Multi-stage dental visual detection based on YOLOv8: Dental 3D CBCT

Zhihao Zheng[1,2,⋆ [0009−0001−4205−6614]] , Dongdong Ma[1,2,⋆ [0009−0001−4606−6131]] , Lian He[1,2,3,† [0009−0005−6785−9041]], Miao Cheng[1,2,3 [0009−0002−6809−7833]], Jianhao Liu[1,2 [0009−0003−2845−9905]]

[1] .Chengdu Institute of Computer Application, Chinese Academy of Science, Chengdu Sichuan 610041, China
[2] School of Computer Science and Technology, University of Chinese Academy of Science, Beijing 100049, China
[3] Shenzhen CBPM-KEXIN Banking Technology Company Limited, Shenzhen Guangdong 518206, China
†Corresponding author: helian@cbpm-kexin.com

Abstract. With the popularity of panoramic X-ray imaging (PXI) and cone beam computed tomography (CBCT) in dental treatment planning and comprehensive prognostic assessment, the demand for effective tooth segmentation is growing. Although 2D panoramic X-rays can effectively detect invisible caries, impacted teeth and supernumerary teeth in children, 3D CBCT has more significant advantages in orthodontic and endodontic treatment, mainly due to its low radiation dose and accurate three-dimensional images. Although 3D images provide more detailed anatomical information, which helps accurate diagnosis and treatment planning, they also face challenges such as difficult data acquisition, radiation exposure, increased processing complexity and metal artifacts. In the field of tooth segmentation, although deep learning methods have made significant progress, the tooth segmentation task still faces many challenges. Compared with tooth semantic segmentation, tooth instance segmentation requires more refined boundary information and accurate capture of tooth morphology, which is particularly difficult in irregular gingival tissue and complex tooth morphology. In addition, the quality of tooth segmentation depends largely on detailed manual annotation, which is not only time-consuming and labor-intensive, but also prone to errors and subjectivity. Recently, STS MICCAI 2024 launched a 3D tooth segmentation challenge, which aims to explore tooth segmentation algorithms based on semi-supervised learning. By participating in this competition, this study explored semi-supervised learning methods through yolov8 [11], aiming to reduce dependence on a large amount of labeled data and train the model by utilizing a large amount of unlabeled data. Finally, among 77 teams, it achieved a score of 0.77 on the validation set, ranking 8th, and a score of 0.77 on the test set.

Keywords: Artificial intelligence · deep learning · semi-supervision · 3D tooth segmentation · medical image processing.

⋆ These authors contributed equally to this work.

1 Introduction

There is a close relationship between oral health and a variety of systemic diseases. For example, oral diseases, especially periodontal diseases, are related to systemic diseases such as cardiovascular disease, diabetes, respiratory diseases, digestive diseases, and certain types of cancer. In addition, if diabetic patients have periodontitis, they will increase the risk of diabetic complications. Conversely, high blood sugar will also increase the body's susceptibility to periodontal pathogens. Long-term tooth loss not only affects chewing function, but may also lead to alveolar bone loss, oral dysfunction, adjacent tooth displacement, and increase the risk of caries and periodontal disease. In terms of overall health, tooth loss may increase the risk of chronic diseases and affect immunity. In terms of mental health, tooth loss may reduce self-confidence and social skills, and even cause psychological problems such as anxiety and depression. The prevention and timely treatment of oral diseases are essential for maintaining overall health.

In modern dental diagnosis and treatment, three-dimensional tooth segmentation technology is essential to improve the accuracy of dental disease diagnosis, the formulation of treatment plans, and basic research on oral health. With the development of digital dental technology, panoramic X-ray imaging (PXI) and cone beam computed tomography (CBCT) are increasingly widely used, which can provide detailed three-dimensional information of teeth and gums. However, due to the complexity of tooth anatomy, differences in imaging protocols, and limited public access to data, the development of automated tooth analysis algorithms faces significant challenges. To overcome these challenges, STS MICCAI 2024 recently launched a challenge for 3D tooth segmentation, which aims to solve the problems of tooth positioning, segmentation, and labeling through deep learning methods.

The tooth segmentation task involves identifying and separating the complex morphology and position of individual teeth from oral 3D scans. This task is essential for the formulation of dental treatment plans such as orthodontic treatment, dental implants, and dental corrections. Although a variety of deep learning-based methods have been proposed in previous studies, many existing methods are based on a large amount of labeled data, but the labeling of these data requires a large number of people with professional knowledge to label, which is time-consuming and labor-intensive, and these methods are still insufficient in dealing with complex tooth boundaries and the variability of tooth morphology between different patients. In order to solve the above problems, we used yolov8 [11] and combined it with semi-supervised learning methods to conduct segmentation experiments on 3D tooth data and achieved excellent results in the competition, thereby alleviating the problem of requiring a large number of people to perform medical tooth labeling.

2 Related Works

Three-dimensional tooth segmentation technology is of great significance for improving the accuracy of diagnosis, formulating treatment plans, and studying

oral health. With the development of computer vision and deep learning technology, the research on automatic tooth segmentation algorithms has received widespread attention.

2.1 Based on traditional method

Early studies mainly relied on manually extracted geometric features, such as surface curvature [1,2], contours [3], and harmonic fields [4], to segment 3D dental scan images. These methods have certain limitations in dealing with tooth and gum boundaries, are sensitive to noise, and are difficult to fully automate.

2.2 Deep learning based methods

With the development of deep learning technology, researchers have begun to explore the use of convolutional neural networks (CNNs) to automatically extract features and perform tooth segmentation. For example, Rao et al. [6] (2020) proposed a symmetric fully convolutional residual network for tooth segmentation. In recent years, the Transformer architecture [8] has demonstrated powerful capabilities in processing sequence data. Hao et al. (2024) [10] proposed the T-Mamba model, which integrates shared position encoding and frequency-based features, as well as gated selection units to improve the accuracy of tooth segmentation.

2.3 Point cloud based methods

With the advancement of 3D scanning technology, the application of point cloud data in tooth segmentation has gradually increased. Zanjani et al. (2019) [9] proposed an end-to-end deep learning system based on PointNet for segmenting teeth and gums from point cloud representations and used a secondary neural network in an adversarial learning setting to refine tooth labels. Lian et al. [7]improved the PointNet architecture and introduced a series of graph-constrained learning modules to extract multi-scale local contextual features for tooth segmentation and labeling in 3D oral scan images.

2.4 Multi-stage approach

Some studies have adopted a multi-stage approach to tackle tooth segmentation tasks, including preliminary segmentation, refined segmentation, and post-processing stages.For example, Cui et al. (2019) [5] proposed a two-stage deep supervised neural network architecture for automatic tooth instance segmentation and recognition. This method usually combine deep learning models with traditional image processing techniques to improve the accuracy and robustness of segmentation.

3 Methods

3.1 YOLOv8

YOLOv8 [11] is the next major update of YOLOv5 [12], which Ultralytics opened on January 10, 2023. It currently supports image classification, object detection, and instance segmentation tasks. YOLOv8 [11] builds on the success of previous YOLO versions and introduces new features and improvements to further improve performance and flexibility. It can run on a variety of hardware platforms from CPUs to GPUs. However, Ultralytics did not directly name the open source library YOLOv8 [11], but directly used the word Ultralytics. The reason is that Ultralytics positions this library as an algorithm framework rather than a specific algorithm. One of the main features is scalability. It hopes that this library will not only be used for the YOLO series models, but will also be able to support non-YOLO models and various tasks such as classification, segmentation, and pose estimation.

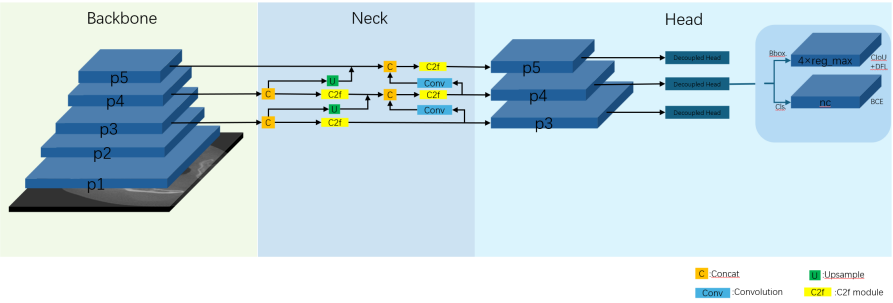


Fig. 1. YOLOv8 Architecture

YOLOv8's [11] architecture adopts novel improvements to achieve excellent detection accuracy while maintaining high speed and efficiency. YOLOv8's [11] backbone architecture is improved on the basis of YOLOv5 [12], including the following key modifications:

1. Backbone & Neck: The backbone network and Neck part may refer to the design concept of YOLOv7 ELAN. The C3 structure of YOLOv5 [12] is replaced with the C2f structure with richer gradient flow, and the number of channels is adjusted for models of different scales, which greatly improves the model performance.
2. Head: The Head part has been significantly changed compared to YOLOv5 [12]. It has been replaced with the current mainstream decoupled head structure, separating the classification and detection heads. It has also been changed from Anchor-Based to Anchor-Free.

3. Data augmentation: The data augmentation part of training introduces the operation of turning off Mosiac augmentation in YOLOX for the last 10 epochs, which can effectively improve accuracy.

3.2 Multi-stage training

We used it in this competition. The specific workflow can be seen in Figure 1. We first pass the labeled data in the training set through the data processing module to obtain data that conforms to the input format of yolov8 [11], and then send it to the yolov8 [11] network as input. After training with these small amounts of labeled samples, we will get a model1 with certain predictive ability on dental data. In order to combine the semi-supervised learning method to make full use of the 300 unlabeled 3D CBCT sample data in the training set, we used the model1 model that was just trained to predict these unlabeled data, and obtained a large amount of pseudo-label data to make up for the lack of labeled data. Then, we introduced a multi-stage learning method, which specifically includes a rough training stage and a fine training stage. In the rough training stage, we reinitialized a new model and used the 300 data and their pseudo-labels as the input part of the new model. After multiple iterations, we will get a rough training model weight model2. The implementation shows that only using these 300 pseudo-labels, after submitting the verification results in the verification stage, we can get a score of 72% on the comprehensive evaluation index of the competition. Based on the model weights of the rough training, we collected 30 labeled high-quality training data for fine training to allow the model to learn better data distribution. Experiments show that after the fine training stage, the obtained model model3 predicted the validation set again and obtained a score of 77%, an increase of about 6 percentage points, verifying the effectiveness of multi-stage training.

4 Experiment

4.1 Dataset

The 3D CBCT dataset used in this data experiment is provided by the MICCAI STS 2024 competition. The training set includes 330 CBCT data, of which 30 are officially annotated and 300 are unlabeled. The validation set includes 20 CBCT data. The annotated data of the validation set and the test set data are not made public by the official competition to ensure the fairness of the competition.

4.2 Evaluation Metric

This challenge uses multiple evaluation metrics to measure segmentation quality, including Dice coefficient (DSC), intersection over union (IoU), normalized surface distance (NSD), and instance accuracy (IA). The following is a detailed explanation and calculation formula of these metrics:

Dice coefficient (DSC) The Dice coefficient is a commonly used indicator to measure the segmentation quality in binary classification problems. The calculation formula is:

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{1}$$

Wherein, X is the predicted segmentation result, Y is the ground truth, $|X \cap Y|$ represents the size of the intersection of X and Y, and $|X|$ and $|Y|$ represent the size of X and Y respectively.

Intersection over Union (IoU) The intersection-over-union ratio is also called the Jaccard index, which is used to measure the overlap between the predicted segmentation result and the true annotation. The calculation formula is:

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \tag{2}$$

The definitions of X and Y are the same as those in the Dice coefficient.

Normalized Surface Distance (NSD) Normalized surface distance is a metric that measures the surface distance between the predicted segmentation result and the true annotation, and is usually used in conjunction with the Dice coefficient to tolerate a certain distance error. The calculation formula is:

$$\text{NSD} = \frac{\text{surface\_dice\_at\_tolerance } (X, Y, \text{ tolerance })}{\text{tolerance}} \tag{3}$$

Where, tolerance is the specified tolerance distance threshold.

Instance Accuracy (IA) 1. Instance accuracy is a metric that measures the quality of segmentation in instance segmentation tasks. It calculates the ratio of correctly classified instances with an intersection-over-union ratio greater than 0.5 to all instances:

$$\text{IA} = \frac{TP}{|\ \text{unique\_classes } (X) \cup \ \text{unique\_classes } (Y)\ | - 1} \tag{4}$$

Where TP is the number of correctly classified instances, unique_classes(X) and unique_classes(Y) are the unique sets of classes in X and Y respectively.

4.3 Experimental Results

This competition is divided into the verification phase and the testing phase. In the verification phase, we conducted two phases of coarse training and fine training on the model. In the testing phase, we submitted the fine training model for official testing.

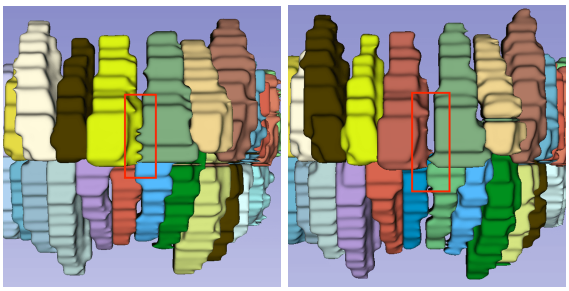
Table 1. This table shows the two prediction results obtained in the verification phase, where in stands for instance, im stands for image, IA stands for Identification Accuracy, stage 1 stands for the rough training phase, and stage 2 stands for the fine training phase.

	all	dice_in	dice_im	NSD_in	NSD_im	mIoU_in	mIoU_im	IA
stage 1	0.7152	0.7201	0.8052	0.6370	0.7300	0.5914	0.6742	0.8486
stage 2	0.7745	0.7534	0.8523	0.7304	0.8473	0.6368	0.7429	0.8582

Table 2. This table is the predicted results obtained during the testing phase, and the predicted results are given by the competition officials.

	all	dice_in	dice_im	NSD_in	NSD_im	mIoU_in	mIoU_im	IA
stage 1	0.7728	0.7787	0.8357	0.7363	0.8087	0.6498	0.7187	0.8818

From Table 1, we can see that the model obtained after only the rough training stage achieved a score of 72% on the validation set. After the fine training stage, we can find that it is better than the indicators of the rough training stage in all evaluation indicators, and the comprehensive score is 6 percentage points higher than the rough training result. The results in Table 2 show that the model obtained in the fine training stage also has stable results on the test set.



(a) stage1 demo (b) stage2 demo

Fig. 2. Comparison of the results of the two stages

From Figure 2, we [1] can see the rendering effects of the predictions of these two stages on the 3D slicer software. Figure a is the result of coarse training, and Figure b is the result of fine training. From the red marked area, we can see that the segmentation boundary effect of the fine training prediction is clearer

than that of the coarse training prediction, which proves the effectiveness of multi-stage training.

References

1. Zhao, M., Ma, L., Tan, W., & Nie, D. (2006). Interactive tooth segmentation of dental models. In IEEE Engineering in Medicine and Biology Conference (IMBC'06) (pp. 654–657)
2. Yuan, T., Liao, W., Dai, N., Cheng, X., & Yu, Q. (2010). Single-tooth modeling for 3D dental model
3. Sinthanayothin, C., & Tharanont, W. (2008). Orthodontics treatment simulation by teeth segmentation and setup. In International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON'08) (pp. 81–84)
4. Zou, B.-j., Liu, S.-j., Liao, S.-h., Ding, X., & Liang, Y. (2015). Interactive tooth partition of dental mesh base on tooth-target harmonic field. Computers in biology and medicine, 56, 132–144
5. Cui, Z., Li, C., & Wang, W. (2019). Toothnet: Automatic tooth instance segmentation and identification from cone beam ct images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6368–6377)
6. Rao, Y., Wang, Y., Meng, F., Pu, J., Sun, J., & Wang, Q. (2020). A symmetric fully convolutional residual network with dcrf for accurate tooth segmentation. IEEE Access, 8, 92028–92038
7. Lian, C., Wang, L., Wu, T.H., Liu, M., Durán, F., Ko, C.C., & Shen, D. (2019). Meshsnet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 (pp. 837–845). Springer.
8. Ashish, V., Noam, S., Niki, P., Jakob, U., Llion, J., Aidan N., G., Lukasz, K., & Illia, P. (2017) Attention is All You Need, Advances in neural information processing systems, 30: 5998-6008.
9. Zanjani, F. G., Moin, D. A., Verheij, B., Claessen, F., Cherici, T., Tan, T., et al. (2019). Deep Learning Approach to Semantic Segmentation in 3D Point Cloud Intra-oral Scans of Teeth. In International Conference on Medical Imaging with Deep Learning (MIDL).
10. Hao, J., Liao, W., Zhang, Y., Peng, J., Zhao, Z., Chen, Z., Zhou, B., Feng, Y., Fang, B., Liu, Z., Zhao, Z. (2024). Toward clinically applicable 3-dimensional tooth segmentation via deep learning. Journal of Dental Research, 101(3), 304–311.
11. Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics
12. Jocher, G. (2020). YOLOv5 by Ultralytics (Version 7.0) [Computer software]. https://doi.org/10.5281/zenodo.3908559