Medical Image Analysis 69 (2021) 101949

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

TSegNet: An efficient and accurate tooth segmentation network on 3D dental model

Zhiming Cui[a,*], Changjian Li[b,a], Nenglun Chen[a], Guodong Wei[a], Runnan Chen[a], Yuanfeng Zhou[c], Wenping Wang[a]

[a] Department of Computer Science, The University of Hong Kong, Hong Kong, China
[b] Department of Computer Science, University College London, London, UK
[c] Department of Software Engineering, Shandong University, Jinan, China

ARTICLE INFO

Article history:
Received 29 June 2020
Revised 6 November 2020
Accepted 12 December 2020
Available online 19 December 2020

Keywords:
Dental model segmentation
Tooth centroid prediction
Confidence-aware cascade segmentation
3D point cloud

ABSTRACT

Automatic and accurate segmentation of dental models is a fundamental task in computer-aided dentistry. Previous methods can achieve satisfactory segmentation results on normal dental models; however, they fail to robustly handle challenging clinical cases such as dental models with missing, crowding, or misaligned teeth before orthodontic treatments. In this paper, we propose a novel end-to-end learning-based method, called *TSegNet*, for robust and efficient tooth segmentation on 3D scanned point cloud data of dental models. Our algorithm detects all the teeth using a distance-aware tooth centroid voting scheme in the first stage, which ensures the accurate localization of tooth objects even with irregular positions on abnormal dental models. Then, a confidence-aware cascade segmentation module in the second stage is designed to segment each individual tooth and resolve ambiguities caused by aforementioned challenging cases. We evaluated our method on a large-scale real-world dataset consisting of dental models scanned before or after orthodontic treatments. Extensive evaluations, ablation studies and comparisons demonstrate that our method can generate accurate tooth labels robustly in various challenging cases and significantly outperforms state-of-the-art approaches by 6.5% of Dice Coefficient, 3.0% of F1 score in term of accuracy, while achieving 20 times speedup of computational time.

1. Introduction

Computer-aided design (CAD) has been widely used in orthodontics for diagnosis, dental restoration and treatment planning. CAD systems in dentistry require dental models as input to assist dentists to delete, extract or rearrange the teeth for treatment procedures. In this regard, segmenting 3D tooth models in different image modalities such as CBCT images (Cui et al., 2019; Lechuga and Weidlich, 2016) and dental models (Hajeer et al., 2004a; 2004b; Lian et al., 2019; Zanjani et al., 2019), is of great importance. As dental model scanners are free of X-ray radiation, they are widely used to acquire high-precision dental models of crown shapes in surface representation, compared with CBCT using 3D volumetric representation. Because it is laborious to manually label teeth from the dental model, the development of automatic and accurate 3D tooth segmentation methods for dental models has attracted tremendous research attention.

Although considerable efforts have been put into improving 3D tooth segmentation performance, developing an automatic method for robustly extracting the individual tooth from dental models is still a challenging task, due to following factors. First, some patients suffer from complex abnormalities such as the teeth crowding, missing and misalignment problems (Fig. 1(a) and (d)). Thus adjacent teeth are often irregular and hard to be separated. Second, the lack of pronounced shape variation at the boundaries between teeth and gum bring difficulties to segmentation methods based on geometric features (Fig. 1(c)). Lastly, dental models may have artifacts from the model-making process or dental braces worn by patients (Fig. 1(b)). All this may greatly affect the tooth shape appearance, thus making segmentation error-prone.

To address these challenges, many previous works exploited handcrafted geometric features for dental model segmentation, such as surface curvature (Yuan et al., 2010), geodesic information (Sinthanayothin and Tharanont, 2008) and harmonic field (Zou et al., 2015). However, these methods typically rely on domain-specific knowledge and lack the robustness required to represent intricate tooth shape appearances. Recently, with the advance in deep learning, more learning-based methods employing

* Corresponding author.
E-mail address: zmcui@cs.hku.hk (Z. Cui).

https://doi.org/10.1016/j.media.2020.101949
1361-8415/

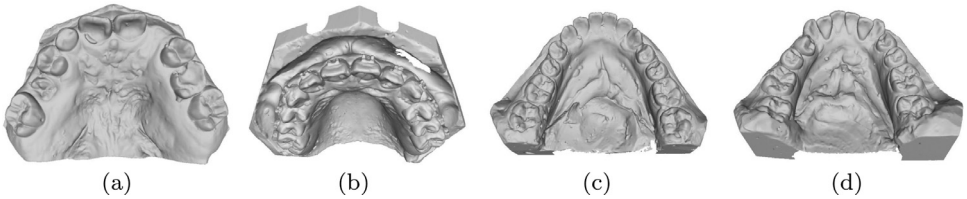


Fig. 1. Four typical examples with extreme appearance, including (a) missing teeth, (b) additional braces, (c) blurred boundary signals between incisors and the gum, and (d) crowding teeth.

convolutional neural networks (CNNs) (Xu et al., 2018; Tian et al., 2019) or mesh-based graph neural networks (Lian et al., 2019) have been proposed with compelling performance. Unfortunately, most of these methods make a strongly restrictive assumption that the dental models consist of a complete set of natural teeth, which is difficult to be satisfied, for example nearly 70% of the patients in orthodontic clinics are at the tooth exfoliation time so they they often do not have a fixed number of teeth (Cobourne and DiBiase, 2015). Mask-MCNet (Zanjani et al., 2019) transforms the dental model into point cloud data and uses a volumetric anchor-based region proposal network for tooth detection and segmentation. However, the proposal generation module results in the resolution deduction and requires huge memory resources.

Another line of deep learning methods, including PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b) and PointCNN (Li et al., 2018), directly take 3D point cloud data (e.g., mesh vertices) as input and learn deep geometric features to make a classification and segmentation for general geometric processing tasks. A major limitation of these methods, when applied to our tooth segmentation task, is that it is difficult to accurately separate neighboring teeth with similar shape appearances such as incisors, premolars and molars, especially on dental models with missing teeth (Fig. 9, PointNet++).

To tackle these issues, we present a novel end-to-end learning-based method for automatic tooth segmentation on 3D dental models. The core of our method is a two-stage neural network which firstly detects all the teeth and then segments each detected tooth accurately. In the tooth detection stage, instead of the traditional approach that utilizes bounding boxes to crop the detected objects (He et al., 2017; Hou et al., 2019; Zhou and Tuzel, 2018), we exploit the centroid (i.e. the center of mass) of a tooth to identify each tooth object based on our observation that regardless of the tooth shape, position and orientation, the centroid point is a stable feature point inside the tooth shape. Therefore it is a more reliable signal than the bounding box especially when the teeth are relatively small and packed tightly. In this way, the tooth detection problem is naturally converted to a tooth centroid prediction problem. To predict all the tooth centroids reliably, we design a distance-aware voting scheme that generates the tooth centroids from subsampled points with reliable learning local context. In the second stage of individual tooth segmentation, we first crop the points and corresponding features with the guidance of the predicted tooth centroid, and combine them as one tooth proposal. Subsequently, all the tooth proposals are sent to the segmentation module to generate individual tooth labels. Moreover, to improve segmentation accuracy, especially for tooth boundaries with blurring signals, we introduce a point-wise confidence map based on a cascade network to enhance the label learning with an attention mechanism. The newly proposed novel components and loss functions efficiently produce an accurate tooth segmentation and boost the usability of our algorithm in the real-world clinical scenario.

Our main contributions are summarized as follows:

- We propose a novel pipeline that formulates the dental model segmentation as two sub-problems: robust tooth centroids prediction and accurate individual tooth segmentation on point cloud data.
- We design a distance-aware voting scheme to efficiently predict all tooth centroids. Besides, a confidence-aware attention mechanism is introduced to improve segmentation in noisy areas.
- Extensive evaluations and ablation studies are conducted on a dataset collected from dental clinics. Compared with the state-of-the-art methods, the proposed framework achieves superior results both qualitatively and quantitatively by a significant margin.

The rest of the paper is organized as follows. In Section 2, we briefly review the existing methods for dental model segmentation and point cloud learning. Section 3 describes the proposed methodology in detail. In Section 4, we present the quantitative and qualitative results of our method and compare with the state-of-the-art methods. We also discuss the effectiveness of different components of the network and the limitations of our approach in this section. Section 5 provides the conclusion of our study.

2. Related works

2.1. Dental model segmentation

Non-learning based methods Many traditional methods based on the handcrafted geometric features have been proposed to segment 3D dental models. These methods can roughly divided into three categories: surface curvature based methods, surface contour line based methods and harmonic field based methods.

The surface curvature based methods aim to identify the tooth boundaries. For example, Yuan et al. (2010) calculated the minimum curvatures of the teeth surface and extracted the boundary between the tooth and soft tissues. Zhao et al. (2006) proposed an interactive method based on the curvature values of the triangle mesh. Kumar et al. (2011) developed a system in which users can set a certain curvature threshold via an intuitive slider. In addition, Li et al. (2007) integrated fast marching watersheds and manual threshold regulating to improve segmentation accuracy and reduce processing time. Kronfeld et al. (2010) minimized user annotation by positioning a snake around cusp points of each tooth. Wu et al. (2014) proposed to take advantage of the morphological technique to facilitate effective tooth separation. However, these methods based on the surface curvature are very sensitive to the variation of tooth surfaces and appearances.

The methods based on surface contour lines are more reliable in generating tooth boundaries because the contour lines are manually annotated. Specifically, these methods (Sinthanayothin and Tharanont, 2008; Yaqi and Zhongke, 2010) allowed users to manually select tooth boundary landmarks on a dental model. Then, the contour lines computed from the geodesic information of neighboring landmarks are formed as the desired tooth boundaries. Although achieving the good performance, these methods require users to translate or rotate the 3D model multiple times to select the particular landmarks carefully, which is tedious and time-consuming.

As for the third category (Zou et al., 2015), the framework allowed users to annotate a limited number of surface points as priors and employed a harmonic field to segment the tooth successfully. Compared to other interactive methods, this method presented a more efficient and smarter user interfaces with minimum interactions. However, the input models is assumed to manifold, which requires a sophisticated preprocessing step.

Another group of methods that aimed to effectively segment 3D dental models are based on 2D images. For example, Yamany and El-Bialy (1999) built a 2D image representation using the curvature and surface normal information, and extracted the structures of high/low curvatures as the segmentation results. Kondo et al. (2004) proposed to detect the tooth features both on the plane-view and panoramic-view images. Similarly, some works (Wongwaen and Sinthanayothin, 2010; Grzegorzek et al., 2010) developed systems to find the contour or cutting points on the 2D sectional images and then converted it back to the 3D space for separating individual tooth. Unfortunately, these methods often fail when dental models have severe malocclusion.

Learning based methods Recently, with the development of deep learning techniques, many studies leverage neural networks on 2D images, meshes and point clouds to extract teeth from a dental model. Specifically, Xu et al. (2018) used a 2D CNN to classify the image produced from the pre-defined handcrafted features of each mesh face. Tian et al. (2019) employed a 3D CNN and a sparse voxel octree for tooth segmentation. In addition, Lian et al., 2020; Lian, Wang, Wu, Liu, Durán, Ko, Shen; Sun et al. (2020) integrated a series of graph-constrained learning modules to hierarchically extract multi-scale contextual features for automatically labeling on raw dental surface. However, since these methods typically group points or faces into pre-defined clusters, they usually fail to process the data with missing teeth, which is common for real-world clinical scenarios. In addition, Zanjani et al. (2019) extended the Mask R-CNN (He et al., 2017) to a 3D point cloud extracted from the dental model, it suffered from low efficiency and segmentation artifacts.

2.2. 3D point cloud learning

3D understanding is an essential task in computer vision. State-of-the-art methods take as input all kinds of 3D data to perform tasks such as 3D shape segmentation, detection and classification. Among the input data, 3D point cloud representation is becoming more popular since it is flexible and memory efficient. Qi et al. (2017a) designed a novel network to take as input an unstructured point cloud and learn translation-invariant geometric features. Some state-of-the-art methods (Qi et al., 2017b; Li et al., 2018; Wu et al., 2019) improved the framework by recursively applying neural networks on a nested partitioning of the input point cloud, which had the ability to learn local features with increasing contextual scales and achieved state-of-the-art performance on many segmentation and classification tasks. However, their performance is limited in our specific task, because the tooth is very small compared with the whole dental model.

3. Methods

In this section, we present a novel framework for tooth segmentation on 3D dental model. As shown in Fig. 2, our approach takes as input the 3D point cloud extracted from the input dental model, and aims to assign every point a unique label. Specifically, we first introduce the distance-aware tooth centroid prediction module that generates a set of candidate points for the tooth centroids (Section 3.1). Then, we propose a confidence-aware attention mechanism to segment each tooth guided by the predicted tooth centroid (Section 3.2). At the testing stage, we utilize a tooth centroid clustering algorithm to speed up the segmentation, and directly transfer the point cloud labels back to the dental model (Section 3.3).

3.1. Distance-aware tooth centroid prediction

To identify a tooth object properly, we formulate it as the tooth centroid prediction problem. Formally, given an input dental model, we first extract the mesh vertices and uniformly downsample it to obtain the input point cloud P with dimension $N \times 6$, where $N = 16,000$ is the number of sampled input points and each point is described by a 6-*D* vector. Specifically, other than the 3D coordinates (*3-dims*), we also acquire the normal vector (*3-dims*) at each point from the dental mesh as an additional feature to provide auxiliary information. Having the input point cloud P, we first normalize it within a unit ball, and extract the geometric features utilizing PointNet++ as the backbone encoder, which includes three blocks of multi-layer perceptrons (MLPs) followed by a batch normalization layer and a ReLU nonlinearity layer. The output of the backbone encoder is a set of subsampled points F with dimension $M \times (3 + 256)$, where $M = 256$ is the number of subsampled points. For each point, in addition to the 3D coordinates, there are another 256-*D* features encoding the local contextual information around it.

For the dental model of an upper or lower jaw, we have the ground truth tooth centroid set $C = \{c_1, c_2, \ldots, c_k\}$, and the goal is to predict all tooth centroids from the subsampled points F using the learned local features. Therefore, we design a displacement function to learn the offset of each subsampled point to its corresponding tooth centroid c_i. This is feasible because if a subsampled point appears around a tooth, the encoded features capturing the tooth shape have the ability to predict the centroid of the nearby tooth. Specifically, as shown in Fig. 2, the MLPs take as input the subsampled points F with the learned local features to output a set of M displacement vectors $\Delta C = \{(\Delta x_i, \Delta y_i, \Delta z_i)\}$. Finally, the set of regressed centroid points, $\hat{C} = \{(x_i + \Delta x_i, y_i + \Delta y_i, z_i + \Delta z_i)\}$ $(i \in [1, M])$, are generated to approximate the ground truth set C, where (x_i, y_i, z_i) denotes the 3D coordinate of the ith subsampled point F_i.

However, since the subsampled points F are uniformly sampled from the input point cloud by the farthest sampling operation, we observe that some subsampled points may be far from any tooth, e.g., on the dental palate (Fig. 4(a)), which encode little information of any tooth object and cannot predict reliable tooth centroids. To filter such points automatically, we exploit another distance estimation branch to regress a distance value for each subsampled point, measuring the closeness of the point to its nearest ground truth tooth centroid, as shown in Fig. 2.

To train the network, we propose three novel loss terms to supervise the prediction of all tooth centroids.

Distance estimation To remove the subsampled points that are far way from any tooth, we first measure the distance between each subsampled point and its nearest tooth centroid, and set it as the ground truth of the distance estimation. Then we utilize smooth L1 loss to calculate the regression error. Let $F^{(3)}$ denote the 3D coordinates of the subsampled points F. Then the loss function of the distance estimation is defined as follows:

$$\mathcal{L}_D = \sum_{f_i^{(3)} \in F^{(3)}} L_1^{\text{smooth}}(\hat{d}_i - \min_{c_k \in C} \left\| f_i^{(3)} - c_k \right\|_2), \tag{1}$$

where

$$L_1^{\text{smooth}}(x) = \begin{cases} 0.5x^2 & \text{if} |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \tag{2}$$

where $\hat{d}_i$ refers to the predicted distance value from the subsampled point F_i to its nearest tooth centroid. With this distance esti-

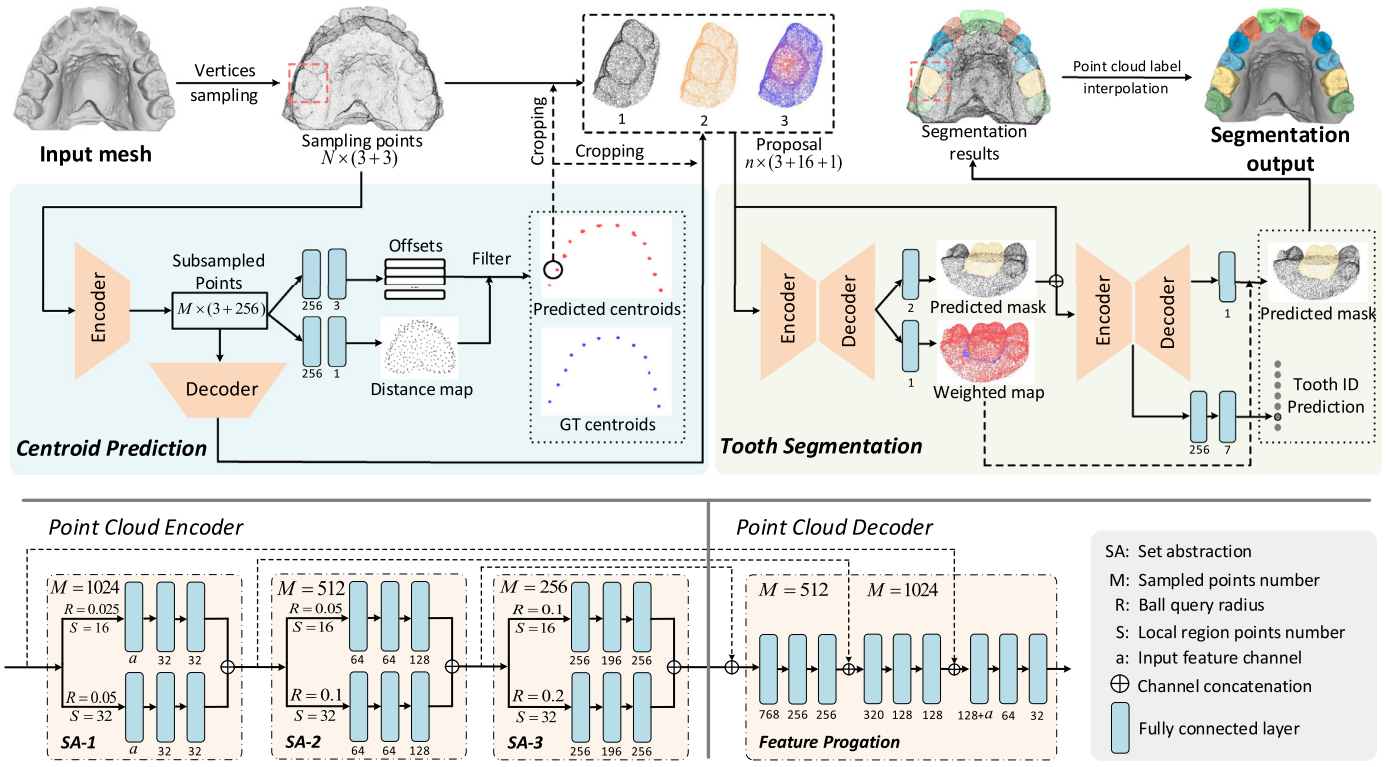


Fig. 2. The two-stage network architecture and the algorithm pipeline. The dental mesh is first fed into the centroid prediction network in stage one, then the cropped features based on the regressed points go through the tooth segmentation network in stage two. Finally, we derive the accurately segmented tooth objects. The numbers 1, 2, 3 in the proposal box, represent the input signals for the segmentation network, i.e., cropped *coordinate feature*, *propagated point feature* and dense *distance field feature* respectively. See algorithm details in Section 3.

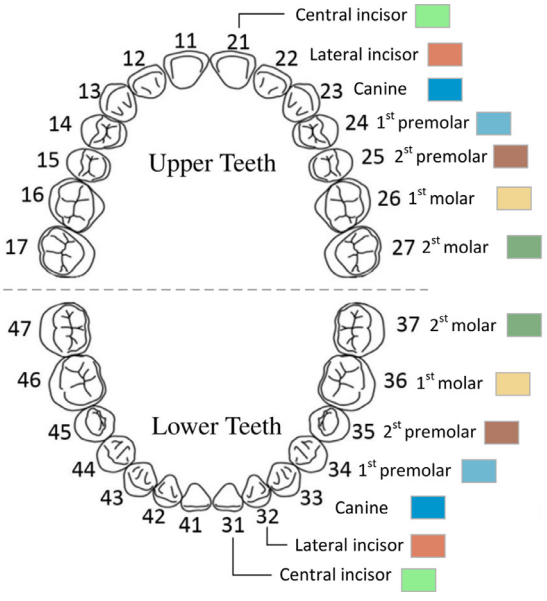


Fig. 3. The ISO standard tooth numbering system and the corresponding color coding.

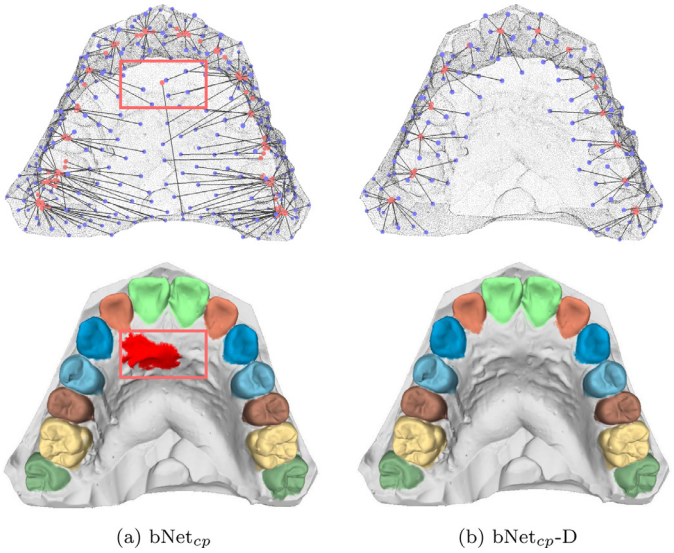
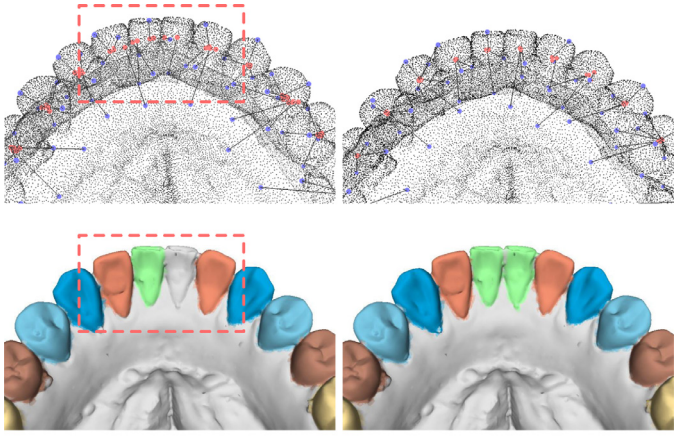


Fig. 4. The qualitative comparison of the centroid prediction results with (b) or without (a) the robust filter. The first row shows centroid point prediction results with paired purple and red points indicating the start and end positions, while the second row shows the corresponding segmentation results using bNet$_{seg}$. The wrongly predicted points lead to incorrect tooth segmentation (red color). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mation module in the framework, we filter the subsampled points that have a relative large predicted distance both at the training and testing stage. The threshold α is set to 0.2 on the normalized point sets, which is consistent with the receptive field of the last set abstraction layer in the encoder.

Chamfer distance In the tooth centroid prediction branch, we train the network by minimizing the distance between the regressed centroid set $\hat{C}$ and ground truth centroid set C, which is formulated to consider the following two factors: (1) every tooth centroid in C should correspond to at least one regressed centroid in $\hat{C}$ (surjection function); (2) every regressed centroid in $\hat{C}$ should



(a) bNet$_{cp}$-D-CD (b) bNet$_{cp}$-D-CD-S

Fig. 5. The qualitative comparison of the centroid points prediction. The first row shows the predicted centroid points, while the second row shows the segmentation results using bNet$_{seg}$. Without the separation loss, bCNet$_{cp}$-D-CD outputs ambiguous points that cheat the clustering algorithm in the testing stage to miss some teeth, as highlighted in the dotted boxes.

correspond to exactly one tooth centroid in C (injection function). It is a bidirectional distance minimization and we use Chamfer distance to supervise the tooth centroid prediction. The loss function $\mathcal{L}_{CD}$ of the two sets of centroids is formulated as:

$$\mathcal{L}_{CD} = \sum_{\hat{c}_i \in \hat{C}, \hat{d}_i < \alpha} \min_{c_k \in C} ||\hat{c}_i - c_k||_2^2 + \sum_{c_k \in C} \min_{\hat{c}_i \in \hat{C}, \hat{d}_i < \alpha} ||c_k - \hat{c}_i||_2^2, \tag{3}$$

where $\alpha = 0.2$ is introduced in the distance estimation term.

Separation loss The tooth centroid prediction with distance estimation and chamfer distance supervision already achieves excellent performance. But we still observe that a few predicted centroids are located near the boundary between two adjacent teeth, especially for the incisors of the lower jaw as shown in Fig. 5, which are relatively small and closely packed. This happens because these ambiguous centroids receive little penalization from Chamfer distance loss. To tackle this issue, we add a separation loss, defined as:

$$\mathcal{L}_S = \sum_{\hat{c}_i \in \hat{C}, \hat{d}_i < \alpha} \frac{\Delta d_1}{\Delta d_2}, \tag{4}$$

where Δd_1 and Δd_2 are the distances of the predicted tooth centroid $\hat{c}_i$ to its first and second closest centroids in C, respectively. This term encourages each predicted centroid to be as close as possible to a correctly corresponding tooth centroid in the ground truth set C.

Finally, our training loss function $\mathcal{L}_{cp}$ for the robust centroid point prediction is obtained by combining the three loss terms as follows:

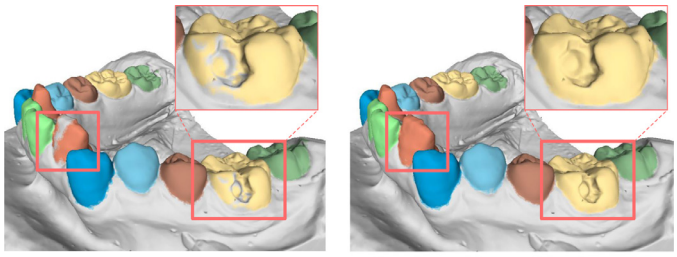
$$\mathcal{L}_{cp} = \mathcal{L}_D + \mathcal{L}_{CD} + \beta \mathcal{L}_S, \tag{5}$$

Where β is the balancing weight and is empirically set to 0.1 for all experiments.

3.2. Confidence-aware tooth segmentation

Now we discuss how to use the accurately predicted tooth centroids as the guidance information to perform individual tooth segmentation.

Tooth proposal generation Since each tooth is found by at least one predicted centroid, we first generate tooth proposals according to the predicted centroids. Instead of utilizing a bounding box to crop a tooth object, we crop the nearest $n = 4096$ points in the input point cloud data based on the Euclidean distance to the predicted tooth centroid, which are roughly a quarter of the points of a input dental model (16,000) and ensures a complete tooth is included in the proposal. As highlighted in the top row of Fig. 2 with red dotted box, a tooth proposal is represented by three components. The first is the cropped points coordinates (*3-dims*), and the second is the cropped points propagation features (*32-dims*). The last component is a dense distance field $df_{(i)}$ (*1-dim*) for the $i-th$ proposal, defined as:



(a) bNet$_{seg}$ (b) bNet$_{seg}$-C

Fig. 6. The qualitative comparison of tooth segmentation with (b) or without (a) the cascaded refinement. With the refinement module, bNet$_{seg}$-C generates results without artifacts in the tooth body part.

$$df_{(i)}^j = \exp\left(-4 \left\| \hat{c}_i - \hat{p}_{(i)}^j \right\|_2\right), \tag{6}$$

where $\hat{c}_i$ is the predicted centroid of proposal i, while $\hat{p}_{(i)}^j$ is the 3D coordinate of point j in the cropped points. By proposing the distance field, the foreground tooth corresponding to the predicted centroid will have a higher value compared to other teeth in the cropped points, which is regarded as a guidance map for the segmentation sub-network.

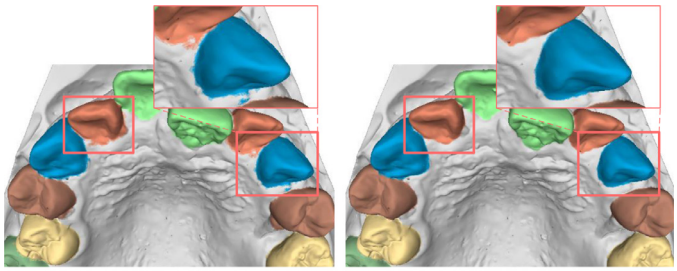
At last, we directly concatenate the three individual features and feed them into the segmentation network to segment the foreground tooth shape.

Confidence-aware cascaded segmentation The segmentation network building upon PointNet++, takes as input the concatenated feature of dimension $n \times (3 + 32 + 1)$ and outputs the binary label of each point belonging to the tooth shape or the background. Although PointNet++ demonstrates excellent performance in point cloud segmentation, it is hard to separate the tooth shape clearly from the surrounding gum due to the blurred geometric signals near the tooth boundary and large variations of tooth shapes (Fig. 6). Thus we first design our network using a cascaded segmentation scheme with two segmentation sub-networks S_1 and S_2. The cascaded scheme that S_2 takes as input both the proposal features and the 1-dimensional segmentation result from S_1. In addition, to further improve segmentation accuracy near the boundary of complicated tooth shapes (Fig. 7(a)), we propose a novel confidence-aware attention mechanism for tooth segmentation and the details are given below.

In the first segmentation sub-network S_1, in addition to predicting the segmentation results of the proposals, we introduce another branch to estimate the point-wise confidence value λ, measuring the accuracy of the segmentation, defined as:

$$\mathcal{L}_{S_1} = \frac{1}{n} \sum_j^n (\mathcal{L}_{bce_j^{S_1}} \times \lambda_j)^2 + \left(1 - \lambda_j\right)^2, \tag{7}$$

where $\mathcal{L}_{bce_j^{S_1}}$ indicates the point-wise binary cross-entropy (BCE) loss between the predicted point label and the ground truth label and λ is trained in an unsupervised manner to measure the ambiguity of the predicted label. That is, the higher is the value, the more accurate is the prediction results. A visual illustration of the point-wise confidence map is presented in Fig. 2 (the weight map).



(a) bNet$_{seg}$-C (b) TSegNet

Fig. 7. The qualitative comparison of tooth segmentation with (b) or without (a) the confidence-aware refinement. The tooth boundaries are highlighted in red boxes and TSegNet generates a more accurate boundary. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Clearly, the boundary area with blurred geometric signals tends to have lower confidence value.

In the second segmentation sub-network S_2, we convert the confidence map into a normalized weight map that emphasizes the segmentation of the area with lower λ in S_2, e.g., the boundary area. The training loss is:

$$\mathcal{L}_{S_2} = \frac{1}{n}\sum_{j}^{n}(1.0 + W_j) \times \mathcal{L}_{bce_j^{S_2}}, \tag{8}$$

where $W_j = 1.0 - \lambda_j$ is a point-wise value on the weighted map and $\mathcal{L}_{bce_j^{S_2}}$ refers to the point-wise BCE loss in S_2.

In addition, to identify the foreground tooth ID in each proposal, we utilize the global feature extracted in S_2 to make a classification and calculate the cross entropy loss $\mathcal{L}_{ID}$ to supervise the task. Finally, we train the cascaded segmentation network using the loss function:

$$\mathcal{L}_{seg} = \mathcal{L}_{S_1} + \mathcal{L}_{S_2} + \mathcal{L}_{ID}. \tag{9}$$

3.3. Centroid clustering and label prediction

In the previous step, the predicted tooth centroids exhibit the clustering tendency as shown in Figs. 4, 5, 8. To remove redundant tooth centroids and speed up processing, in both training and testing phases, we first apply the DBSCAN (Ester et al., 1996) clustering algorithm to all the predicted centroids controlled by the distance threshold l. Here, l is empirically set to 0.015, which is relatively small compared to the tooth size in the normalized point cloud data. For every cluster, we calculate the representative average centroid point and derive the corresponding proposal for segmentation.

During the testing phase, after the individual tooth extraction on the generated proposals, the next step is to produce the labels for the input point cloud data. To this end, we first calculate the foreground point overlap of each two proposals. If the Intersection over Union (IoU) is higher than the threshold 0.35, the two proposals are regarded to contain the same tooth. In this case, we average the point-wise label probability to fuse the overlapped points. At last, the point cloud labels are directly transferred back to the dental surface based on the trilinear interpolation.

In implementation, we first train the centroid prediction network for 500 epochs, then connect the single tooth segmentation network and jointly train the framework for 100 epochs. We utilize Adam's solver with a fixed learning rate of 1×10^{-3}. Generally, using one Nvidia GeForce 1080Ti GPU, it takes about 4 h for the centroid prediction network training and 18 h for the joint training.

4. Experiments and results

In this section, we evaluate our algorithm on a dataset collected from the real-word clinics, including upper and lower jaws. The tooth identification is based on the dental notation system (ISO-3950) (Grace, 2000) (as shown in Fig. 3), which is consistent with the color coding of our segmentation results. The teeth subgroups for evaluation purpose in this section, i.e., incisor, canine, premolar, molar (in Tables 2 and 3), are set according to the types marked in Fig. 3 as well. All experiments are performed on a computer with a Intel(R) Xeon(R) V4 1.9 GHz CPU, a 1080Ti GPU, and 32 GB RAM.

4.1. Dataset and evaluation metrics

To train the network, we collected a set of dental models from some patients before or after orthodontics, which include many cases with abnormal tooth shapes, such as crowded teeth, missing teeth and additional braces. The dataset includes a total of 2000 dental models (1000 upper jaws and 1000 lower jaws), where each dental surface contains about 150,000 faces and 80,000 vertices. To train the network, we randomly split it into three subsets, 1500 models for training, 100 models for validating and 400 models for testing. To obtain the ground truth, we manually annotated the tooth-level label, and the centroid of each tooth is calculated based on the labeled mask. To quantitatively evaluate the performance of our method, we use the mean distance (MeanD) and max distance (MaxD) metrics to validate the performance of the tooth centroid prediction, defined as:

$$\text{MeanD}(R^1, R^2) = \frac{1}{|R^1|}\sum_{r_i^1 \in R^1}\min_{r_j^2 \in R^2}||r_i^1 - r_j^2||_2^2, \tag{10}$$

$$\text{MaxD}(R^1, R^2) = \max_{r_i^1 \in R^1}\min_{r_j^2 \in R^2}||r_i^1 - r_j^2||_2^2, \tag{11}$$

where R^1 and R^2 represent two point sets. The two metrics are computed by the predicted tooth centroids set and the ground truth tooth centroids set in a bidirectional manner (Table 1). For the segmentation task, we utilize the dice similarity coefficient (DSC) metric to validate on the point cloud and the dental surface respectively, that are calculated as:

$$\text{DSC}_{point} = 2 \times \frac{|L_{GT} \cap L_P|}{|L_{GT} + L_P|}, \tag{12}$$

$$\text{DSC}_{surface} = 2 \times \frac{Area(L_{GT} \cap L_P)}{Area(L_{GT}) + Area(L_P)}, \tag{13}$$

where L_{GT} and L_P denote the ground truth tooth labels, and the corresponding predicted labels, respectively. Note that the DSC of the dental surface is calculated in a manner weighted by the face area. Besides, macro F1-score (F1) is used to measure the tooth identification accuracy (Opitz and Burst, 2019). In the following quantitative results, other than Fig. 10, we report the average values computed on the testing subset.

4.2. Ablation analysis of key components

We conduct extensive experiments to validate the effectiveness of our network components and loss functions. First, we build baseline networks for both tooth centroid prediction and individual tooth segmentation tasks, which are denoted as bNet$_{cp}$ and bNet$_{seg}$, respectively. For the tooth centroid prediction baseline network, we directly supervise all the subsampled points to translate to their nearest tooth centroids, while the segmentation baseline network is the single PointNet++ segmentation module without the confidence-aware cascade mechanism. All the alternative

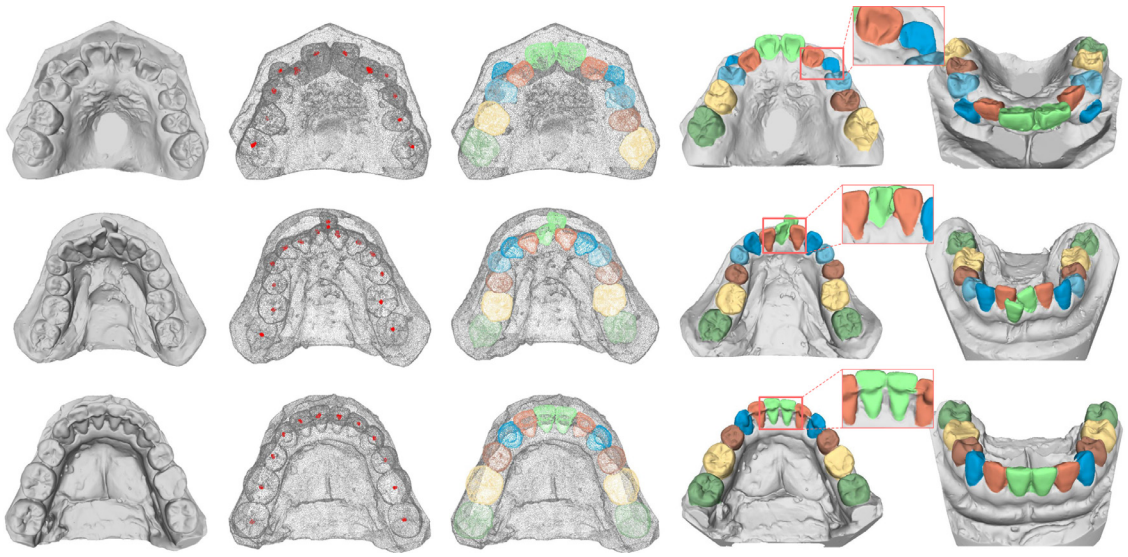


Fig. 8. Representative segmentation results. From left to right: input, predicted centroid points, tooth segmentation on the point cloud, tooth segmentation on dental models with two different views. The accurate segmentation boundary is highlighted in the boxes.

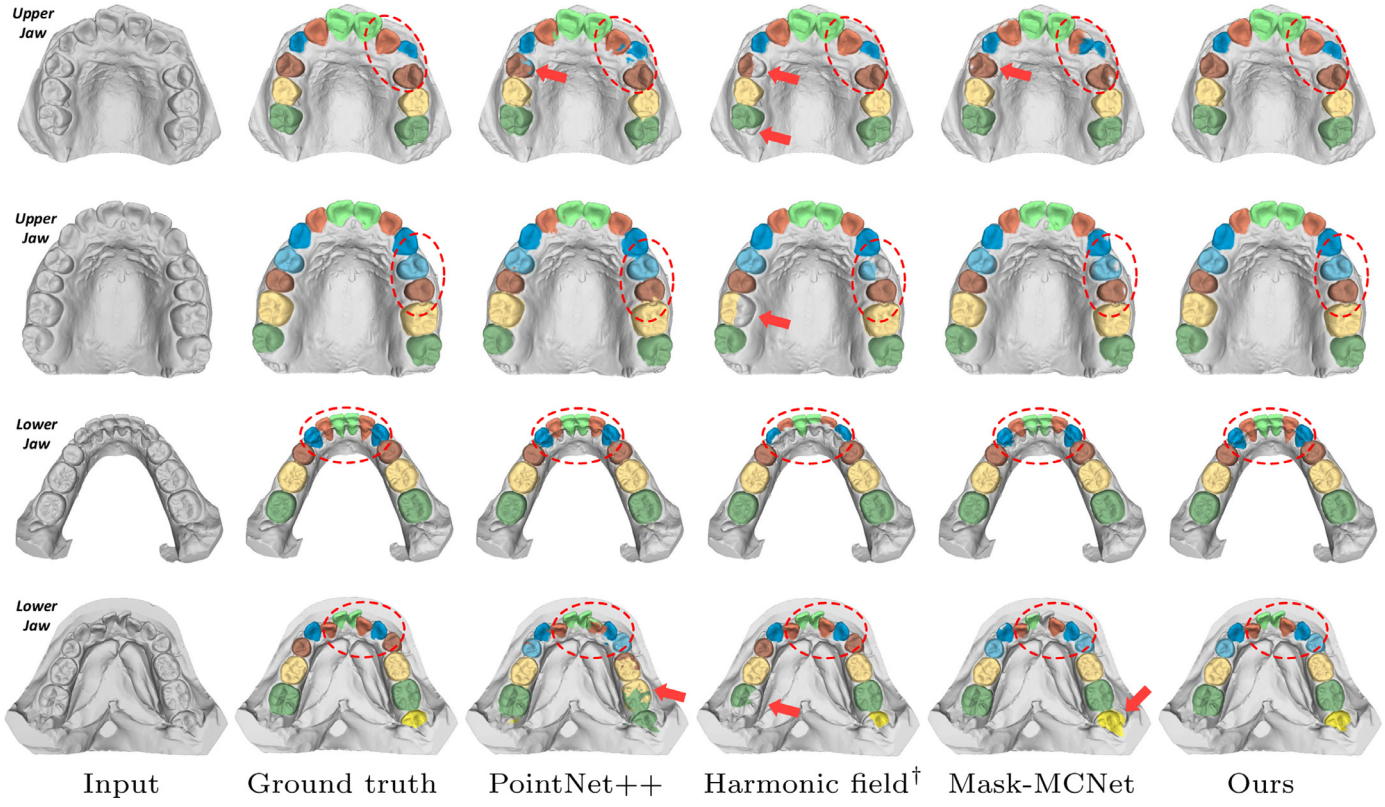


Fig. 9. The visual comparison of dental model segmentation results produced by different methods, with each row corresponding to a typical example of the upper or lower jaw. From left to right are the scanned dental surface, the ground truth result, results of other methods (3rd-5th columns) and result of our method (last column). Red dotted circles and arrows represent some segmentation details. '†' denotes the method is a semi-automatic method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Statistical performance of the tooth centroid prediction and segmentation with different tooth centroid prediction loss terms. Here, the distance metric is calculated on the point cloud data normalized within a unit ball, i.e., the ratio referencing to the unit length.

Network	Tooth centroid prediction [$\times 10^{-3}$]				Segmentation [%]	
	MeanD$(C, \hat{C})$	MaxD$(C, \hat{C})$	MeanD$(\hat{C}, C)$	MaxD$(\hat{C}, C)$	DSC_{point}	$DSC_{surface}$
$bNet_{cp}$	3.679 ± 1.850	16.553 ± 18.135	10.650 ± 10.662	86.245 ± 67.863	94.3 ± 2.4	95.2 ± 2.1
$bNet_{cp}$-D	2.998 ± 1.123	10.091 ± 12.584	9.350 ± 3.576	60.065 ± 18.760	95.5 ± 1.7	96.5 ± 1.4
$bNet_{cp}$-D-CD	2.673 ± 1.135	9.893 ± 12.159	6.857 ± 3.908	42.991 ± 19.836	95.9 ± 1.4	96.8 ± 1.1
$bNet_{cp}$-D-CD-S	**2.565 ± 0.880**	**8.785 ± 11.899**	**6.961 ± 3.490**	**14.883 ± 10.652**	**96.1 ± 1.2**	**96.9 ± 0.9**

Table 2
Numerical performance of segmentation accuracy for different segmentation network variants. The F1 scores are also included.

Methods	DSC_{point}[%]					$DSC_{surface}$[%]					F1[%]
	Incisor	Canine	Premolar	Molar	Mean	Incisor	Canine	Premolar	Molar	Mean	
$bNet_{seg}$	95.1	96.5	96.3	96.6	96.1	96.4	96.9	97.0	97.2	96.9	92.5
$bNet_{seg}$-C	97.1	98.1	97.4	97.2	97.4	97.6	98.1	97.6	97.7	97.8	93.4
TSegNet	**97.9**	**98.2**	**98.1**	**97.9**	**98.0**	**98.3**	**98.5**	**98.6**	**98.8**	**98.6**	**94.2**

Table 3
The qualitative comparisons with state-of-the-art methods on tooth detection, segmentation and running time metrics. '†' denotes the method is a semi-automatic method and '' means the metric is non-applicable.

Methods	DSC_{point}(%)					$DSC_{surface}$[%]					F1[%]	Time[s]
	Incisor	Canine	Premolar	Molar	Mean	Incisor	Canine	Premolar	Molar	Mean		
PointNet+	90.5	91.9	73.0	88.8	86.1	91.4	92.6	76.5	90.2	87.7	87.4	**0.3**
Harmonic Field†	92.2	90.0	95.0	95.4	93.2	92.5	90.1	95.1	96.0	93.4	–	30.0
Mask-MCNet	91.3	92.6	90.1	91.9	91.5	92.4	93.0	91.6	93.0	92.5	91.2	18.1
Ours (TSegNet)	**97.9**	**98.2**	**98.1**	**97.9**	**98.0**	**98.3**	**98.5**	**98.6**	**98.8**	**98.6**	**94.2**	0.8

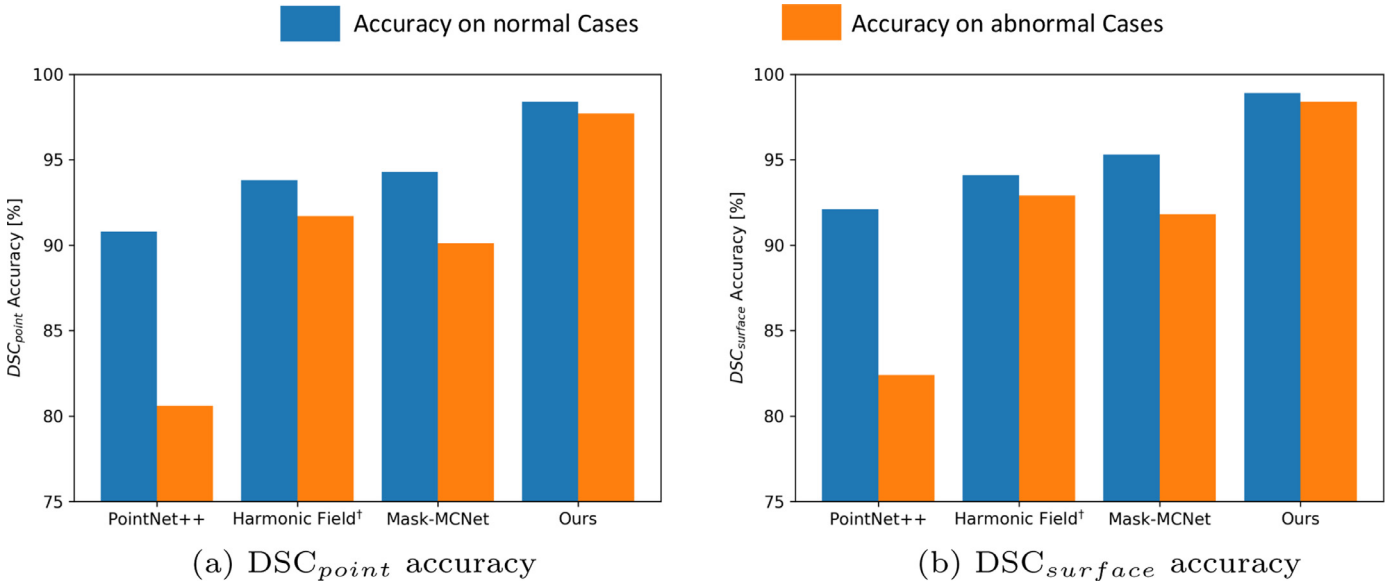


Fig. 10. The segmentation performance of different methods on normal and abnormal cases. (a) DSC accuracy on the point cloud; (b) DSC accuracy on the dental surface.

networks are derived by augmenting the baseline network with different network components or loss terms, and are trained on the same training dataset. We describe the details and present quantitative and qualitative results in the following section.

Benefits of distance estimation The distance estimator in the tooth centroid prediction module serves as a robust filter to remove the subsampled points that are far away from their nearest tooth centroids. To validate its benefits, we augment the baseline network $bNet_{cp}$ with the distance-aware filter ($bNet_{cp}$-D) and compare tooth centroid prediction results of both networks as shown in Table 1 and Fig. 4. Quantitatively, $bNet_{cp}$-D consistently improve tooth centroid prediction results of all metrics. Specifically, the max distances $MaxD(\hat{C}, C)$ from the predicted tooth centroids set $\hat{C}$ to the ground truth tooth centroids set C are reduced by a large margin (86.245×10^{-3} vs. 60.065×10^{-3}), which demonstrates that the filtered subsampled points encode little tooth shape information and usually produce unreliable tooth centroid predictions. Correspondingly, the segmentation performance is improved accordingly (1.2% and 1.3% improvements of DSC_{point} and $DSC_{surface}$, respectively).

In addition, to analyze the effectiveness of the proposed distance-aware filter more comprehensively, we visualize the displacement vectors in the point cloud (the first row of Fig. 4) and their corresponding dental model segmentation results (the second row of Fig. 4). On the one hand, we have efficiently filtered sample points that are far away from any centroid points and less likely to find an optimal position. Usually, these points will result in wrongly regressed centroid points and segmentation results, as highlighted using the red color in Fig. 4. On the other hand, with the learned filter, the predicted points tend to lie close to the target points, which demonstrates the clustering effect benefiting the proposal generation.

Chamfer distance loss To supervise the tooth centroid prediction, instead of using the intuitive way that directly forces the subsampled point to move to its nearest tooth centroid, we utilize the Chamfer distance to calculate bidirectionally distances ($bNet_{cp}$-D-CD). Compared to $bNet_{cp}$-D, the mean distances $MeanD(C, \hat{C})$ and $MeanD(\hat{C}, C)$ are reduced from 2.998×10^{-3} to 2.673×10^{-3} and 9.350×10^{-3} to 6.857×10^{-3}, demonstrating that the Chamfer distance is the key to the success of our tooth detection component. And not surprisingly, it improves the segmentation accuracy DSC_{point} from 95.5% to 95.9% and $DSC_{surface}$ from 96.5% to 96.8%, respectively.

Separation loss To validate the effectiveness of the separation loss in the accurate tooth centroids prediction, especially for incisors that are crowding and packing together, we explore the al-

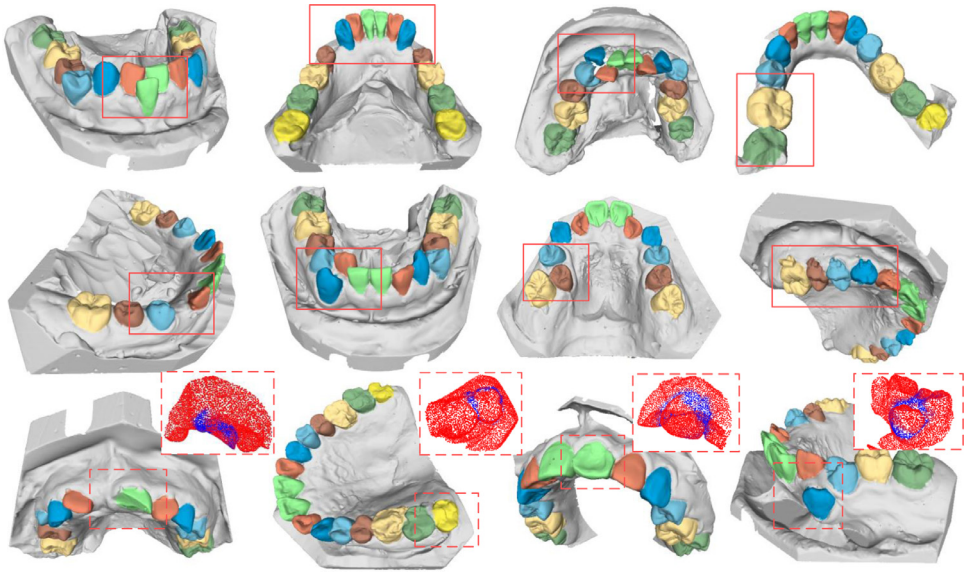


Fig. 11. Segmentation results of dental models with complex appearances, including teeth missing, crowding and irregular shapes highlighted by red boxes. Four attention maps of abnormal cases in the last row are also presented, and the red color indicates higher segmentation confidence while the blue color indicates lower segmentation confidence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ternative loss combination by augmenting the bNet$_{cp}$-C-CD with a separation loss, that is denoted as bNet$_{cp}$-C-CD-S. Statistically, with the separation loss, the MaxD$(\hat{C}, C)$ gains about a remarkably 28×10^{-3} reducing, as shown in Table 1. Although only little improvement is achieved in the other three metrics, the importance of the separation loss is presented more clearly in the visual comparison in Fig. 5. For the lower jaw dental model, bNet$_{cp}$-C-CD-S successfully predicts all correct centroids, while bNet$_{cp}$-C-CD misses one incisor since the predicted centroids around incisors are clustered into one group so as to miss one proposal in the proposal generation stage. Considering the small crown area of incisors, it contributes slightly to the segmentation metrics, 0.2% improvement of DSC$_{point}$ and 0.1% improvement of DSC$_{surface}$.

It is interesting to find that the distance metrics along two directions differ a lot. The reason is that under the supervision of the centroid detection losses, every ground truth centroid receives at least one regressed centroid, and most of them are close enough to the target, but a few regressed centroids are a little bit far away from the nearest tooth centroid. Thus, from $\hat{C}$ to C, the errors are bigger. The bidirectional constraints reveal that the two metrics share equal importance. Higher error of $(\hat{C}, C)$ indicates that some regressed centroids are far away from any tooth, which would lead to over-detection; while higher error of $(C, \hat{C})$ implies that some ground truth centroids are not detected by any regressed centroid point, which usually leads to miss-detection.

Cascaded segmentation refinement The cascading mechanism usually has a beneficial influence on the image segmentation problem. To validate the efficacy of the cascading scheme in our specific task, we first use the bNet$_{cp}$-C-CD-S as the tooth centroid prediction network and augment the base segmentation network (bNet$_{seg}$) with another sub-module to refine the preliminary results (denoted as bNet$_{seg}$-C). The quantitative results are listed in Table 2 for comparison. It can be seen that the cascading network bNet$_{seg}$-C consistently improves the segmentation performance of all four teeth types with higher DSC$_{point}$ and DSC$_{surface}$ values, especially for the tooth with irregular shape. One typical example in Fig. 6 presents the visual comparison. As highlighted in brown boxes, the special case with extreme appearance can be robustly handled by the network bNet$_{seg}$-C. In addition, the cascade scheme also brings 0.9% F1 score improvement for the tooth identification coming from the correctly predicted labels.

Confidence-aware segmentation refinement To validate the effectiveness of the confidence-aware cascading mechanism, we further augment bNet$_{seg}$-C with the confidence map as our final network (TSegNet), which encourages the segmentation network to pay more attention to the area with relative low confidence, especially near the tooth boundary and ambiguous regions with blurred geometric signals. As illustrated in Table 2, compared to the common cascade mechanism (bNet$_{seg}$-C), the confidence-aware segmentation network improves the average DSC$_{point}$ and DSC$_{surface}$ with 0.6% and 0.8% rising, respectively. The qualitative results in Fig. 7 also show that TSegNet can produce more reliable segmentation results without artifacts. More representative and challenging segmentation results of TSegNet are presented in Figs. 8 and 11.

4.3. Comparison with state-of-the-art methods

We compare our framework with several state-of-the-art point or mesh segmentation approaches, including PointNet++ (Qi et al., 2017b), harmonic field (Zou et al., 2015) and Mask-MCNet (Zanjani et al., 2019). The first one directly takes the 3D point cloud as input and achieves the state-of-the-art performance in many public segmentation datasets. The last two are specialized methods for dental model segmentation. Specifically, Zou et al. (2015) presents a semi-automatic method based on geometric surface features and outperforms other traditional methods. Zanjani et al. (2019) extends Mask-RCNN to 3D point clouds and achieves the leading performance in automatic dental model segmentation. For a fair comparison, we train PointNet++, Mask-MCNet and our method with the same point cloud input (i.e., coordinates and normals). The statistic and visual comparisons are shown in Table 3 and Fig. 9, respectively.

Quantitative comparison The overall tooth segmentation, identification and running time results are summarised in Table 3, where our framework significantly outperforms other state-of-the-art methods by a large margin. Concretely, compared with the backbone network PointNet++, our method leads to 11.9%, 10.9% and 6.8% improvements of DSC$_{point}$, DSC$_{surface}$ and F1 score, which demonstrates the effectiveness of the network architecture and loss design. Moreover, our framework inherits the efficiency of PointNet++ as it shows comparable running time (0.8 s vs. 0.3 s). Although (Zou et al., 2015) proposed a semi-automatic method that

employs harmonic field of the crown surface and high-level semantic information manually provided by users, our full-automatic framework still outperforms it in terms of segmentation accuracy and running time. Note that harmonic field based method is unable to predict tooth identification automatically. At last, it is observed that our approach achieves better results than Mask-MCNet, that is a state-of-the-art learning based method in this specific task. In particular, our method boosts the segmentation accuracy from 91.5% to 98.0% (DSC_{point}), 92.5% to 98.6% ($DSC_{surface}$), and F1 score of tooth identification from 91.2% to 94.2%. In the meanwhile, because Mask-MCNet is an anchor-based method that has to crop the dental model into several patches, our anchor-free method is more efficient and nearly 25 times faster.

We also quantify segmentation results per tooth type in Table 3. It can be seen that PointNet++ only obtains 73.0% DSC_{point} and 76.5% $DSC_{surface}$ for premolar teeth, which is much lower compared to other types. The reason is that most patients seeking orthodontic treatment are in the tooth exfoliation period, and usually have unfixed number of premolars. In addition, young children do not have premolars because these teeth do not grow until they are around 10 years old. Thus, these clustering-based learning methods, such as PointNet++ (Qi et al., 2017b), MeshSegNet (Lian et al., 2020; Lian, Wang, Wu, Liu, Durán, Ko, Shen) and TGCNN (Xu et al., 2018), cannot robustly handle the cases with missing teeth even though it is a common situation in real-world clinics.

To further demonstrate the robustness of our proposed method, we construct two testing subsets containing the abnormal (206 dental models) and normal cases (194 dental models) based on our testing dataset (400 dental models). As shown in Fig. 10, our method is robust to handle the abnormal cases and the two DSC metrics change mildly on the normal and abnormal subsets. However, the performance of PointNet++ and Mask-MCNet drops rapidly on abnormal cases, due to their lack of ability to handle abnormal cases with teeth crowding, missing and misalignment problems. It is also worth noting that Harmonic Field is a semi-automatic method, where additional human input would help process the abnormal cases to some extent, but it is still hard to find the accurate tooth boundaries.

Qualitative comparison The visual comparison results are shown in Fig. 9 for upper and lower jaws. It can be observed that segmentation results produced by our method match better with the ground truth, especially for extreme cases, such as additional braces (the third row) or crowding teeth (the fourth row). Notably, PointNet++ and Mask-MCNet, usually produce lots of artifacts in the tooth body and boundary areas. This shows that high-level features extracted by such methods are not reliable when the dental model has blurred geometric signals. Besides, the harmonic field based method heavily depends on human interactions and is sensitive to variations of tooth shape appearances. For example, it fails to extract a complete tooth body when the tooth surface is complicated (as highlighted by red arrows in the fourth column of Fig. 9). The qualitative results shown in Fig. 9 are consistent with the quantitative comparison, which further demonstrates the effectiveness and efficiency of our framework for automatic tooth segmentation and identification on dental models.

4.4. Discussions

In clinical practice, automatic dental model segmentation is an essential yet challenging problem in computed-aided orthodontics. Many algorithms, including traditional and deep learning based methods, are explored to extract the tooth individually from the dental model. However, these methods cannot robustly handle typical cases with extreme appearance before orthodontics treatments. In this paper, we propose a two-stage framework with the distance-aware centroid prediction module and the confidence-aware cascade segmentation module to successfully extract all teeth from dental models with large variations.

Parameter analysis Totally, there are five core hyper parameters used in our method. To analyze the robustness of our method to these parameters, we conduct five experiments, as shown in Fig. 12, with different parameter settings and report the statistics in terms of the segmentation accuracy. Specifically, our method achieves comparable segmentation results when changing the number ($N = 16,000$) of input sampled points (Fig. 12(a)). The reason is that after the first set abstraction layer of the point cloud encoder, 1024 points are sampled via farthest sampling to encode the local context, which is far smaller and the distribution is similar given different numbers of input points. Overall, our network is insensitive to the choice of N. As for the number ($n = 4096$) of cropped points within a tooth proposal, it should ensure a complete foreground tooth is included without too many background points. Then, when varying n to a smaller or bigger value than 4096, the slight performance degradation appears (Fig. 12(b)). The similar curve tendency can be observed in Fig. 12(a), since the two parameters n and N are tightly coupled. In our configuration, $N = 16,000$ and $n = 4096$ achieved the best performance. In the tooth centroid detection stage, $M = 256$ candidate points are regressed for indicating tooth objects. Thus, if M is small, some teeth with small crown area, e.g. the lower central incisor, may be miss-detected. As illustrated in Fig. 12(c), the performance drops rapidly when M is set to 64. Instead, when M is set to 256 or bigger, it is sufficient to capture all tooth objects in the dental model and no performance fluctuation occurs. Meanwhile, in the distance-aware filter, the distance threshold α is set to 0.2, which is consistent with the receptive field of the last set abstraction layer in the encoder. A smaller α would filter more candidate points and leads to miss-detection, while a bigger α usually takes some points far from any tooth into consideration and leads to over-detection, as shown in Fig. 12(d). The last core parameter is the distance threshold $l = 0.015$ in DBSCAN clustering algorithm. Due to the strong clustering effect achieved by our method in the tooth centroid detection stage, it is insensitive to the choice of l (Fig. 12(e)).

*Centroid vs. bounding box*In the first stage of our network, instead of utilizing the bounding box, an intuitive way to indicate an object in 2D or 3D images, we design a centroid voting method to detect and represent each tooth. To investigate the effectiveness, we conduct an experiment to compare the two representations by only replacing the centroid prediction module to a bounding box regression module in the TSegNet. As illustrated in Fig. 13, a premolar tooth is failed to be detected by any bounding box. This is because the predicted bounding box of the premolar tooth has a relative large overlap with the bounding box of the neighboring molar tooth and is filtered by the non-max suppression (NMS) operator that is designed to remove redundant boxes. In addition, there is an under-segmentation in the canine tooth because the corresponding bounding box does not cover the tooth appropriately. Generally, the centroid representation has two main advantages compared to the bounding box representation in this task. The first is that the centroid voting and detection is more efficient and accurate than the NMS operator to filter redundant bounding boxes. The second is that the tooth centroid is more stable information and less sensitive to the tooth shape boundary, while the bounding box is mainly decided by the tooth shape with blurred geometric signals.

*Teeth missing problem*Missing teeth is a common problem in dental clinics. As shown in the Fig. 14, there are two typical teeth missing cases referencing the normal tooth case in Fig. 14(a). In Fig. 14(b), it misses a cuspid tooth in the left half and a premolar in the right half, but visually there is no vacant position. While in Fig. 14(c), it misses a cuspid tooth in the left with a corresponding vacant position. Generally, it is challenging for semantic-based

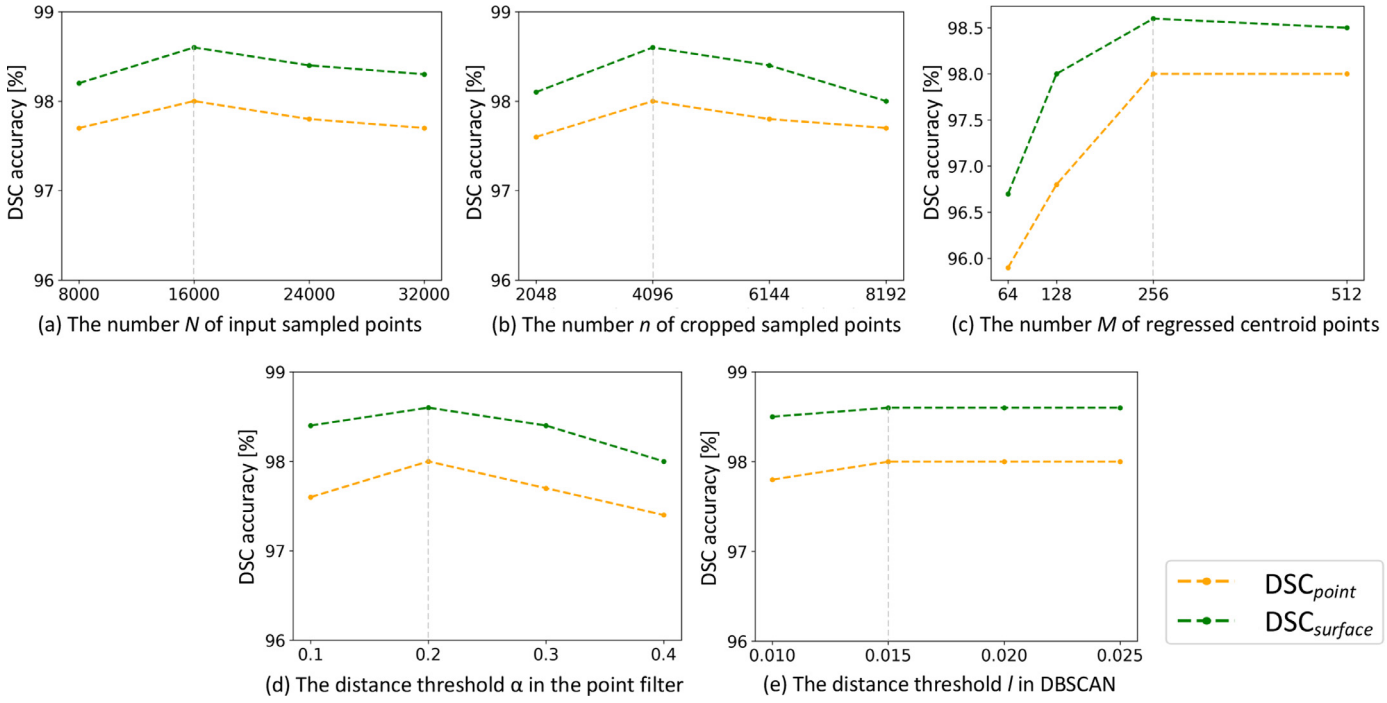


Fig. 12. The tooth segmentation performance of our TSegNet, when changing the value of different parameters.

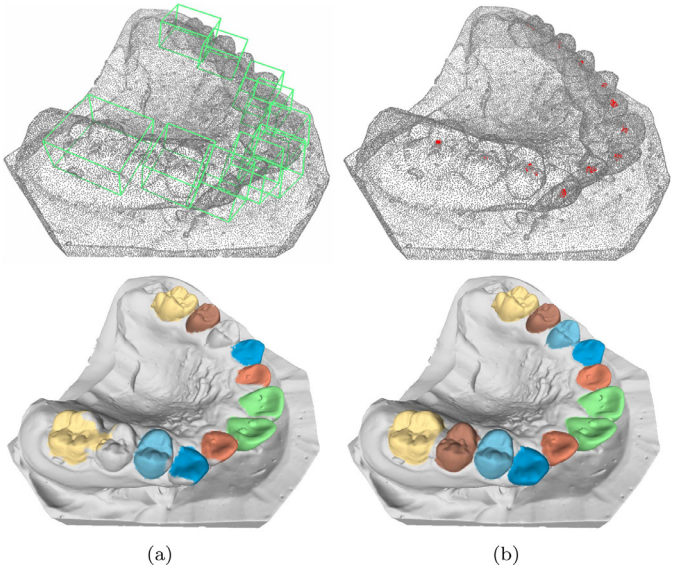


Fig. 13. Visual comparison between the bounding box and tooth centroid representations. (a) The predicted bounding boxes and corresponding segmentation results. (b) The predicted tooth centroids and corresponding segmentation results.

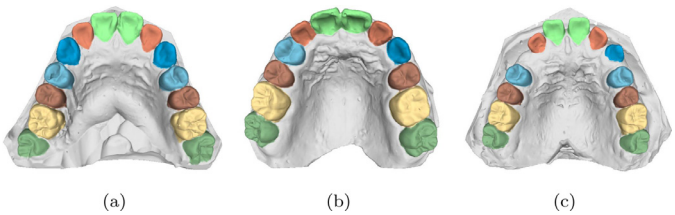


Fig. 14. Normal and two typical examples with missing teeth. (a) A normal case; (b) Missing a cuspid tooth in the left half and a premolar in the right half; (c) Missing a cuspid tooth in the left half.

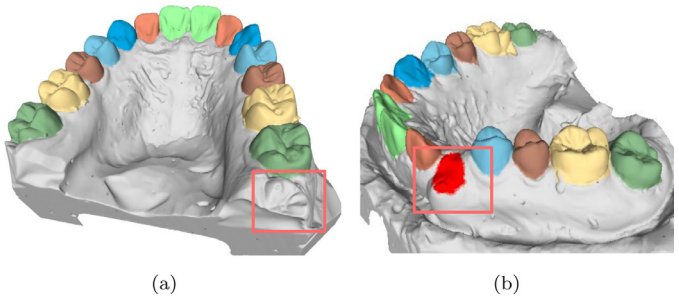


Fig. 15. Failure cases. (a) The wisdom tooth with abnormal shape is failed to be detected. (b) The rudimentary tooth is hard to be segmented accurately due to the small crown area.

learning methods, e.g., PointNet++, to successfully process the abnormal case in Fig. 14(b). However, instead of relying on the global semantic context encoded in the input point cloud, we regress a tooth centroid from the local points around the target tooth. Thus, with the proposed distance filter, the Chamfer distance loss and the separation loss, our method can accurately regress the centroid points to indicate the tooth object confidently.

Limitation Although our proposed framework has achieved outstanding tooth segmentation results and outperforms many state-of-the-art methods, it presents some limitations that are worth considering. One typical example is that it tends to yield incomplete tooth segmentation in some cases such as the wisdom tooth and the rudimentary tooth. One possible reason is that these cases are quite rare and seldom seen by the network during the training phase. Specifically, the wisdom tooth is a special case for humans, because it has large variations and usually a small part of the crown is appeared on the dental model. As shown in Fig. 15(a), we fail to detect the wisdom tooth marked in the brown box. Another case is the rudimentary tooth, it shares similar situation that the seen crown part has small area and is quite different from other teeth. Thus, some background area is likely to be treated as part of a rudimentary tooth, as illustrated in Fig. 15(b). In the fu-

ture, we would like to explore more effective method on datasets with imbalanced tooth type distribution.

5. Conclusion

In this work, we develop a novel fully automatic algorithm to segment tooth on 3D dental models guided by the tooth centroid information. The algorithm builds upon a two-stage neural network containing a robust tooth centroid prediction subnetwork and a single tooth segmentation subnetwork with our novel components and loss functions. We have evaluated our algorithm both qualitatively and quantitatively, and compared it with the state-of-the-art learning and non-learning based methods, where our method produces superior results and significantly outperforms others.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Zhiming Cui: Methodology, Software, Writing - original draft. **Changjian Li:** Methodology, Writing - original draft, Software. **Nenglun Chen:** Methodology, Writing - original draft. **Guodong Wei:** Methodology, Data curation. **Runnan Chen:** Methodology. **Yuanfeng Zhou:** Data curation, Writing - original draft. **Wenping Wang:** Supervision, Writing - original draft.

References

Cobourne, M.T., DiBiase, A.T., 2015. Handbook of Orthodontics. Elsevier Health Sciences.

Cui, Z., Li, C., Wang, W., 2019. 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.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, pp. 226–231.

Grace, M., 2000. Dental notation. Br. Dent. J. 188, 229.

Grzegorzek, M., Trierscheid, M., Papoutsis, D., Paulus, D., 2010. A multi-stage approach for 3D teeth segmentation from dentition surfaces. Springer. International Conference on Image and Signal Processing, 521–530.

Hajeer, M., Millett, D., Ayoub, A., Siebert, J., 2004. Applications of 3D imaging in orthodontics: part i. J. Orthod. 31, 62–70.

Hajeer, M., Millett, D., Ayoub, A., Siebert, J., 2004. Applications of 3D imaging in orthodontics: part ii. J. Orthod. 31, 154–162.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969.

Hou, J., Dai, A., Nießner, M., 2019. 3D-SIS: 3D semantic instance segmentation of RGB-D scans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4421–4430.

Kondo, T., Ong, S.H., Foong, K.W., 2004. Tooth segmentation of dental study models using range images. IEEE Trans. Med. Imaging 23, 350–362.

Kronfeld, T., Brunner, D., Brunnett, G., 2010. Snake-based segmentation of teeth from virtual dental casts. Comput. Aided Des. Appl. 7, 221–233.

Kumar, Y., Janardan, R., Larson, B., Moon, J., 2011. Improved segmentation of teeth in dental models. Comput. Aided Des. Appl. 8, 211–224.

Lechuga, L., Weidlich, G.A., 2016. Cone beam CT vs. fan beam CT: a comparison of image quality and dose delivered between two differing ct imaging modalities. Cureus (Palo Alto, CA) 8 (9), E778.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. Pointcnn: convolution on x-transformed points. In: Advances in Neural Information Processing Systems, pp. 820–830.

Li, Z., Ning, X., Wang, Z., 2007. A fast segmentation method for STL teeth model. IEEE. 2007 IEEE/ICME International Conference on Complex Medical Engineering, 163–166.

Lian, C., Wang, L., Wu, T. H., Liu, M., Durán, F., Ko, C. C., Shen, D., 2019. MeshsNet: deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces. Springer. International Conference on Medical Image Computing and Computer-Assisted Intervention, 837–845,

Opitz, J., Burst, S., 2019. Macro f1 and macro f1. arXiv preprint arXiv:1911.03347.

Lian, C., Wang, L., Wu, T.H., Wang, F., Yap, P.T., Ko, C.C., Shen, D., 2020. Deep multi--scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners. IEEE Trans. Med. Imaging 39 (7), 2440–2450.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: deep learning on point sets for 3Dclassification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems, pp. 5099–5108.

Sinthanayothin, C., Tharanont, W., 2008. Orthodontics treatment simulation by teeth segmentation and setup. IEEE. 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 81–84.

Sun, D., Pei, Y., Song, G., Guo, Y., Ma, G., Xu, T., Zha, H., 2020. Tooth segmentation and labeling from digital dental casts. IEEE. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 669–673.

Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., Cheng, X., 2019. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access 7, 84817–84828.

Wongwaen, N., Sinthanayothin, C., 2010. Computerized algorithm for 3D teeth segmentation. IEEE. 2010 International Conference on Electronics and Information Engineering, V1–277.

Wu, K., Chen, L., Li, J., Zhou, Y., 2014. Tooth segmentation on dental meshes using morphologic skeleton. Comput. Graph. 38, 199–211.

Wu, W., Qi, Z., Fuxin, L., 2019. Pointconv: deep convolutional networks on 3D point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9621–9630.

Xu, X., Liu, C., Zheng, Y., 2018. 3D tooth segmentation and labeling using deep convolutional neural networks. IEEE Trans. Vis. Comput.Graph. 25, 2336–2348.

Yamany, S. M., El-Bialy, A. M., 1999. Efficient free-form surface representation with application in orthodontics. International Society for Optics and Photonics. Three-Dimensional Image Capture and Applications II, 115–124.

Yaqi, M., Zhongke, L., 2010. Computer aided orthodontics treatment by virtual segmentation and adjustment. IEEE. 2010 International Conference on Image Analysis and Signal Processing, 336–339.

Yuan, T., Liao, W., Dai, N., Cheng, X., Yu, Q., 2010. Single-tooth modeling for 3D dental model. J. Biomed. Imaging 2010, 9.

Zanjani, F. G., Moin, D. A., Claessen, F., Cherici, T., Parinussa, S., Pourtaherian, A., Zinger, S., et al., 2019. Mask-MCNet: instance segmentation in 3D point cloud of intra-oral scans. Springer. 22nd International Conference on Medical Image Computing and Computer Assisted Intervention, (MICCAI2019).

Zhao, M., Ma, L., Tan, W., Nie, D., 2006. Interactive tooth segmentation of dental models. IEEE. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 654–657.

Zhou, Y., Tuzel, O., 2018. VoxelNet: end-to-end learning for point cloud based 3D object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4490–4499.

Zou, B.j., Liu, S.j., Liao, S.h., Ding, X., Liang, Y., 2015. Interactive tooth partition of dental mesh base on tooth-target harmonic field. Comput. Biol. Med. 56, 132–144.