

Hierarchical Morphology-Guided Tooth Instance Segmentation from CBCT Images

Zhiming Cui^{1,6,7}, Bojun Zhang², Chunfeng Lian³, Changjian Li^{4(✉)},
Lei Yang¹, Wenping Wang^{1,5}, Min Zhu^{2(✉)}, and Dinggang Shen^{6,7(✉)}

¹ Department of Computer Science, The University of Hong Kong, Hong Kong, China

² Shanghai Ninth People's Hospital, Shanghai JiaoTong University, Shanghai, China

ZHUM1612@sh9hospital.org.cn

³ School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

⁴ Department of Computer Science, University College London, London, UK

changjian.li@ucl.ac.uk

⁵ Department of Visualization, Texas A&M University, College Station, TX, USA

⁶ School of Biomedical Engineering, ShanghaiTech University, Shanghai, China

dgshen@shanghaitech.edu.cn

⁷ Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. Automatic and accurate segmentation of individual teeth, i.e., tooth instance segmentation, from CBCT images is an essential step for computer-aided dentistry. Previous works typically overlooked rich morphological features of teeth, such as tooth root apices, critical for successful treatment outcomes. This paper presents a two-stage learning-based framework that explicitly leverages the comprehensive geometric guidance provided by a hierarchical tooth morphological representation for tooth instance segmentation. Given a 3D input CBCT image, our method first learns to extract the tooth centroids and skeletons for identifying each tooth's rough position and topological structures, respectively. Based on the outputs of the first step, a multi-task learning mechanism is further designed to estimate each tooth's volumetric mask by simultaneously regressing boundary and root apices as auxiliary tasks. Extensive evaluations, ablation studies, and comparisons with existing methods show that our approach achieved state-of-the-art segmentation performance, especially around the challenging dental parts (i.e., tooth roots and boundaries). These results suggest the potential applicability of our framework in real-world clinical scenarios.

1 Introduction

Computer-aided design (CAD) has been widely used in digital dentistry for diagnosis, restoration, and orthodontic treatment planning. In these processes, 3D tooth models, typically segmented from cone beam computed tomography

Z. Cui and B. Zhang—Equal contribution.

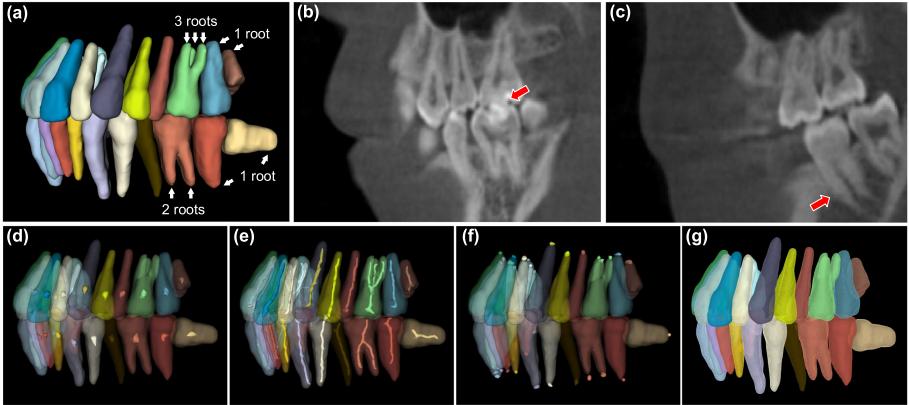


Fig. 1. The first row shows three typical examples, including (a) teeth with large shape variations; (b) touching boundaries of maxillary and mandibular teeth during close bite; (c) blurred signals between tooth roots and the surrounding alveolar bones. The second row illustrates different components of the hierarchical morphological representations, including the points (i.e., the tooth centroid (d) and the root landmarks at root apices (f)), the tooth skeleton (e), and the tooth boundary surface (g).

(CBCT) images [4,5], are essential to assist dentists in extracting, implanting, or rearranging teeth. In clinical practice, dentists need to manually label each tooth slice-by-slice from the CBCT images, which is laborious and time-consuming, and also highly depends on an operator's experience. Thus, it is practically demanded of accurate and fully automatic methods to segment individual teeth from dental CBCT images.

However, automatic segmentation of individual teeth is still a challenging task as teeth exhibit large variations in their geometry. For example, maxillary molars usually have three roots, while mandibular molars usually have two roots [8] (see Fig. 1(a)). Beyond the general rules, special cases where molars have one root can also be found (Fig. 1(a)) and such disparities are fairly commonplace in the real-world clinics. Even state-of-the-art learning-based methods [4,5,14] often fail to handle such complicated cases. This is mainly because such methods employ only simple representations (e.g., the tooth centroid or the bounding box) for teeth and thus cannot capture detailed shape variations in each tooth. This is even worse at the regions where image contrast is low, such as the common boundary of touching teeth during close bite (Fig. 1(b)) and the interface between tooth roots and their surrounding alveolar bone (Fig. 1(c)). Without any prior knowledge of the tooth structure, either traditional methods [1,2,6,7,9,11,15] or learning-based networks [4,5,14] cannot properly segment the tooth from the background tissue at these regions, although the tooth root information is critical in orthodontic treatment to ensure apices cannot penetrate the surrounding alveolar bone during tooth movement.

In this paper, we propose a hierarchical morphological representation for capturing the complicated tooth shapes and important tooth features. Specifically, this hierarchical morphological representation consists of the tooth centroid and root apices (i.e., points), skeleton, boundary surface, and volume (Fig. 1(d)–(g), (a)). Based on this hierarchical representation, we design a coarse-to-fine learning-based framework for automatic and accurate tooth instance segmentation. Given a 3D input CBCT image, to capture the positions and varying topological structures of all individual teeth, especially at the multi-root areas, a neural network at the 1st-stage (at the coarse level) is designed to predict tooth centroids and skeletons, respectively. Then, a multi-task network is further proposed at the 2nd-stage to simultaneously predict the detailed geometric features, i.e., root landmarks (or apices), boundary surface, and volumetric mask of each tooth using the tooth skeletons estimated at the 1st-stage as guidance. Since the three tasks are intrinsically related from a geometric perspective, regressing the root landmarks and boundary surface of each tooth can intuitively boost the segmentation performance at the important and challenging regions (e.g., the tooth boundary and root apices). The performance of our method was evaluated through extensive experiments on a CBCT dataset collected from real-world clinics. The corresponding results showed that our method significantly outperformed other state-of-the-art approaches, suggesting the efficacy of the hierarchical morphological representations designed in this study for tooth instance segmentation.

2 Methods

The proposed framework consists of two stages. At the 1st-stage, a prediction network is designed to extract the coarse-level morphological representations, i.e., the centroid and skeleton of each tooth, to represent the tooth structure. At the 2nd-stage, a segmentation network with the coarse-level morphological guidance (the tooth skeleton) is trained with a multi-task learning mechanism to generate a detailed tooth volume, boundary, and root landmarks. The schematic diagrams of these two steps are shown in Figs. 2 and 3, respectively, with the details elaborated below.

2.1 Tooth Centroid and Skeleton Extraction Network

As the centroid and skeleton of a tooth define its spatial location and topological structure, respectively, the network in this step aims to achieve the following goals: 1) localize individual teeth by identifying their centroids, and then 2) capture their topological structures by predicting their skeletons.

Given the 3D input CBCT image I , two sub-networks are designed to this end, as shown in Fig. 3. Each sub-network contains two output branches to produce a binary segmentation map B and a 3D offset map O . Specifically, the binary segmentation map B indicates whether a voxel of the input CBCT image belongs to a foreground tooth object or the background tissue (denoted as B_c

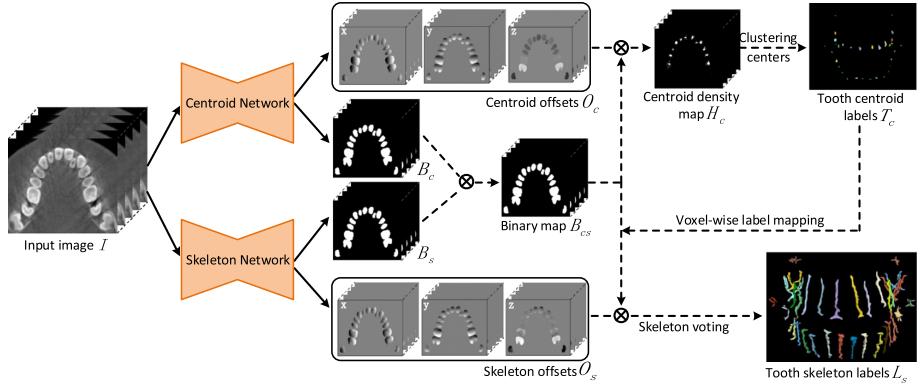


Fig. 2. The pipeline of our 1st-stage network for tooth skeleton extraction. The CBCT scan is first fed into both the centroid network and the skeleton network to generate the offsets and binary maps, respectively. Then, the tooth centroids and skeletons are detected and predicted by the later steps (dotted lines).

and B_s of the centroid sub-network and the skeleton sub-network, respectively). The 3D offset map O indicates the 3D vector pointing from each foreground voxel to its target point. Here, for each voxel, its target point in centroid offset map O_c refers to a vector pointing to the centroid of the corresponding tooth, while in skeleton offset map O_s the target point is defined as a vector pointing to the nearest point on the skeleton of the corresponding tooth.

With the outputs of the two sub-networks, we detect the tooth centroids and skeletons as follows. First, the common binary map B_{cs} is produced by the element-wise product of B_c and B_s , which masks out the foreground voxels shared by both centroid and skeleton offsets. Then, we generate a tooth centroid density map H_c by counting the frequency of a voxel being pointed by other voxels according to the 3D centroid offset map. Finally, we adopt a fast search clustering method [13] to localize the peaks in H_c as the predicted tooth centroids, denoted as T_c . The rationale is that the clustering centers usually have relatively high density values (i.e., frequency) and large distance to the nearest point with a higher density value, defined as:

$$T_c = (H_c^i > \delta) \cap (DT_c^i > \lambda), \quad (1)$$

where DT_c^i refers to the distance between voxel i and its nearest voxel with higher density value than H_c^i . The scalars $\delta = 20$ and $\lambda = 10$ are the density and distance thresholds, respectively. Moreover, we assign each foreground voxel in B_{cs} with different instance labels based on the minimum distance from its predicted candidate tooth centroid to the clustered tooth centroids in T_c .

Although a tooth centroid is stable to distinguish and localize a tooth, a single point is insufficient to capture its geometric and topological properties, compared to the skeletal representation. To obtain the tooth skeleton, we add the skeleton offsets on the coordinates of corresponding foreground voxels in B_{cs} .

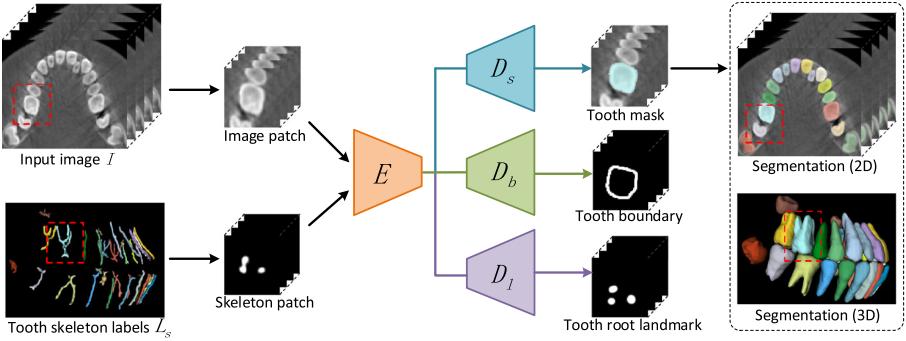


Fig. 3. The pipeline of our 2nd-stage multi-task network for tooth instance segmentation guided by skeletons. The inputs are the cropped input image and skeleton patches cropped from the 3D input image and tooth skeleton label map, and the outputs are the tooth segmentation volume, boundary, and root landmarks.

After frequency counting, the skeleton density map H_l is obtained in the same process as the generation of H_c . Finally, we filter voxels in H_l with the lower frequency to produce the target tooth skeleton map. Notably, as the foreground voxels already have the instance labels after the centroid clustering, we can generate the instance-level tooth skeleton label map L_s in a straightforward manner, as shown in Fig. 3.

The two sub-networks are trained independently with the same loss functions.

$$\mathcal{L}_{CS} = \mathcal{L}_{seg}^b + \eta \mathcal{L}_{reg}^{smoothL1}, \quad (2)$$

where η is the balancing weight empirically set as 10 in our experiments. Specifically, the smooth L1 loss is employed to calculate the offset regression error ($\mathcal{L}_{reg}^{smoothL1}$) on the voxels belonging to tooth objects. And the binary cross-entropy loss is utilized to compute the binary segmentation error (\mathcal{L}_{seg}^b).

2.2 Multi-task Learning for Tooth Segmentation

Guided by the instance-level tooth skeleton label map L_s , we further extract individual teeth. To improve the segmentation accuracy, especially near the tooth boundary and root areas, we introduce a multi-task learning mechanism that can efficiently employ the intrinsic relatedness between the tooth volume, boundary, and root landmarks.

To train the multi-task learning network, we process the 3D CBCT image I and tooth skeleton label map L_s at a finer scale where individual tooth patches are extracted and processed. In Step one, we select a tooth skeleton instance and crop two patches (with the same size, and centered at the mass of the skeleton) from the original image I (the image patch) and the skeleton instance label map L_s , respectively. In Step two, the selected tooth skeleton instance is further converted to a Gaussian map centered at the skeleton voxels with a small

standard deviation $\delta_1 = 3$ voxel-size, serving as one of the inputs (the skeleton patch). As shown in Fig. 3, the skeleton patch is concatenated with the original image patch, yielding a two-channel input of the tooth segmentation network.

Using the two-channel input, we design the segmentation network by leveraging a multi-task learning mechanism, i.e., simultaneously predicting the tooth volume, boundary, and corresponding root landmarks. Figure 3 presents an overview of the network that consists of a shared encoder (E) and three task-specific decoders (D_s , D_b and D_l) with the skip connections combining features of different levels. The three individual branches output the tooth segmentation, boundary, and root landmarks, respectively. Notably, each tooth's ground-truth boundary and root landmark are defined as the 3D Gaussian heatmaps centered at the surface and point location with a standard deviation δ_2 set as 3 voxel-size in our experiments. The loss function \mathcal{L}_{MT} of this multi-task network is defined as:

$$\mathcal{L}_{MT} = \mathcal{L}_{seg} + \lambda(\mathcal{L}_b + \mathcal{L}_l), \quad (3)$$

where \mathcal{L}_{seg} , \mathcal{L}_b and \mathcal{L}_l refer to the tooth volume segmentation, boundary, and landmark prediction losses, respectively. For \mathcal{L}_{seg} , we combine the Dice loss and binary cross-entropy loss in our experiment, while for \mathcal{L}_b and \mathcal{L}_l , we use the L2 error. The hyper-parameter λ is empirically fixed as 0.2 to balance the loss terms.

2.3 Implementation Details

We employed 3D V-Net [12] as the network backbone of our two-stage framework. All CBCT images were converted to have the same input size of $256 \times 256 \times 256$ in the 1st-stage. The cropped patch size of the 2nd-stage was set as $96 \times 96 \times 96$ to ensure that the whole foreground tooth object is included. The framework was implemented in PyTorch, which was trained using Adam optimizer with a fixed learning rate of $1e^{-4}$. The networks were trained in $50K$ iterations in both two stages. Generally, the training time was around 5 h (1st-stage) and 8 h (2nd-stage) on a Linux server with one Nvidia GeForce 1080Ti GPU.

3 Experimental Results

3.1 Dataset and Evaluation Metrics

We have extensively evaluated the proposed framework on 100 CBCT scans collected from patients before or after orthodontic treatments in dental clinics. The dataset contains many abnormal cases with teeth crowding, missing or malocclusion problems. The resolution of the dataset is 0.4 mm. We manually crop the tooth area on the 3D CBCT image, resize it to $256 \times 256 \times 256$, and then normalize the CBCT image intensity to the range of $[0, 1]$. To obtain the ground truth, the segmentation labels and tooth root landmarks are manually annotated by dentists. The corresponding tooth skeletons and boundaries are generated using morphological operations [10] based on the annotated segmentation labels, and

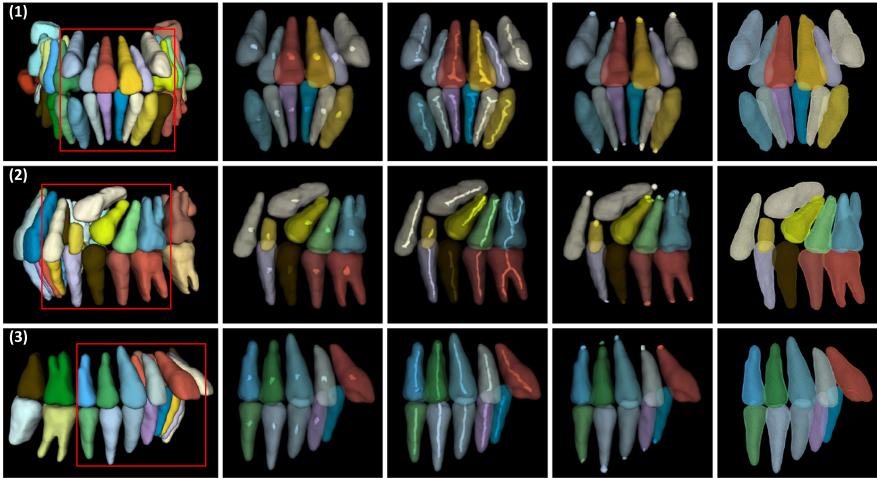


Fig. 4. Typical results of the hierarchical tooth morphological representation. From left to right: 3D segmentation results, predicted tooth centroids, skeletons, root apices, and boundaries. The last four columns are the partial results of the first column within the red boxes. (Color figure online)

the tooth centroids are directly computed based on the labelled mask. To train the network, the dataset is randomly split into three subsets, i.e., 50 scans for training, 10 scans for validation, and the remaining 40 scans for testing.

To quantitatively evaluate the performance of our framework, we employ different metrics to measure the tooth detection and segmentation accuracy. Specifically, we measure the tooth detection accuracy (DA) by $DA = \frac{|GT \cap P|}{|P|}$, where GT and P refer to two sets of the ground-truth and the predicted teeth. For the tooth segmentation, four metrics, including Dice, Jaccard, the average surface distance (ASD), and Hausdorff distance (HD), are utilized to evaluate the performance. Since the Hausdorff distance is the maximum of the minimum distances between the predicted and the ground-truth tooth surfaces, it is the key metric to especially measure the segmentation error around the tooth root area with only a tiny percentage of foreground voxels.

3.2 Evaluation and Comparison

We conduct extensive experiments to demonstrate the effectiveness of our tooth instance segmentation framework guided by hierarchical morphological components, including skeleton representation, and multi-task learning for joint prediction of tooth boundary and root landmarks. In Table 1, we present segmentation results of four configurations: (1) we build our baseline network (bNet) by directly utilizing the tooth centroid to detect and represent each tooth in the 1st-stage network, and also a single-task segmentation network without tooth boundary and root landmark predictions in the 2nd-stage network; (2) we add

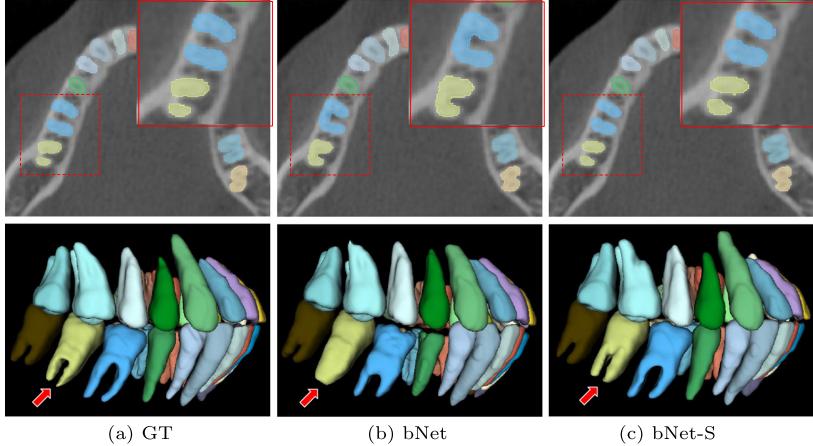


Fig. 5. The qualitative comparison of tooth segmentation with (c) or without (b) tooth skeleton representation. bNet-S achieves better results especially near the tooth root area highlighted by red boxes and arrows, compared to the ground truth (a). (Color figure online)

Table 1. Quantitative results of ablation analysis of different morphological components.

Methods	Dice [%]	Jaccard [%]	ASD [mm]	HD [mm]
bNet	92.1 ± 1.5	84.7 ± 2.4	0.33 ± 0.10	2.42 ± 0.88
bNet-S	93.1 ± 1.1	85.3 ± 1.7	0.31 ± 0.05	2.30 ± 0.77
bNet-S-L	94.3 ± 0.6	88.4 ± 1.0	0.26 ± 0.02	1.63 ± 0.45
FullNet	94.8 ± 0.4	89.1 ± 0.9	0.18 ± 0.02	1.52 ± 0.28

only one tooth morphological information, i.e., the tooth skeleton, to the baseline network to better represent each tooth object, which is denoted as bNet-S; (3) compared to bNet-S, we add the tooth root landmark detection as a separate branch in the 2nd-stage network for multi-task learning, denoted as bNet-S-L; (4) we further argument bNet-S-L with tooth boundary prediction branch in our 2nd-stage network as the final network (FullNet). Note that all of the four configurations utilize the tooth centroid point to detect the tooth object in the 3D CBCT image, thus the detection accuracy is the same and not listed in Table 1.

Benefits of Tooth Skeleton Representation. Compared with the tooth centroid, the tooth skeleton provides richer and more faithful geometric and topological information to guide the subsequent tooth segmentation, especially for handling the molars with multi-roots. To validate its effectiveness, we add the tooth skeleton detection component (bNet-S) to the baseline network (bNet) in the 1st-stage, and show the quantitative results in Table 1. It can be seen that bNet-S consistently improves the segmentation performance in terms of all

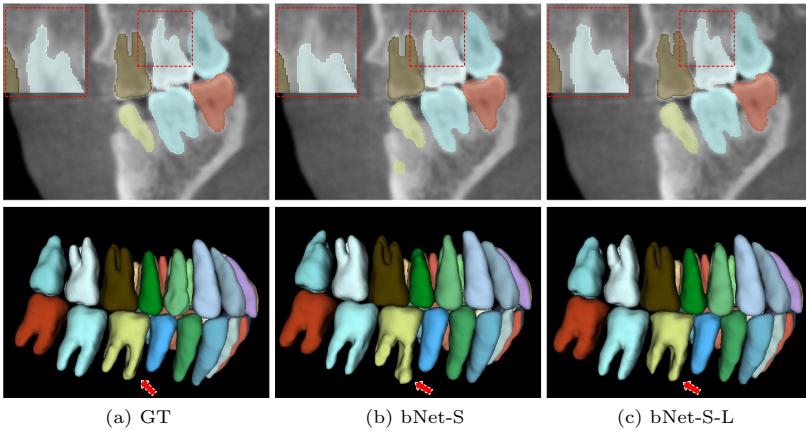


Fig. 6. The qualitative comparison of tooth segmentation with (c) or without (b) tooth root landmark detection branch. Compared to the ground truth (a), bNet-S-L avoids under- or over-segmentation around the tooth root apices highlighted by red boxes and arrows. (Color figure online)

metrics (e.g., 1.0% Dice improvement and 0.02 mm ASD improvement, respectively). Additionally, a typical visual comparison is shown in Fig. 5, which indicates that, with the guidance of tooth skeleton, the 2nd-stage segmentation network can accurately separate different roots of a molar. This demonstrates that the tooth skeleton, with clear tooth shape information, brings significant benefits to capture complicated tooth shapes.

Benefits of Tooth Root Landmark Detection. In our 2nd-stage network, instead of only generating the segmentation mask, bNet-S-L adds another branch to predict the tooth root landmarks by a multi-task learning mechanism. As shown in Table 1, compared with bNet-S, the Hausdorff distance of bNet-S-L significantly drops from 2.30 mm to 1.63 mm. Note that the HD metric measures the maximum of the minimum surface distances between the ground-truth and predicted tooth surfaces, such that the under- or over-segmentation near the tooth root apices usually leads to the large error. This indicates the multi-task learning of tooth segmentation prediction and landmark detection assists the network to capture the intrinsic relatedness from a geometric perspective and then benefits the segmentation task. To further analyze the effectiveness, we also provide a visual example in Fig. 6, where bNet-S-L efficiently addresses the under- or over-segmentation problem of the tooth roots (highlighted by the red boxes) even with limited intensity contrast.

Benefits of Tooth Boundary Prediction. In our FullNet, a third branch, tooth boundary prediction, is added in the 2nd-stage network, which encourages the network to pay more attention to the tooth boundary area with limited intensity contrast. Statistically, the FullNet obtains the best segmentation

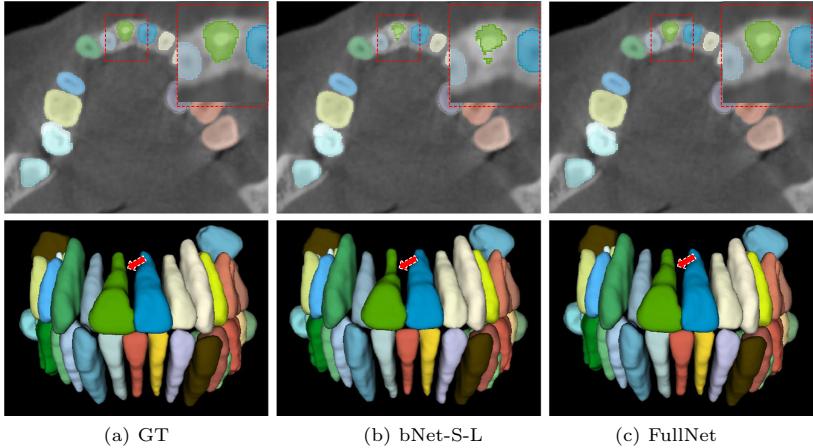


Fig. 7. The qualitative comparison of tooth segmentation with (c) or without (b) tooth boundary prediction branch. Compared to the ground truth (a), FullNet produces more accurate segmentation results especially on the boundaries with metal-artifacts.

Table 2. Quantitative comparison with the state-of-the-art methods in terms of the segmentation and detection accuracy.

Methods	Dice [%]	Jaccard [%]	ASD [mm]	HD [mm]	DA [%]
MWTNet [3]	89.6 ± 1.3	82.5 ± 1.9	0.36 ± 0.14	4.82 ± 1.68	98.1 ± 0.8
ToothNet [5]	91.6 ± 1.4	84.2 ± 1.8	0.30 ± 0.11	2.82 ± 1.02	98.6 ± 1.1
CGDNet [14]	92.5 ± 1.1	85.2 ± 1.6	0.27 ± 0.03	2.21 ± 0.69	98.9 ± 1.5
Ours	94.8 ± 0.4	89.1 ± 0.9	0.18 ± 0.02	1.52 ± 0.28	99.7 ± 0.6

performance and boosts the average Dice score and the ASD error to 94.8% and 0.18 mm, respectively. The qualitative results in Fig. 7 also show that the FullNet can segment more accurate tooth boundaries even with metal artifacts in CBCT images. More representative segmentation results of the FullNet are presented in Fig. 4 and Fig. 8.

3.3 Comparison with the State-of-the-Art Methods

We implement and compare our framework with several state-of-the-art deep learning based tooth segmentation methods, including the region proposal based network (ToothNet) [5], the center-guided network (CGDNet) [14], and the semantic-based method (MWTNet) [3]. Note that we utilized the same network backbone (V-Net) in all methods for fair comparison. As shown in Table 2, compared with MWTNet [3] that directly utilizes tooth boundaries to simultaneously detect and segment individual teeth in a single step, our method leads to remarkable improvement of 5.2% Dice score and 3.30 mm HD error, demonstrating the advantage of the two-stage detect-then-segment framework. Although

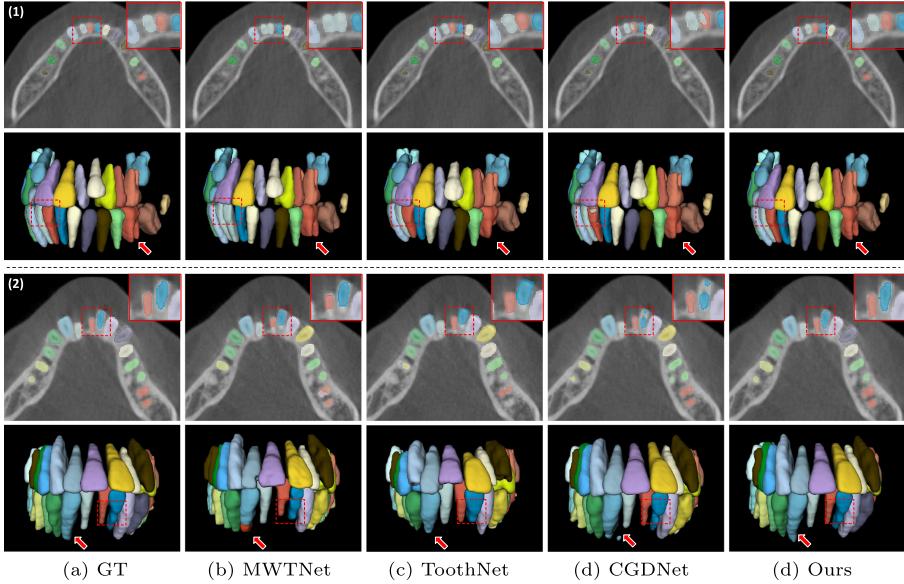


Fig. 8. The visual comparison of tooth segmentation results by four different methods. Two typical examples are presented, each being shown by two rows with 2D segmentation masks and corresponding 3D reconstruction results, respectively.

ToothNet [5] is a two-stage network, it only utilizes bounding boxes to represent individual teeth and our method still outperforms it in terms of segmentation and detection performances by a large margin. At last, it is also observed that our approach consistently achieves higher accuracy than CGDNet [14], which achieves the state-of-the-art performance in this specific task. Particularly, the segmentation accuracy (Dice) is increased from 92.5% to 94.8%, and the detection accuracy (DA) is improved from 98.9% to 99.7%. It is worth noting that all these competing methods pay little attention to the segmentation around tooth root apices with limited intensity contrast, which usually leads to under- or over-segmentation of the tooth roots and a higher HD error, even if the root information is an important consideration in orthodontic treatment.

To further demonstrate the advantage of our method, we provide a quantitative comparison of two typical examples in Fig. 8. It can be found that the segmentation results generated by our approach (in the last column) match better with the ground truth (in the 1st column), especially near the tooth root apices and occlusion planes with blurred boundary signals. Notably, MWTNet (the 2nd column) is more likely to lead to failure in tooth separation. For example, two incisors are regarded as the same object in the 1st case, and a cuspid is broken into two parts in the 2nd case. This shows that the tooth boundaries alone are not stable signals for segmenting adjacent teeth due to limited intensity contrast between these teeth. Besides, ToothNet [5] (the 3rd column) and CGDNet [14] (the 4th column), respectively representing each tooth by a bound-

ing box or a center point, produce lots of artifacts near the tooth boundary and root areas, since most tooth topological features are overlooked by the simple representation of the bounding box or the center point. The visual results shown in Fig. 8 are consistent with the quantitative comparison, indicating the effectiveness and advantages of the hierarchical morphology-guided tooth instance segmentation framework.

4 Conclusion

In this paper, we present a novel tooth instance segmentation network from CBCT images guided by the hierarchical morphological representations of each tooth, including its centroid and root landmarks at apices (i.e., points), skeleton, boundary surface, and volumetric mask. Specifically, the tooth centroid and skeleton are first utilized to detect and represent each tooth. Then, a multi-task learning mechanism is presented to achieve high segmentation accuracy especially around tooth boundaries and tooth root apices. Comprehensive experiments have validated the effectiveness of our method, showing it can outperform the state-of-the-art methods. This gives the potential of our method to be widely used in the real-world clinics.

References

1. Akhoondali, H., Zoroofi, R.A., Shirani, G.: Rapid automatic segmentation and visualization of teeth in CT-scan data. *J. Appl. Sci.* **9**(11) (2009)
2. Barone, S., Paoli, A., Razionale, A.V.: CT segmentation of dental shapes by anatomy-driven reformation imaging and B-spline modelling. *Int. J. Numer. Methods Biomed. Eng.* (2015)
3. Chen, Y., et al.: Automatic segmentation of individual tooth in dental CBCT images from tooth surface map by a multi-task FCN. *IEEE Access* **8**, 97296–97309 (2020)
4. Chung, M., Lee, M., Hong, J., Park, S., Shin, Y.G.: Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. *Comput. Biol. Med.* **120**, 103720 (2020)
5. Cui, Z., Li, C., Wang, W.: ToothNet: automatic tooth instance segmentation and identification from cone beam CT images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6368–6377 (2019)
6. Gan, Y., Xia, Z., Xiong, J., Li, G., Zhao, Q.: Tooth and alveolar bone segmentation from dental computed tomography images. *IEEE J. Biomed. Health Inform.* **22**(1), 196–204 (2018)
7. Gao, H., Chae, O.: Individual tooth segmentation from CT images using level set method with shape and intensity prior. *Pattern Recogn.* **43**(7), 2406–2417 (2010)
8. Ibsen, O., Phelan, J.: Oral Pathology for the Dental Hygienist. Saunders/Elsevier
9. Keyhaninejad, S., Zoroofi, R.A., Setarehdan, S.K., Shirani, G.: Automated segmentation of teeth in multi-slice CT images. In: IET International Conference on Visual Information Engineering. VIE 2006 (2006)
10. Lee, T.C.: Building skeleton models via 3-D medial surface/axis thinning algorithms. *Graph. Models Image Process.* **56**(6), 462–478 (1994)

11. Li, W., et al.: Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Med. Phys.* **43**, 336–346 (2015)
12. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
13. Rodriguez, A., Laio, A.: Clustering by fast search and find of density peaks. *Science* **344**(6191), 1492–1496 (2014)
14. 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)
15. Xia, Z.: Towards accurate tooth segmentation from computer tomography images using a hybrid active contour model. *Med. Phys.* (2014)