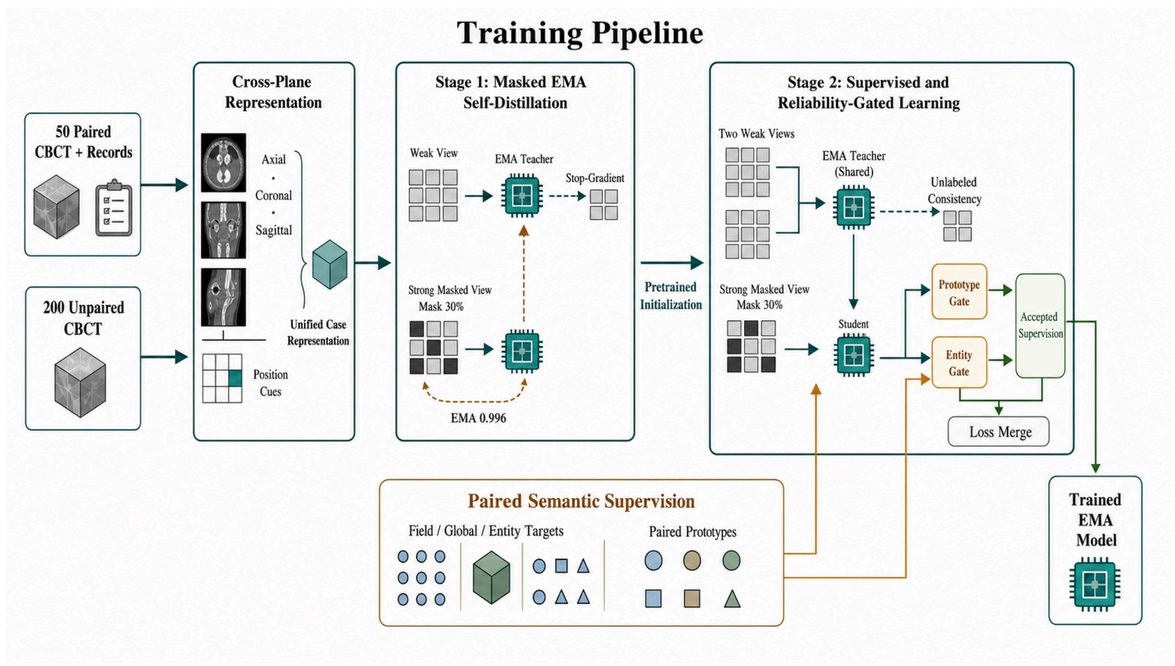
MICCAI 2026 · STS 2026 CHALLENGE · MMDENTAL TASK 3

CPMT: Cross-Plane Masked Prototype-Gated Mean Teacher for CBCT-to-Clinical Record Prediction

Jie Huang ✉ · **Haonan Wu**

School of Communication Engineering, Hangzhou Dianzi University · huang_jie@hdu.edu.cn · wu_haonan@hdu.edu.cn

Method Overview



Training flow of CPMT. Paired and unpaired volumes share one cross-plane representation and masked EMA self-distillation (Stage I). Stage II adds paired semantic targets and gates unpaired prototype/entity targets in parallel; only accepted targets reach the loss, and the final EMA model is kept for retrieval.

Clinical Problem

Dental CBCT shows teeth, roots, bone and treatment materials in three dimensions. A clinical record adds seven fields that the volume alone does not state, including chief complaint, history, diagnosis, treatment plan and advice.

- Paired volumes and records are scarce; unpaired scans are far more common.
- Several record fields depend on evidence that is not visible in the scan.
- Free-text generation returns fluent records that the anatomy does not support.
- Similar tooth structures recur across positions, confusing image-only models.

Task and Data

Given one CBCT volume, predict all seven record fields. Training used **50 paired volumes with records** and **200 unpaired volumes**. Scoring is on **50 public validation volumes** with hidden records, via Codabench.

Method in Four Steps

1. **Cross-plane representation.** 24 observations per volume (8 positions × axial/coronal/sagittal 2.5D slices) are encoded by a partly frozen ResNet-18, tagged with a plane embedding and a position code, then fused by a two-layer Transformer into one 13.1M-parameter case token.
2. **Stage I: masked EMA self-distillation.** All 250 volumes train without records: a strong student view with 30% of tokens masked matches a weak EMA teacher view by stopped-gradient cosine distance, at case level and per plane, with a variance floor against a collapsed representation (teacher EMA 0.996).
3. **Stage II: paired alignment and gated teacher learning.** Records become hashed n-gram field/record targets plus a 100-entity vector. For unpaired volumes, a prototype target needs two-view agreement and margin δ; an entity target needs confidence τ_c and two-view stability τ_s. Rejected targets contribute nothing to the loss.
4. **Fieldwise retrieval.** The final EMA model scores every labeled record by field, whole-record and entity similarity plus a global-anchor bonus (0.65/0.20/0.15/0.05), then copies the best nonempty value per field — the output never leaves the observed record vocabulary.

Implementation Snapshot

- Representation: 3 planes × 8 locations, 512-D ResNet-18 feature → 192-D token.
- Stage I: 12 epochs, batch 4, mask ratio 0.30, EMA 0.996.
- Stage II: 24 epochs, labeled/unpaired batch 2/2, AdamW cosine schedule.
- Entity vocabulary of 100 labels; retrieval weights 0.65/0.20/0.15/0.05.
- Software: PyTorch 2.1.2, RTX 3090 24GB, Ubuntu 20.04.

Challenge Results

0.21	0.51	0.31	1.00
Weighted Score	Treatment-plan recall	Treatment action F1	Prediction completion

Metric (50 public validation cases)	Score
Weighted Score	**0.21**
BLEU-1 / 2 / 3 / 4	0.17 / 0.06 / 0.03 / 0.02
METEOR / ROUGE-L / CIDEr	0.20 / 0.20 / 0.05
Prediction / reference completion	1.00 / 0.90
Tooth notation P / R / F1	0.11 / 0.23 / 0.10
Diagnosis code P / R / F1	0.07 / 0.09 / 0.07
Treatment action P / R / F1	0.28 / 0.45 / 0.31
Medication P / R / F1	0.29 / 0.28 / 0.28
Diagnosis exact / partial	0.04 / 0.09
Treatment-plan recall	0.51

A complete seven-field record was returned for every case. Treatment scores highest; tooth location, diagnosis and long-phrase agreement remain weak (BLEU-4 0.02).

Qualitative Example

Predicted report	Case	Ground truth report
Main appeal: Pain in left lower back tooth for several days **Diagnosis:** Chronic apical periodontitis Dental caries Impacted tooth **Treatment plan:** It is recommended to remove the restoration and perform root canal retreatment and crown restoration. Filling treatment.		**Main appeal:** The left upper back tooth cracked for several days after biting something hard. **Diagnosis:** Longitudinal fracture of tooth **Treatment plan:** Remove and elective repair.

Internal labeled validation case (not an official public case): vocabulary is correct, but location and diagnosis are wrong — matching the low tooth/diagnosis scores above.

Limitations

- Sparse 24-view sampling can miss small, tooth-specific findings.
- Complaint, history and advice often lack any imaging correlate.
- The 50-record bank cannot express concepts it never contains.
- Independent field retrieval can combine incompatible values.
- Two-view agreement does not rule out shared systematic error.

Conclusion

- Anatomical representation learning stays separate from uncertain clinical supervision, via masked EMA self-distillation on all 250 volumes.
- Two reliability gates admit only stable unlabeled targets, and constrained retrieval keeps every output inside the observed record vocabulary.
- Remaining tooth-location and diagnosis errors call for stronger spatial evidence and record-level consistency constraints.

Jie Huang · huang_jie@hdu.edu.cn · Hangzhou Dianzi University · Thanks to the MMDental data owners and the STS 2026 organizers.