

A Public Dataset for Tooth Segmentation in Multi-View Intraoral Photographs

Yibeltal A. Zelelew*, Lorenzo Borghi*, Kevin Marchesini,
Matteo Lugli, Costantino Grana, and Federico Bolelli[✉]

University of Modena and Reggio Emilia, Italy
{name.surname}@unimore.it

Abstract. Intraoral photographs (IOPs) provide a low-cost record of tooth appearance, alignment, soft tissue, and occlusal relationships. We present IOP-Compass, a dataset, benchmark, and annotation resource built on 1,000 patients from the Bite2Text collection. It comprises 5,000 standardized clinical photographs, five views per patient, with view labels and expert-verified tooth-instance masks carrying FDI numbers, together with frozen patient-disjoint splits. The dataset was produced through a browser-based human-in-the-loop annotation platform that we also released. Using IOP-Compass, we benchmark a pipeline representative of the current literature for view classification, region-of-interest extraction, and FDI-aware tooth instance segmentation, providing reference results and ablations across alternative pipeline components. View classification is near-saturated on the benchmark, while FDI-aware instance segmentation remains the main challenge. Both the dataset¹ and code² are publicly released.

Keywords: Tooth Instance Segmentation · FDI · View Classification

1 Introduction

A standard orthodontic acquisition contains five intraoral photographs (IOPs): frontal, left- and right-buccal, and upper- and lower-occlusal views. They are routinely acquired, require no ionizing radiation or additional appointment, and they capture what neither radiographs nor intraoral scans expose in the same way: gingival color and texture, plaque, enamel appearance, appliances, and the occlusal relationship as the clinician sees it. Bite2Text pairs this modality with intraoral scans and clinician-authored reports for 1,000 patients [5,19].

What this modality lacks is identity. Almost every clinical question—which tooth is crowded, which surface is discolored, how a specific tooth moved between

*Equal contribution. Authors are allowed to list their names first on their CVs.

✉ Corresponding author: federico.bolelli@unimore.it

¹ <https://ditto.ing.unimore.it/iop-compass/>

² <https://github.com/AImageLab-zip/IOP-Compass>

appointments—is posed per tooth, and answering it automatically needs an instance mask with an FDI number [10]. No public resource provides that combination for real clinical photographs. The largest openly released photographic corpus with a five-view protocol annotates only caries as bounding boxes [1]; the largest “open-sourced” intraoral collection claims roughly 800 real photographs within a 150,000-image corpus dominated by renderings of scans and plaster models [33], but the dataset and annotations were not publicly accessible at the time of submission³ and the corpus matching our protocol and scale, used to build SAT, is not publicly released [22].

Our contribution is threefold: (i) 5,000 standardized clinical photographs from 1,000 patients with view labels, expert-verified FDI tooth-instance masks, and frozen patient-disjoint splits, released publicly for download; (ii) a reproducible benchmark for multi-view IOP analysis, with FDI-aware instance evaluation and reference baselines for view classification, ROI extraction, and tooth segmentation; and (iii) a semi-automatic annotation workflow, with its browser application, which tracks how each mask was generated.

2 Related Work

Photographic Dental Resources. Several recent datasets have advanced learning-based IOP analysis, but none combines all the properties required for tooth-level instance segmentation and identification in standardized clinical imagery. SegmentAnyTooth [22], or SAT for short, comprises five-view IOPs from 953 Vietnamese subjects and follows a protocol similar to ours, but the dataset is not publicly released. IO150K [33] includes real intraoral photographs alongside rendered and plaster-model imagery, although its photographic component is comparatively limited. AlphaDent [26] provides intraoral photographs from 295 patients, but does not release the corresponding tooth-level annotations. More broadly, recent reviews of publicly available dental datasets report no photographic resource providing both per-tooth instance masks and FDI labels [27,32].

Beyond photographic imaging, public CBCT benchmarks have demonstrated the value of shared data, standardized annotations, and common protocols for systematically comparing dental segmentation methods [2,3,4,8,20,30,31]. These resources, however, address volumetric dental imaging and do not provide tooth-level instance masks with FDI identification for intraoral photographs.

Thus, despite growing interest in photographic dental analysis, an openly available dataset combining standardized multi-view clinical photographs with tooth-level segmentation and FDI identification remains lacking. Table 1 compares existing datasets along the characteristics most relevant to our setting: standardized multi-view acquisition, tooth-level instance segmentation, FDI identification, and public availability.

View Classification and Tooth Segmentation. Prior work has addressed the main components of automated IOP analysis largely as separate tasks.

³ <https://zoubo9034.github.io/TeethSEG/> does not provide access to either the source code or the data. We contacted the authors but had not received a response.

Table 1: Comparison with photographic dental datasets. *5 views* denotes a standardized five-view IOP protocol. *Type* distinguishes real clinical photographs from datasets that also contain synthetic or non-photographic imagery. *Public* indicates whether the corresponding images or tooth-level annotations are openly accessible. “–” means not disclosed/unverifiable.

Dataset	Dataset			Task Characteristics			Public access	
	Subjects	Photos	Type	5 Views	Masks	FDI	Images	Labels
<i>Closest to our setting</i>								
IOP-Compass (ours)	1,000	5,000	Real	✓	✓	✓	✓	✓
SAT [22]	953	5,000	Real	✓	✓	✓	✗	✗
<i>Other photographic datasets</i>								
IO150K [33]	–	150,000 [†]	Mixed	Partial	✓	–	✗	✗
Caries IOP [1]	631	6,313	Real	✓	✗	✗	✓	✓
AlphaDent [26]	295	1,200	Real	✗	✓	✗	✓	✗

[†]IO150K reports a mixed corpus containing 800 real intraoral photographs together with rendered and plaster-model imagery; its total image count is therefore not directly comparable with the number of real clinical photographs reported for the other datasets.

View classification has typically been formulated as image-level recognition using convolutional neural networks [9,14,25]. Tooth segmentation, in contrast, has been approached with region-based detection and segmentation architectures [28,33], as well as by adapting promptable foundation models to dental imagery [15,16,21,29]. More recently, SAT [22] combines tooth detection, enumeration, and segmentation using view-specific YOLO11 detectors to localize and identify teeth and Light HQ-SAM to refine the resulting boxes into masks.

These methods are primarily evaluated with aggregate classification, detection, or overlap metrics. Such measures characterize average predictive accuracy, but provide limited information about whether an individual image has been processed completely and consistently. In annotation-oriented workflows, however, discrete failure modes—such as a missing tooth, a spurious mask, merged adjacent teeth, or an incorrect tooth identifier—can be more consequential than small differences in aggregate overlap [23]. This motivates evaluation at the level of image completeness and clinically meaningful segmentation and enumeration errors, in addition to conventional accuracy metrics.

3 Dataset

Source Data and Annotations. IOP-Compass augments the 5,000 intraoral photographs of 1,000 patients released with Bite2Text [19], comprising the standard five orthodontic views (frontal-, left-/right-buccal, and upper-/lower-occlusal; Fig. 1). Cohort selection, acquisition, de-identification, and ethical approval are described in the original publication; no additional clinical data were collected for this work. We annotate every visible permanent tooth with a polygonal instance mask and its FDI identifier, covering the 32 permanent-dentition



Fig. 1: Two sample patients across the five standardized intraoral views. For each view, the raw photograph is shown above the corresponding tooth-instance segmentation, with a different color assigned to each tooth instance.

codes. Deciduous teeth, implants, and pontics are not treated as separate classes but are assigned to the corresponding permanent-tooth position. This convention preserves a fixed 32-position FDI label space and reflects our focus on tooth-position identification rather than differentiation of prosthetic type. Fig. 2 reports the resulting distribution of annotated instances over the 32 FDI codes for each view, characterizing the class balance and the tooth-per-view visibility.

Verification and Splits. For all the experiments involving our dataset, we use three fixed train-validation-test splits of 700/200/100 patients (i.e., 3,500/1,000/500 images each), keeping all five views of each patient in the same set. Validation and test partitions are mutually exclusive across splits, so that a patient appears in an evaluation partition for at most one split, and the three test partitions are disjoint. The union of the three test sets contains 300 patients.

Annotations for the 700 patients outside the union of the three test sets are generated with

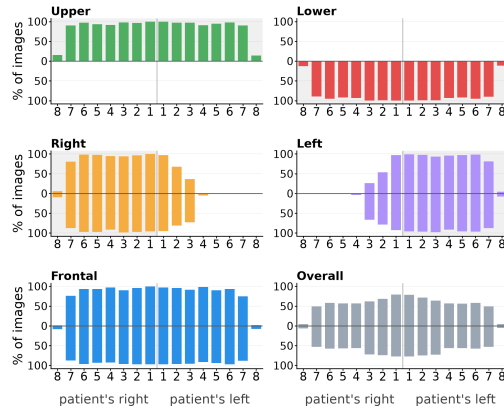


Fig. 2: Distribution of annotated tooth instances across the 32 FDI codes, separately for each of the five views and overall.

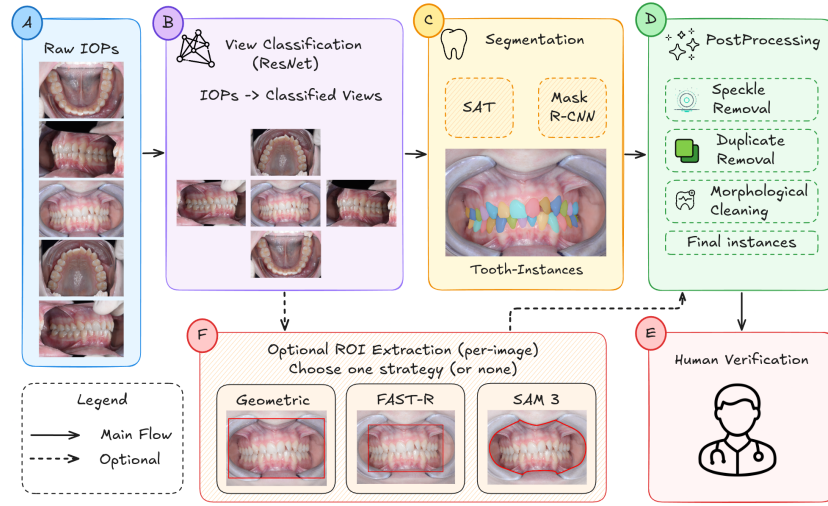


Fig. 3: Overview of the IOP-Compass pipeline, from raw IOPs (A) through standardized view assignment (B), optional ROI extraction (F), tooth-instance segmentation (C), post-processing (D), and expert verification (E).

the proposal-and-review pipeline of Sec. 4. After the initial automatic proposal, every instance and FDI label is reviewed and, where necessary, added, removed, split, merged, relabeled, or boundary-corrected; the release therefore contains expert-verified rather than raw model annotations.

The reference masks of the 300 test patients (1,500 images) were annotated *de novo* from blank masks, without any model proposal being displayed, loaded, or otherwise available in the interface. All the results reported in Sec. 5 are computed against these independently drawn masks, so no model—SAT included—is evaluated against a reference derived from its own predictions.

4 Methods

Fig. 3 summarizes the complete IOP-Compass pipeline. Starting from the five raw intraoral photographs of each patient, a view-classification stage assigns the standardized orthodontic views, after which tooth instances are segmented using either SAT or Mask R-CNN. An optional per-image ROI extraction step—geometric cropping, FAST-R [24], or SAM 3 [6]—can be used to restrict processing to the intraoral region. The predicted instances are then refined through deterministic post-processing to remove artifacts and resolve inconsistencies before undergoing final expert verification.

View Classification and Five-View Assignment (B). We fine-tune an ImageNet-pretrained ResNet18 [7,12] to classify the five standardized views from 256×256 images. Training augmentations include geometric and photometric

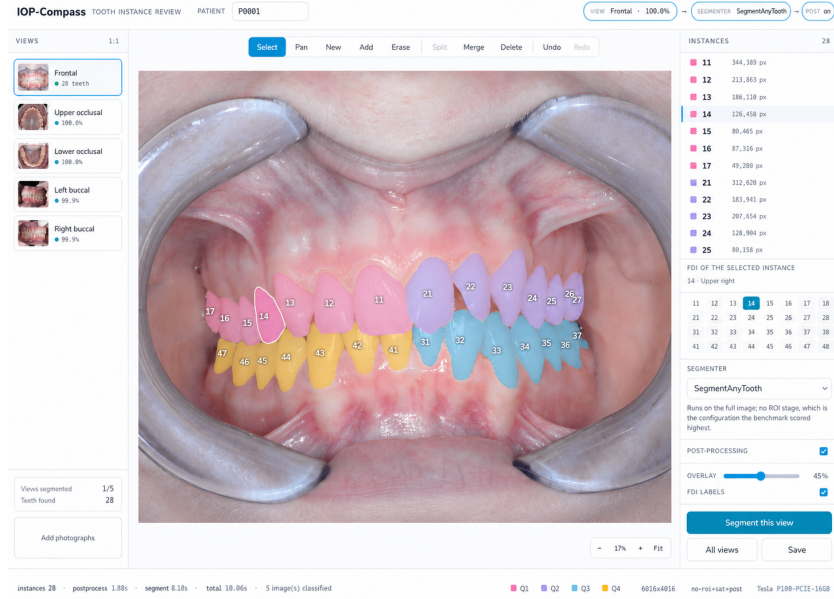


Fig. 4: Browser-based expert correction of tooth-instance masks and FDI labels.

perturbations, blur, noise, and JPEG compression. Horizontal flips are coupled with the corresponding left-right label remapping, whereas vertical flips are excluded. Since each patient record contains exactly one image per view, we additionally perform patient-level assignment by maximizing the summed log-probability under a one-to-one view constraint, solved with the Hungarian algorithm [17].

Tooth Instance Segmentation (C). We evaluate two segmentation pipelines. As an external baseline, we run SegmentAnyTooth (SAT) [22] with its published weights and inference procedure, combining view-specific YOLO11 detection with Light HQ-SAM [15]. Additionally, we fine-tune a COCO-pretrained Mask R-CNN with a ResNet50-FPN backbone [11,18], replacing the prediction heads for the 32 FDI classes.

Post-Processing (D). Predicted instances are processed by a fixed sequence of nine deterministic label-wise cleanup rules that remove small or out-of-ROI components, repair fragmented masks, suppress buccal-view border artifacts, and resolve duplicate predictions carrying the same FDI label. The same rules are applied to all configurations, with parameters fixed using training/validation data; their complete implementation is included in the released code.

Human Verification (E). Model predictions serve only as annotation proposals. An expert orthodontist reviews the resulting instances in the browser interface of Fig. 4, where masks can be added, deleted, split, merged, relabeled, or boundary-corrected. The application preserves the remaining views of the

same patient during correction and records annotation provenance. Only the expert-verified annotations are included in the released dataset.

ROI Extraction (F). For this optional pipeline step, we evaluate three strategies for restricting each image to the intraoral region. These ROI strategies are evaluated on the completed dataset and are not required to generate the expert-verified annotations. *Geometric* uses a fixed, view-specific crop derived from the expert-verified tooth extents of the training split: for each view, we compute the 2nd–98th percentile envelope of the tooth-union boxes and enlarge it by 5%. *FAST-R* predicts the tooth-union ROI with a Faster R-CNN MobileNetV3-Large-320-FPN detector [13]. *SAM 3* obtains an intraoral region mask from a fixed set of text prompts and converts the highest-ranked valid mask into a crop, falling back to the full image when no plausible mask is returned. All parameters are selected using the training and validation data.

5 Experiments and Results

Cohort, Reference Standard, and Protocol. All experiments reported in this section, except those involving annotation time, use the complete cohort of 1,000 patients (5,000 images) and the three fixed patient-level splits defined in Sec. 3. To bootstrap view classification, the five standardized views were manually identified for 100 patients and used to train the view classifier, which was then applied to the remaining cohort. For tooth-instance annotation, the pretrained SAT model generated initial FDI-labeled mask proposals for the 700 patients outside the union of the three test partitions. These proposals were subsequently reviewed and corrected by the expert annotator as described in Sec. 4. Annotations were verified by a single expert; inter-observer variability is therefore not quantified. The 300 patients forming the union of the three test partitions were excluded from this proposal-assisted procedure and annotated from blank masks, with the annotator blinded to any model output, so that the reference standard used for evaluation is independent of both segmenters under comparison. Model fitting and configuration selection use only the training and validation partitions.

Metrics. Reference and predicted instances are first matched one-to-one using the Hungarian algorithm to maximize total mask IoU, independently of FDI labels; assignments with $\text{IoU} < 0.50$ are discarded. A retained match is counted as an FDI-aware true positive only when its FDI number is correct; a geometrically matched tooth with an incorrect FDI number therefore contributes one false positive and one false negative. Counts are micro-pooled over the evaluation split before computing F1. Matched-instance Dice was also computed but was close to saturation across methods and therefore provided little discrimination; we consequently use FDI-aware instance F1 as the primary benchmark metric.

View Classification. The fine-tuned ResNet18 reaches 99.87 % accuracy and macro F1, with 99.33 % of patients having all five views correctly classified after constrained assignment. Accordingly, view classification is not a major perfor-

Table 2: Tooth instance segmentation on the held-out test splits (100 patients per split), FDI-aware instance F1 (%) at mask IoU ≥ 0.50 , reported as mean $\pm$ standard deviation over the three splits; label-wise post-processing is applied in every row. F1 up/low: upper-/lower-occlusal; F1 front: frontal; F1 L/R: pooled left and right buccal views. Time/pat.: wall-clock time per patient, including the ROI stage. **Bold**: best; underlined: second-best.

Seg.	ROI	F1 Avg.	F1 Up	F1 Low	F1 Front	F1 L/R	Time/pat.
SAT	No ROI	95.09 ± 0.27	<u>98.33</u> ± 0.17	<u>98.16</u> ± 1.07	<u>94.27</u> ± 0.31	<u>93.31</u> ± 0.14	17.3 s
	Geometric	<u>94.74</u> ± 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

mance bottleneck in the proposed benchmark: within this standardized acquisition setting, view classification is close to saturation.

Tooth Instance Segmentation. Table 2 reports the eight ROI $\times$ segmenter configurations on the held-out test partitions, with the same label-wise post-processing applied throughout. The validation-selected configuration, SAT without ROI extraction, also performs best on test (95.09 ± 0.27 FDI-aware F1). ROI extraction does not improve average FDI-aware F1, while SAT consistently outperforms Mask R-CNN, particularly on buccal views, at substantially higher computational cost.

Region-of-Interest Extraction. Assessed on the validation split before segmentation, all three strategies preserve nearly all reference tooth pixels: the geometric prior and SAM 3 retain 100% of them, and FAST-R 99.95%. They differ in how aggressively they crop, with FAST-R discarding 34.4% of the frame on average, against 10.0% for the geometric prior and 6.7% for SAM 3, with SAM 3 falling back to the full image on 3.9% of images because the returned mask is empty or implausible. FAST-R is also the faster learned option, at 0.165 s per image against 2.52 s for SAM 3, while the geometric prior is a fixed lookup and adds no measurable cost. None of these properties translates into a segmentation gain (Table 2): every ROI strategy has a lower F1 Avg. than the no-crop configuration. The ROI stage is therefore retained in the released pipeline but disabled by default (Fig. 3).

Annotation Effort. As an exploratory measurement of annotation efficiency, we recorded annotation time and editing operations over a session covering 10 patients (50 images). From-scratch annotation required approximately 9.1 min per patient. Correcting Mask R-CNN proposals reduced this to 2.3 min per patient, with 2.3 edits and 0.2 boundary corrections per image, while SAT proposals

required 1.6 min per patient, with 1.4 edits and 0.1 boundary corrections per image. These timings are illustrative rather than a population-level estimate.

6 Conclusion

Our experiments show that restricting the image to the intraoral region does not improve FDI-aware instance segmentation, suggesting that the remaining errors are not primarily driven by irrelevant image context.

IOP-Compass turns routine five-view intraoral photography into a benchmarked, tooth-resolved resource: standardized clinical photographs with expert-verified FDI instance masks, frozen patient-disjoint splits, a human-in-the-loop annotation application, and a shared evaluator. View classification is close to saturation on this benchmark, while FDI-aware instance segmentation remains the main challenge and now has a reproducible baseline to build on. By reporting FDI-aware instance quality per view rather than aggregate overlap alone, IOP-Compass provides both a measurable target and the dataset, tool, and evaluator needed to support further progress.

Acknowledgments. This project received funding from the EU co-funded Future-Data4EU project, Grant No. 101126733. The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. This project also received funding from Fondazione di Modena, through the FAR 2024 (E93C24002080007), and from MUR, under the NRRP “Fit4MedRob-Fit for Medical Robotics” (PNC0000007).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmed, S.M.F., Ghori, M.H., Khalid, A., Nooruddin, A., Adnan, N., Lal, A., Umer, F.: Annotated intraoral image dataset for dental caries detection. *Scientific Data* **12** (2025)
2. Ben-Hamadou, A., Neifar, N., Rekik, A., Smaoui, O., Bouzguenda, F., Pujades, S., Boyer, E., Ladoit, E.: Teeth3DS+: An Extended Benchmark for Intraoral 3D Scans Analysis. *arXiv preprint arXiv:2210.06094* (2022)
3. Bolelli, F., Lumetti, L., van Nistelrooij, N., Vinayahalingam, S., Di Bartolomeo, M., Marchesini, K., Pellacani, A., Candeloro, E., Rosati, G., Xi, T., Isensee, F., Kirchhoff, Y., Krämer, L., Rokuss, M., Ulrich, C., Maier-Hein, K., Jiang, Y., Liu, Y., Wang, L., Wang, H., Chen, S., Cui, Z., Shi, P., Pan, Z., Liang, X., Ma, Q., Konukoglu, E., Wodzinski, M., Müller, H., Mai, H., Dang, X., Bhandary, S., Grosu, R., Bergé, S., Anesi, A., Grana, C.: Multi-Structure Segmentation in CBCT Volumes: the ToothFairy2 Challenge. *Medical Image Analysis* (2026)
4. Bolelli, F., Lumetti, L., Vinayahalingam, S., Di Bartolomeo, M., Pellacani, A., Marchesini, K., van Nistelrooij, N., van Lierop, P., Xi, T., Liu, Y., Xin, R., Yang, T., Wang, L., Wang, H., Xu, C., Cui, Z., Wodzinski, M., Müller, H., Kirchhoff,

- Y., Rokuss, M., Maier-Hein, K., Han, J., Kim, W., Ahn, H.G., Szczepański, T., Grzeszczyk, M.K., Korzeniowski, P., Caselles Ballester, V., Burgos-Artizzu, X.P., Prados Carrasco, F., Bergé, S., van Ginneken, B., Anesi, A., Grana, C.: Segmenting the Inferior Alveolar Canal in CBCT Volumes: the ToothFairly Challenge. *IEEE Transactions on Medical Imaging* (2024)
5. Borghi, L., Lumetti, L., Cremonini, F., Rizzo, F., Grana, C., Lombardo, L., Bolelli, F.: Bits2Bites: Intra-Oral Scans Occlusal Classification. In: *International Workshop on Oral and Dental Image Analysis* (2025)
6. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris Coll-Vinent, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., et al.: SAM 3: Segment Anything With Concepts. In: *International Conference on Learning Representations* (2026)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2009)
8. Di Bartolomeo, M., Pellacani, A., Bolelli, F., Cipriano, M., Lumetti, L., Negrello, S., Allegretti, S., Minafra, P., Pollastri, F., Nocini, R., Colletti, G., Chiarini, L., Grana, C., Anesi, A.: Inferior alveolar canal automatic detection with deep learning cnns on cbcts: Development of a novel model and release of open-source dataset and algorithm. *Applied Sciences* (2023)
9. Eum, J., Seo, H., Jeong, T., Lee, E., Park, S., Shin, J.: Automated processing of intraoral clinical photographs using deep learning techniques. *Journal of Clinical Pediatric Dentistry* **50** (2026)
10. Fédération Dentaire Internationale: Two-digit system of designating teeth. *International Dental Journal* **21**(1) (1971)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *IEEE International Conference on Computer Vision (ICCV)* (2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)
13. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for MobileNetV3. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 1314–1324 (2019)
14. Kartbak, S.B.A., Özel, M.B., Kocakaya, D.N.C., Çakmak, M., Sinanoğlu, E.A.: Classification of Intraoral Photographs with Deep Learning Algorithms Trained According to Cephalometric Measurements. *Diagnostics* **15**(9) (2025)
15. Ke, L., Ye, M., Danelljan, M., Liu, Y., Tai, Y.W., Tang, C.K., Yu, F.: Segment Anything in High Quality. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2023)
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment Anything. In: *IEEE/CVF International Conference on Computer Vision (ICCV)* (2023)
17. Kuhn, H.W.: The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly* **2**(1-2) (1955)
18. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: *European Conference on Computer Vision (ECCV)* (2014)
19. Lumetti, L., Rizzo, F., Cremonini, F., Candeloro, E., Lombardo, L., Grana, C., Bolelli, F.: Do Multimodal LLMs Understand Intraoral Dental Data? Dataset, Platform, and Baselines. In: *European Conference on Computer Vision (ECCV)* (2026)

20. Lumetti, L., Tan, Z.Q., Borghi, L., van Nistelrooij, N., Rosati, G., Addison, O., Li, Y., Vinayahalingam, S., Grana, C., Bolelli, F.: ToothFairy3: Scaling CBCT Maxillofacial Segmentation to 77 Classes with U-Mamba2. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2026 (2026)
21. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. *Nature Communications* **15** (2024)
22. Nguyen, K.D., Hoang, H.T., Doan, T.P.H., Dao, K.Q., Wang, D.H., Hsu, M.L.: SegmentAnyTooth: An open-source deep learning framework for tooth enumeration and segmentation in intraoral photos. *Journal of Dental Sciences* **20**(2) (2025)
23. van Nistelrooij, N., Krämer, L., Kempers, S., Beyer, M., Bolelli, F., Xi, T., Bergé, S., Heiland, M., Maier-Hein, K.H., Vinayahalingam, S., Isensee, F.: ToothSeg: Robust Tooth Instance Segmentation and Numbering in CBCT using Deep Learning and Self-Correction. *IEEE Journal of Biomedical and Health Informatics* (2025)
24. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems* **28** (2015)
25. Ryu, J., Lee, Y.S., Mo, S.P., Lim, K., Jung, S.K., Kim, T.W.: Application of deep learning artificial intelligence technique to the classification of clinical orthodontic photos. *BMC Oral Health* **22** (2022)
26. Sosnin, E.I., Vasilev, Y.L., Solovyev, R.A., Stempkovskiy, A.L., Telpukhov, D.V., Vasilev, A.A., Amerikanov, A.A., Romanov, A.Y.: AlphaDent: A dataset for automated tooth pathology detection. *Computer Optics* **49**(6) (2025)
27. Uribe, S.E., Issa, J., Sohrabniya, F., Denny, A., Kim, N.N., Dayo, A.F., Chaurasia, A., Sofi-Mahmudi, A., Büttner, M., Schwendicke, F.: Publicly Available Dental Image Datasets for Artificial Intelligence. *Journal of Dental Research* **103**(13) (2024)
28. Wang, J., Suo, R., Zhou, Y.: ML R-CNN: A Location-Aware Network for Tooth Instance Segmentation. In: 15th International Conference on Digital Image Processing (ICDIP) (2023)
29. Wang, P., Gu, H., Sun, Y.: Tooth segmentation on multimodal images using adapted segment anything model. *Scientific Reports* **15** (2025)
30. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., et al.: MICCAI 2023 STS Challenge: A retrospective study of semi-supervised approaches for teeth segmentation. *Pattern Recognition* **170** (2026)
31. Wang, Y., Li, Z., Wu, C., Liu, J., Zhang, Y., Ni, J., Luo, Q., Chen, J., Zhang, H., Liu, J., Han, C., Fu, K., Ji, C., Cai, X., Hao, J., Zheng, Z., Xu, S., Chen, J., Yu, X., Zhang, Q., Qian, D., Wang, S., Zhou, H.: MICCAI STS 2024 challenge: Semi-supervised instance-level tooth segmentation in panoramic X-ray and CBCT images. *Medical Image Analysis* **110** (2026)
32. Zhou, Z., Zhu, J., Zhang, Y., Guan, X., Wang, P., Li, T.: Deep Learning in Dental Image Analysis: A Systematic Review of Datasets, Methodologies, and Emerging Challenges. *CAAI Artificial Intelligence Research* (2026)
33. Zou, B., Wang, S., Liu, H., Sun, G., Wang, Y., Zuo, F., Quan, C., Zhao, Y.: TeethSEG: An Efficient Instance Segmentation Framework for Orthodontic Treatment Based on Multi-Scale Aggregation and Anthropic Prior Knowledge. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)