Robust Tooth Segmentation Under Orthodontic CBCT: A Metal Artifact-Aware Approach

SeungKwan Jo[1], Jeonglok Park[1], So Hyun Kim[1], and Minyoung Chung[2†]

[1] Soongsil University, Seoul, Republic of Korea
[2] Osstem Implant, Seoul, Republic of Korea
[†]Corresponding author: chungmy.contact@gmail.com

Abstract. Tooth instance segmentation in Cone-Beam Computed Tomography (CBCT) is essential for clinical diagnosis, treatment planning, and prognosis prediction. Practical pipelines commonly adopt a two-stage detection–segmentation framework, where accurate bounding boxes are critical: incorrect boxes either incompletely cover the target tooth or include adjacent teeth, increasing input variability and destabilizing subsequent segmentation. However, orthodontic CBCT suffers from severe metal artifacts that corrupt structural cues, hinder robust detection, and degrade final segmentation. To address this, we propose a two-stage framework that leverages metal artifacts as informative cues for robust tooth localization while suppressing their residual influence during boundary delineation. Specifically, we improve bbox localization by minimizing the intensity-weighted center-of-mass discrepancy between predicted and ground-truth patches, and stabilize boundary delineation by regressing a normalized distance map to suppress artifact responses near the tooth surface. Our method achieves robust localization and accurate instance segmentation even under severe orthodontic metal artifacts. Our code is publicly available at https://github.com/joosk3R/MAM.

Keywords: Dental CBCT · Orthodontics · Orthodontic brackets · Metal artifact-aware segmentation · Tooth detection · Tooth segmentation

1 Introduction

Cone-Beam Computed Tomography (CBCT) is a core imaging modality in orthodontics, providing high-resolution three-dimensional information on individual teeth and surrounding anatomical structures [1]. Such volumetric data are essential for tooth-level analysis required in diagnosis, treatment planning, and prognosis prediction, particularly when extraction or continuous tooth movement is involved [2]. In clinical workflows, tooth instance segmentation is therefore a prerequisite for quantifying positional and morphological changes of each tooth. However, orthodontic brackets and wires inevitably induce strong metal artifacts, which cause severe distortion and information loss, blur tooth boundaries, and hinder reliable boundary estimation [3]. Consequently, tooth-level annotation in metal artifact–embedded CBCT (MA-CBCT) becomes costly and labor-intensive, motivating robust tooth segmentation methods tailored to orthodontic MA-CBCT.

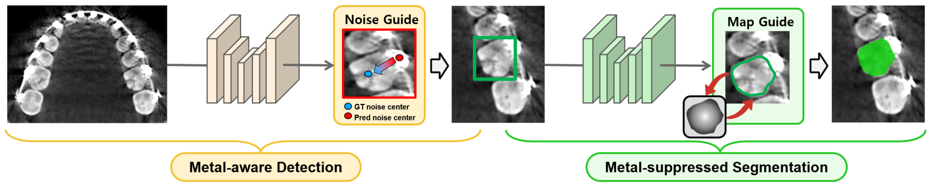

Fig. 1. Overview of artifact-guided detection and artifact-suppressed segmentation.

Many practical pipelines adopt a two-stage (detection-then-segmentation) design for tooth instance segmentation [3–11]. In such pipelines, segmentation is only as good as the detected bounding box (bbox): if the box cuts off part of the target tooth or includes neighboring teeth, the cropped input becomes noisy and inconsistent, directly degrading the downstream segmentation. More accurate bboxes provide cleaner tooth-centered crops and improve the stability of the subsequent segmentation.

To improve detection, prior work has primarily exploited structural or morphological cues—such as tooth boundaries [5, 6], skeleton representations [6, 9], or tooth centers/landmarks [6, 8, 9, 11]—by explicitly learning these auxiliary signals. While effective under limited artifact conditions, orthodontic MA-CBCT often exhibits severe boundary corruption and shape degradation, making such cues unreliable and difficult to model consistently. As a result, detection strategies driven mainly by explicit structural cues frequently degrade when the image evidence required by these cues is corrupted, leading to inaccurate bboxes and unstable downstream segmentation.

In our work, we propose a two-stage framework that leverages metal artifacts to improve detection robustness while mitigating their residual influence during segmentation. In the detection stage, we introduce an artifact-aware localization loss that penalizes discrepancies between the intensity distributions of tooth patches cropped by the predicted and ground-truth bboxes. Specifically, we summarize the intensity distribution inside each patch using an intensity-weighted center of mass and improve bbox localization under metal noise by reducing the centroid discrepancy between the predicted and ground-truth patches. In the segmentation stage, we take the bbox-cropped tooth patch and regress a distance map to stabilize boundary delineation and suppress false positives caused by responses near the tooth surface. By handling metal artifacts differently across detection and segmentation, the proposed framework is designed to improve the robustness of tooth instance segmentation under orthodontic MA-CBCT.

In summary, the main contributions of this study are as follows:

- We achieve robust tooth instance segmentation on a real clinical orthodontic MA-CBCT dataset under severe metal artifacts.
- Artifact-robust tooth detection that leverages metal artifacts as informative cues for bounding box localization, rather than treating them as noise.
- A two-stage design that decouples artifact-guided detection from artifact-suppressed segmentation for reliable instance segmentation.

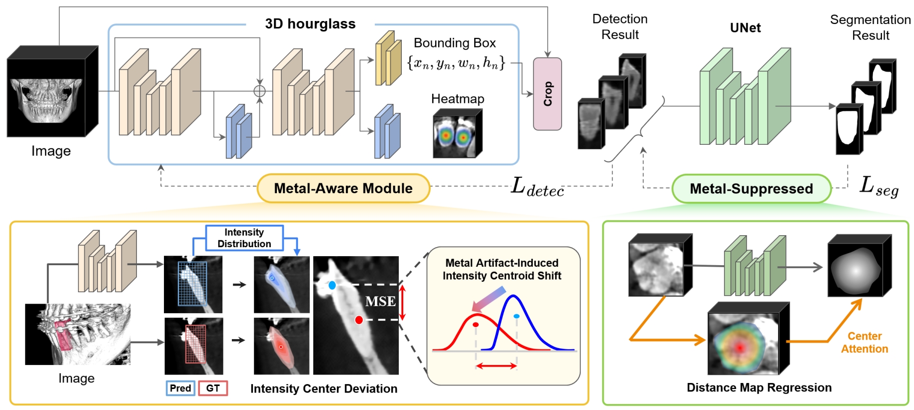

Fig. 2. A heatmap-based 3D hourglass detector predicts per-tooth center heatmaps and axis-aligned bounding boxes (bboxes) [8]. For metal-aware localization, tooth patches cropped by the predicted and GT bboxes are used to compute an intensity-weighted center of mass C_{mass} from voxel samples (x, y, z, I) in absolute coordinates, and the centroid discrepancy (MSE) is minimized to improve bbox localization under severe artifacts. The detected bboxes are then used to crop tooth patches for a 3D U-Net [12]. The segmentation network jointly predicts the tooth mask and a normalized Chamfer distance map (1 at the tooth center, decreasing toward 0 at the boundary/background), which de-emphasizes ambiguous boundary-adjacent voxels and reduces boundary false positives.

2 Method

2.1 Overview of Framework

As illustrated in Fig. 2, we propose a two-stage framework for orthodontic CBCT that performs individual tooth detection followed by tooth instance segmentation, leveraging metal artifacts as localization cues while suppressing their influence during boundary delineation. Given an input CBCT volume, we first apply a CNN-based Volume of Interest (VOI) cropping module to retain only the tooth-containing region [3], then use a heatmap-based detector to predict per-tooth bounding boxes (bboxes) and improve localization robustness via an artifact-aware localization loss. Finally, each detected tooth patch is segmented by a U-Net-based model to produce the final instance masks [12].

2.2 Metal Artifact-Aware Tooth Detection

In orthodontic CBCT, bracket-induced metal artifacts often appear as high-intensity structures and can be confused with tooth boundaries, which may enlarge or shift predicted bboxes. Consequently, the intensity distribution inside a predicted tooth patch can differ from that inside the corresponding ground-truth (GT) tooth patch. We exploit this discrepancy by summarizing each tooth patch with a single representative *intensity centroid* and penalizing the discrepancy between the predicted and GT centroids, thereby regularizing bbox localization under severe artifacts.

For each tooth, the detector predicts a 3D bbox parameterized as

$$B = (x, y, z, w, h, d), \tag{1}$$

where (x, y, z) denotes the minimum (corner) coordinate of the box and (w, h, d) are box sizes along each axis. This bbox defines a tooth patch $P(B)$ obtained by cropping the original CBCT volume within the box extent.

Let V denote the original (pre-crop) CBCT volume. For each voxel inside $P(B)$, we form a sample

$$(x_i, y_i, z_i, v_i), \quad v_i = V(x_i, y_i, z_i), \tag{2}$$

where (x_i, y_i, z_i) are absolute voxel coordinates in V (not patch-local coordinates), and v_i is the (preprocessed) voxel intensity. Let N be the number of voxels in $P(B)$.

We treat voxel intensities within the cropped tooth patch as a discrete mass distribution and compute an intensity-weighted center of mass (intensity centroid):

$$C_{\text{mass}}(B) = \frac{1}{\sum_{i=1}^{N} \tilde{v}_i} \sum_{i=1}^{N} \tilde{v}_i \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}, \qquad \tilde{v}_i = \max(v_i, 0). \tag{3}$$

Because (x_i, y_i, z_i) are absolute coordinates in the original volume, $C_{\text{mass}}(B)$ is expressed in the same coordinate system, enabling direct comparison between predicted and GT tooth patches. The rectification $\tilde{v}_i$ avoids negative/background intensities that can destabilize the denominator.

Let B_{pred} and B_{GT} be the predicted and GT bboxes, respectively. We define the artifact-aware localization loss as

$$L_{\text{artifact}} = \left\| C_{\text{mass}}(B_{\text{pred}}) - C_{\text{mass}}(B_{\text{GT}}) \right\|_2^2, \tag{4}$$

which penalizes metal artifact-induced centroid discrepancies between the predicted and GT tooth patches, thereby improving bbox localization robustness under severe artifacts.

2.3 Metal Artifact-Suppressed Tooth Segmentation

Due to the cuboidal nature of extracted bboxes, residual metal-induced noise may remain near tooth boundaries (Fig. 1(b)). To alleviate this issue, we regress a normalized distance map of the GT mask computed using a Chamfer distance transform [3, 13], where values are 1 at the tooth center and decrease toward 0 at the tooth boundary (and background), scaled to $[0, 1]$. This yields a *boundary-ignore* effect: ambiguous boundary-adjacent voxels receive small regression targets, while reliable interior voxels provide larger targets. As a result, the network is encouraged to rely on stable interior tooth evidence and suppress spurious responses near metal-contaminated boundaries, reducing over-segmentation and boundary false positives.

The distance map is computed from the GT mask using a Chamfer distance transform on the voxel grid [13]. The distance map loss is defined as

$$L_{\text{dist}} = \|D_{\text{pred}} - D_{\text{true}}\|_2^2 , \tag{5}$$

where D_{true} is the GT distance map and D_{pred} is the predicted distance map.

2.4 Learning Model

Learning tooth detection For tooth detection training, we supervise the predicted axis-aligned bbox B and the center C using GT annotations. To ensure robust center localization under severe artifact noise, we adopt a heatmap-based keypoint detection formulation that stably models the spatial uncertainty of tooth centers [8].

Each tooth center is encoded as a Gaussian heatmap $\hat{H}$. We apply a focal loss between the predicted heatmap H and the GT heatmap $\hat{H}$ to emphasize hard samples during heatmap learning [14]:

$$L_{\text{heatmap}} = -\sum_i \left(\hat{H}_i (1 - H_i)^\gamma \log(H_i) + (1 - \hat{H}_i) H_i^\gamma \log(1 - H_i) \right), \tag{6}$$

where H_i and $\hat{H}_i$ denote the predicted and GT heatmap values at location i, respectively, and γ is a hyperparameter controlling the weighting of difficult samples.

In addition, we regress the bbox size and position using a Smooth L_1 loss [15]:

$$L_{\text{bbox}} = \sum_i L_{\text{smooth}}\left(b_i - \hat{b}_i\right), \qquad L_{\text{smooth}}(x) = \begin{cases} \frac{1}{2}x^2, & |x| < 1, \\ |x| - \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{7}$$

Finally, the detection-stage objective combines the heatmap loss, artifact-aware localization loss, and bbox regression loss:

$$L_{\text{detec}} = \alpha L_{\text{heatmap}} + \beta L_{\text{artifact}} + \lambda L_{\text{bbox}}, \tag{8}$$

where α, β, and λ are coefficients controlling the contribution of each loss term.

Learning tooth segmentation For training and inference, we extract a 3D tooth patch for each tooth using the detected bbox and feed it into a 3D U-Net to obtain voxel-wise instance segmentation [12]. The 3D U-Net follows an encoder–decoder architecture and leverages multi-resolution features to recover fine tooth morphology [12]. The final segmentation objective combines a Dice loss [16] and the distance map loss:

$$L_{\text{seg}} = \lambda_{\text{DSC}} L_{\text{DSC}} + \lambda_{\text{dist}} L_{\text{dist}}, \tag{9}$$

where λ_{DSC} and λ_{dist} control the importance of each term.

3 Results

Dataset, Implementation, and Metrics We collected 20 de-identified clinical 3D CBCT scans acquired from patients wearing orthodontic appliances (40 jaws; 520 teeth) and performed 5-fold cross-validation with an 80:20 train:test split in each fold, reporting mean $\pm$ standard deviation over the five folds. Each volume was windowed with a window level of 1500 and a window width of 4000, normalized to $[0, 1]$ to facilitate model training, VOI-cropped for background removal and orientation alignment [3], processed separately for the maxilla and mandible, and resampled to $128 \times 128 \times 64$. Following [8], we performed tooth detection using tooth-size-adaptive Gaussian heatmaps (32 bboxes over the full dentition), assigned anatomical tooth labels to the detected candidates using a ResNet-based tooth-number classifier [17], cropped each labeled tooth region using the crop module in Fig. 2, and segmented teeth with a U-Net that regresses a chamfer distance map [12, 13]. Models were implemented in PyTorch and trained on an NVIDIA A6000 GPU using Adam [18] (lr = 0.001) for 150/200 epochs with batch sizes of 1/32 (detection/segmentation). We evaluated detection using Average Precision at IoU 0.5 (AP_{50}) [19], the over-identified rate (OIR) [3], and mean Intersection-over-Union (mIoU) [3], and assessed segmentation using the Dice Similarity Coefficient (DSC) for overlap and the average symmetric surface distance (ASD) and Hausdorff distance (HD) for surface errors [16].

3.1 Comparison with State-of-the-Art Methods

We conducted quantitative and qualitative comparisons against CNN-based baselines for tooth detection and instance segmentation within the scope of representative two-stage detection–segmentation pipelines: (i) ToothNet (3D-RPN; boundary cues) [5], (ii) PATRCNN (pose readjustment to reduce bbox overlap) [3], and (iii) CGDNet (DenseASPP-UNet [20]; tooth-center guidance) [11]. This established two-stage paradigm provides a fair and practically relevant setting for assessing the applicability of our method within existing tooth detection and instance segmentation workflows. By evaluating against these methods, we demonstrate the challenges of MA-CBCT segmentation and verify that our metal guidance and suppression strategy across the two-stage pipeline yields quantitative improvements in detection and segmentation accuracy.

Tooth Detection Performance Table 1 reports AP_{50}, OIR, and mIoU as mean $\pm$ standard deviation. ToothNet [5] and CGDNet [11] show limited performance under metal artifacts, indicating that structural degradation caused by artifact noise corrupts structural cues and degrades detection performance. PATRCNN [3] improves detection bboxes via pose readjustment, but exhibits limited performance due to the lack of a direct mechanism for handling artifact noise. In contrast, the proposed method achieves $AP_{50} = 91.1$, OIR = 0.946, and mIoU = 0.742, demonstrating superiority over other methods. This suggests that our approach, which uses distribution differences of metal artifacts as guidance, enables robust and precise detection under MA-CBCT conditions.

Table 1. Comparison of detection and segmentation performance (mean ± std). Best results are highlighted in **bold**.

Method	Detection			Segmentation		
	AP_{50} ↑	**OIR ↑**	**mIoU ↑**	**DSC ↑**	**HD ↓**	**ASD ↓**
ToothNet [5]	78.1 (±1.03)	0.826 (±0.098)	0.595 (±0.063)	83.3 (±0.84)	3.38 (±0.91)	0.497 (±0.108)
CGDNet [11]	83.2 (±1.19)	0.873 (±0.106)	0.634 (±0.054)	87.1 (±0.91)	3.26 (±1.02)	0.384 (±0.118)
PATRCNN [3]	87.4 (±0.91)	0.923 (±0.082)	0.712 (±0.059)	90.4 (±0.67)	2.12 (±0.67)	0.271 (±0.051)
Ours	**91.1 (±0.78)**	**0.946 (±0.078)**	**0.742 (±0.038)**	**94.6 (±0.52)**	**1.45 (±0.44)**	**0.164 (±0.032)**

Table 2. Tooth-group-wise segmentation performance (mean only). Anterior includes incisors and canines. Best results are highlighted in **bold**.

Method	DSC (%) ↑			HD (mm) ↓			ASD (mm) ↓		
	Ant.	**Pre.**	**Mol.**	**Ant.**	**Pre.**	**Mol.**	**Ant.**	**Pre.**	**Mol.**
ToothNet [5]	81.1	85.1	83.3	3.65	3.22	3.27	0.522	0.477	0.492
CGDNet [11]	85.2	88.4	87.7	3.43	3.08	3.27	0.421	0.361	0.370
PATRCNN [3]	88.1	91.9	91.2	2.31	2.06	1.99	0.294	0.261	0.258
Ours	**94.5**	**94.9**	**94.4**	**1.45**	**1.47**	**1.43**	**0.166**	**0.169**	**0.157**

Tooth Segmentation Performance The proposed method achieves the best scores across all metrics, indicating that stable tooth localization benefits downstream segmentation. In particular, superior performance on boundary metrics (HD/ASD) shows that the distance map regression strategy substantially reduces false positives near boundaries. This trend is also observed in tooth-group-wise segmentation performance. Anterior teeth are more susceptible to structural degradation due to differences in the tooth–metal artifact size ratio, and accordingly, existing models exhibit lower performance in the order of anterior, premolar, and molar groups. In contrast, the proposed distance map regression efficiently suppresses boundary-adjacent metal noise, leading to consistently similar performance across the anterior, premolar, and molar groups. Specifically, the proposed method achieves DSC = 94.6%, HD = 1.45 mm, and ASD = 0.164 mm, demonstrating that our segmentation strategy is robust even under residual noise near boundaries.

Qualitative results show that existing methods exhibit failure cases such as tooth misdetection/misclassification and partially missing regions. Methods that embed structural cues are particularly vulnerable, which is evident in the relatively challenging anterior cases discussed above. ToothNet [5] infers metal artifacts as tooth segmentation labels, whereas CGDNet [11] fails to reliably

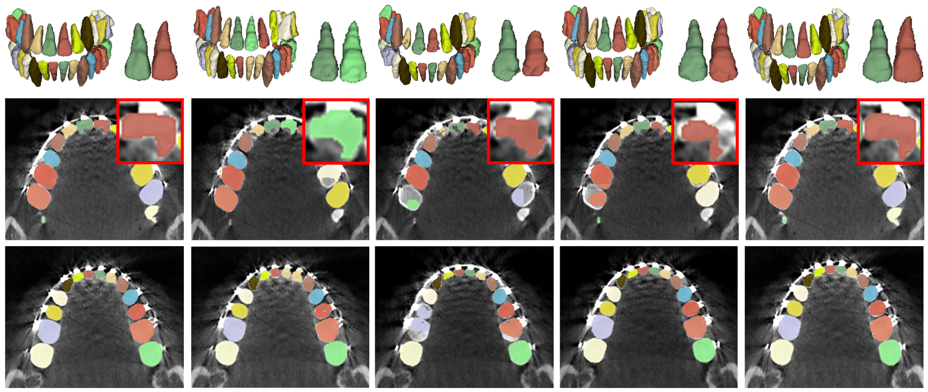

Fig. 3. Qualitative comparison of segmentation results produced by different methods along with the ground truth.

Table 3. Ablation study on metal-aware components. Best results are highlighted in **bold**. Detection metrics for MAM+DM are omitted since DM is applied only to segmentation. "O" and "X" indicate enabled and disabled components, respectively.

Ablation setting		Detection			Segmentation		
MAM	DM	AP_{50} ↑	OIR ↑	mIoU ↑	DSC ↑	HD ↓	ASD ↓
X	X	84.1	0.881	0.654	87.3	3.44	0.402
O	X	**91.1**	**0.946**	**0.742**	91.9	2.28	0.303
O	O	–	–	–	**94.6**	**1.45**	**0.164**

localize anterior centers, resulting in low-quality label predictions. In contrast, the proposed method largely prevents misclassification and partially missing regions, demonstrating that our metal guidance and suppression strategy across the two-stage pipeline produces high-quality segmentation labels.

3.2 Ablation Study

To evaluate performance accurately, we conducted an ablation study by sequentially adding the metal artifact-aware module (MAM) for robust detection and the distance map (DM) regression strategy to suppress boundary-adjacent false positives. Adding MAM, which leverages metal artifacts as guidance, improves detection to $AP_{50} = 91.1$, mIoU $= 0.742$, and OIR $= 0.946$, and yields a $+4.6$ DSC-point gain while reducing HD/ASD by 33.7%/24.6%, suggesting more reliable tooth-centered patches for downstream segmentation. DM further improves DSC by $+2.7$ points and reduces HD/ASD to 1.45 mm/0.164 mm (additional 36.4%/45.9% reduction), confirming substantial suppression of boundary false positives.

4 Conclusion

We propose a two-stage processing framework for segmentation in MA-CBCT. In particular, we introduce a robust detection scheme that leverages metal artifacts as localization cues and integrate a patch-wise segmentation strategy to suppress residual noise near tooth boundaries, thereby demonstrating robust and accurate segmentation performance. However, CBCT segmentation for patients with orthodontic appliances remains underexplored, with limited large-scale standardized datasets. Further validation on larger, more diverse datasets and extensions to various orthodontic conditions are needed.

Acknowledgments. This work was supported by the Materials and Components Technology Development Program funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and Korea Evaluation Institute of Industrial Technology (KEIT) (Project Number: RS-2023-KT224280).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. van Ginneken, B., Romeny, B.T.H., Viergever, M.A.: Computer-aided diagnosis in chest radiography: a survey. IEEE Transactions on Medical Imaging **20**(12), 1228–1241 (2001)
2. Khanagar, S.B., Al-Ehaideb, A., Maganur, P.C., Vishwanathaiah, S., Patil, S., Baeshen, H.A., Sarode, S.C., Bhandi, S.: Developments, application, and performance of artificial intelligence in dentistry–a systematic review. Journal of dental sciences **16**(1), 508–522 (2021)
3. Chung, M., Lee, M., Hong, J., Park, S., Lee, J., Lee, J., Yang, I.H., Lee, J., Shin, Y.G.: Pose-aware instance segmentation framework from cone beam ct images for tooth segmentation. Computers in Biology and Medicine **120**, 103720 (2020)
4. Cui, Z., Li, C., Chen, N., Wei, G., Chen, R., Zhou, Y., Shen, D., Wang, W.: Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model. Medical Image Analysis **69**, 101949 (2021)
5. Cui, Z., Li, C., Wang, W.: Toothnet: automatic tooth instance segmentation and identification from cone beam ct images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6368–6377 (2019)
6. Cui, Z., Zhang, B., Lian, C., Li, C., Yang, L., Wang, W., Zhu, M., Shen, D.: Hierarchical morphology-guided tooth instance segmentation from cbct images. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27. pp. 150–162. Springer (2021)
7. 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), 6562–6568 (2021)
8. Lee, J., Chung, M., Lee, M., Shin, Y.G.: Tooth instance segmentation from cone-beam ct images through point-based detection and gaussian disentanglement. Multimedia Tools and Applications **81**(13), 18327–18342 (2022)

9. Jamal, M.A., Chao, G., Lee, B.: Dentseg-eda: A 3d tooth segmentation framework for cbct images using enhanced dual attention. Neurocomputing p. 131437 (2025)
10. Su, S., Jia, X., Zhan, L., Gao, S., Zhang, Q., Huang, X.: Automatic tooth periodontal ligament segmentation of cone beam computed tomography based on instance segmentation network. Heliyon **10**(2) (2024)
11. Wu, X., Chen, H., Huang, Y., Guo, H., Qiu, T., Wang, L.: Center-sensitive and boundary-aware tooth instance segmentation and classification from cone-beam ct. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 939–942. IEEE (2020)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
13. Borgefors, G.: Distance transformations in digital images. Computer vision, graphics, and image processing **34**(3), 344–371 (1986)
14. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: Proceedings of the European conference on computer vision (ECCV). pp. 734–750 (2018)
15. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
16. Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al.: Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE transactions on medical imaging **28**(8), 1251–1265 (2009)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. 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: Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. pp. 740–755. Springer (2014)
20. Yang, M., Yu, K., Zhang, C., Li, Z., Yang, K.: Denseaspp for semantic segmentation in street scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3684–3692 (2018)