Efficient nnU-Net for Tooth and Root Canal Segmentation in CBCT

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

School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
{changkaiji, lswang}@sjtu.edu.cn

Abstract. Accurate segmentation of teeth and root pulp canals from cone-beam computed tomography (CBCT) images is essential for clinical applications such as treatment planning, root canal therapy, and prosthetics. Manual segmentation is time-consuming, subjective, and impractical for routine use, motivating the need for automated approaches. In this work, we propose a solution based on nnU-Net for multi-class dental structure segmentation. Our pipeline incorporates customized preprocessing, efficient training, and lightweight post-processing. Furthermore, we introduce inference acceleration strategies, including the removal of redundant augmentations and optimized interpolation, which reduce inference time by nearly fourfold with only marginal performance degradation. Experimental results on the MICCAI STSR 2025 Challenge Task 1 demonstrate that our approach achieves competitive segmentation accuracy across multiple metrics, achieving a top-three ranking in the competition. These findings highlight the effectiveness of nnU-Net and our acceleration strategies in achieving a favorable balance between accuracy and efficiency, underscoring the potential of our method for clinical deployment. Our codes are available at: https://github.com/duolawa/MICCAI-2025-STSR-Task-1

Keywords: CBCT segmentation · nnU-Net · Root canal segmentation · Inference acceleration

1 Introduction

The MICCAI STSR 2025 Challenge Task 1 focuses on the segmentation of teeth and root pulp canals from cone-beam computed tomography (CBCT) images. CBCT is a widely used imaging modality in dental diagnostics, providing high-resolution 3D reconstructions of hard tissues with relatively low radiation doses [2]. The task involves accurately segmenting teeth and their corresponding root pulp canals from a set of labeled and unlabeled CBCT images. The challenge aims to push the boundaries of automated segmentation for dental structures, which are critical for clinical applications such as treatment planning, root canal

therapy, and dental prosthetics [12,8]. The competition is designed to evaluate both segmentation accuracy and efficiency, with the goal of developing robust methods suitable for clinical settings.

Manual segmentation of dental structures in CBCT images is a labor-intensive and highly specialized task that relies heavily on the expertise of experienced radiologists and dental professionals [15,5]. This process requires meticulous attention to detail and considerable time, as the accurate delineation of complex anatomical structures like teeth and root pulp canals can be challenging. Furthermore, manual segmentation is inherently subjective, with variations in results between different practitioners, which can lead to inconsistencies and affect the reliability of clinical decision-making. The time-consuming nature of manual segmentation also makes it impractical for routine clinical workflows, where rapid and reliable decisions are crucial.

In response to these limitations, deep learning approaches have gained considerable attention in recent years for automating medical image segmentation tasks [1,6,9,11,13,10,7,19]. These methods have shown remarkable promise in reducing the time and effort involved in manual segmentation, while maintaining or even surpassing the accuracy of junior clinicians. Deep learning models are capable of learning complex patterns from large-scale datasets, enabling them to automatically identify and segment anatomical structures with high consistency and precision [14]. This has made them particularly attractive for dental and maxillofacial imaging, where accurate segmentation is critical for diagnosis and treatment planning.

Despite the successes of deep learning methods, challenges remain in applying these techniques to CBCT images. The segmentation of fine structures such as teeth and their root pulp canals requires a high level of detail and precision [3]. Additionally, the high computational cost associated with deep learning models can pose a challenge, especially in clinical settings where real-time processing and low resource consumption are crucial [20]. Balancing the trade-off between segmentation accuracy and computational efficiency is therefore a key challenge in the development of practical solutions for dental image segmentation.

To address these challenges, we propose a solution based on nnU-Net [4], a deep learning framework that has shown robust performance in medical image segmentation tasks. nnU-Net automatically adapts to the specific characteristics of a given dataset, making it a strong choice for multi-class segmentation problems such as this one. In our approach, we incorporate a variety of strategies to optimize the trade-off between segmentation quality and computational efficiency. The performance of our method was validated by achieving top-three results in the MICCAI STSR 2025 Challenge Task 1, demonstrating its effectiveness and potential for real-world applications.

- Our method achieved satisfactory segmentation accuracy, effectively identifying both teeth and root pulp canals across various validation cases.
- We optimized and accelerated the inference runtime, ensuring that the model could perform efficiently even with large-scale datasets.

- Our approach secured a top-three finish in the MICCAI STSR 2025 Challenge Task 1, demonstrating the effectiveness and competitiveness of our solution in the field of dental image segmentation.

2 Proposed Method

2.1 Framework Overview

As shown in Fig. 1, we propose a solution based on the nnU-Net framework. The overall architecture of our model is specifically designed to handle the segmentation of teeth and root pulp canals in CBCT images effectively.

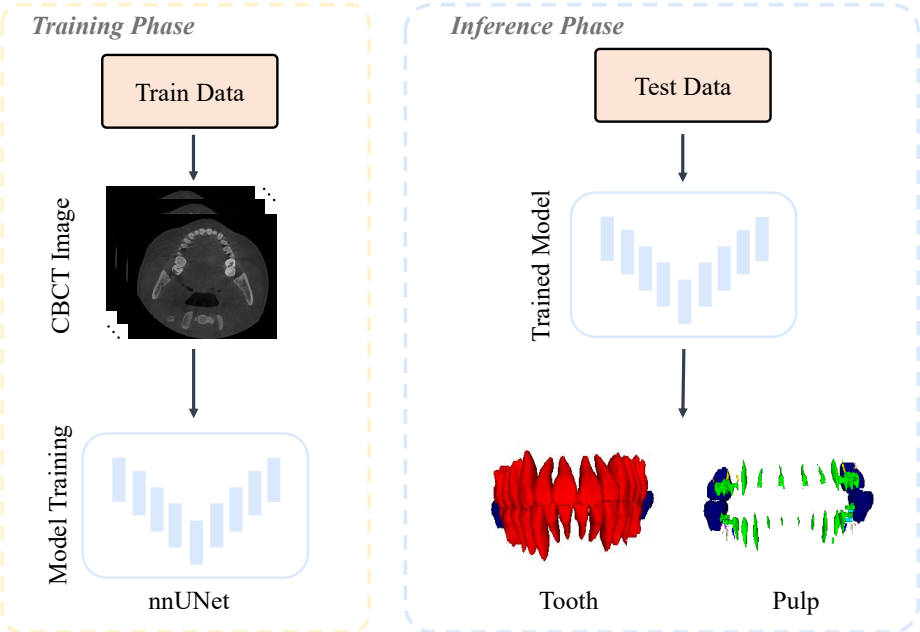

Fig. 1. The overall structure of our nnU-Net-based model for teeth and root pulp canal segmentation.

2.2 Data Preprocessing

In our approach, we perform a series of preprocessing steps to ensure the integrity and consistency of the dataset before training the model. After confirming the integrity of the dataset, we normalize the images to a consistent scale and intensity range. This step helps to standardize the data, making it easier for the model to learn meaningful features. Additionally, we apply necessary transformations,

such as resampling the images to a uniform resolution, ensuring that the data is suitable for the nnU-Net framework. These preprocessing steps allow the model to efficiently process the CBCT images and perform accurate segmentations of the teeth and root pulp canals.

2.3 Model Training

For training the model, we utilize the nnU-Net framework, which is specifically designed to handle a variety of medical image segmentation tasks. The training process begins by using a 3D full-resolution approach, allowing the model to learn spatially rich features from the original high-resolution data. We leverage the entire training dataset, ensuring that the model learns from a comprehensive set of images representing a diverse range of cases. During training, the nnU-Net model adapts to the specific characteristics of the dataset, automatically adjusting its architecture to achieve optimal segmentation performance.

2.4 Post-processing

During the post-processing phase, we address the label mapping required by the nnU-Net framework, which expects labels to be sequentially incremented starting from 1. Specifically, we map the segmentation output label values, which are sequentially incremented to meet nnU-Net's requirements, back to their original label values. The label mapping is applied through a lookup table, which is constructed to accommodate all possible label values in the dataset, with the specific labels being mapped according to the predefined mapping dictionary. This approach provides a fast and effective method for label mapping, reducing computational overhead by directly modifying the label values within the image data without additional redundant processing steps.

3 Experiments and Results

3.1 Dataset and Assessment Metrics

The dataset consists of both labeled and unlabeled CBCT images. The training set includes 30 labeled images with segmentation masks for teeth and root canal structures, alongside 300 unlabeled images for model development. The validation set contains 40 images [18,16,17].

The evaluation of segmentation algorithms considers both segmentation accuracy and computational efficiency. For accuracy, the following metrics are used: Dice Similarity Coefficient (DSC) for overlap evaluation, Normalized Surface Distance (NSD) for surface proximity, mean Intersection-over-Union (mIoU) for region overlap, and Identification Accuracy (IA) for correct identification of anatomical structures. In terms of efficiency, the algorithm's running time and GPU memory consumption are assessed during inference.

3.2 Implementation details

Environments and Requirements. The specific details of the computational environment and dependencies are provided in Table 1. Our model was trained for 1000 epochs.

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	Python 3.9.19
Deep learning framework	PyTorch (torch 1.12.1, torchvision 0.19.1)
Codes available at	https://github.com/duola-wa/MICCAI-2025-STSR-Task-1

Inference Acceleration. To improve the inference speed of our model, we implemented two key optimizations.

Firstly, we simplified the inference pipeline by disabling certain computationally expensive operations. Specifically, we omitted the use of applying multiple augmentations (e.g., rotations, flips) during inference and averaging the results. While this opration can improve segmentation accuracy, it significantly increases inference time due to the additional computations required for each augmentation.

Secondly, we refined the interpolation step in handling multi-class predictions. Rather than using traditional, computationally expensive integer-based resampling techniques, we opted for a more efficient method utilizing floating-point tensors and the interpolate function from PyTorch. This approach maintains the precision of the predictions while enhancing throughput. Together, these optimizations allowed for faster and more efficient inference across the test set.

3.3 Results and Analysis

Quantitative Performance. The segmentation results are summarized in Table 2. The results indicate that the model with speed-up achieved slightly lower accuracy metrics across all evaluation criteria, such as Dice, mIoU, and NSD. However, the time required for inference in the speed-up version was significantly reduced, with the total time for inference dropping from 3257 seconds to 852 seconds.

Although there was a slight drop in segmentation performance, the substantial improvement in processing time makes the accelerated model particularly advantageous for clinical scenarios where rapid results are essential. In situations where time-sensitive decisions are needed, such as in real-time diagnostic

Table 2. Segmentation Results with and without Speed-Up

Model	Dice		mIoU		NSD		IA	Time (s)
	Instance	Image	Instance	Image	Instance	Image		
w/ Speed-Up	0.6691	0.9659	0.5535	0.9343	0.8655	0.9979	0.6910	852
w/o Speed-Up	0.6738	0.9671	0.5593	0.9365	0.8672	0.9980	0.6953	3257

systems, this speed-up could provide significant benefits. For the final test submission, we chose the accelerated version of the model, prioritizing efficiency while maintaining segmentation accuracy.

We further designed a two-stage segmentation framework in addition to the one-stage nnU-Net baseline. In the first stage, a tooth instance segmentation model was trained using external data, enabling simultaneous delineation of individual teeth and assignment of their FDI indices. This model was then applied to the 300 unlabeled cases to generate pseudo-labels. Unreliable labels were filtered out before using the remaining labels to retrain and enhance the tooth instance segmentation network. After obtaining a refined tooth-level segmentation model, we applied it to the labeled pulp dataset to extract accurate regions of interest (ROIs) for each tooth, thereby providing localized inputs for the second stage.

In the second stage, we introduced task-specific segmentation networks for different tooth categories. A binary segmentation model was trained to handle impacted teeth, while a six-class model was designed for other teeth, incorporating five root–pulp-related classes and one tooth-level class. This hierarchical design aimed to leverage tooth-level localization in the first stage to support finer anatomical segmentation in the second stage. Moreover, by decomposing each case into multiple tooth-level ROIs, the framework effectively transformed a single annotated scan into dozens of tooth-specific training samples. This substantially alleviated the data scarcity problem, as the original per-case annotations were expanded into a richer set of per-tooth instances for supervision. The quantitative performance of this two-stage framework is reported in Table 3, along with the results of the one-stage baseline model.

Table 3. Comparison of one-stage and two-stage segmentation models

Model	Dice		mIoU		NSD		IA
	Instance	Image	Instance	Image	Instance	Image	
One-stage nnU-Net	0.6738	0.9671	0.5593	0.9365	0.8672	0.9980	0.6953
Two-stage nnU-Net	0.626	0.9207	0.511	0.8682	0.8354	0.9753	0.5896

Although the two-stage segmentation framework was conceptually designed to leverage hierarchical localization and fine-grained classification, its perfor-

mance lagged behind the one-stage nnU-Net baseline. A key limitation lies in the dependency of the second stage on the quality of the first-stage predictions. If the tooth instance segmentation is inaccurate or fails to properly delineate certain teeth, the subsequent ROI extraction becomes unreliable, thereby compromising the downstream segmentation of roots and pulp canals. In contrast, the one-stage model avoids this error propagation by directly performing holistic multi-class segmentation. Nevertheless, we believe that with more robust tooth-level segmentation in the first stage, the two-stage framework has the potential to achieve competitive performance, as the modular design is inherently well-suited for capturing both global and localized anatomical structures.

Qualitative Results. As shown in Fig. 2, our model effectively segments teeth and root canals across various cases. For (a) normal cases, (b) cases with multiple missing teeth, and (c) cases with a small field of view, the model consistently produces accurate segmentation results.

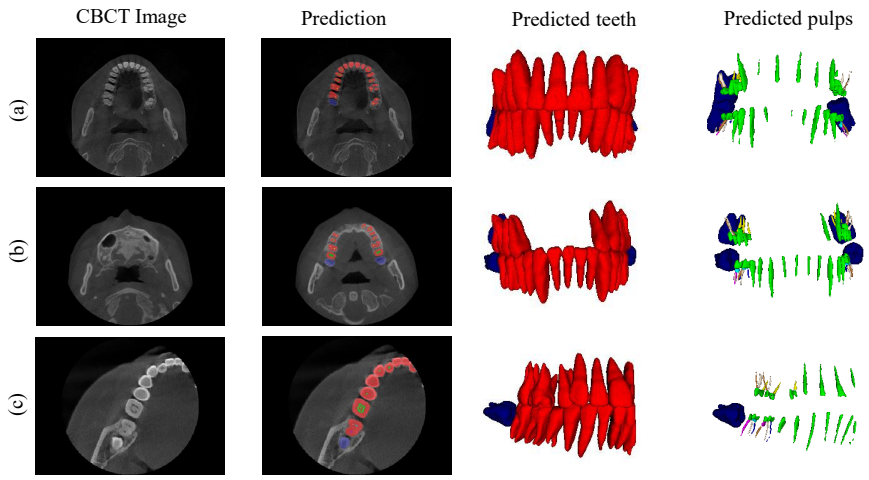

Fig. 2. Qualitative results of our model on different test cases: (a) normal case, (b) case with multiple missing teeth, and (c) case with a small field of view.

As shown in Fig. 3, we compare the segmentation results with and without inference acceleration. The results using the speed-up method still maintain high quality, demonstrating that the acceleration does not significantly compromise accuracy. The visual comparison highlights that, despite a slight drop in precision during testing, the accelerated model maintains satisfactory segmentation performance for both teeth and root canals. This suggests that the trade-off between speed and accuracy is favorable for real-time clinical applications.

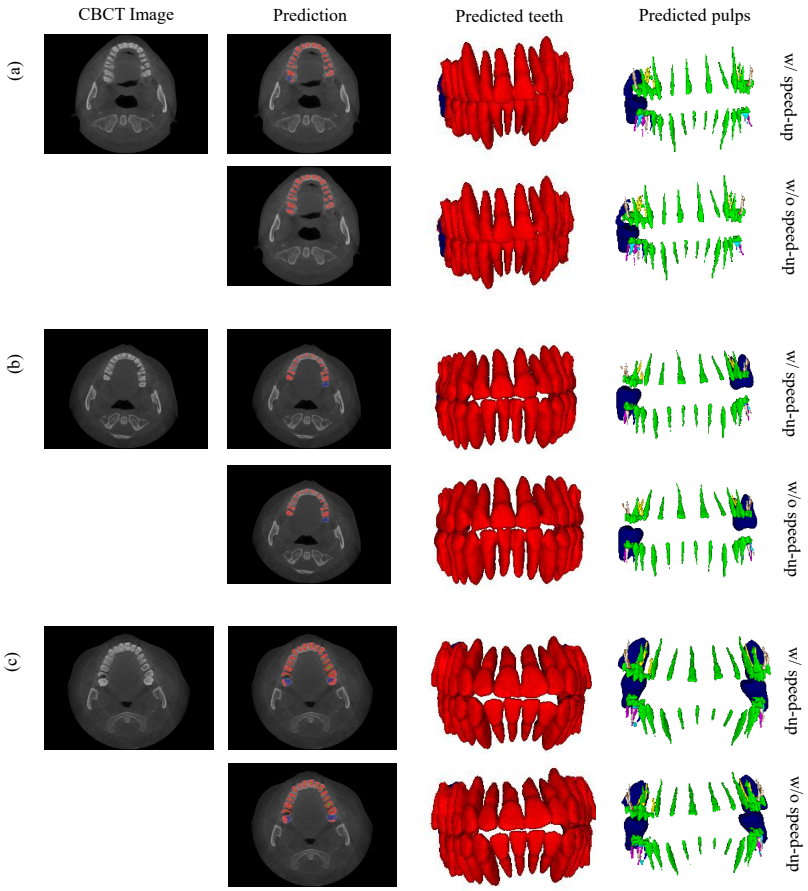

Fig. 3. Comparison of segmentation results with and without acceleration. Each case shows the segmentation performance before and after applying the inference speed-up.

4 Conclusion

In this paper, we presented an nnU-Net-based framework for the segmentation of teeth and root pulp canals from CBCT images. Our method integrates effective preprocessing, robust training using nnU-Net, and efficient post-processing to perform accurate multi-class segmentation. To address the challenge of high computational costs during inference, we designed acceleration strategies that simplified the inference pipeline and optimized interpolation operations. The proposed optimizations reduced inference time by nearly fourfold while maintaining comparable segmentation accuracy. Both quantitative and qualitative results confirm that the method provides reliable and efficient segmentation, making it suitable for clinical scenarios where rapid decision-making is crucial.

References

1. Bolelli, F., Lumetti, L., Vinayahalingam, S., Di Bartolomeo, M., Pellacani, A., Marchesini, K., Van Nistelrooij, N., Van Lierop, P., Xi, T., Liu, Y., et al.: Segmenting the inferior alveolar canal in cbcts volumes: the toothfairy challenge. IEEE Transactions on Medical Imaging (2024)
2. De Vos, W., Casselman, J., Swennen, G.: Cone-beam computerized tomography (cbct) imaging of the oral and maxillofacial region: a systematic review of the literature. International journal of oral and maxillofacial surgery **38**(6), 609–625 (2009)
3. Duan, W., Chen, Y., Zhang, Q., Lin, X., Yang, X.: Refined tooth and pulp segmentation using u-net in cbct image. Dentomaxillofacial Radiology **50**(6), 20200251 (2021)
4. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
5. Ji, C., Du, C., Zhang, Q., Wang, S., Ma, C., Xie, J., Zhou, Y., He, H., Shen, D.: Mammo-net: Integrating gaze supervision and interactive information in multi-view mammogram classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 68–78. Springer (2023)
6. Ji, C., Liu, Y., He, L., Jiang, Y., Huang, C., Wang, L.: Two-stage semi-supervised nnu-net framework for tooth segmentation in cbct images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 100–109. Springer (2024)
7. Ji, C., Liu, Y., He, L., Jiang, Y., Huang, C., Wang, L.: A two-stage semi-supervised nnu-net model for automated tooth segmentation in panoramic x-ray images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 91–99. Springer (2024)
8. Jiang, Y., Liu, Y., Ji, C., Wang, L.: Enhanced multi-structure segmentation in cbct images with adaptive structure optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 30–40. Springer (2024)
9. Jiang, Y., Wang, X., Yang, T., Ji, C., He, L., Liu, Y., Wang, W., Liu, M., Shi, J., Guo, H., et al.: Morphology prior enhanced teeth segmentation for high-resolution oral scans. IEEE Journal of Biomedical and Health Informatics (2025)
10. Lin, Z., Liu, Y., Wu, J., Wang, D.H., Zhang, X.Y., Zhu, S.: Multi-modal pre-post treatment consistency learning for automatic segmentation and evaluation of the circle of willis. Computerized Medical Imaging and Graphics **122**, 102521 (2025)
11. Liu, Y., Xin, R., Yang, T., Wang, L.: Inferior alveolar nerve segmentation in cbct images using connectivity-based selective re-training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12. Springer (2024)
12. Liu, Y., Zhang, S., Wu, X., Yang, T., Pei, Y., Guo, H., Jiang, Y., Feng, Z., Xiao, W., Wang, Y.P., et al.: Individual graph representation learning for pediatric tooth segmentation from dental cbct. IEEE Transactions on Medical Imaging (2024)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

14. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**(1), 221–248 (2017)
15. Wang, S., Ouyang, X., Liu, T., Wang, Q., Shen, D.: Follow my eye: using gaze to supervise computer-aided diagnosis. IEEE Transactions on Medical Imaging **41**(7), 1688–1698 (2022)
16. Wang, Y., Chen, X., Qian, D., Ye, F., Wang, S., Zhang, H.: Semi-supervised Tooth Segmentation: First MICCAI Challenge, SemiToothSeg 2023, Held in Conjunction with MICCAI 2023, Vancouver, BC, Canada, October 8, 2023, Proceedings, vol. 14623. Springer Nature (2024)
17. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., et al.: Miccai 2023 sts challenge: A retrospective study of semi-supervised approaches for teeth segmentation. Pattern Recognition **170**, 112049 (2026)
18. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Zhang, Q., Wu, C., et al.: Sts miccai 2023 challenge: grand challenge on 2d and 3d semi-supervised tooth segmentation. arXiv preprint arXiv:2407.13246 (2024)
19. Zhang, Y., Ye, F., Chen, L., Xu, F., Chen, X., Wu, H., Cao, M., Li, Y., Wang, Y., Huang, X.: Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. Scientific Data **10**(1), 380 (2023)
20. Zhong, T., Ning, Y., Wu, X., Ye, L., Li, C., Zhang, Y., Du, Y.: Tips: Tooth instance and pulp segmentation based on hierarchical extraction and fusion of anatomical priors from cone-beam ct. Artificial Intelligence in Medicine p. 103247 (2025)