

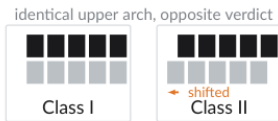
Seeing the Bite: Learned Cross-Arch Occlusion Grounding for Orthodontic Report Generation from Intraoral Scans and Photographs

Gagneet Singh¹ Adinath Madhavrao Dukre² Imran Razzak^{1,2}
¹MedOS, Abu Dhabi, UAE ²MBZUAI, Abu Dhabi, UAE ODIN 2026 Bite2Text Challenge — Team **GenMI**

Orthodontic findings live **between** the two arches. We teach a small network to read the bite — and to **stay silent when it is unsure**.

1 One tooth, two answers

The **same tooth** reads Class I or Class II depending on how the *opposing* arch meets it — so one arch, or one photograph, **cannot express** molar class, overjet, overbite or crossbite.



These findings are **relational**. Paired scans carry the bite; photographs carry dental health — caries, restorations, gingiva — invisible in untextured geometry. Prior dental VLMs occupy only the single-modality corners.

2 Research question

Can a model **learn the bite** from paired-arch geometry — and does injecting what it learns make the report **more factual**?

3 Task, data and scoring

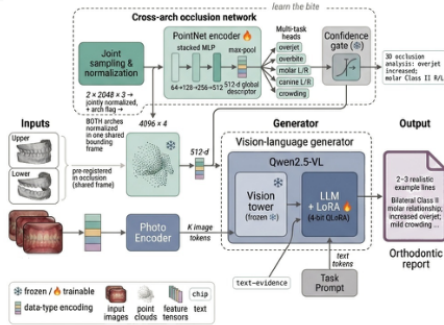
Upper + lower intraoral-scan meshes, *pre-registered in occlusion*, plus intraoral photographs → the clinician's English report.

Final = 0.8 RadFact + 0.2 BLEU-4, METEOR

RadFact is an entailment metric: **hallucinated and omitted findings cost the same**. Factuality dominates the score.

Data. 872 report-bearing patients, split 698 / 174 train / validation by a fixed hash.

4 The system



A small point-cloud network states only the occlusal findings it is **sure of**; a fine-tuned VLM reads the photographs and writes the report, treating those statements as evidence.

References

1. Singh, Dukre, Razzak. *Seeing the Bite*. ODIN, MICCAI 2026.
2. Lumetti et al. *Do Multimodal LLMs Understand Intraoral Dental Data?* ECCV 2026. 3. Bannur et al. *MAIRA-2 / RadFact*. 2024. 4. Qi et al. *PointNet*. CVPR 2017. 5. Dettmers et al. *QLoRA*. NeurIPS 2023.

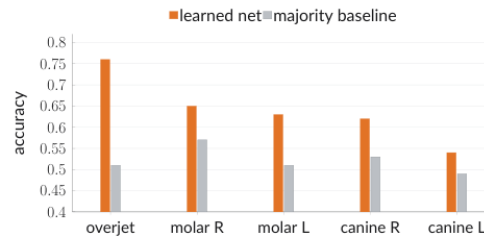
5 Normalize the arches together

Centring each arch on its own **destroys the bite**. We stack both and normalize jointly, so the relative pose survives:

$$\tilde{P} = \frac{P - c}{\max_i \|P_i - c\|_\infty + \epsilon}, \quad X = [\tilde{P} \parallel f] \in \mathbb{R}^{4096 \times 4}$$

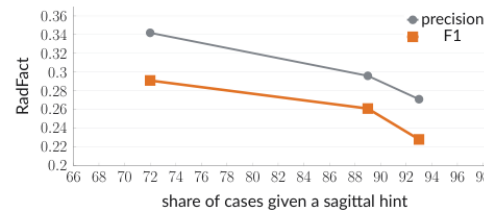
A small **PointNet** (2 × 2048 points, 512-d descriptor, one head per occlusal slot) — **trains in minutes**. The generator is **Qwen2.5-VL-7B** with 4-bit QLoRA, one epoch on a single 24 GB GPU.

6 Sagittal class is learnable



Molar class and right canine are clearly learnable. **Left canine sits at baseline**; overbite and crowding never beat it and are **gated out**.

7 Looser gate ⇒ worse reports



More coverage, **monotonically worse factuality**. **Precision falls first**: a wrong class is asserted and crowds out correct content.

The best configuration is the one that stays silent most often.

8 Three warnings for this field

1. **The judge picks the winner**. A weak judge ranks a data-scaled model *above* ours; the real judge **reverses** it.
2. **The reward picks the target**. A cheap proxy reward chooses a different best-of-*K* sample in **half** of cases.
3. **Retrieval cuts both ways**. It rescues a weak 3B, **degrades** a well-fit 7B.

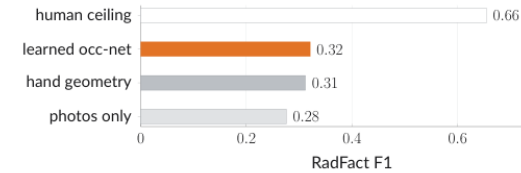
9 Official hidden test set

0.358 RadFact F1
50 hidden patients

0.354 blended final score

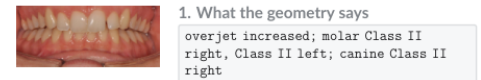
Scored by the organizers with *their* judge — and **higher** than our own judge predicted (0.358 vs. 0.321).

10 Grounding geometry lifts factuality



Validation, 100 cases, same judge throughout — **+0.045** from grounding. The last step sits inside the ±0.02 judge variance. Two clinicians agree only at **0.655**: that, not 1.0, is the ceiling.

11 One case, end to end



2. What the model writes

The patient presents **correct transverse relationships**, while from a vertical standpoint there is a deep bite. Sagittally, there is a full bilateral Class II molar and canine relationship with **increased overjet**. The dental midlines are deviated. The Curve of Spee and the Curve of Wilson are increased. There is mild crowding in the upper arch and **moderate-to-severe crowding in the lower**.

3. What the clinician wrote

"...maxillary constriction... full bilateral Class II molar and canine... mild crowding in both arches."

The **bold sentence** is exactly what the occlusion net supplied. **Red** marks what the model got wrong.

12 Limitations

Single 100-case run, no seeds; gating trades wrong facts for *omissions*. Transverse is where recall collapses (cross-bite 0.07).

Take-home

Model the **pair** of arches — then let the model **decline to answer** when unsure. Silence beats a confident wrong finding.