A Two-Stage Semi-Supervised nnU-Net Model for Automated Tooth Segmentation in Panoramic X-ray Images

Changkai Ji[1[0009−0007−7090−7360]], Yusheng Liu[1[0009−0004−2624−9223]], Lanshan He[1[0009−0009−6093−8803]], Yuxian Jiang[1[0009−0002−7689−5333]], Chuanyi Huang[1[0009−0009−3223−0082]], and Lisheng Wang[1[0000−0003−3234−7511]]

Institute of Image Processing and Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
{changkaiji, lswang}@sjtu.edu.cn

Abstract. Automatic tooth segmentation in 2D panoramic X-ray images is crucial for various applications. Task 1 of the MICCAI STS 2024 Challenge aims to advance automated tooth segmentation techniques by providing datasets comprising labeled and unlabeled panoramic X-ray images. This paper addresses the challenge of limited labeled data by framing the problem as a semi-supervised learning task. We propose a two-stage deep learning model based on nnU-Net. The method first performs quadrant segmentation using one nnU-Net, followed by fine tooth segmentation using another nnU-Net within each quadrant. To effectively utilize unlabeled data, we implement a selective stability-based retraining strategy to generate reliable pseudo-labels. We further enhance model performance through post-processing methods such as connected domain analysis. Quantitative evaluation on the STS 2024 validation set demonstrates that our method achieves strong performance across several metrics (Dice_instance = 79.82%, Dice_image = 94.02%). In the competition's validation phase, the method was awarded second place, validating its efficacy in automated dental segmentation of panoramic X-ray images.

Keywords: Panoramic X-ray analysis · Semi-supervised Learning · nnU-Net

1 Introduction

Panoramic radiographs play a crucial role in dentistry, offering clinicians a comprehensive view of the oral and maxillofacial structures. This imaging technique is widely utilized for various purposes, including diagnosis, treatment planning, orthodontic assessment, and forensic applications. Despite their indispensability in dental practice, panoramic radiographs have inherent limitations, such as the lack of three-dimensional information, potential image distortion, and insufficient measurement accuracy in certain cases [5]. These constraints may impact diagnostic precision and treatment planning efficacy.

Traditionally, dental professionals have manually segmented dental images from panoramic radiographs, a process that is time-consuming and heavily reliant on operator expertise. Given the increasing workload in dental practices [8], this manual method fails to meet current efficiency demands. Consequently, the introduction of automated systems for these tasks has become imperative. Automation not only significantly enhances efficiency but also ensures consistency in analysis results and minimizes human error [9].

In recent years, artificial intelligence techniques, particularly convolutional neural networks (CNNs), have achieved significant breakthroughs in medical image analysis. These techniques have demonstrated exceptional performance across various medical imaging tasks, offering new possibilities for automated analysis of panoramic radiographs [6]. However, numerous challenges persist in multi-category detection and segmentation tasks for panoramic radiographs, necessitating further research and development.

A notable challenge is the simultaneous segmentation of permanent and deciduous tooth [4]. While most current research focuses on a single tooth type, combining both significantly increases task complexity. This complexity arises from two main factors [2]: firstly, the significant imbalance in the number of permanent and deciduous tooth may cause the model to favor the category with the larger sample size during training; secondly, the inclusion of deciduous tooth increases the number of categories requiring recognition and segmentation, placing higher demands on the model's learning capacity. Additionally, the inference time of these models must be considered, as prolonged processing times may limit their clinical applicability. Striking a balance between model accuracy and computational efficiency is crucial for the successful integration of these automated systems into dental practice workflows.

These challenges underscore the necessity of developing novel algorithms for dental image processing. The ideal solution should significantly enhance processing speed while ensuring high-precision segmentation and accurate Federation Dentaire Internationale (FDI) numbering [3]. This necessitates a balanced approach in algorithm design, considering accuracy, degree of automation, and operational efficiency. To address these challenges in Task 1 of the MICCAI STS 2024 Challenge, we propose a semi-supervised model based on nnU-Net [1], aimed at achieving efficient dental instance segmentation while considering algorithm runtime. The contributions of our work can be summarized as follows:

- We designed a self-training framework based on nnU-Net, enhancing model performance through selective iterative training.
- Our algorithm improves segmentation accuracy while maintaining operational efficiency, striving for an optimal balance between precision and performance.
- At the competition's validation phase, our method secured second place, further demonstrating its efficacy in dental segmentation tasks.

2 Proposed Method

2.1 Framework Overview

As shown in Fig. 1, we proposed a tooth segmentation method based on nnU-Net, incorporating semi-supervised learning techniques [7], particularly self-training strategies, to enhance segmentation efficacy.

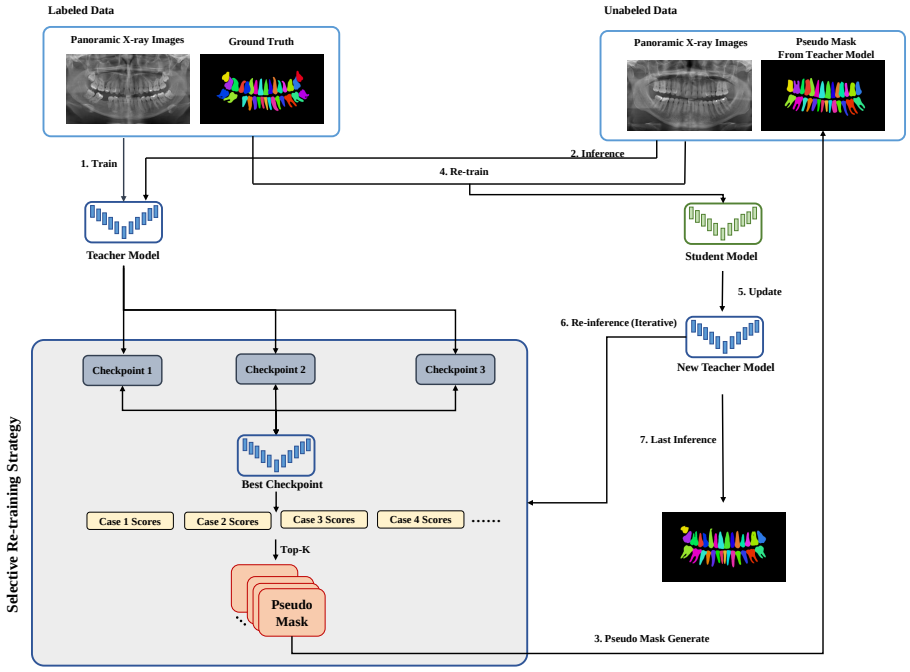

Fig. 1. The framework comprises a teacher model and a student model. The teacher model, initially trained on a limited set of labeled data, generates pseudo-labels for unlabeled data. These pseudo-labels undergo a selective retraining strategy to filter out low-quality labels. Subsequently, the student model is trained using both the original labeled data and the filtered pseudo-labeled data. In each iteration, the student model is then promoted to become the new teacher model.

2.2 Stability-Based Data Screening Mechanism

We have developed a data filtering mechanism to expand the effective training dataset. This mechanism evaluates the reliability of unlabeled samples by assessing the stability of their pseudo-labels during the training process. The procedure is as follows: (1) Multiple teacher models are saved at various training stages.

These models generate predictions on unlabeled data, comparing the consistency of predictions across different stages. (2) The stability of each unlabeled sample is quantified by calculating the average Dice coefficient between early and final predictions. Unlabeled data are sorted and filtered according to their stability scores. (3) High-quality pseudo-labeled data are combined with the original labeled data for student model training. This iterative process gradually improves model performance.

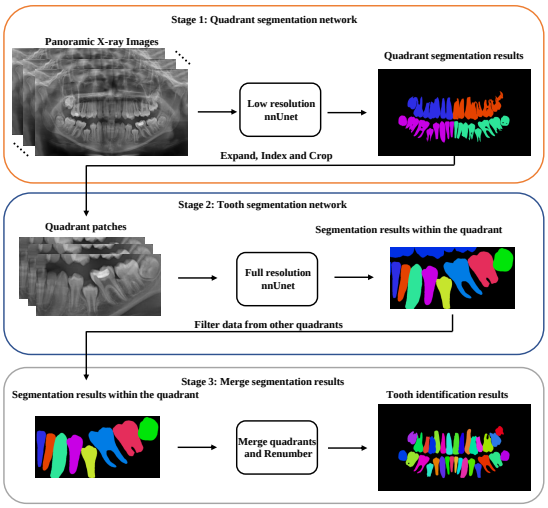

Fig. 2. Schematic diagram of the two-stage segmentation method. The first stage divides all tooth into four quadrants. The second stage identifies and segments each tooth in the quadrants. The third stage combines the results from all four quadrants to reconstruct the segmentation in the original image space.

2.3 Two-Stage nnU-Net Architecture

As shown in Fig. 2, to address the specific challenges of medical image segmentation, we employ a two-stage segmentation strategy based on nnU-Net. The strategy comprises the following key steps:

- nnU-Net segments the input dental panoramic image into four categories, identifying the four dental quadrants.
- Based on the first stage results, tooth in each quadrant undergo fine segmentation. A 14-class nnU-Net segmentation model is applied, including eight classes of permanent tooth, five classes of deciduous tooth, and one class for tooth in other quadrants.
- Segmentation results from all four quadrants are combined to produce a comprehensive tooth segmentation map.

This two-stage approach ensures segmentation accuracy while maintaining computational efficiency.

3 Experiments and Results

3.1 Dataset and Assessment Metrics

This paper utilized the dataset provided by Task 1 of the MICCAI STS Challenge 2024. The dataset comprises three sections: a labeled set of 30 panoramic dental images with precise individual tooth annotations and corresponding FDI tooth numbers, an unlabeled set of 2,350 panoramic dental images, and a validation set of 20 panoramic dental images.

Evaluation metrics encompassed both segmentation accuracy and operational efficiency. Accuracy metrics included instance-level and image-level Dice Similarity Coefficient (DSC), Normalized Surface Distance (NSD), Mean Intersection over Union (mIoU), and Identification Accuracy (IA). Additionally, algorithm runtime and GPU memory consumption were evaluated during the challenge's testing phase to comprehensively assess the segmentation algorithm's performance.

3.2 Implementation details

Training Procedure. In implementing the two-stage nnU-Net model, we adopted different strategies for each stage. For the quadrant segmentation stage, we disabled the symmetric data augmentation in the nnU-Net model to mitigate the impact of tooth symmetry. When generating regions of interest (ROIs), we applied a 10-pixel edge extension to each quadrant boundary, ensuring the inclusion of critical periodontal regions and facilitating more accurate tooth boundary delineation in subsequent stages. For the intra-quadrant tooth segmentation stage, we re-enabled symmetric data augmentation to fully leverage tooth structural features. The model was trained for 150 epochs to ensure sufficient learning.

Post-processing. Post-processing of segmentation results was a crucial step in our approach. For the first-stage results, we implemented several processes: filtering 5% of image boundaries to remove potential interfering information, eliminating connectivity domains smaller than 2,000 pixels, and generating ROIs using the processed masks. After the second-stage intra-quadrant tooth segmentation, we addressed cases of tooth misidentification using a connectivity domain analysis-based post-processing method. For connected domains smaller than 2,000 pixels, we calculate the number of pixels corresponding to each FDI category number in the bounding box, and convert the connected domain to the tooth category with the highest number of pixels.

Selective Re-training Strategy. In the selective retraining strategy, we utilized three uniformly saved checkpoints (at 1/3, 2/3, and 3/3 of the total epochs) to assess image reliability. We performed two iterations of pseudo-label updates. To maintain balance between permanent and deciduous tooth, the final model

was trained using 30 labeled datasets, 101 pseudo-labeled datasets for permanent tooth, and 101 pseudo-labeled datasets for deciduous tooth.

Environments and Requirements. The proposed method's environments and requirements are detailed in Table 1.

Table 1. System Configuration

Ubuntu version	Ubuntu 24.04 LTS
CPU	Intel(R) Xeon(R) Platinum 8352S CPU @ 2.20GHz
RAM	503 GB
GPU	1 NVIDIA GeForce RTX 4090 (24G)
CUDA version	12.4
Programming language	3.9.19
Deep learning framework	PyTorch (torch 1.12.1, torchvision 0.19.1)
Code will available at	After the release of the test rankings

Inference Acceleration. Given that algorithm runtime was assessed during the competition's testing phase, we optimized computational efficiency. The challenge involved instance segmentation across multiple classes, where traditional interpolation methods can introduce substantial computational overhead. To mitigate this, we applied interpolation to floating-point tensors using PyTorch's interpolate function. For more details, please refer to our code. This method preserves floating-point precision and significantly enhances computational efficiency in large-scale multi-class segmentation tasks.

3.3 Results and Analysis

Quantitative Results. The two-stage segmentation model proposed in this paper demonstrates good performance in dental image segmentation tasks. As shown in Table 2, the two-stage nnU-Net model achieves a Dice coefficient of 79.82% at the instance level and 94.02% at the image level.

Table 2. Comparison of One-stage and Two-stage nnU-Net Models

Model	Dice (instance)	Dice (image)	NSD (instance)	NSD (image)	mIoU (instance)	mIoU (image)	IA
One-stage	54.49%	90.92%	77.14%	94.06%	67.60%	83.52%	74.16%
Two-stage	79.82%	94.02%	84.14%	96.74%	75.66%	88.77%	81.99%

To further validate the effectiveness of our method, we conducted a comparison experiment between the one-stage and two-stage nnU-Net models. The one-stage nnunet model does not go through the step of segmenting quadrants, but directly segments all categories of tooth. The experimental results show that the two-stage nnU-Net model significantly outperforms the one-stage model in

the instance-level evaluation metrics. We believe that this performance difference mainly stems from the following factors.

Firstly, the tooth segmentation task has its own special characteristics. The order of arrangement between tooth is crucial for correctly discriminating tooth numbers, which requires the model to be able to effectively capture the spatial relationships between tooth. The one-stage nnU-Net model has limitations in dealing with such complex spatial relationships. Specifically, the patch size of the model is set to 896×1536, while the size of the validation set input image is 2000×942 or 1991×1127. This mismatch makes it difficult for the model to fully capture the spatial information in the image.

While increasing the patch size to 2048×1024 to match the input size seems to be an intuitive solution, it would cause nnU-Net to take too long in the data preprocessing stage, which is not favourable for clinical applications. On the other hand, decreasing the input image size affects the segmentation accuracy, creating a performance-efficiency trade-off dilemma.

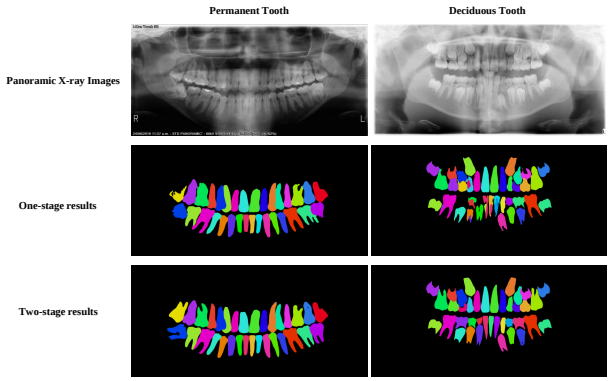

Fig. 3. Visualization results of one-stage nnU-Net and two-stage nnU-Net.

In contrast, our proposed two-stage nnU-Net model cleverly solves this problem. In the first stage, the model only needs to divide the tooth into four quadrants, which greatly reduces the dependence on the spatial location relationship between tooth. Subsequently, the second stage performs tooth segmentation in each quadrant. Since the size of the processed image is already smaller than the patch size, the model can more effectively use the spatial structure relationship between tooth to achieve more accurate segmentation.

Based on the above analysis, we believe that the two-stage nnU-Net segmentation model is not only better than the one-stage model in terms of performance, but also has a wider range of application scenarios. Especially in tasks such as tooth segmentation, where complex spatial structural relationships need to be considered, the two-stage model shows obvious advantages. This approach pro-

vides an effective solution for medical image segmentation, especially in the field of dental image processing, and has the potential to significantly improve the efficiency and accuracy of clinical diagnosis and treatment planning.

Qualitative Results. As illustrated in Fig. 3, there is a significant disparity in segmentation effectiveness between the one-stage and two-stage nnU-Net models. The one-stage nnU-Net tends to misassign FDI numbers due to its inability to fully capture the spatial relationships among all tooth in the panoramic image. This phenomenon explains the one-stage model's significantly lower performance in instance-level Dice coefficients, despite minimal differences in image-level Dice coefficients. These results underscore the importance of accurately modeling spatial information in tooth segmentation tasks.

Figure 4 presents a the segementation results of the two-stage nnU-Net model's effectiveness in segmenting permanent tooth and deciduous tooth. The results demonstrate that our method significantly improves segmentation accuracy.

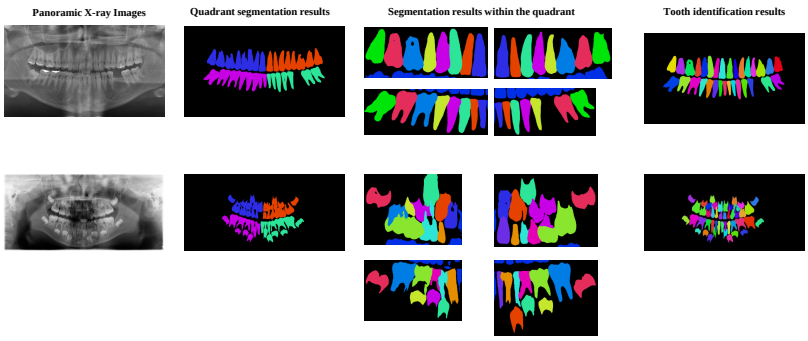

Fig. 4. Visualization results of permanent tooth and deciduous tooth.

4 Conclusion

In this paper, we propose a two-stage semi-supervised learning framework based on nnU-Net for automatic tooth segmentation in 2D panoramic X-ray images. Our method first uses a 2D nnU-Net model for quadrant segmentation, followed by another 2D nnU-Net model for tooth segmentation in each quadrant. This strategy significantly improves computational efficiency while maintaining segmentation accuracy. We utilize a stability-based selective retraining strategy to obtain reliable pseudo-labels from limited labelled data, effectively expanding the training dataset. To further optimise the model performance, we apply post-processing methods such as connected domain analysis. The experimental results show that our method exhibits good performance in segmenting both permanent

and deciduous tooth instances, and achieves the second place in the validation phase of the MICCAI STS 2024 Task 1 Challenge.

References

1. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
2. Li, P., Liu, Y., Cui, Z., Yang, F., Zhao, Y., Lian, C., Gao, C.: Semantic graph attention with explicit anatomical association modeling for tooth segmentation from cbct images. IEEE Transactions on Medical Imaging **41**(11), 3116–3127 (2022)
3. Mei, L., Fang, Y., Cui, Z., Deng, K., Wang, N., He, X., Zhan, Y., Zhou, X., Tonetti, M., Shen, D.: Hc-net: Hybrid classification network for automatic periodontal disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 54–63. Springer (2023)
4. Pinheiro, L., Silva, B., Sobrinho, B., Lima, F., Cury, P., Oliveira, L.: Numbering permanent and deciduous teeth via deep instance segmentation in panoramic x-rays. In: 17th International Symposium on Medical Information Processing and Analysis. vol. 12088, pp. 95–104. SPIE (2021)
5. Şekerci, A.E., Şişman, Y.: Comparison between panoramic radiography and cone-beam computed tomography findings for assessment of the relationship between impacted mandibular third molars and the mandibular canal. Oral Radiology **30**, 170–178 (2014)
6. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**(1), 221–248 (2017)
7. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine learning **109**(2), 373–440 (2020)
8. 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. IEEE (2020)
9. Zhao, Z., Wang, S., Gu, J., Zhu, Y., Mei, L., Zhuang, Z., Cui, Z., Wang, Q., Shen, D.: Chatcad+: Towards a universal and reliable interactive cad using llms. IEEE Transactions on Medical Imaging (2024)