



STS 2026 CHALLENGE

Semi-Supervised CBCT Tooth and Root Canal Segmentation with Tri-Directional Mamba and a High-Confidence Foreground Mean Teacher

**Liangyu Chen[†], Yetong Shen[†], Dongping Zhang*, Hui Lin, Yawen Zhu,
Xiameng Qiu**

China Jiliang University · Sir Run Run Shaw Hospital, Zhejiang University · Zhejiang Wanli University

[†]Equal contribution *Corresponding author



1. Background & Motivation

CBCT dental segmentation, limited annotations, and the STSR 2026 Task 1 setting

2. Method

Tri-Directional Mamba Bottleneck + High-Confidence Foreground Mean Teacher + LMG-BDice loss

3. Experimental Setup

Dataset802, protocol, implementation details and metrics

4. Results & Analysis

Controlled ablations, official validation results, qualitative examples

5. Conclusion

Key numbers, limitations and future work

Background & Motivation



- **CBCT guides dental diagnosis and treatment**

3D depiction of teeth, roots and the pulp-root canal system for endodontic treatment, implant planning and orthodontic analysis.

- **Annotation is expensive, labels are scarce**

STSR 2026 Task 1 (Dataset802): only 30 labeled volumes vs. 299 unlabeled; metallic restorations cause artifacts that blur tooth / canal boundaries.

- **64 foreground labels - small, thin, imbalanced**

Each tooth instance and root canal is a separate class; thin canals and adjacent teeth create ambiguous boundaries and extreme class imbalance.

- **Semi-supervised learning is a natural fit**

Exploit the 299 unlabeled scans, but naive pseudo-labeling propagates errors into training.



Axial CBCT slice (labeled)



Ground-truth 3D instances (teeth + canals)

Three Gaps, Three Components



MICCAI2026

Limited long-range modeling

Convolutions are local; Mamba is linear-time but a fixed flattening order introduces directional bias.

① TDMB

Tri-directional Mamba bottleneck: three axis orders (D-H-W / H-W-D / W-D-H) modeled in parallel and fused.

Pseudo-label noise in Mean Teacher

Dominant background, class imbalance and confidence bias let erroneous pseudo-labels propagate.

② HCF-MT

High-Confidence Foreground MT: foreground-only, class-adaptive thresholds, patch-quality gating, class-balanced weights.

Ambiguous thin-structure boundaries

Boundary loss improves surfaces but does not separate thin canals from adjacent teeth.

③ LMG-BDice

Label-wise morphological-gradient boundary Dice: probability constraints inside GT boundary bands.



杭州电子科技大学
HANGZHOU DIANZI UNIVERSITY



上海交通大学
Shanghai Jiao Tong University

杭州钱塘牙科医院
HANGZHOU QIANTANG DENTAL HOSPITAL

杭州牙科医院
HANGZHOU DENTAL HOSPITAL



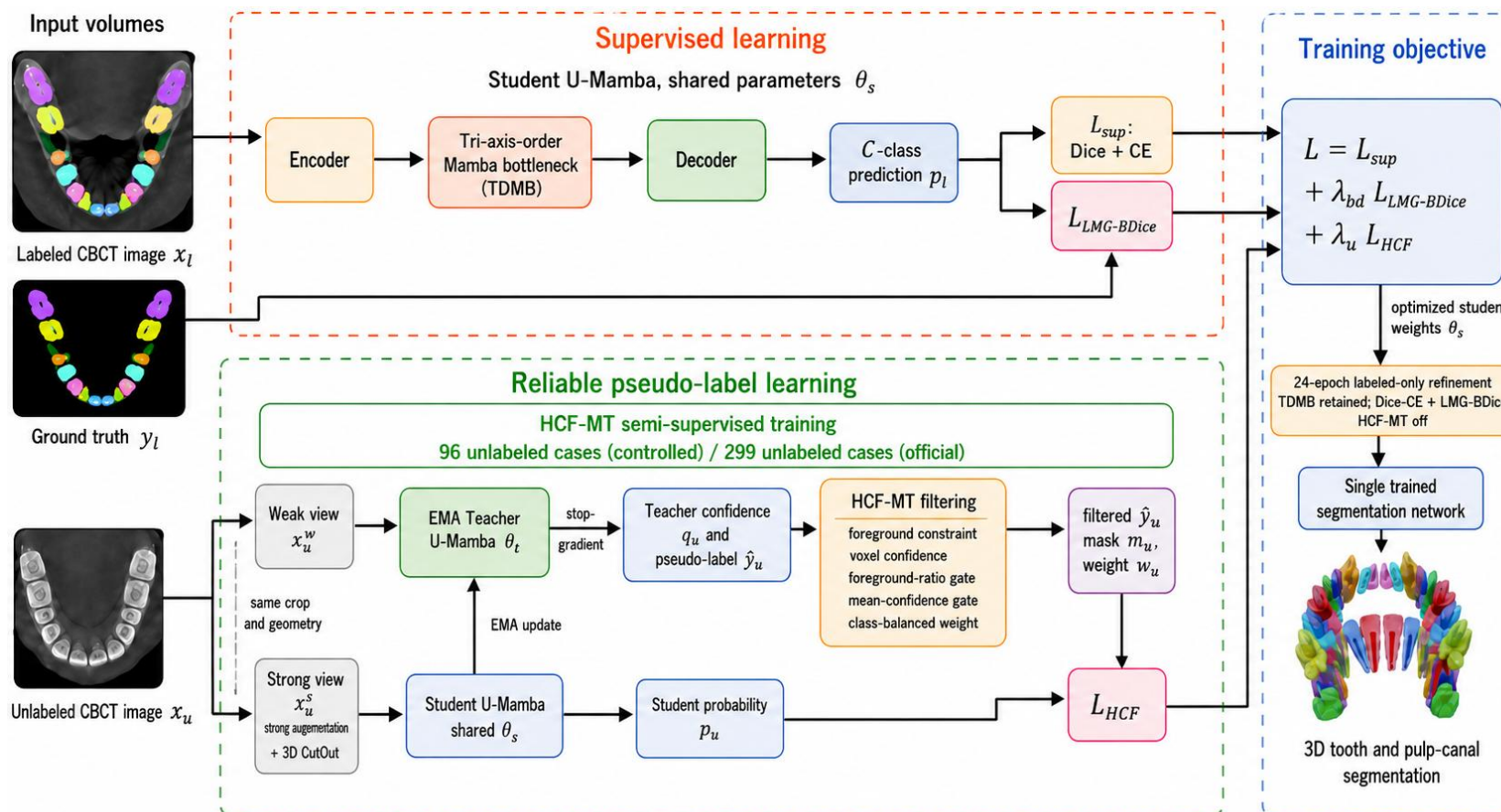
DataCanvas
九章云极

Method Overview

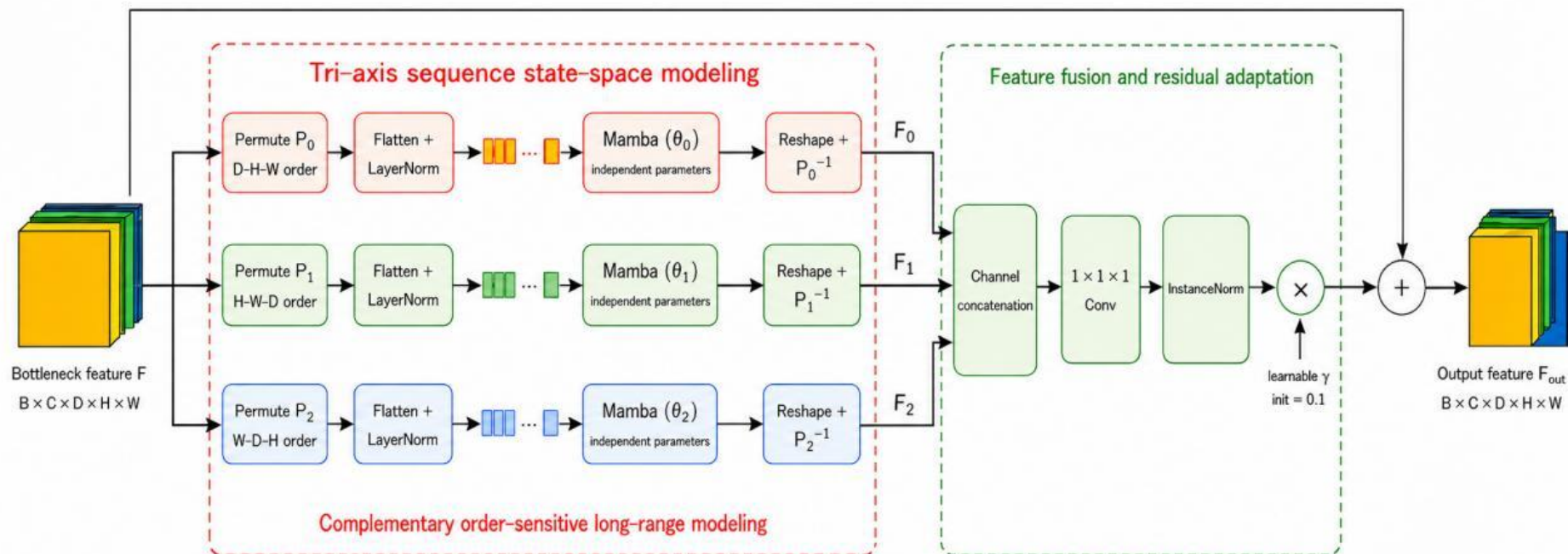


$$L_{\text{Dice-CE}} + \lambda_{bd}(k) \cdot L_{\text{LMG-BDice}} + \lambda_u(k) \cdot L_{\text{HCF}}$$

λ_{bd} : 0 $\rightarrow$ 0.08 linear over 80 epochs · λ_u : 0 $\rightarrow$ 0.25 exponential over 160 epochs · teacher EMA $\alpha = 0.9995$ · single-network inference



Component ① — Tri-Directional Mamba Bottleneck (TDMB)



$$F_{\text{out}} = F + \gamma \cdot \text{IN}(\text{Conv}_{1 \times 1 \times 1}(\text{Concat}(F_0, F_1, F_2))), \quad \gamma \text{ init} = 0.1$$

Three complementary serializations

D-H-W, H-W-D, W-D-H orders; independent permutation, LayerNorm and Mamba parameters per branch.

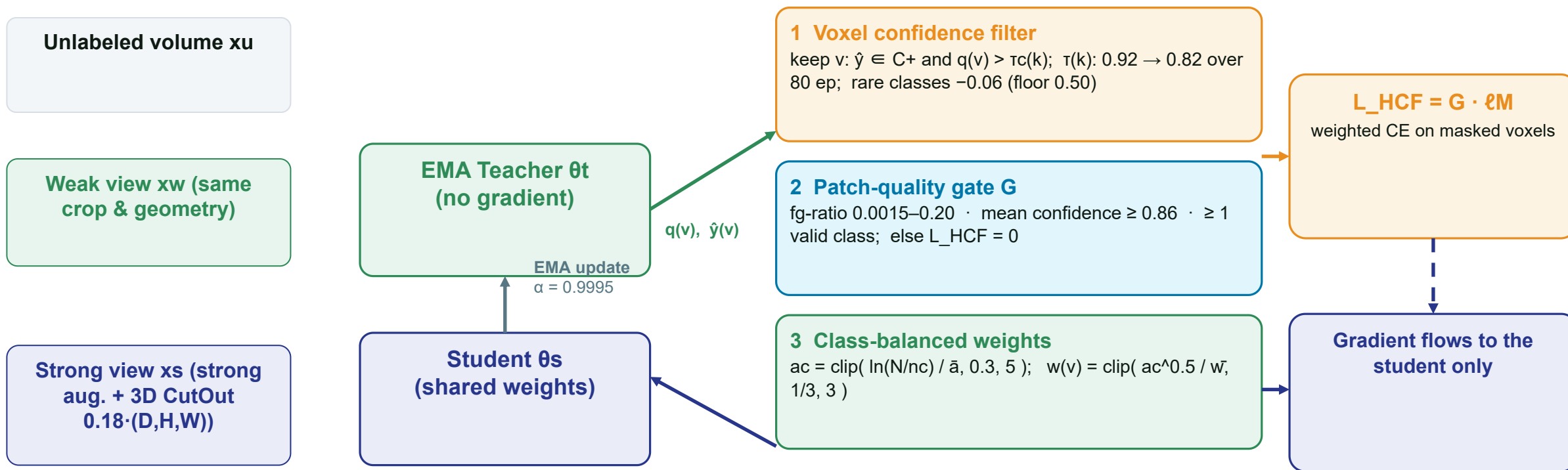
Channel fusion

Concatenate the three outputs, fuse with $1 \times 1 \times 1$ conv + InstanceNorm.

Residual calibration & efficiency

Learnable γ (init 0.1) scales the residual; placed only at the low-resolution bottleneck to bound sequence length and GPU memory.

Component ② — High-Confidence Foreground Mean Teacher (HCF-MT)



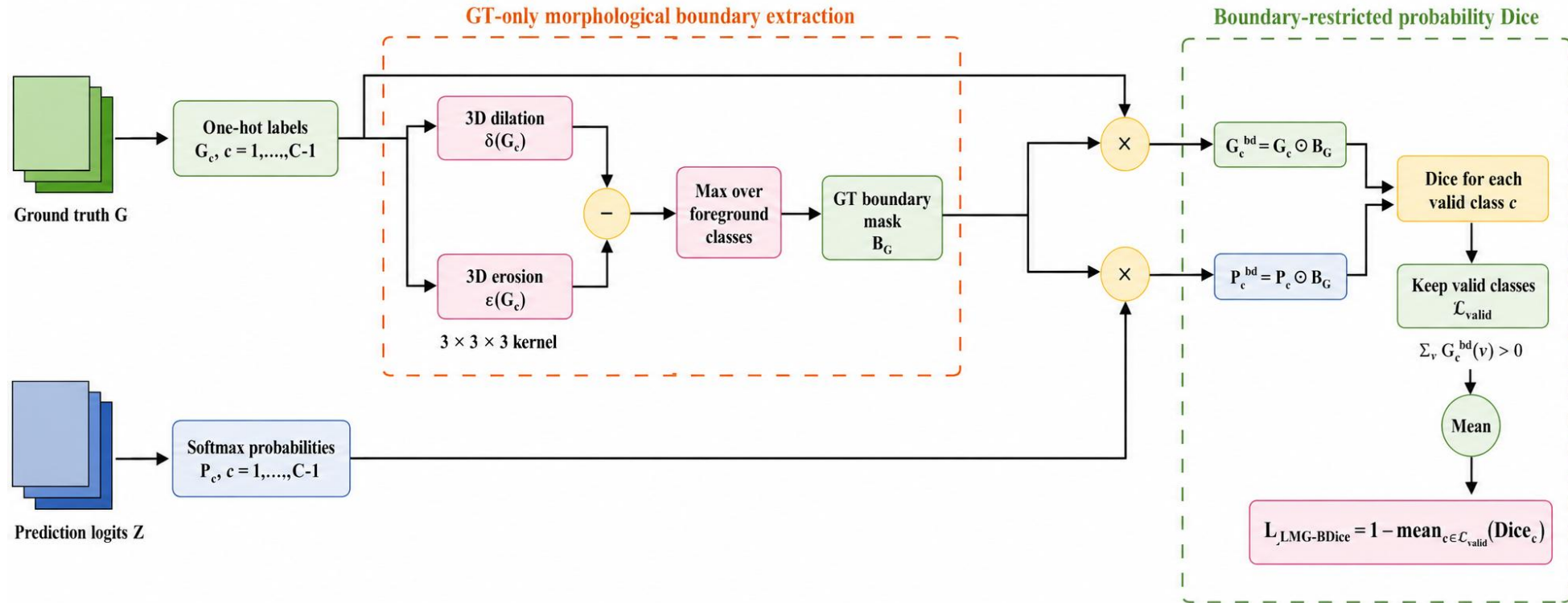
$$\ell_M = \sum w(v) \cdot \text{CE}(z_s(v), \hat{y}(v)) / \sum w(v) \quad \text{— computed only over voxels passing all three gates}$$

p_u (unlabeled-batch activation): 0.25 $\rightarrow$ 0.50 over 80 ep · temperature $T = 1.0$ · $\rho = 0.5$ · $\kappa = 3$ · min mean confidence 0.86

Component ③ — Label-wise Morphological-Gradient Boundary Dice (LMG-BDice)



MICCAI2026



$$L_{\text{LMG-BDice}} = 1 - \text{mean}_{c \in \mathcal{C}_{\text{valid}}} \text{Dice}(P_c \times B_G, G_c \times B_G)$$

$$B_G = \max_c \text{clip}(\delta(G_c) - \epsilon(G_c), 0, 1) \quad \text{— } 3 \times 3 \times 3 \text{ morphological gradient, GT only}$$

- GT-only boundary band: no reliance on predicted surfaces
- Probability Dice restricted to the boundary region
- Strengthens thin canals and adjacent-tooth interfaces

Experimental Setup



Dataset & Splits

- STSR 2026 Dataset802: 30 labeled / 299 unlabeled / 20 hidden validation
- Foreground labels 1–64 (per-tooth instances + root canals)
- Ablations: fold-0 $\rightarrow$ 24 train / 6 internal validation; 96 unlabeled subset (seed 20260522)
- Official: 24 labeled + all 299 unlabeled; hidden cases never used for tuning

Preprocessing

- nnU-Net 3D full-resolution pipeline
- Resample to $0.25 \times 0.25 \times 0.25 \text{ mm}^3$ (labels: discrete interpolation)
- Intensity clip at foreground p0.5–p99.5 (144–3944), then z-score ($\mu=1748.60$, $\sigma=762.51$)
- Patch $112 \times 128 \times 128$; rotation / scaling / intensity augmentation

Training

- 3 $\times$ RTX 4090 D (24 GB), PyTorch 2.0.1, CUDA 11.8, nnU-Net v2.1.1
- SGD, Nesterov 0.99, lr 0.01, poly decay, wd $3e-5$; global batch 6
- ≤ 420 epochs, EMA validation, patience 45; teacher EMA 0.9995
- Official: semi-supervised training + 24 labeled-only refinement epochs (HCF-MT off)

Metrics & Protocol

- Internal: case-label macro Dice / IoU over all 64 foreground labels
- NaN case-class pairs omitted when absent on both sides
- Official: image- & instance-level Dice, mIoU, NSD, identification accuracy (IA), overall score
- Single checkpoint, no mirroring, no ensembling, no extra post-processing

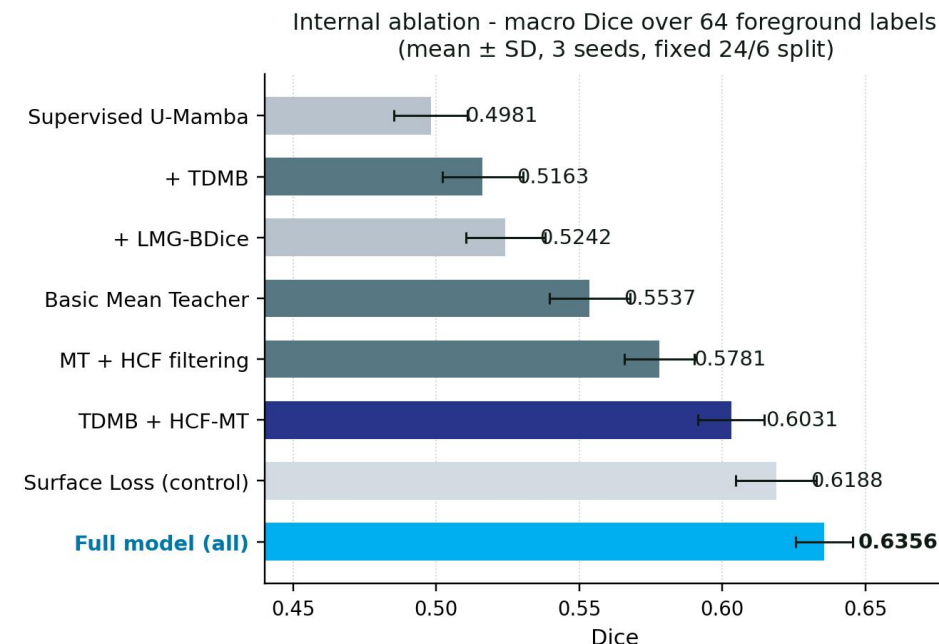


Controlled Ablation — Every Component Counts



MICCAI2026

Configuration (Dataset802, labels 1–64)	Dice (mean $\pm$ SD)	IoU
Supervised U-Mamba	0.4981 $\pm$ 0.0129	0.3957
+ TDMB	0.5163 $\pm$ 0.0142	0.4113
Basic Mean Teacher	0.5537 $\pm$ 0.0140	0.4451
MT + HCF filtering	0.5781 $\pm$ 0.0122	0.4683
Baseline + LMG-BDice	0.5242 $\pm$ 0.0138	0.4171
TDMB + HCF-MT	0.6031 $\pm$ 0.0116	0.4928
Full model (TDMB + MT + HCF + LMG)	0.6356 $\pm$ 0.0099	0.5319
Surface-Loss control (TDMB + MT + HCF + Surface)	0.6188 $\pm$ 0.0140	0.5111



- +13.75 Dice points (27.6% relative) over supervised U-Mamba
- HCF filtering adds +2.4 pts over basic Mean Teacher
- LMG-BDice beats Surface-Loss control by +1.7 pts
- 3 random seeds, fixed 24/6 patient-level split

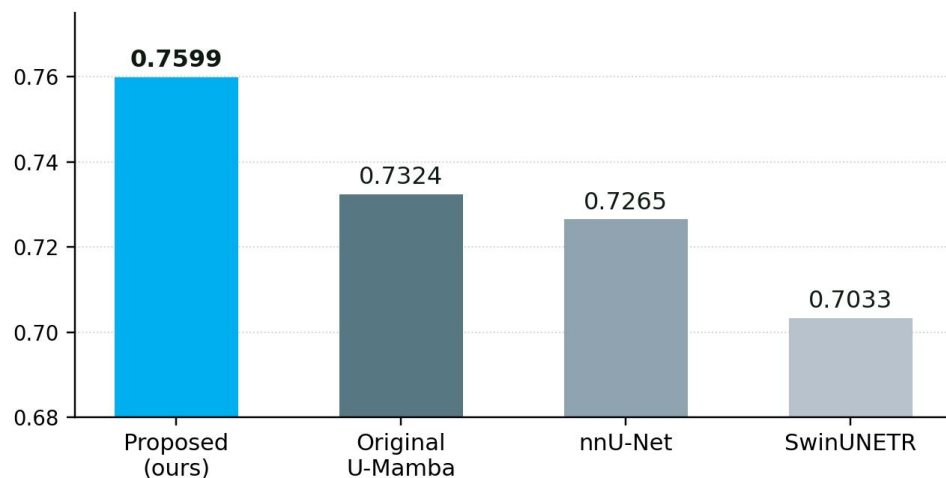
Official Validation Results (STSR 2026 Task 1)



Submission	Score	DiceImg	mIoUImg	NSDIImg	DiceInst	mIoUInst	NSDInst	IA
Proposed method (ID 851931)	0.7599	0.9616	0.9263	0.9955	0.7135	0.6452	0.8217	0.6098
Original U-Mamba	0.7324	0.9603	0.9210	0.9952	0.7001	0.6313	0.7907	0.6002
nnU-Net	0.7265	0.9585	0.9128	0.9949	0.6832	0.6127	0.7798	0.6009
SwinUNETR	0.7033	0.9450	0.8979	0.9896	0.6529	0.5745	0.7495	0.5630

Values quoted from official platform records; historical submissions serve as context (full configurations unavailable).

Official STSR 2026 validation - overall score



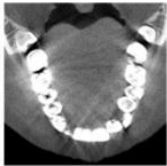
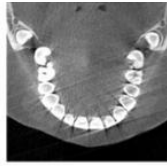
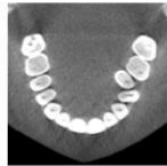
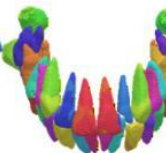
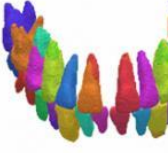
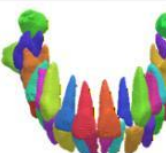
Reading the numbers

- Overall score 0.7599 — best among listed references (+2.75 pts vs original U-Mamba)
- Image-level Dice 0.9616 and NSD 0.9955: arch-level anatomy essentially solved
- Gains concentrate at instance level: DiceInst 0.7135, NSDInst 0.8217, IA 0.6098 — exactly where thin canals and adjacent teeth live

Qualitative Results



MICCAI2026

Input				
Ground Truth				
SwinUNETR				
nnU-Net				
Original U-Mamba				
Proposed method				

■ All methods recover the dental arch

Overall tooth arrangement and arch shape are captured by SwinUNETR, nnU-Net and U-Mamba as well.

■ Differences hide in small structures

Adjacent instances and thin root canals are where predictions diverge.

■ Proposed method: cleaner instances

Fewer fragmentations and smoother canal surfaces; instance separation closer to ground truth.

Rows (top → bottom): axial CBCT input · ground truth · SwinUNETR · nnU-Net · original U-Mamba · proposed method. Four internal-validation cases (columns).

Conclusion & Future Work



MICCAI2026

Takeaways

- **TDMB** tri-axis serialization removes directional bias at the bottleneck.
- **HCF-MT** three-level gating turns noisy pseudo-labels into reliable supervision.
- **LMG-BDice** boundary-band Dice strengthens thin structures.

Key numbers

- **Internal macro Dice: 0.6356 vs 0.4981 (+13.75 pts, +27.6%)**
- **Official overall score: 0.7599 (submission 851931)**
- **Official DiceImg 0.9616 · NSDInst 0.8217**

Limitations & future work

- One fixed 24/6 patient-level split × 3 seeds: estimates training variability, not significance or cross-center generalization
- HCF-MT thresholds / batch gate / class weights not individually ablated
- Next: mechanism-level ablations, case-level statistics, cross-center validation, subgroup analyses



3D tooth + pulp-canal segmentation result



MICCAI2026

Q & A

Thank you for your attention!