

The Dataset

Expert-verified tooth masks, each with its FDI number

Built on the 5,000 Bite2Text photographs. Every visible permanent tooth has a polygonal mask and its FDI number; deciduous teeth, implants and pontics map to the corresponding permanent position.



3 frozen splits, patient-disjoint (700 / 200 / 100).

300 test patients annotated from blank masks, blind to any model; the other 700 are SAT proposals corrected by an orthodontist.

Each patient contributes five linked views. Tooth masks and FDI labels support segmentation and identity evaluation across the full photo set.

The only public five-view set with FDI masks

Dataset	5 Views	FDI Masks	Public
IOP-Compass	✓	✓	✓
SAT	✓	✓	✗
IO150K	part	–	✗
Caries IOP	✓	✗	✓
AlphaDent	✗	✗	✗

The Pipeline

Weights on request = our trained models

Model proposals speed up annotation; an orthodontist verifies every mask before release.

A Raw Photos



5 Views

Standard orthodontic set per patient

Download available

B View Assignment

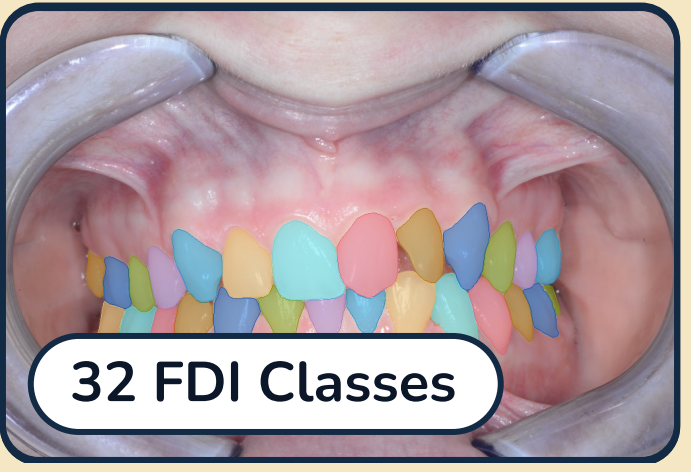


Upper Occlusal

ResNet18 + Hungarian matching

Weights on request

C Segmentation



32 FDI Classes

SAT or fine-tuned Mask R-CNN

Weights on request

D Post-Processing



9 deterministic cleanup rules

In the code

E Expert Check



✓ Verified

Orthodontist corrects in the browser app

App released

F Optional ROI Crop

Geometric · FAST-R · SAM 3

FAST-R

Off by default: cropping never improved FDI-aware F1.

View assignment. A ResNet18 classifies each photo; since each patient has exactly one image per view, labels are assigned jointly under a one-to-one constraint (Hungarian algorithm).

Segmentation and cleanup. SAT (YOLO11 + Light HQ-SAM, published weights) or Mask R-CNN fine-tuned on 32 FDI classes; nine fixed rules remove fragments and duplicate FDI labels.

Results

Held-out test, 100 patients × 3 splits

No ROI crop improves average F1, even though all three keep ≥ 99.95% of tooth pixels: remaining errors are not driven by image context. SAT beats Mask R-CNN on every view, most on buccal ones, at ~12× the compute.

95.1%

FDI-aware Tooth F1

SAT, no ROI crop

99.9%

View Classification

99.33% of patients fully correct

5.7×

Faster Annotation

1.6 min/patient on average

Seg.	ROI	F1 Avg.	F1 Up	F1 Low	F1 Front	F1 L/R	Time/pat.
SAT	No ROI	95.09±0.27	98.33±0.17	98.16±1.07	94.27±0.31	93.31±0.14	17.3 s
	Geometric	94.74±0.33	98.35±0.21	98.18±0.81	93.89±0.26	92.69±0.42	14.1 s
	FAST-R	93.99±0.44	97.96±0.24	97.57±1.09	92.73±0.28	92.05±0.57	11.3 s
	SAM 3	94.56±0.25	96.10±0.60	95.16±1.37	94.48±0.32	93.81±0.20	25.8 s
R-CNN	No ROI	91.45±0.32	96.40±0.43	95.56±1.16	90.23±0.47	88.82±0.53	1.4 s
	Geometric	91.42±0.31	96.40±0.43	95.56±1.16	90.11±0.48	88.84±0.48	1.2 s
	FAST-R	90.75±0.53	96.51±0.22	95.33±1.19	88.89±0.66	88.05±0.75	<u>1.3 s</u>
	SAM 3	90.49±0.59	94.28±0.88	91.63±2.74	90.19±0.48	88.79±0.54	7.1 s

FDI-aware instance F1 (%) at mask IoU ≥ 0.50, mean ± std over 3 splits, post-processing in every row. L/R: pooled buccal views. Time includes ROI stage. Bold: best · underlined: second-best.