UNIVERSITY MEDICAL CENTER
HAMBURG–EPPENDORF
Institute for Applied Medical Informatics

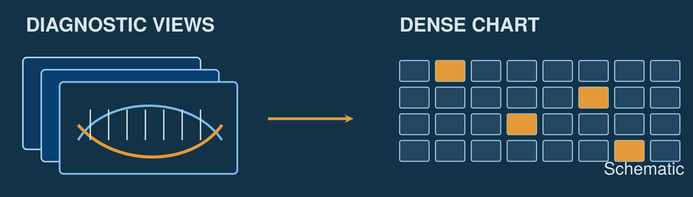
MICCAI 2026
29TH INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING & COMPUTER ASSISTED INTERVENTION
Sept. 27- Oct. 1, 2026
Strasbourg Convention Center, FRANCE

RGBite: Vision-Language Report Generation for the ODIN 2026 Challenge

Sören Spiegel* **Louis Bellmann*** Maximilian Nielsen

Institute for Applied Medical Informatics, UKE • Hamburg, Germany

*Equal contribution



TWO TASKS / ONE CHART-FIRST PIPELINE

All reports. One structured target.
Images → charts → reports.

+0.0315

Richer views at fixed supervision

95% CI [0.0125, 0.0503]; $p = 0.001$

Task 1 validation FinalScore

01 Image preprocessing: make 3D findings visible

TOOTHFAIRY4 / TASK 1

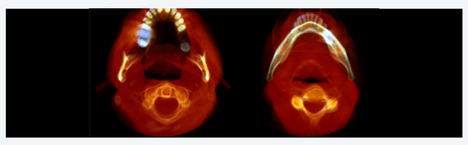
CBCT → maxillofacial report

LAS reorient • 0.3 mm isotropic • fitted arch

RGB panoramic



Occlusal projection



R: soft tissue **G:** roots / bone **B:** enamel / metal

Stills L/W (HU): 150/1400 | 1100/2200 | 2200/2600

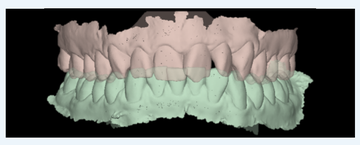
64 arch-cross-section frames + 5 stills

BITE2TEXT / TASK 2

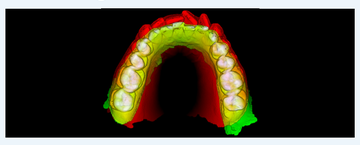
Scans + photos → orthodontic report

Shared inter-arch frame preserves bite registration

Frontal



Occlusogram



Photograph



R: upper crown height **G:** lower crown height **B:** gap

Fixed ranges: 15 mm height, 3 mm gap; 0.15 mm/px

7 orthographic renders + up to 7 photos; no video

Selected inputs from manuscript Figs. 1–2, not the complete image sets. One preprocessing path is used for training and inference.

02 Merge all reports, then learn the structured chart

A TARGET PREPROCESSING / GROUP REPORTS BY CASE

All reports
Same case / image set

→

Report charts
Structured findings

→

Hierarchical merge
One ground-truth (GT) chart per case

Weighted chart loss (training only)

ALL REFERENCES REUSED
Every image set uses all reports belonging to its case.

B TRAIN IMAGE → CHART; RENDER REPORTS AT INFERENCE

Image set
Task-specific views

→

Qwen2.5-VL-7B-Instruct[1]
LoRA; same recipe for both tasks

→

Predicted chart
Dense, multilabel findings

parse →

Report
Deterministic template

Learn structured findings, not free-text reports; prose is rendered from the predicted chart at inference.

TASK 1 TARGET / DENSE MULTILABEL CHART

32 FDI slots (Q1–Q4) + FOV / IAN / Sinuses / Bone

All slots present; healthy ≠ n/a (outside the field of view).

CLASS-IMBALANCE NORMALISATION

$w_c \propto 1/f_c, \quad \bar{w} = 1$ Inverse-frequency slot weights

Preserve loss scale; n/a : root remnant = 124:1.

DATA + AUGMENTATION / TASK 1

- **622 CBCTs / 1,000 reports.**
- Train / val / test: **498 / 62 / 62**.
- Family-stratified 80/10/10 case split; resampling.
- Left–right mirror ($p = 0.5$); 3 mm z-jitter.
- **64 frames** per input.

LoRA ADAPTATION[2]

- LM projections + vision attention / MLP.
- $r = 16$, $\alpha = 32$, dropout 0.05.
- 51.8M / 8.34B trainable (**0.62%**).
- **Completion-only loss.**
- bf16 + FlashAttention-2.

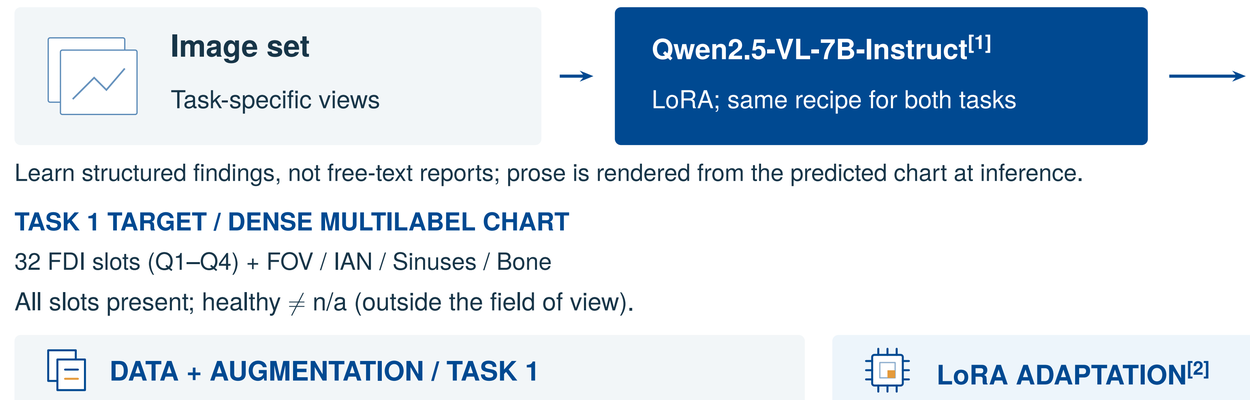
OPTIMISATION + RUN

- AdamW; LR 10^{-4} / 5×10^{-5} (vision).
- Cosine schedule; warmup 0.03.
- Batch **1** × accumulation **8**.
- **8 epochs / 496 steps**; greedy decoding.
- One NVIDIA RTX PRO 6000.

03 Representation gains; chart weighting trades accuracy

Task 1 / local validation ($n = 62$)

FinalScore $= 0.8$ LogicalF1[3] $+0.2$ CaptioningScore



†Also changes target canonicalisation; not a chart-only test.

Official challenge evaluation

Unseen centres; single-annotator references

TASK / INPUT	BLEU-4	METEOR
Task 1 / 2 panoramics Tooth-only weights	0.0833	0.2321
Task 1 / 5 stills + video Tooth + non-tooth weights	**0.1021**	**0.2661**
Task 2 / scans + photos **7th among teams**	0.233	0.435

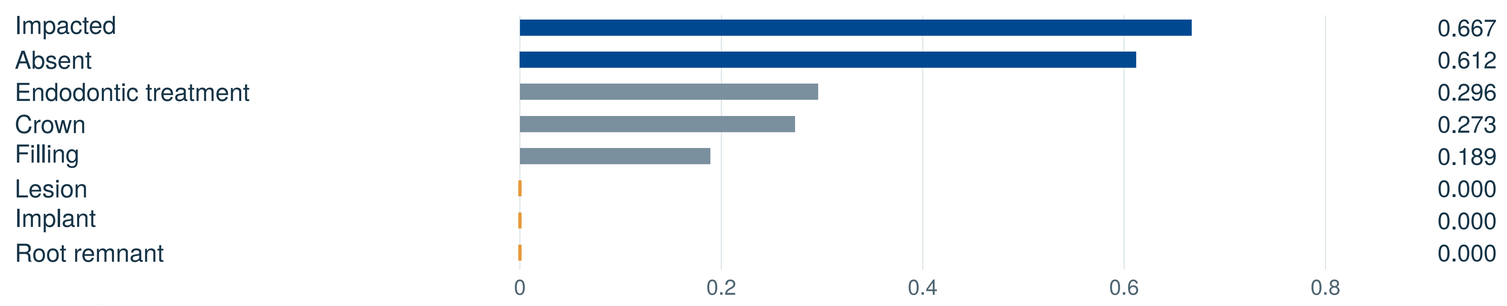
Task 1 test: $n = 50$. Task 2: one run, no per-task tuning.

Non-tooth weighting: −0.0295 FinalScore

Non-tooth macro-F1: 0.388 → 0.425; tooth: 0.310 → 0.271

04 Where the model still fails

Task 1 • judge-free macro-F1 0.255 • 100% chart parsing



Healthy F1 (not plotted): 0.744; ∼4 mm of anatomy per patch.

TAKE-HOME

Use all reports; learn charts.
Render prose; audit findings.
Fine detail remains unresolved.

VALIDATION BOUNDARY

Single seed; 3 of 4 ablation cells.
No per-stage or Task 2 ablation.
Validation-selected epochs: optimistic.
Video: percentile normalisation.

CONTACT

s.spiegel@uke.de • l.bellmann@uke.de

m.nielsen@uke.de

REFERENCES / SOURCE

[1] Bai et al. *Qwen2.5-VL Technical Report* (2025). [2] Hu et al. *LoRA*. ICLR (2022).

[3] Bannur et al. *MAIRA-2 / RadFact* (2024). Results and images: accompanying RGBite manuscript.