

Action-Aware Field Retrieval for Structured Dental CBCT Reporting

Rutong Peng · Suhang Yang · Jianxiang Zhao*
School of Cyberspace Security, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China

Method Overview

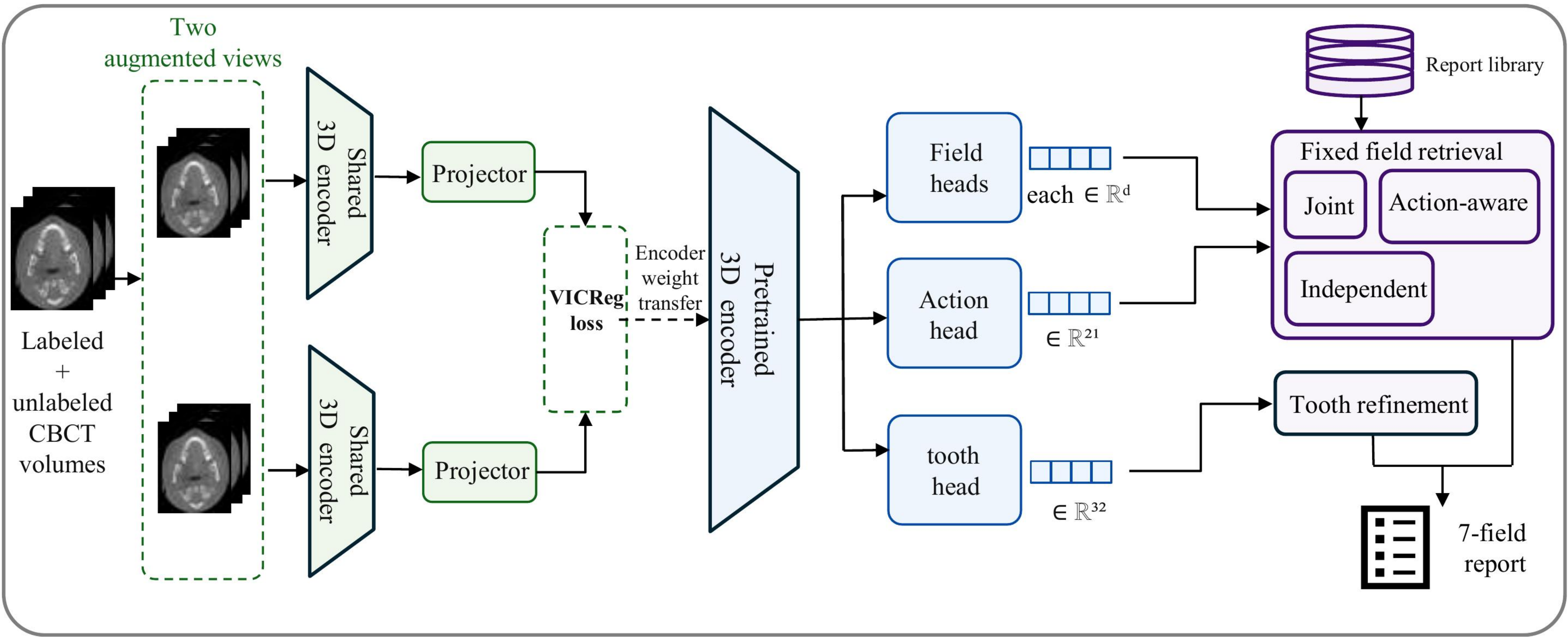


Fig. 1. DentSAFER. VICReg initializes the 3D encoder. Supervised learning predicts seven field embeddings, 32 FDI tooth positions and 21 treatment actions. Action-aware retrieval and image-conditioned tooth-code refinement assemble the final seven-field report.

Clinical Problem

Seven report fields. Main appeal, present medical history, oral check, diagnosis, treatment plan, handle and doctor advices must be constructed from a full CBCT volume.

Limited supervision. Only 50 labeled CBCT volumes and 200 unlabeled volumes are available.

Patient-specific consistency. Findings must connect to tooth positions and treatment actions; unrestricted generation can produce unsupported content.

Method — DentSAFER

1 Self-supervised 3D encoder. VICReg learns from two augmented views. A compact residual encoder produces a 256-dimensional volume representation.

2 Joint field and entity learning. Seven field heads learn text targets from word and character n-grams. Auxiliary heads predict 32 FDI positions and 21 treatment actions.

3 Supervised objective. Field cosine loss is combined with class-weighted tooth and action classification losses.

$$L_{sup} = L_{field} + 0.35 L_{tooth} + 0.20 L_{act}$$

Action-Aware Field Retrieval

Joint diagnosis and treatment plan. Both fields share one candidate report selected by field similarity and predicted action compatibility.

$$\text{Joint score} = 0.70 \times \text{diagnosis similarity} + 0.30 \times \text{plan similarity} + 0.30 \times \text{action compatibility}$$

Other fields. Handle also uses action compatibility. The four remaining fields retrieve their highest-scoring nonempty candidate independently.

Image-conditioned tooth codes. Remove FDI codes from all retrieved fields. Append at most five predicted codes (probability ≥ 0.50) to oral check.

Ensemble: candidate scores average over four eligible models; entity probabilities average over all five.

Training Protocol

COMPONENT	SETTING
Input / batch	80 × 128 × 128; batch 8 SSL / 4 supervised
SSL split	225 training / 25 validation images
Supervised folds	5 folds; 40 training / 10 held-out cases
Retrieval library	Only the 40 training reports in each fold
Optimization	AdamW; 80 epochs; cosine schedule; AMP
Learning rates	2 × 10 ⁻⁴ SSL; 5 × 10 ⁻⁶ encoder / 3 × 10 ⁻⁴ heads
Model / hardware	5.258 M parameters per model; RTX 4090 (24 GB)
Inference time	114 s / 50 cases; 2.28 s per case

Results — STS 2026 Official Test Set

0.2193
Weighted Score

1.0000
Field completeness

0.3094
Treatment-action F1

CLINICAL METRIC	PRECISION	RECALL	F1 / MATCH
Tooth notation	0.0897	0.1912	0.1093
Diagnosis code	0.0733	0.1540	0.0930
Treatment action	0.2693	0.4137	0.3094
Medication	—	—	0.6100
Exact / partial match	—	—	0.0200 / 0.1340
Treatment plan recall	—	0.4737	—

Text metrics. BLEU-1 0.1957 · BLEU-2 0.0735 · BLEU-3 0.0434 · BLEU-4 0.0315
ROUGE-L 0.2171 · METEOR 0.2417 · CIDEr 0.0602

Seven fields for every test case; completeness does not ensure clinical accuracy.

Qualitative Prediction

Diagnosis and treatment excerpts from the two cases in Fig. 2

Case 1	GROUND TRUTH	PREDICTION
	Diagnosis: acute pulpitis; impacted teeth; dental caries. Treatment: root canal and crown restoration; removal; filling.	Diagnosis: dental caries. Treatment: filling treatment; removal.
Case 2	GROUND TRUTH	PREDICTION
	Diagnosis: impacted teeth; dentition defect; chronic periodontitis. Treatment: extraction; implant restoration; full-mouth scaling.	Diagnosis: impacted teeth. Treatment: extraction recommended; risks explained; patient willing to proceed.

Fig. 2 (adapted excerpts). Predictions capture common findings and broad actions but omit some case-specific diagnoses and procedures. Axial slices show anatomical context, not prediction attribution.

Conclusion & Limitations

DentSAFER constructs complete seven-field reports with VICReg pretraining, action-aware retrieval and image-conditioned tooth-code refinement.

Clinical accuracy remains limited. Diagnosis F1 is 0.0930 and tooth-position F1 is 0.1093. A small report library, global volume features and fixed thresholds limit case specificity.

Evidence gaps. The study lacks controlled ablations, external validation and clinician-adjudicated error analysis.

Future work. Tooth-local representations, constrained diagnosis–tooth–treatment linking and calibrated uncertainty.