CBCT Tooth and Pulp Segmentation via Watershed Fusion and Lightweight Dual-Branch nnU-Net

Rusheng Xu[1], **Zhou Yang**[1], **Ting Yan**[2], **Bin Wang**[1,*]

[1]College of Computer Science and Technology, Taiyuan University of Technology, China

[2]Department of Pathology, Shanxi Medical University, China

1. Background & Key Contributions

CBCT Tooth and Pulp Segmentation Challenges:

Cone-beam computed tomography provides critical 3D structural detail for dental procedures, but automated processing faces three primary hurdles:

- **Touching Tooth Instances:** Narrow interdental gaps and low local contrast cause frequent instance merging.
- **Unclear Pulp Boundaries:** Fine root canal morphology requires labor-intensive annotation and high spatial resolution.
- **High 3D Computational Burden:** Processing full-resolution 3D volumes imposes severe GPU memory limits.

Main Contributions:

1. **Watershed Fusion:** A spatial topology marker-controlled watershed algorithm cleanly splits touching instances with zero extra network parameters.
2. **Lightweight Semantic Branch:** Resampling input spacing to 1.0 mm expands the receptive field for global FDI numbering while lowering GPU memory requirements.
3. **Optimized Inference Pipeline:** Parallel branch execution and unified spatial resampling shorten per-volume execution time.
4. **Semi-Supervised Learning:** Dual-level anatomical bounding-box filtering leverages unannotated volumes effectively.

2. Dual-Branch nnU-Net Architecture

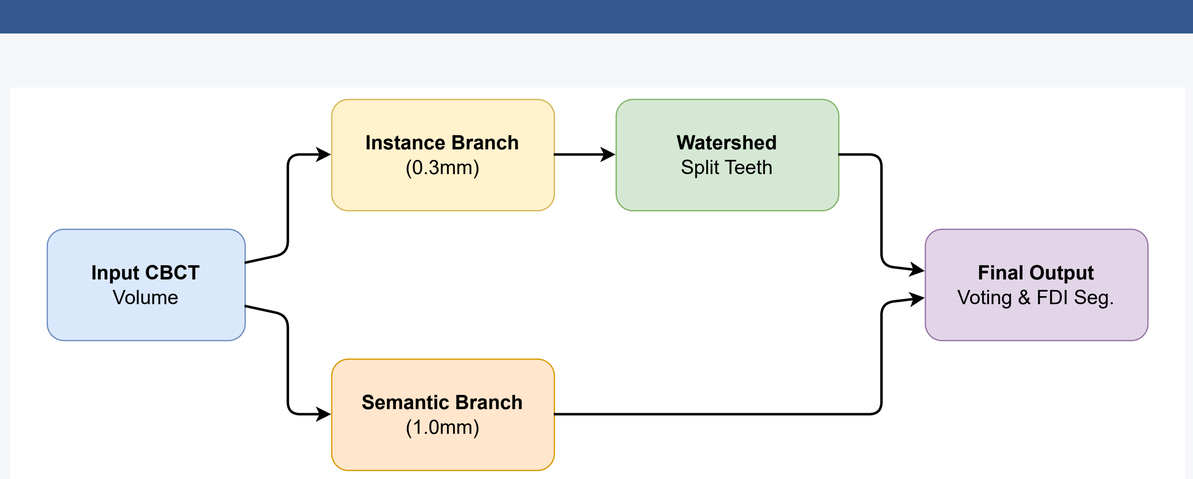

Figure 1: Overview of the proposed lightweight dual-branch nnU-Net framework. The high-resolution instance branch extracts fine structures, while the coarse-resolution semantic branch captures global FDI context, fused via watershed separation and majority voting.

- **Instance Branch:** Operates at native resolution to predict background, tooth, and pulp region masks, supplying spatial markers for watershed separation.
- **Semantic Branch:** Operates at 1.0 mm voxel spacing to capture whole-arch symmetry and jaw layout, with mirror augmentation disabled during training to prevent label confusion.

3. Semi-Supervised Training with Anatomical Filtering

To leverage unannotated data when manual annotations are limited to 30 fully annotated volumes, a teacher model generates pseudo-labels for 299 unannotated scans. These scans undergo a rigorous two-level bounding-box filtering strategy:

- **Single-Structure Constraint:** Individual tooth or pulp components are retained only if their 3D bounding box dimensions satisfy $\max(D, H, W) \leq 100$ voxels.
- **Arch-Block Constraint:** Contiguous 8-tooth maxilla or mandible blocks are retained only if $W_{\text{block}} \leq 200$ voxels and $D_{\text{block}} \leq 150$ voxels.
- **Final Training Dataset:** Exactly 247 unannotated volumes passed filtering and were merged with the 30 annotated scans for final training.

4. Watershed Separation & Decision Fusion

Marker-Controlled Watershed:

- Computes Euclidean distance transform on tooth probability masks.
- Extracts local intensity peaks as tooth center markers.
- Grows regions outward to physically split touching tooth boundaries cleanly.

Instance-Semantic Fusion:

- Separated instance masks act as localized spatial boundaries.
- Aggregates semantic predictions independently within each instance boundary.
- Applies majority voting to assign precise FDI labels to individual teeth.

5. Quantitative Results & Ablation Study

MICCAI STSR2026 Challenge Task 1 Achievement:

Secured **2nd Place** on the official hidden test set, achieving a final score of **0.7308** with $DSC_{\text{img}} = 0.8495$, $DSC_{\text{inst}} = 0.7122$, and IA $= 0.5976$.

Ablation Study on Validation Set:

Method	Inf.	Opt.	Par.	Light.	$DSC_{\text{img}} \uparrow$	$DSC_{\text{inst}} \uparrow$	VRAM (GB)$\downarrow$	Time (s)$\downarrow$
Baseline (0.3mm)					0.9619	0.7532	40.01	72.28
+ Inf. Opt.	✓				0.9609	0.7527	40.01	26.27
Ours (Full Model)	✓		✓	✓	**0.9609**	**0.7525**	**10.84**	**12.58**

- **GPU Memory Reduction:** Peak VRAM dropped from 40.01 GB to 10.84 **GB**, enabling execution on standard consumer GPUs.
- **Inference Speedup:** Per-volume inference execution time was accelerated from 72.28 s down to 12.58 **s**.

6. Qualitative Segmentation Results

Axial Slice View | 3D Tooth Surface (FDI) | Pulp Cavity View

Figure 2: Qualitative outputs on Validation Case 018. Panel A shows marker-controlled watershed cleanly separating adjacent touching molars. Panel B demonstrates precise FDI tooth identification across the dental arch. Panel C presents high-fidelity 3D reconstruction of pulp cavities and root canals.

7. Conclusion & Future Work

- **Conclusion:** We proposed a lightweight dual-branch nnU-Net framework with watershed fusion that balances high-precision tooth and pulp instance segmentation with real-time inference efficiency.
- **Limitations & Future Work:** Metal restoration artifacts can alter CT intensity values and distort watershed markers. Future work will explore artifact-resistant topological priors to handle restored teeth.