Global-to-Local Tooth and Pulp Instance Segmentation in Dental CBCT

Fuyu Zhang[1][0009−0001−2606−8375] and Wei Liu[1][0009−0004−3971−2854] ⋆

School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, People's Republic of China
liuwei@xupt.edu.cn, leguanfyzz@outlook.com

Abstract. Accurate segmentation of teeth and their pulp structures in cone-beam computed tomography (CBCT) is clinically important for digital treatment planning, root canal therapy, and prosthetic restoration. However, manual segmentation is time-consuming and labor-intensive and depends heavily on the expertise of experienced clinicians, motivating the development of fully automatic methods. We propose an nnU-Net-based global-to-local framework for tooth and pulp instance segmentation in the MICCAI STSR 2026 Challenge Task 1. The provided annotations are processed to construct solid-tooth instance targets for global localization and nested tooth–pulp region targets for local segmentation. A low-resolution nnU-Net first predicts up to 32 tooth instances from the full CBCT volume and generates tooth-wise regions of interest. A full-resolution nnU-Net then jointly segments the tooth and pulp within each region. The local predictions are subsequently mapped back to the full-volume grid, where overlapping voxels are resolved deterministically using the corresponding global localization mask. The method achieved an official validation score of 0.7643 and ranked among the top three teams on the hidden test set. These results demonstrate the effectiveness of the proposed framework for fully automatic tooth and pulp instance segmentation in dental CBCT images. The source code is publicly available at https://github.com/LeguanFy/MICCAI-STSR-2026-Task1.

Keywords: Cone-beam computed tomography · Tooth and pulp segmentation · Instance segmentation · nnU-Net · Global-to-local segmentation

1 Introduction

Accurate delineation of individual teeth and their pulp structures in cone-beam computed tomography (CBCT) supports digital treatment planning, root canal therapy, and prosthetic restoration. Manual delineation, however, requires clinicians to inspect numerous cross-sectional slices and depends strongly on specialist expertise. The procedure is therefore time-consuming, labor-intensive, and difficult to integrate into time-sensitive workflows [11,14].

⋆ Corresponding author.

Fully automatic tooth–pulp instance segmentation remains challenging. Pulp structures are small and may contain thin, branching regions with weak boundaries [11,14]. Adjacent and contralateral teeth can appear similar, although the task requires both accurate delineation and correct instance identification [4,8]. Full CBCT volumes require global anatomical context, whereas pulp delineation requires fine local detail. This scale conflict limits methods operating at a single resolution [2,11].

nnU-Net automatically configures preprocessing, network architecture, and training for medical segmentation tasks [1]. Previous STS challenges have advanced tooth segmentation and identification in CBCT [10,7]. In STS 2024 Task 2, Ji et al. [2] used low-resolution quadrant localization before full-resolution tooth segmentation and reported sensitivity to spatially symmetric augmentation. Wang et al. [11] first localized tooth instances and then segmented pulp and root canals within tooth-wise regions of interest (ROIs). These studies motivate global-to-local processing. However, Task 1 additionally requires fixed tooth identities and consistent reconciliation of overlapping local predictions in the reconstructed volume.

We therefore propose an nnU-Net-based global-to-local framework for MICCAI STSR 2026 Task 1. The annotations are converted into solid-tooth global targets and nested tooth–pulp local targets. A low-resolution NoDA model predicts up to 32 fixed tooth identities and generates tooth-wise ROIs. A shared full-resolution model then segments the tooth and pulp within each ROI. Finally, localization-guided deterministic reconstruction maps the local predictions to the full-volume grid and resolves overlaps. Inference is fully automatic.

Our contributions are threefold:

- The first stage segments individual teeth from the full image, while the second stage segments the tooth and pulp within each single-tooth ROI.
- Second, the two-stage framework combines global NoDA training with full-resolution local segmentation.
- Third, deterministic reconstruction resolves overlapping ROI predictions, and the complete system achieved a top-three ranking in the challenge.

2 Method

2.1 Framework Overview

Given a full CBCT volume, the task is to predict an instance map containing background, tooth labels 1–32, and the corresponding pulp labels 33–64. As shown in Fig. 1, the proposed framework consists of two nnU-Net models [1] and three stages. First, a low-resolution nnU-Net predicts global tooth instances from the full CBCT volume. connected- component cleanup is applied to the prediction. Tooth-wise ROIs are then cropped from the original volume, while their spatial coordinates and tooth identities are retained. Second, a full-resolution nnU-Net segments the tooth and pulp regions within each ROI. Finally, the local predictions are mapped back to the original image space and combined using the global tooth identities to generate the final tooth–pulp instance map.

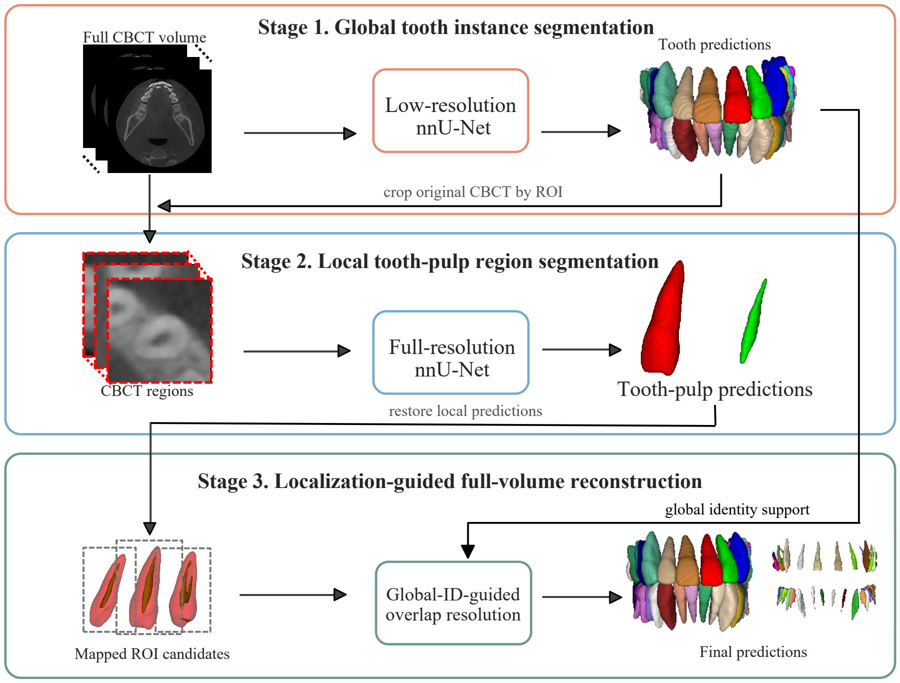

Fig. 1. Overview of the proposed two-model, three-stage global-to-local framework. A low-resolution nnU-Net predicts global tooth instances and constructs tooth-wise ROIs, a shared full-resolution nnU-Net segments the tooth and pulp regions within each ROI, and the mapped local predictions are reconstructed in the original CBCT space using the global tooth identities.

2.2 Task-Specific Target Construction

Each tooth is assigned an index $t \in \{1, \ldots, 32\}$, and its pulp is represented by $t + 32$ in the reconstructed volume. Complementary targets are constructed for the two models.

Global solid-tooth targets.

For each tooth index t, three-dimensional hole filling is applied to its binary mask within a bounding box enlarged by 3 voxels. Original and filled voxels receive the same index.At this stage, pulp labels are not considered, as the objective is to construct a solid representation of each tooth. The resulting solid-tooth targets provide complete identity support during reconstruction.

Local tooth–pulp targets.

For each tooth with a consistently matched pulp annotation, an ROI is extracted from the ground-truth tooth bounding box with a 10-voxel margin. Labels are remapped to background (0), tooth (1), and pulp (2). The region formulation uses Tooth = $\{1, 2\}$ and Pulp = $\{2\}$, preserving the nested anatomy. Training ROIs with fewer than 50 pulp voxels or failed annotation-consistency checks are excluded without redrawing retained masks. This quality control is confined to target construction and never selects inference cases.

2.3 Global Tooth Instance Segmentation and ROI Generation

The global model uses the low-resolution 3D nnU-Net configuration and predicts $B : \Omega \to \{0, \dots, 32\}$, where each nonzero label denotes a fixed tooth identity. Because spatial transformations may decouple anatomical position from tooth index, nnUNetTrainerNoDA is used to disable the default augmentation pipeline[2,4]. Other planning and preprocessing settings are determined by nnU-Net [1].

At inference, 26-connected components are analyzed separately for each tooth label. The largest component is retained, and an additional component is kept only when its minimum distance from the main component is at most 3 mm. The cleaned mask defines a bounding box enlarged by 10 voxels and clipped to the image boundary. Its tooth index, coordinates, and spatial metadata are recorded for reconstruction; the resulting ROIs may overlap.

2.4 Local Tooth–Pulp Region Segmentation

The local model uses the full-resolution 3D nnU-Net configuration with default augmentation. One model is shared across tooth positions and predicts $C_i \in \{0, 1, 2\}$ for ROI i. It jointly learns the complete tooth and nested pulp regions, while tooth identity is inherited from the global ROI record rather than re-estimated locally.

Every automatically generated ROI is processed, and its raw prediction is retained for reconstruction. Geometry, label range, connected components, boundary contact, and volume are recorded for quality control.

During training, tooth-wise ROIs were derived from the annotations of the 30 labeled volumes; during validation and test inference, they were generated automatically from the global predictions.

2.5 Localization-Guided Full-Volume Reconstruction

For each ROI, the local prediction is restored to its original coordinates in the CBCT volume. The predicted tooth and pulp regions inherit the tooth identity t_i associated with the ROI and are relabeled as t_i and $t_i + 32$, respectively. Because tooth-wise ROIs may overlap, they are processed in ascending order of tooth identity. A candidate is written directly to an unassigned voxel. At an occupied voxel, a candidate supported by the global tooth identity map replaces an unsupported candidate, whereas equal-priority candidates retain the existing assignment. Consequently, ties are resolved deterministically by the processing order, and each candidate-covered voxel is assigned to a single tooth identity.

3 Experiments

3.1 Dataset and evaluation metrics

The official STSR 2026 Task 1 dataset contains 30 labeled and 299 unlabeled CBCT images for training, together with 20 labeled CBCT images for validation. The 299 unlabeled volumes were not used in this study. The evaluation metrics include Dice Similarity Coefficient (DSC), Normalized Surface Distance (NSD), Mean Intersection over Union (mIoU), and Identification Accuracy (IA) to evaluate the segmentation region overlap and boundary distance.

3.2 Implementation details

Preprocessing The labeled training set was converted into the solid-tooth global targets and 10-voxel-margin tooth–pulp ROIs described in Section 2.2.

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

System	Ubuntu 22.04.5 LTS
CPU	Intel Xeon Platinum 8470Q @ 2.10 GHz
RAM	90 GB
GPU (number and type)	1 $\times$ NVIDIA GeForce RTX 4090 (24 GB)
CUDA version	12.8
Programming language	Python 3.10.20
Deep learning framework	torch 2.5.1+cu121, torchvision 0.20.1+cu121

Training protocols Both models were trained from scratch for 500 epochs using the complete labeled training set (fold=all). The global model used low-resolution NoDA training for full-volume tooth identification, whereas the local model used full-resolution training with default nnU-Net augmentation. Other settings are listed in Tables 2 and 3.

Table 2. Training protocols for Low-resolution global tooth segmentation model.

Pre-trained Model	None
Batch size	2
Patch size	128×96×160
Total epochs	500
Optimizer	SGD with Nesterov momentum
Initial learning rate (lr)	0.01
Lr decay schedule	Polynomial decay, power = 0.9
Training time	4.47 hours
Loss function	Dice loss + Cross-Entropy loss with deep supervision
Number of model parameters	30.81M[1]
Number of flops	451.36G[2]

Table 3. Training protocols for Full-resolution local tooth–pulp region segmentation model.

Pre-trained Model	None
Batch size	18
Patch size	56×88×48
Total epochs	500
Optimizer	SGD with Nesterov momentum
Initial learning rate (lr)	0.01
Lr decay schedule	Polynomial decay, power = 0.9
Training time	3.68 hours
Number of model parameters	5.60 M[3]
Number of flops	50.72 G[4]

Inference Settings Inference was fully automatic with nnU-Net's default test-time augmentation. Each CBCT image was processed independently, and global probability maps were resampled on the GPU in channel-wise chunks to limit memory and resampling overhead. Local ROI predictions were then reconstructed on the original grid as described in Section 2.

4 Results and discussion

4.1 Quantitative results on validation set

Table 4. Quantitative comparison on the official validation set of the **coreset track**. Identity reassignment modifies only automatic post-processing.

Method	Level	DSC↑	NSD↑	mIoU↑	IA↑	Score↑
Ours	Image	0.9672	0.9964	0.9369	0.6391	**0.7643**
	Instance	**0.7330**	**0.7804**	**0.6870**		
+ identity reassignment	Image	0.9672	0.9964	0.9369	**0.6392**	0.7635
	Instance	0.7321	0.7792	0.6862		

Table 4 reports the official validation results. The proposed system achieved an overall score of 0.7643. Its image-level DSC, NSD, and mIoU were 0.9672, 0.9964, and 0.9369, respectively. Lower instance-level scores and an IA of 0.6391 indicate that the remaining errors were concentrated in tooth-specific delineation and identity assignment.

Automatic identity reassignment left the image-level metrics unchanged because it modified identities without changing foreground geometry. Although IA increased from 0.6391 to 0.6392, all three instance-level metrics decreased and the overall score fell to 0.7635. We therefore retained the simpler deterministic reconstruction in the final system.

4.2 Qualitative results on validation set

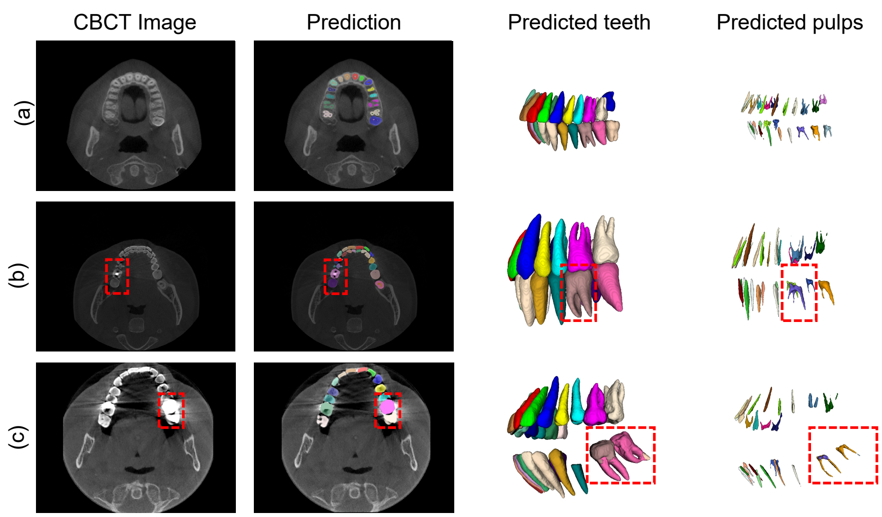

Fig. 2. Qualitative examples from the validation set: (a) a typical case, (b) a case with mild-to-moderate metal artifacts, and (c) a case with severe metal artifacts. Columns show the CBCT image, prediction overlay, tooth instances, and pulp instances; red dashed boxes mark artifact-affected regions.

Fig. 2 presents representative qualitative results under different levels of metal-induced artifacts. In the successful case shown in Fig. 2(a), the proposed method reconstructs complete and anatomically coherent tooth and pulp structures. In Fig. 2(b), visually mild-to-moderate localized artifacts are present within the highlighted region, while the tooth-instance labels and pulp predictions remain spatially coherent, with no evident artifact-related degradation in this case. By contrast, the severe streak artifacts in Fig. 2(c) obscure the boundaries between adjacent teeth and cause neighboring-tooth leakage and tooth-instance identity misassignment during global segmentation. Since ROI construction and final identity mapping depend on the global prediction, this upstream error propagates through the cascade: the resulting ROI contains structures from adjacent teeth, and two pulp components are ultimately assigned to a single tooth identity [3]. These qualitative observations suggest that the proposed method tolerates limited artifact-related degradation but remains sensitive to severe metal-induced streak artifacts.

The qualitative examples showed that mild-to-moderate metal artifacts did not visibly affect the segmentation results. Under severe artifacts, however, the global tooth segmentation model produced an incorrect tooth localization and identity assignment, resulting in a spatially displaced ROI. Because the local tooth–pulp segmentation model operates on the predicted ROI and inherits the tooth identity from the global stage, this upstream error propagated through the cascade to the final reconstruction. Consequently, two distinct pulp regions were assigned to a single tooth instance, and the corresponding tooth label was spatially misassigned. These qualitative observations suggest that global tooth localization and identity assignment remain a bottleneck under severe metal artifacts.

4.3 Results on final testing set

The proposed system ranked third on the hidden final test set of MICCAI STSR 2026 Task 1 under the official evaluation protocol.

4.4 Limitation and future work

The main observed limitation is sensitivity to severe metal artifacts. Metal restorations can introduce streak artifacts that distort dental structures in CBCT images and complicate tooth identification [14]. Our qualitative analysis suggests that, under severe artifacts, errors primarily arise during global tooth localization and identity assignment and subsequently propagate through ROI construction to local segmentation. Future work will therefore prioritize metal artifact reduction and artifact-robust feature learning for the global tooth segmentation model. After improving the global predictions, stage-wise evaluation stratified by artifact severity will determine whether residual errors also require further optimization of the local tooth–pulp segmentation model.

A second direction is to strengthen global spatial modeling through an arch-level representation. Previous methods have exploited inter-tooth topology or dental-arch localization to provide spatial context for tooth instance segmentation [4,13]. We will further explore whether a binary whole-dentition target, followed by explicit tooth localization and identity assignment, provides a more stable global representation than direct multi-class tooth-instance prediction. This hypothesis remains to be systematically evaluated.

5 Conclusion

This paper presented a fully automatic global-to-local framework for tooth–pulp instance segmentation in dental CBCT. By combining global tooth identification, high-resolution tooth-wise segmentation, and deterministic full-volume reconstruction, the framework preserves global tooth-identity information while restricting full-resolution processing to compact tooth-wise ROIs. It achieved an overall validation score of 0.7643 and ranked third on the official final test set for Task 1 of the MICCAI STSR 2026 Challenge. These results support the effectiveness of the proposed framework under the official challenge protocol.

Acknowledgements This work was developed and evaluated within the ODIN 2026 challenge cluster, including the STS, ToothFairy, and Bite2Text challenges [9,5,6,8]. We thank all the data owners for making the medical images publicly available and Codabench [12] for hosting the challenge platform.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

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, 4
2. 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: Wang, Y., Qian, D., Wang, S., Ben-Hamadou, A., Pujades, S., Lumetti, L., Grana, C., Bolelli, F. (eds.) Supervised and Semi-supervised Multi-structure Segmentation and Landmark Detection in Dental Data. pp. 100–109. Springer Nature Switzerland, Cham (2025) 2, 4
3. Ji, C., Liu, Y., Jiang, Y., Wang, L.: Efficient nnu-net for tooth and root canal segmentation in cbct. In: International Workshop on Oral and Dental Image Analysis. pp. 171–180. Springer (2025) 8
4. 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). https://doi.org/10.1109/TMI.2022.3179128 2, 4, 8
5. Lumetti, L., Di Bartolomeo, M., Pellacani, A., Anesi, A., Grana, C., Bolelli, F., et al.: Ontology-grounded structured prediction for dental cbct reporting. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2026 (2026) 9
6. Lumetti, L., Rizzo, F., Cremonini, F., Candeloro, E., Luca, L., Grana, C., Bolelli, F., et al.: Do multimodal llms understand intraoral dental data? dataset, platform, and baselines. In: Computer Vision–ECCV 2026 (2026) 9
7. 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. Springer Nature (2024) 2
8. Wang, Y., Li, Z., Wu, C., Liu, J., Zhang, Y., Ni, J., Luo, Q., Chen, J., Zhang, H., Liu, J., Han, C., Fu, K., Ji, C., Cai, X., Hao, J., Zheng, Z., Xu, S., Chen, J., Yu, X., Zhang, Q., Qian, D., Wang, S., Zhou, H.: Miccai sts 2024 challenge: Semi-supervised instance-level tooth segmentation in panoramic x-ray and cbct images. Medical Image Analysis **110**, 103986 (2026). https://doi.org/https://doi.org/10.1016/j.media.2026.103986 2, 9
9. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., Huang, X., Huang, Z., Liu, J., Zheng, Z., Wu, C., Li, Y., Li, Z., Ma, Z., Cui, W., Luo, S., Wang, C., Chen, Y., Li, T., Liu, Y., Feng, X., Ni, J., Liu, D., Wang, Q., Lin, Z., Song, W., Li, Y., Wang, B., Wang, C., Chen, Q., Li, M., Zhou, H., Jin, Q.: Miccai 2023 sts challenge: A retrospective study of semi-supervised approaches for teeth segmentation. Pattern Recognition **170**, 112049 (2026). https://doi.org/https://doi.org/10.1016/j.patcog.2025.112049, https://www.sciencedirect.com/science/article/pii/S0031320325007095 9
10. 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) 2
11. Wang, Y., Xia, W., Yan, Z., Zhao, L., Bian, X., Liu, C., Qi, Z., Zhang, S., Tang, Z.: Root canal treatment planning by automatic tooth and root canal segmentation in dental cbct with deep multi-task feature learning. Medical Image Analysis **85**, 102750 (2023). https://doi.org/https://doi.org/10.1016/j.media.2023.102750 1, 2

12. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022) 9
13. Zhang, Y., Zhang, M., Ye, J., Zhou, X., Qin, Z., Tong, S., Deng, S., Wang, Y.: Tapseg: An open-source deep learning tool for instance-level tooth and pulp segmentation in cbct. Journal of Dentistry **171**, 106643 (2026). https://doi.org/https://doi.org/10.1016/j.jdent.2026.106643 8
14. 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 **169**, 103247 (2025). https://doi.org/https://doi.org/10.1016/j.artmed.2025.103247 1, 2, 8