of calcified tooth recognition. Backbone 2 is the calcified tooth recognition of the whole module of tooth instance segmentation+classifier. w/o ICT is to remove the tooth instance correlation block, and w/o L_t is to remove the discriminator. We can find that the proposed two modules can effectively improve the performance of calcification recognition.

The accuracy of the model is only 74.62% when only swin transformer is used to classify tooth samples, while our proposed model can improve the performance of pulp calcification recognition by 3.85%. Especially, in the ablation experiment, when the ICT module is removed, the model performance drops obviously, which proves that our proposed ICT module can effectively learn the relationship between dental examples. In addition, after the loss Lt of the discriminant module is removed, the accuracy of the model decreases by about 1.16%, which proves that this method can effectively reduce the distance between similar samples and increase the distance between different samples. (See the supplementary materials for more visualization results)

Table 1. Performance comparison of instance segmentatio)

Method	Dice (%) ↑	Jaccard (%) ↑	HD95 ↓	ASD ↓
ToothNet [4]	90.84 ± 1.10	82.67 ± 1.59	3.13 ± 1.31	0.36 ± 0.17
CGDNet [5]	92.07 ± 0.91	84.93 ± 1.34	2.61 ± 1.011	0.31 ± 0.13
ToothSeg [6]	92.68 ± 0.86	87.39 ± 0.88	2.13 ± 0.47	0.27 ± 0.11
Backbone1	93.90 ± 0.51	88.67 ± 0.82	1.66 ± 0.26	0.25 ± 0.09
w/o RSC	93.99 ± 0.46	88.77 ± 0.93	1.72 ± 0.31	0.25 ± 0.05
Proposed	94.23 ± 0.61	89.64 ± 0.86	1.50 ± 0.27	0.23 ± 0.06

Table 2. Performance of pulp calcification recognition

Method	Accuracy (%) ↑	Precision (%) ↑	Recall (%) ↑	F1 (%) ↑
Backbone2	74.62 ± 0.41	72.67 ± 0.34	77.31 ± 0.43	76.79 ± 0.37
w/o ICT	75.31 ± 0.29	74.48 ± 0.36	78.13 ± 0.31	78.31 ± 0.33
w/o L_t	77.31 ± 0.36	$75{,}81 \pm 0.27$	82.06 ± 0.27	81.15 ± 0.41
Proposed	78.47 ± 0.25	77.31 ± 0.37	82.08 ± 0.23	82.41 ± 0.32

4 Conclusion

In this study, we proposed a calcified tooth recognition method based on transformer, which can detect calcified teeth in high-resolution CBCT images while

achieving tooth instance segmentation and identification. Specifically, we proposed a coarse-to-fine processing method to make it possible to process high-resolution CBCT with deep network for calcification recognition. In addition, the design of instance correlation and triple loss further improved the accuracy of calcification detection. The validation of clinical data showed the effectiveness and advantages of the proposed method. We believe that this research will bring help to the intellectualization of oral imaging diagnosis and the navigation of oral surgery.

Acknowledgement. This research is supported by the school-enterprise cooperation project (No.6401-222-127-001).

References

1. Yang, Y.M., et al.: CBCT-aided microscopic and ultrasonic treatment for upper or middle thirds calcified root canals. BioMed Res. Int. **1–9**, 2016 (2016)
2. Patel, S., Brown, J., Pimental, T., Kelly, R., Abella, F., Durack, C.: Cone beam computed tomography in endodontics - a review of the literature. Int. Endodont. J. (2019)
3. Duan, W., Chen, Y., Zhang, Q., Lin, X., Yang, X.: Refined tooth and pulp segmentation using u-net in CBCT image. Dentomaxillofacial Radiol. 20200251 (2021)
4. Cui, Z., Li, C., Wang, W.: Toothnet: automatic tooth instance segmentation and identification from cone beam CT images. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, 16–20 June 2019, pp. 6368–6377. Computer Vision Foundation/IEEE (2019)
5. Wu, X., Chen, H., Huang, Y., Guo, H., Qiu, T., Wang, L.: Center-sensitive and boundary-aware tooth instance segmentation and classification from cone-beam CT. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 939–942 (2020)
6. Cui, Z., et al.: Hierarchical morphology-guided tooth instance segmentation from CBCT images. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 150–162. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_12
7. Cui, Z., et al.: A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat. Commun. **13**, 2096 (2022)
8. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
9. Bhattacharjee, D., Zhang, T., Süsstrunk, S., Salzmann, M.: Mult: an end-to-end multitask learning transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12031–12041 (2022)
10. Xia, L., et al.: 3d vessel-like structure segmentation in medical images by an edge-reinforced network. Med. Image Anal. **82**, 102581 (2022)
11. Shao, Z., et al.: Transmil: transformer based correlated multiple instance learning for whole slide image classification. Adv. Neural Inf. Process. Syst. **34**, 2136–2147 (2021)
12. Hou, Z., Yu, B., Tao, D.: Batchformer: learning to explore sample relationships for robust representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7256–7266 (2022)
13. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)