

Bite2Text Challenge

From Dental Geometry to Clinical Text A Modular Structured Pipeline for Orthodontic Report Generation

Presenter: Changkai Ji

Jingjing Li^{1,2}, Yuxian Jiang¹, Feiyang Wei³, Changkai Ji¹, Lisheng Wang¹, Youren Chen¹

¹ School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai 200240, China

² Shanghai Jiao Tong University Sichuan Research Institute, Chengdu 610213, China

³ Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China



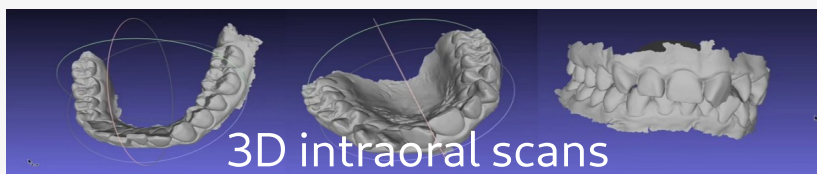
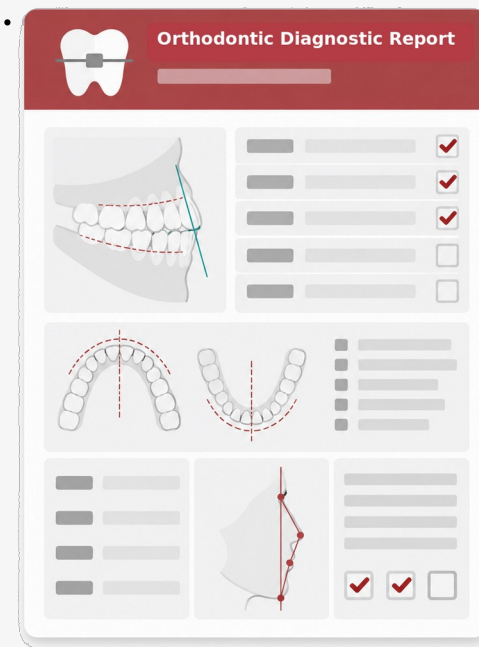
Background

From geometric benchmarks to clinical understanding

The ODIN challenge series has established large-scale benchmarks for tooth segmentation and landmark detection from 3D intraoral scans.

Bite2Text extends this direction toward multimodal orthodontic report generation by combining 3D intraoral scans with 2D photographs.

Orthodontic reporting is time-consuming and subject to inter-observer variability. Bite2Text evaluates structured clinical reports generated from multimodal dental inputs.



Structured orthodontic report



01 From heterogeneous inputs to structured findings

3D scans and 2D photographs provide complementary but heterogeneous evidence. A structured clinical interface is needed to organize them into clinically meaningful fields.

02 Preserving clinical consistency during generation

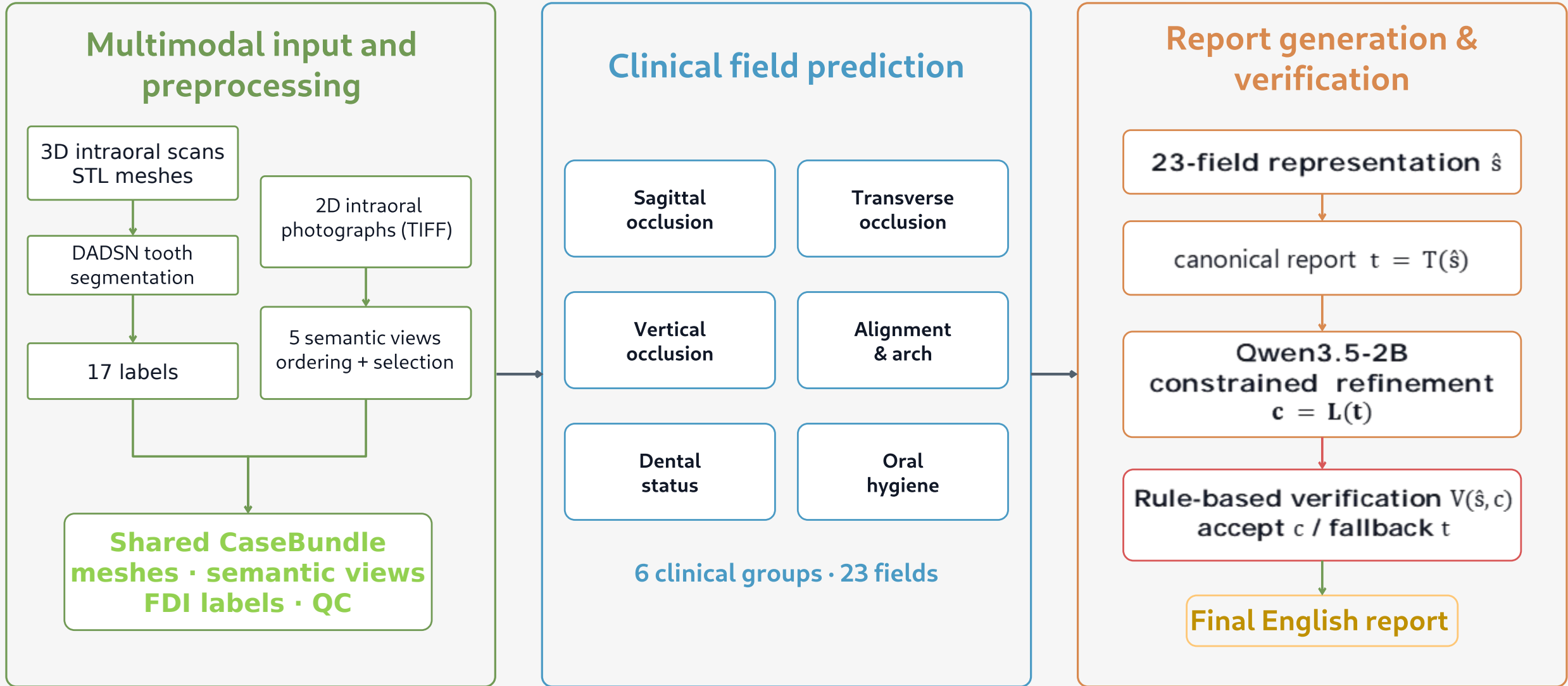
Free-form language generation may omit or contradict predicted findings. Constrained refinement and rule-based verification help keep reports clinically grounded.

03 Handling uncertain or incomplete evidence

Prediction quality varies across cases. Robust handling of uncertain cues and a deterministic fallback improve the reliability of report generation.

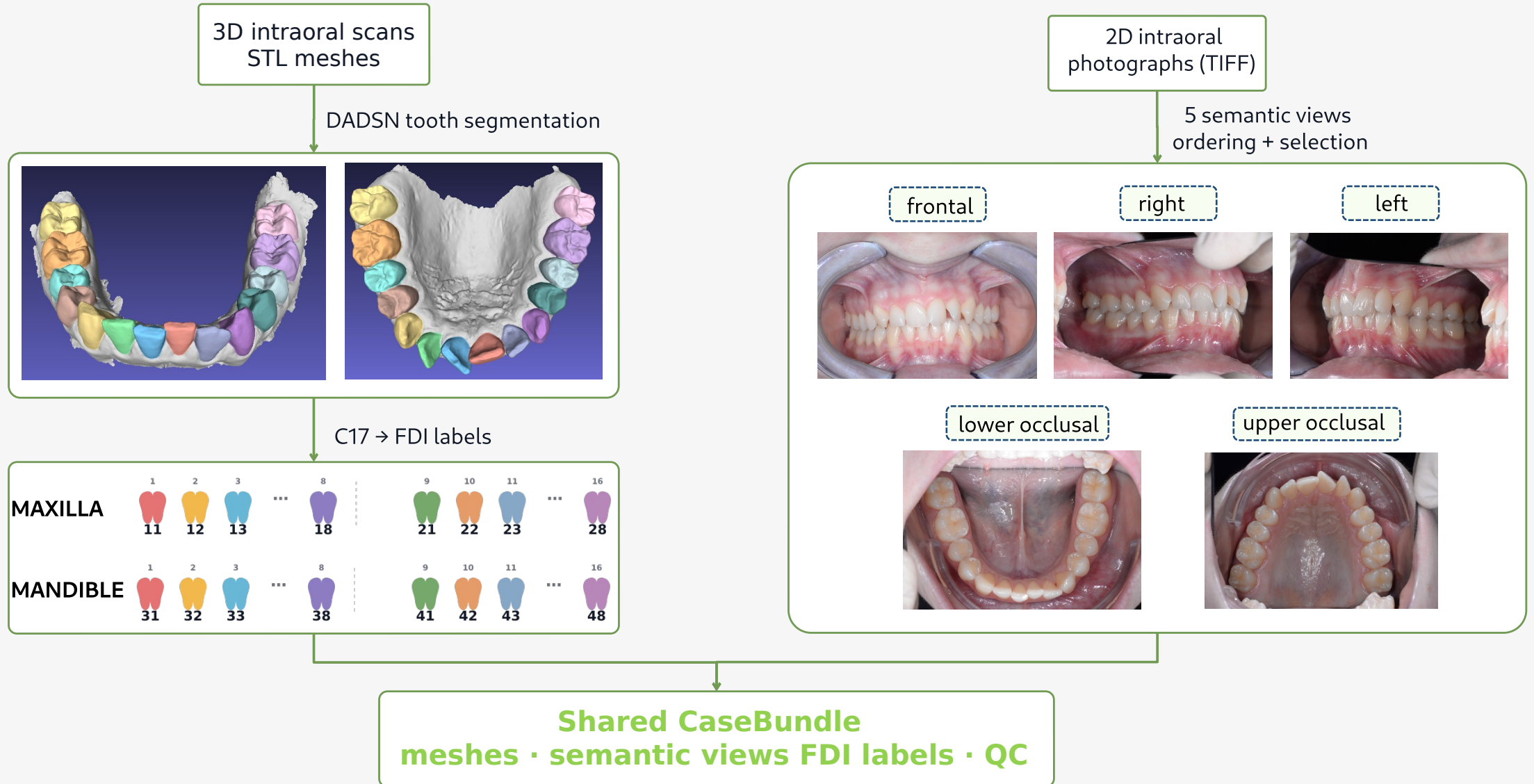


Overall Architecture





Multimodal input and preprocessing





Clinical field prediction



Six clinical groups organize 23 structured fields before report generation.

Clinical group	Structured fields	#	Primary evidence	Route
1 Sagittal occlusion	Right/left molar · right/left canine	4	STL renderings	Visual / CNN
2 Transverse occlusion	Severity · type · crossbite presence/type/location · teeth	6	STL + FDI labels	Geometry / classical
3 Vertical occlusion	Overjet · overbite	2	DADSN + geometry	Hybrid / trees
4 Alignment & arch	Midline · upper/lower crowding · Spee · Wilson · location	6	Geometry + visual cues	Routed / mixed
5 Dental status	Missing teeth · restorations · caries	3	C17 counts + photos	Classical / fixed
6 Oral hygiene	Hygiene status · pigmentation	2	Occlusal photos / fixed	Photo / fixed
Total	Structured clinical fields	23		

This table summarizes the schema fields; detailed category values are omitted for readability.

Three prediction routes

■ **Visual neural · STL/photo + ResNet-18** ■ **Geometry/classical · STL + FDI + rules** ■ **Hybrid/routed · DADSN + geometry + visual cues**



Report generation & verification

23-field representation $\hat{s}$



canonical report $t = T(\hat{s})$



Qwen3.5-2B
constrained refinement
 $c = L(t)$



Rule-based verification $V(\hat{s}, c)$
accept c / fallback t



Final English report

Key control points

01 Canonicalization before language generation

The 23 structured fields become a deterministic report template, preserving explicit clinical findings and their order.

02 Constrained text refinement

Qwen3.5-2B improves fluency from the canonical text only; it does not infer new findings from raw scans or photographs.

03 Verification and fallback

Rule-based checks compare the candidate with the structured fields. If a conflict is detected, the system retains the canonical report.

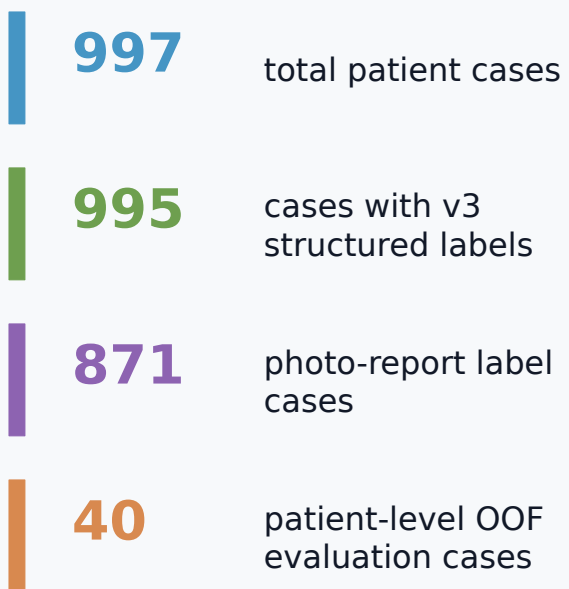
Clinical findings remain grounded in the structured representation.



Evaluation setup and structured accuracy

Patient-level OOF evaluation of the 23 structured clinical fields

Evaluation data

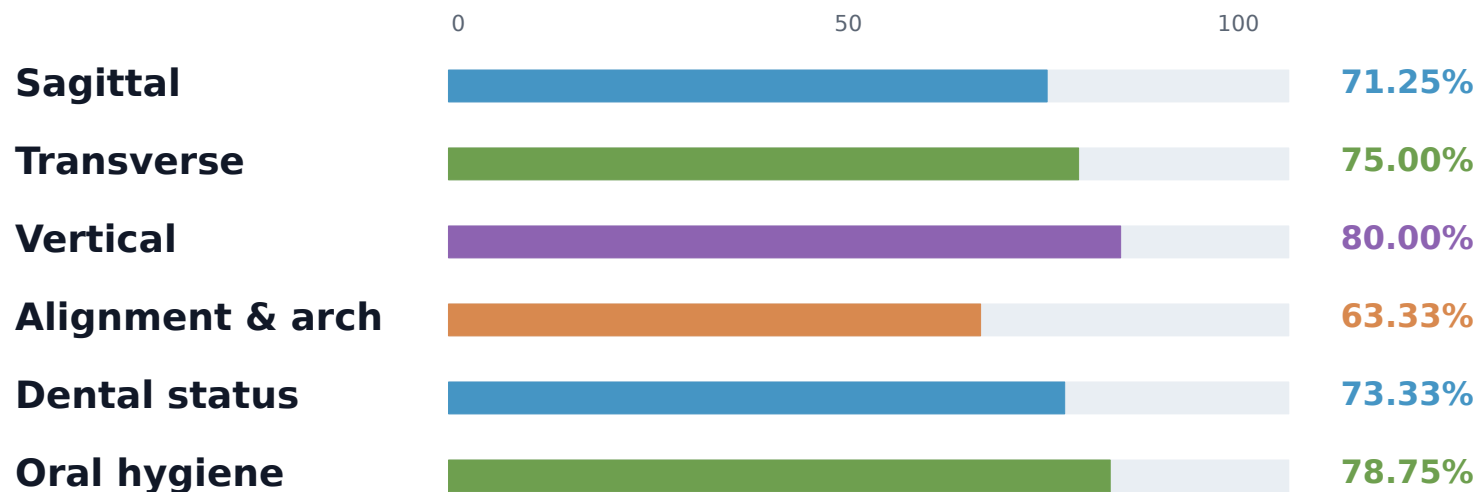


Task-specific cohorts are used. Each case uses a fold-specific OOF prediction.

Structured prediction accuracy

71.85%
23-field OOF accuracy
661 / 920

Mean accuracy by clinical dimension



Strict exact-match accuracy; $n = 40$ for every field.

Patient-level OOF evaluation uses fold models that exclude the evaluated case, with task-specific cohorts.



Table 4. Final report-generation performance of the proposed framework

RadFact	BLEU-4	METEOR	Overall
0.5382	0.1876	0.3943	0.4888

The framework converts structured clinical findings into an English orthodontic report through constrained refinement and rule-based verification.



Visualization

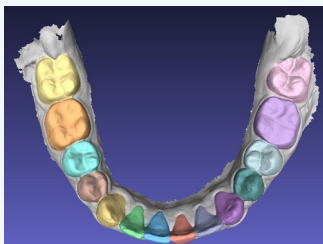
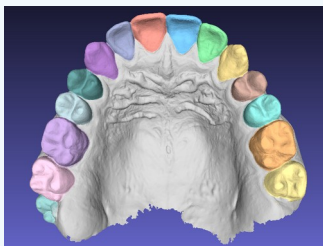


F5264

inputs

STL mesh

TIFF photos



The patient presents normal transverse relationships in the absence of crossbite. Sagittally, on the right, the molar relationship is Class I and the canine relationship is Class III; on the left, the molar relationship is Class I and the canine relationship is Class I. Overjet and overbite are within normal limits. The dental midlines are coincident. No crowding is present in the upper or lower arch. The curve of Spee is increased, while the curve of Wilson is within normal limits.

Evaluation

Structured accuracy

90.28%

BLEU-4

0.3098

METEOR-lite

0.3907

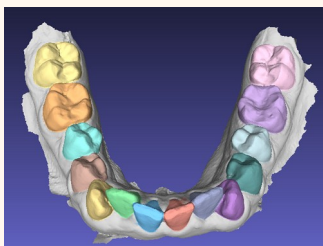
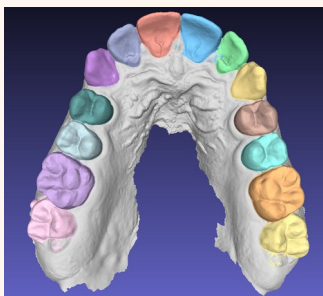
21/23 correct · 2 errors

F485

0
inputs

STL mesh

TIFF photos



The patient presents normal transverse relationships in the absence of crossbite. Sagittally, on the right, the molar relationship is end-to-end Class II and the canine relationship is end-to-end Class II; on the left, the molar relationship is Class I and the canine relationship is Class I. Overjet is increased; overbite is increased. The dental midlines are deviated relative to each other. Mild crowding is present in the upper arch, while moderate crowding is present in the lower arch. The curve of Spee is increased, while the curve of Wilson is increased.

Evaluation

Structured accuracy

81.94%

BLEU-4

0.3385

METEOR-lite

0.4712

18/23 correct · 5 errors



Acknowledgements



- Thanks to the organizers for hosting the competition and providing valuable data for this research.
- Appreciation for the organizers' support throughout the competition, especially for their patience in addressing various questions and issues.



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

Thank you

lijingjing327@sjtu.edu.cn