ToothFairy3: Scaling CBCT Maxillofacial Segmentation to 77 Classes with U-Mamba2

Luca Lumetti[1,⋆], Zhi Qin Tan[2,⋆], Lorenzo Borghi[1], Niels van Nistelrooij[3,4], Gabriele Rosati[1], Owen Addison[2], Yunpeng Li[2], Shankeeth Vinayahalingam[3], Costantino Grana[1], and Federico Bolelli[1,✉]

[1]University of Modena and Reggio Emilia, Italy
[2]King's College London, the United Kingdom
[3]Radboud University Medical Center, the Netherlands
[4]Charité – Universitätsmedizin Berlin, Germany

Abstract. Accurate delineation of maxillofacial anatomy in Cone-Beam Computed Tomography (CBCT) is essential for dental planning, but robust automated segmentation remains challenging due to limited public multi-structure datasets and the high computational burden of 3D deep learning models. We present and release ToothFairy3, a large-scale CBCT benchmark that extends ToothFairy2 with 102 additional fully annotated scans and an expanded taxonomy covering 77 classes, including 32 tooth-specific pulp cavities and small neurovascular structures. ToothFairy3 comprises 582 volumes (over 40 000 annotated objects), 532 released with voxel-level labels and 50 held out for leakage-free, server-side evaluation. We also introduce U-Mamba2, an efficient U-Net-style architecture that inserts a Mamba2 state-space block at the bottleneck to capture global context with favorable computational scaling. Our proposed domain-informed training further improves the learning of maxillofacial anatomies. Across CNN, Transformer, and Mamba baselines, U-Mamba2 achieves competitive Dice/HD95 scores with lower latency and, compared with training on state-of-the-art public CBCT datasets, ToothFairy3-trained models generalize best to the hidden test set, particularly for maxillary structures.

Keywords: Maxillofacial Segmentation · CBCT · Mamba2 · ToothFairy

1 Introduction

Cone-Beam Computed Tomography (CBCT) is the standard imaging modality in dental and maxillofacial practice, providing 3D volumetric information that is essential for tasks such as implant planning, orthognathic surgery, and assessment of neurovascular structures. A central prerequisite for many of these applications is the reliable delineation of multiple structures, including jawbones,

⋆Equal contribution. Authors are allowed to list their names first on their CVs.
✉ Corresponding author: federico.bolelli@unimore.it

teeth, maxillary sinuses, the pharyngeal airway, and canals, directly in the CBCT volume. However, manual voxel-wise annotation is time-consuming, requires expert knowledge, and is particularly error-prone for thin, low-contrast, and highly variable structures (e.g., canals). These challenges motivate automated segmentation methods, yet learning-based approaches critically depend on large, well-curated, and consistently annotated datasets.

Despite increasing interest in CBCT-based maxillofacial segmentation, the dataset landscape remains fragmented. For inferior alveolar canal (IAC) segmentation, several works report sizeable cohorts but do not provide public voxel-level training data [6,17,18,25]. Public releases exist, but they are still limited in scope or rely on mixed supervision, e.g., sparse 2D annotations with only a smaller subset densely annotated in 3D [4,7,8].

A similar situation holds for tooth segmentation in CBCT. While large private cohorts have been collected, public availability is often partial: Cui et al. [10] report thousands of scans, yet only a small portion is released to the community. Other 3D tooth datasets are comparatively small (tens of scans) and/or non-public, limiting generalization across scanners and acquisition protocols [9,12,16,26]. For Intra-Oral Scans (IOS), challenge datasets such as 3DTeethSeg provide a public benchmark [2], but other large IOS collections remain private [27]. Overall, the prevalence of private datasets and reliance on small cohorts hinder fair comparison and slow progress toward clinically reliable models.

Recent benchmarks have significantly advanced the field by releasing multi-structure CBCT datasets and enabling standardized evaluation. ToothFairy2 [3,5] provides a large-scale, densely annotated benchmark with 42 classes, including 480 volumes for training and 50 held-out test volumes evaluated through an online server to avoid train-test leakage and to support fair comparison across methods. In parallel, the Pulpy3D dataset [13] targets internal tooth anatomy, providing 423 segmentation masks with 19 classes focused on pulp structures. While these resources represent major progress, two gaps remain: current public benchmarks still *(i)* provide limited coverage of fine structures like mandibular neurovascular canals, and *(ii)* restrict the field-of-view to the lower jaw, thereby underrepresenting upper-jaw (maxillary) structures.

Alongside data limitations, clinical applicability is increasingly shaped by *efficiency*. Modern 3D segmentation systems often rely on heavy backbones (e.g., attention-based Transformers or large hybrid architectures), which can incur substantial inference time due to a high volumetric resolution, sliding-window processing, and test-time augmentation [20]. This is particularly relevant for multi-class maxillofacial segmentation, where a single inference may involve dozens of structures spanning the full field-of-view. Consequently, improving segmentation accuracy alone is insufficient; reducing latency and computational cost is equally important for integration into clinical workflows.

Contributions. To address these challenges, we introduce *ToothFairy3*,[1] a new dataset that extends the ToothFairy2 release with 102 additional fully annotated CBCT volumes. Beyond scale, we enrich the label taxonomy by adding *32 tooth-*

[1] `https://ditto.ing.unimore.it/toothfairy3/`

Table 1: **Cross-dataset comparison.** Models are trained on Pulpy3D (P3D), ToothFairy2 (TF2), or our ToothFairy3 (TF3) dataset and tested over the private set of TF3. Average DSC (%) is computed over three runs with *std* in [0.5, 3.0] for all the experiments. Blue columns focus on pulp only, green columns represent average performance across all TF2 classes, while orange and red columns report performance on maxillary/mandibular structures from TF2.

Model	**P3D**	**TF3**	$+\Delta$	**TF2**	**TF3**	$+\Delta$	**TF2**	**TF3**	**TF2**	**TF3**
	Pulpy			All			Maxillary		Mandibular	
nnU-Net ResEnc [15]	39.5	73.4	33.9	74.8	78.3	3.5	69.2	73.8	81.0	83.2
MedNeXt [23]	57.5	76.9	19.4	73.8	74.8	1.0	67.4	68.7	80.8	81.6
U-Mamba [21]	36.6	76.3	39.7	75.3	77.6	2.3	69.9	72.3	81.3	83.4

specific pulp classes and *3 small anatomical structures* (i.e., left/right incisive canals, lingual foramen). Overall, ToothFairy3 provides **77** classes across **582** volumes, enabling unified training and evaluation for both maxillofacial structures and internal tooth anatomy.

In addition, we propose *U-Mamba2*,[2] an efficient adaptation of the U-Mamba family [21] tailored for fast, reliable CBCT segmentation. Our design leverages *Mamba2*-style structured state-space modeling (SSM) [11] to capture long-range context with favorable computational scaling, aiming to reduce inference latency without sacrificing accuracy. Moreover, we incorporate dental domain knowledge in model training to enhance the segmentation performance on ToothFairy3.

Finally, we evaluate the proposed dataset and model in a comprehensive experimental protocol, demonstrating *(i)* the benefit of ToothFairy3 as training data compared with the state-of-the-art publicly available alternatives, and *(ii)* that U-Mamba2 achieves competitive Dice/HD95 with improved runtime, moving one step closer to clinically viable automation.

2 The ToothFairy3 Dataset

Overview. ToothFairy3 is a CBCT dataset for multi-structure semantic segmentation of maxillofacial anatomy (Fig. 1). It extends the ToothFairy2 benchmark [5] in both *scale* and *label granularity*, providing 582 fully annotated CBCT volumes. Following a release strategy that supports reproducible research and long-term fair benchmarking, 532 volumes are made available with voxel-level labels, while 50 volumes are kept private for server-side evaluation.[3]

Data sources and acquisition devices. All cases are CBCT scans of the upper and lower jaws. The public portion of the dataset is acquired with a NewTom/NTVGiMK4 CBCT device by Affidea, while the private test set is sourced

[2] https://github.com/zhiqin1998/U-Mamba2
[3] https://toothfairy3.grand-challenge.org/

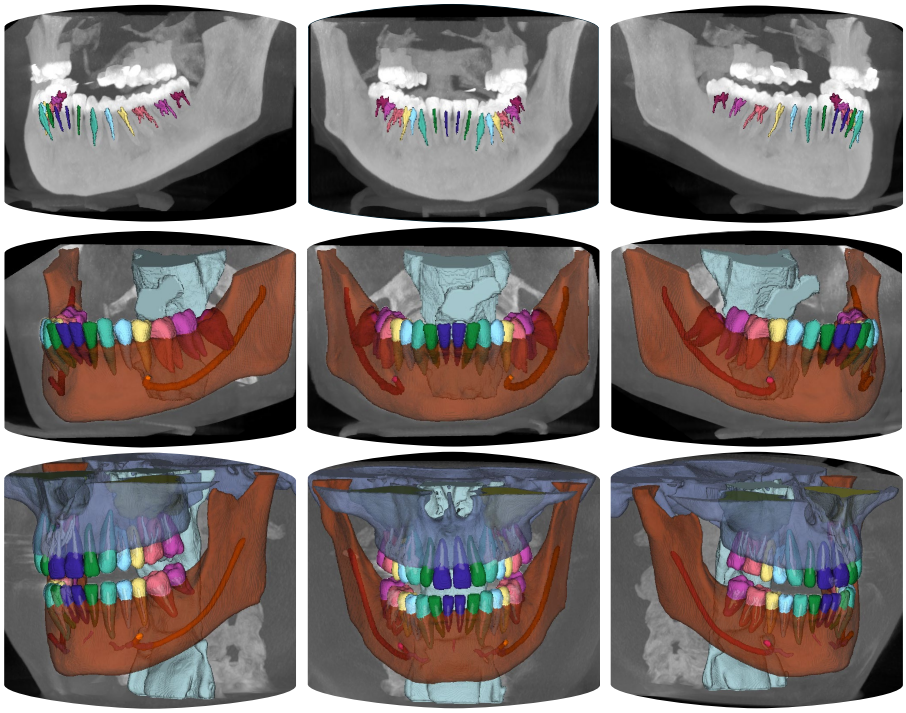

Fig. 1: Sample from **our ToothFairy3 dataset (last row)** in comparison with the Pulpy3D dataset (first row) and the ToothFairy2 dataset (mid row).

from two additional providers to increase heterogeneity in scanners and populations, including Radboud University Medical Center (i-CAT, extended FOV protocol) and 3Shape (PaX-Flex3D and Implagraphy devices by VATECH).

Cohort definition and split principles. A *case* corresponds to a single patient/CBCT volume of the jaws, paired with a dense multi-class label volume. To reduce leakage and to better reflect deployment conditions, the public/private split is designed to separate acquisition devices and patient demographics, discouraging exploitation of scanner- or population-specific cues.

Pre-processing and intensity representation. All volumes are distributed with consistent spatial calibration and intensity representation. Scans are represented in Hounsfield Units (HU) and resampled to a common isotropic voxel size (0.3 mm) to simplify training and evaluation.

Annotation team, workflow, and quality control. Reference segmentations are produced by a team of 7 clinicians with >5 years of experience in the oral and maxillofacial field. To strengthen the independence of the final evaluation, annotators are divided into two disjoint groups: five experts annotate and review the *training* scans, while two experts are reserved exclusively for the *test* set. This separation reduces the risk of information leakage and expectation bias, since

the clinicians who define the reference standard for the held-out test cases have not been exposed to the dataset during development or iterative error analysis. Each scan is annotated by one expert and then *validated* (and corrected when needed) by at least one different expert within the same group. Because only one primary annotation is produced per scan, no label fusion is applied; quality control is ensured through the review-and-correct process.

Label taxonomy and encoding. ToothFairy3 provides a single integer-valued label map per volume. The annotation schema follows ToothFairy2 for the core maxillofacial structures (e.g., jawbones, teeth, sinuses, airway, and major neurovascular canals), ensuring continuity with prior benchmarks. On top of the ToothFairy2 taxonomy, ToothFairy3 introduces additional fine-grained labels across the entire dataset: *(i) tooth pulp cavities*, i.e., 32 tooth-specific classes (one per tooth) and *(ii)* additional *neurovascular structures* including left/right incisive canals and the lingual canal/foramen. The full class list and the corresponding integer encoding are distributed with the dataset metadata, enabling unambiguous evaluation and consistent reporting.

Access model and long-term evaluation. To enable both open research and robust benchmarking over time, ToothFairy3 adopts a dual-release strategy in line with MICCAI standards. The public subset is distributed with voxel-level labels for training and method development. The private subset is not released; instead, evaluation is performed server-side by submitting executable methods (i.e., Docker containers) to compute predictions on the hidden cases and return standardized metrics. This design supports leakage-free evaluation and reduces the risk of overfitting to a fixed public test set.

Ethics. All patient data are pseudonymized prior to distribution. Only minimal non-identifying attributes (e.g., age/sex/year of acquisition) have been retained for cohort balancing and split construction. The use of training data has received approval from "Comitato Etico dell'Area Vasta Emilia Nord (Approval Number 1374/2020/OSS/ESTMO SIRER ID 1275 - NAI-CBCT-D)". The private test cases are not distributed and are used exclusively through server-side evaluation. They are exempt from extensive ethical review by the institutional review board (METC Oost-Nederland, file number 2026-18852).

2.1 ToothFairy3 *vs.* Other Public Datasets

To compare the quality and utility of our ToothFairy3 against existing public CBCT datasets, we conducted a controlled cross-dataset study with ToothFairy2 [5] and Pulpy3D [13]. We selected three representative 3D segmentation architectures spanning different design families: a pure CNN-based model, nnU-Net ResEnc [15], a Transformer-inspired backbone, MedNeXt [23], and a Mamba/SSM-based model, U-Mamba [21]. Each architecture was trained from scratch on *(i)* the full Pulpy3D training set, *(ii)* the full ToothFairy2 training set, and *(iii)* the ToothFairy3 training split, keeping the training recipe fixed across datasets. All models were then evaluated on the private held-out ToothFairy3

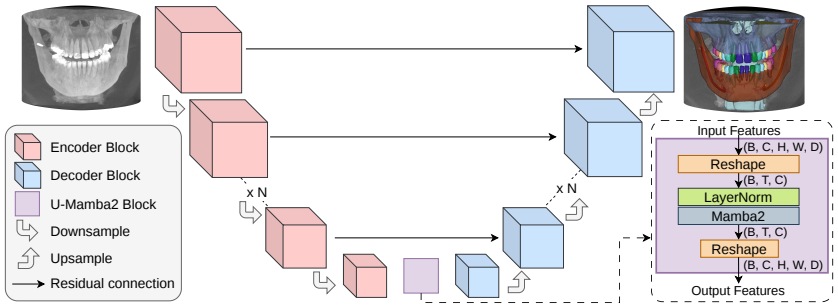

Fig. 2: Overall architecture of the proposed U-Mamba2 model and the details of our U-Mamba2 Block. N: number of stages; B: batch size; C: number of channels; H, W, D: spatial dimensions; T = H × W × D.

test set; since there is no overlap in acquisition devices or annotator groups between ToothFairy3 training and test partitions, the comparison is leakage-free and thus fair. This setup is also practically "forced," as ToothFairy3 is the only dataset that contains the union of structures provided by other benchmarks.

Results are reported in Tab. 1. Because label taxonomies differ across datasets, evaluation is restricted to the overlapping ground truth: for Pulpy3D, we report only the lower-jaw pulp segmentation performance, while for ToothFairy2, we evaluate all ToothFairy2 classes. To reflect the increased field-of-view of ToothFairy3, we additionally report (for ToothFairy2 vs. ToothFairy3) both the overall average and averages confined to upper vs. lower structures. Across architectures, training on ToothFairy3 consistently yields better performance on the ToothFairy3 test set, with the largest gains concentrated in maxillary structures.

3 U-Mamba2

Inspired by [19] and [21], we introduce **U-Mamba2** [24], which combines the strengths of U-Net and Mamba2 to efficiently model global context. As illustrated in Fig. 2, the model adopts a symmetric U-Net encoder-decoder architecture to extract multi-scale features, with residual skip connections between them to fuse low- and high-level image features. The encoder consists of two Residual blocks followed by strided downsampling, while the decoder uses Residual blocks and transposed convolutions for upsampling.

Since convolutions are inherently local, we incorporate Mamba2 to enhance long-range dependency modeling by treating feature maps as sequences. Like Mamba, Mamba2 scales linearly with sequence length, but dramatically improves efficiency by using the state-space duality framework to enforce stronger recurrent structure constraints and replacing selective scan with parallelizable matrix multiplication. In the U-Mamba2 block, features of shape (B, C, H, W, D) are reshaped to (B, T, C) with $T = H \times W \times D$, normalized using Layer Normalization [1], and processed by Mamba2 to capture global context. The output is

Table 2: Results on the test set of the proposed dataset. Classes are grouped by the main anatomical structure to which they belong. The best results are in **bold**, while the second best are underlined.

Model	Average		L/R Canals		Teeth		Jawbones		Pulps	
	DSC	HD95	DSC	HD95	DSC	HD95	DSC	HD95	DSC	HD95
nnU-NetResE. [15]	73.31	78.02	62.16	38.96	80.94	49.25	72.73	86.56	73.43	45.16
nnU-Net [15]	72.28	85.53	63.64	26.08	80.30	56.89	**75.03**	**78.35**	73.31	46.84
MedNeXt [23]	69.77	92.36	59.51	38.41	78.89	59.62	65.58	85.57	**76.87**	23.48
Swin-UNETR [14]	49.54	162.51	43.07	129.20	56.17	129.51	59.80	81.56	50.18	79.41
Primus [28]	60.56	143.55	59.15	32.85	60.63	151.46	73.79	80.19	71.60	**13.15**
U-Mamba [21]	72.68	84.88	**64.96**	26.22	79.97	57.26	73.56	89.30	76.25	36.31
U-Mamba2 [24]	**74.06**	**77.67**	64.91	**23.68**	**81.64**	**49.10**	75.02	78.92	74.28	43.99

reshaped back to (B, C, H, W, D). The U-Mamba2 block is applied only at the bottleneck stage, which yields the best performance for 3D CBCT segmentation [21]. Finally, a Softmax layer produces voxel-wise probabilities.

Dental domain knowledge. Furthermore, to better model anatomical similarities in the maxillofacial region, we replace hard one-hot labels of the ToothFairy3 dataset with *label smoothing* for structurally related classes. As left-right counterparts (e.g., left-right sinuses), neighboring teeth, and closely associated nerves (e.g., inferior alveolar and incisive canals) share morphological and spatial characteristics, each labeled voxel is assigned a softened target distribution: 0.9 probability for the true class and the remaining 0.1 evenly distributed among its predefined related classes. In addition, to address class imbalance for the small left and right incisive canals and the lingual foramen, we apply a higher *loss weight* (10) to these tiny structures so their learning signal is not overpowered by larger anatomies. Finally, although prior findings [5] showed that naïve left-right mirroring can confuse orientation learning due to anatomical symmetry, we leverage this symmetry by explicitly swapping left-right class labels during *left-right mirroring* training augmentation and correspondingly switching predicted logits during test-time augmentation. This strategy expands valid mirroring axes combinations from three to seven while preserving orientation consistency, thereby improving generalization and overall model performance.

4 Experiments

To fully explore the capabilities of state-of-the-art models against our proposed U-Mamba2 on the proposed ToothFairy3 dataset, Tab. 2 and Fig. 3 are reported. For brevity, results are grouped by anatomical structures. All the experiments in this section are repeated with three different seeds and averaged.

Competitors. Experimental analysis has been conducted on recently proposed general-purpose state-of-the-art algorithms for segmenting medical 3D volumes (i.e., CNNs, Transformers, and Mamba-based hybrid solutions). For CNNs, we

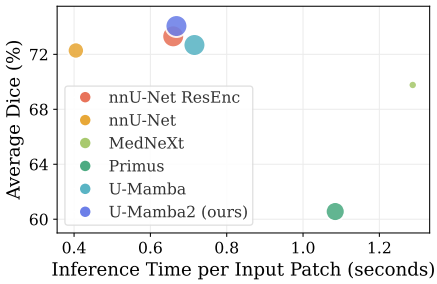

Fig. 3: Inference time *vs* DSC. Larger circles indicate more parameters.

Table 3: Ablation study on dental domain knowledge when introducing *Label Smoothing* (LS), *Weighted Loss* (WL), and/or *Left-Right Mirroring* (LR) introduced in Sec. 3.

LS	WL	LR	DSC	HD95
✓	✗	✗	71.22	87.66
✗	✓	✗	70.25	92.64
✗	✗	✓	72.97	85.02
✓	✓	✓	**74.06**	**77.67**

include both the plain U-Net [22] architecture (nnU-Net [15]), the nnU-Net variations leveraging residual connections in the encoder (nnU-Net ResEnc [15]), and MedNeXt [23]. The considered Transformer-based architectures include Swin-UNETR [14] and Primus [28]. Finally, we include U-Mamba [21] as a representative of state-space model-based solutions.

Experimental setting. We implement all models, including U-Mamba2, in the nnU-Net framework. Specifically, we leverage the planning provided by the nnU-Net framework, modifying only the model configuration for each architecture. All models adopt a patch size of $160 \times 320 \times 320$, except for MedNeXt and Swin-UNETR, which have a patch size of $128 \times 224 \times 224$ due to memory constraints. We kept the training recipe fixed across the different models, i.e., employing the CT Normalization scheme of the nnU-Net framework, a voxel spacing of 0.3 mm, left-right mirroring augmentation as described in Sec. 3, a batch size of 2 (simulated with gradient accumulation), a learning rate of 0.01, SGD with momentum optimizer, and a polynomial learning rate scheduler. The only exception is the Primus model, which follows the original optimization hyperparameters [28] (AdamW and warmup learning rate scheduler). Models are trained from scratch using a combination of cross-entropy and Dice losses on all training samples for 300 epochs with A100 40GB or L40S 48GB GPUs, using PyTorch 2.7.0 and CUDA 12.8.

About dental domain knowledge and time. While the main comparison in Tab. 2 shows that the proposed U-Mamba2 achieves the best overall performance among the evaluated methods with an average Dice score of 74.06 and HD95 of 77.67, it is noteworthy that nnU-Net ResEnc remains the second-best model, further confirming its strength in medical image segmentation. Although U-Mamba2 yields only a modest gain in the overall average Dice/HD95, the improvement is more pronounced for anatomically fine and clinically relevant structures, such as the mandibular neurovascular canals.

In addition to accuracy, a key goal of the proposed approach is to facilitate clinical translation by reducing inference latency without compromising segmentation quality, since runtime and computational cost are critical for integration into routine workflows. Therefore, we measure the average inference time of the

models using an RTX 5090 GPU. The measurement is repeated 100 times on $128 \times 224 \times 224$ input patches, providing a fair comparison regardless of the test-time configuration. Fig. 3 demonstrates the tradeoff between accuracy and inference speed, highlighting the superior performance of U-Mamba2 while maintaining low inference time compared to other methods. Finally, Tab. 3 ablates the effectiveness of dental domain knowledge introduced in Sec. 3.

5 Conclusion

We introduce ToothFairy3, a large-scale 77-class CBCT benchmark with a leakage-free hidden test set, and U-Mamba2, a clinically translation-oriented segmentation model designed for accurate and efficient inference while preserving global context via a Mamba2 bottleneck. Experiments show that U-Mamba2 achieves competitive Dice/HD95 scores and stronger cross-dataset generalization for models trained on the proposed dataset than training on previous state-of-the-art public CBCT datasets, especially for fine maxillary structures.

Acknowledgments. This project has received funding from Fondazione di Modena, through the FAR 2024 (E93C24002080007), and from MUR, under the NRRP "Fit4MedRob-Fit for Medical Robotics" (PNC0000007). Zhi Qin Tan acknowledges the support of an ICASE studentship from the EPSRC.

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

References

1. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer Normalization. arXiv:1607.06450 (2016)
2. Ben-Hamadou, A., Smaoui, O., Rekik, A., Pujades, S., Boyer, E., Lim, H., Kim, M., Lee, M., Chung, M., Shin, Y.G., et al.: 3DTeethSeg'22: 3D Teeth Scan Segmentation and Labeling Challenge. arXiv preprint arXiv:2305.18277 (2023)
3. Bolelli, F., Lumetti, L., van Nistelrooij, N., Vinayahalingam, S., Di Bartolomeo, M., Marchesini, K., Pellacani, A., Candeloro, E., Rosati, G., Xi, T., et al.: 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.M., Müller, H., Kirchhoff, Y., Rokuss, M.R., Maier-Hein, K., Han, J., Kim, W., Ahn, H.G., Szczepanski, 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 CBCTs Volumes: the ToothFairy Challenge. IEEE Transactions on Medical Imaging **44** (2025)
5. Bolelli, F., Marchesini, K., van Nistelrooij, N., Lumetti, L., Pipoli, V., Ficarra, E., Vinayahalingam, S., Grana, C.: Segmenting Maxillofacial Structures in CBCT Volumes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025)

6. Chun, S.Y., Kang, Y.H., Yang, S., Kang, S.R., Lee, S.J., Kim, J.M., Kim, J.E., Kim, K.H., Huh, K.H., Lee, S.S., Lee, M.S., et al.: Automatic classification of 3D positional relationship between mandibular third molar and inferior alveolar canal using a distance-aware network. BMC Oral Health **23**(1) (2023)
7. Cipriano, M., Allegretti, S., Bolelli, F., Di Bartolomeo, M., Pollastri, F., Pellacani, A., Minafra, P., Anesi, A., Grana, C.: Deep Segmentation of the Mandibular Canal: a New 3D Annotated Dataset of CBCT Volumes. IEEE Access **10** (2022)
8. Cipriano, M., Allegretti, S., Bolelli, F., Pollastri, F., Grana, C.: Improving Segmentation of the Inferior Alveolar Nerve through Deep Label Propagation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
9. Cui, W., Wang, Y., Zhang, Q., Zhou, H., Song, D., Zuo, X., Jia, G., Zeng, L.: CTooth: A Fully Annotated 3D Dataset and Benchmark for Tooth Volume Segmentation on Cone Beam Computed Tomography Images. In: International Conference on Intelligent Robotics and Applications (2022)
10. Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., et al.: A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nature Communications **13**(1) (2022)
11. Dao, T., Gu, A.: Transformers are SSMs: Generalized Models and Efficient Algorithms Through Structured State Space Duality. In: Proceedings of the 41st International Conference on Machine Learning (ICML) (2024)
12. Dou, W., Gao, S., Mao, D., Dai, H., Zhang, C., Zhou, Y.: Tooth instance segmentation based on capturing dependencies and receptive field adjustment in cone beam computed tomography. Computer Animation and Virtual Worlds **33**(5) (2022)
13. Gamal, M., Baraka, M., Torki, M.: Automatic Mandibular Semantic Segmentation of Teeth Pulp Cavity and Root Canals, and Inferior Alveolar Nerve on Pulpy3D Dataset. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 (2024)
14. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images (2022), `https://arxiv.org/abs/2201.01266`
15. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2) (2021)
16. Jang, T.J., Kim, K.C., Cho, H.C., Seo, J.K.: A Fully Automated Method for 3D Individual Tooth Identification and Segmentation in dental CBCT. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10) (2021)
17. Jaskari, J., Sahlsten, J., Jarnstedt, J., Mehtonen, H., Karhu, K., Sundqvist, O., Hietanen, A., Varjonen, V., Mattila, V., Kaski, K.: Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes. Scientific Reports **10**(1) (2020)
18. Lahoud, P., Diels, S., Niclaes, L., Van Aelst, S., Willems, H., Van Gerven, A., Quirynen, M., Jacobs, R.: Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT. Journal of Dentistry **116** (2022)
19. Lumetti, L., Pipoli, V., Marchesini, K., Ficarra, E., Grana, C., Bolelli, F.: Taming Mambas for 3D Medical Image Segmentation. IEEE Access (2025)
20. Ma, J., Li, F., Kim, S., Asakereh, R., Le, B.H., Nguyen-Vu, D.K., Pfefferle, A., Wei, M., Gao, R., Lyu, D., et al.: Efficient MedSAMs: Segment Anything in Medical Images on Laptop. arXiv preprint arXiv:2412.16085 (2024)

21. Ma, J., Li, F., Wang, B.: U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. arXiv preprint arXiv:2401.04722 (2024)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (2015)
23. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: MedNeXt: Transformer-Driven Scaling of ConvNets for Medical Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 (2023)
24. Tan, Z.Q., Zhu, X., Addison, O., Li, Y.: U-Mamba2: Scaling State Space Models for Dental Anatomy Segmentation in CBCT. arXiv preprint arXiv:2509.12069 (2025), https://arxiv.org/abs/2509.12069
25. Usman, M., Rehman, A., Saleem, A.M., Jawaid, R., Byon, S.S., Kim, S.H., Lee, B.D., Heo, M.S., Shin, Y.G.: Dual-Stage Deeply Supervised Attention-Based Convolutional Neural Networks for Mandibular Canal Segmentation in CBCT Scans. Sensors **22**(24) (2022)
26. Van Nistelrooij, N., Krämer, L., Kempers, S., Beyer, M., Bolelli, F., Xi, T., Bergé, S., Heiland, M., Maier-Hein, K.H., Vinayahalingam, S., et al.: ToothSeg: Robust Tooth Instance Segmentation and Numbering in CBCT using Deep Learning and Self-Correction. IEEE Journal of Biomedical and Health Informatics (2026)
27. Vinayahalingam, S., Kempers, S., Schoep, J., Hsu, T.M.H., Moin, D.A., van Ginneken, B., Flugge, T., Hanisch, M., Xi, T.: Intra-oral scan segmentation using deep learning. BMC Oral Health **23**(1) (2023)
28. Wald, T., Roy, S., Isensee, F., Ulrich, C., Ziegler, S., Trofimova, D., Stock, R., Baumgartner, M., Köhler, G., Maier-Hein, K.: Primus: Enforcing Attention Usage for 3D Medical Image Segmentation. arXiv preprint arXiv:2503.01835 (2025)