IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 10, 2020, accepted April 28, 2020, date of publication May 6, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991799

Automatic Segmentation of Individual Tooth in Dental CBCT Images From Tooth Surface Map by a Multi-Task FCN

YANLIN CHEN[1], HAIYAN DU[1], ZHAOQIANG YUN[1], SHUO YANG[2], ZHENHUI DAI[3], LIMING ZHONG[1], QIANJIN FENG[1], AND WEI YANG[1]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Stomatological Hospital, Southern Medical University, Guangzhou 510280, China
[3]Department of Radiotherapy, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510120, China

Corresponding author: Wei Yang (weiyanggm@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 81771916.

ABSTRACT Accurate and automatic segmentation of individual tooth is critical for computer-aided analysis towards clinical decision support and treatment planning. Three-dimensional reconstruction of individual tooth after the segmentation also plays an important role in simulation in digital orthodontics. However, it is difficult to automatically segment individual tooth in cone beam computed tomography (CBCT) images due to the blurring boundaries of neighboring teeth and the similar intensities between teeth and mandible bone. In this work, we propose the use of a multi-task 3D fully convolutional network (FCN) and marker-controlled watershed transform (MWT) to segment individual tooth. The multi-task FCN learns to simultaneously predict the probability of tooth region and the probability of tooth surface. Through the combination of the tooth probability gradient map and the surface probability map as the input image, MWT is used to automatically separate and segment individual tooth. Twenty-five dental CBCT scans are used in the study. The average Dice similarity coefficient, Jaccard index, and relative volume difference are 0.936 ($\pm$0.012), 0.881 ($\pm$0.019), and 0.072 ($\pm$0.027), respectively, and the average symmetric surface distance is 0.363 ($\pm$0.145) mm for our method. The experimental results demonstrate that the multi-task 3D FCN combined with MWT can segment individual tooth of various types in dental CBCT images.

INDEX TERMS Individual tooth segmentation, dental CBCT, deep learning, marker-controlled watershed transform.

I. INTRODUCTION

Dental cone beam computed tomography (CBCT), a diagnostic imaging technique, is widely used for dental diseases and dental problems researching [1]. The segmentation of individual tooth in CBCT images facilitates the observation of slices or volumes of the target tooth by dentists, thereby enabling more precise diagnostic decision-making and treatment planning. Moreover, individual tooth segmentation is a necessary step to form a digital tooth arrangement, simulate tooth movement, and build the tooth setup. However, manual segmentation of tooth is tedious, time-consuming and prone to intra- and inter-observer variability. A method to automatically segment individual tooth can eliminate subjective errors in tooth boundary delineation and reduce the workload of dentists.

Several challenges are encountered in the segmentation of individual tooth in dental CBCT images, which are due to the similar intensities between teeth and alveolar bone, close proximity of neighboring teeth, where some are even touching each other. Fig. 1 presents two examples of the complicated dental structures. To address these difficulties, many tooth segmentation methods for dental CBCT images have been proposed. These methods can be divided into two categories: conventional methods which require handcrafted features, and deep learning methods which often need a lot of samples. The conventional methods include graph cut-based methods, template-based fitting methods and level set methods. Evain *et al.* [2] used graph cut methods to segment individual tooth from dental CBCT images and achieved a high Dice score. However, this method needed user input

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Li.

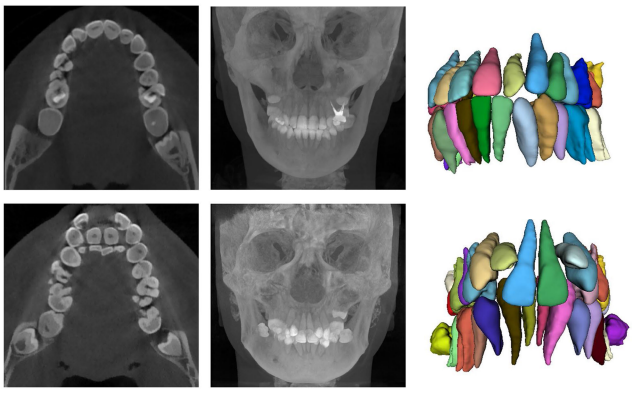


FIGURE 1. Display of the complicated dental structures. From left to right: the transverse plane, a coronal MIP image, and the manual delineation.

to build a statistical shape prior and could not automatically segment the teeth. In addition, graph cut methods are influenced by the definition of the foreground, and the changes of foreground during iterations would affect the final segmentation results. Template-based fitting methods [3] realized the segmentation of anterior or premolar teeth which usually have monoradicular shapes, but they lacked robustness when there were teeth presenting multiradicular anatomies, such as molar teeth. The level set method [4]–[7] is the most popular method for tooth segmentation in dental CBCT images because it can deal with complex tooth topological changes well. Gan *et al.* [4] developed a two-step level set method to segment both tooth and alveolar bone; this method used global level set to extract bony tissues in the first step, and applied the radon transform and local level set to segment individual tooth from alveolar bone in the second step. Although the segmentation results of the two-step level set method were satisfactory, the problem of segmenting a crown from image data which was scanned in a closed bite position or with metal artifacts remained to be solved. In addition, the selection of initial contour curve and the setting of parameters will influence the level set curve evolutions. Therefore, both the contour initialization and the parameters setting should be carefully considered in the two-step level set method. With the development of deep learning [8], data-driven methods have been used in many image processing domains [9]–[13] and have yielded promising results. However, no methods had been presented to use deep learning to segment individual tooth in CBCT images until recently, Cui *et al.* [14] exploited 3D mask R-CNN as a base network to realize automatic tooth segmentation and identification from CBCT images. The method that was proposed by Cui *et al.* focused only on a tooth dataset that excludes the wisdom teeth. Considering that the numbers and classes of teeth vary among patients, the segmentation of individual tooth in the oral environment without ignoring any teeth would be beneficial in the clinical applications.

The current individual tooth segmentation methods performed in dental CBCT images cannot simultaneously handle the following situations: images with metal artifacts, teeth in a natural bite or closed bite position, and special tooth types, such as wisdom teeth and implanted teeth. Many of these methods cannot be implemented automatically. In this work, we propose a method to address these issues and realize individual tooth segmentation in dental CBCT images based on a fully convolutional network (FCN) [15]. However, it is difficult to assign a label to individual tooth, because the tooth categories and indices are extensive. Utilizing multi-class FCN to achieve individual tooth segmentation seems unrealistic. Instead, we use an FCN to predict both tooth region and tooth surface, and then segment individual tooth through marker-controlled watershed transform (MWT) [16]–[18]. Since the dental CBCT images are in the form of 3D high-dimensional structures, we exploit a 3D network structure that can pay more attention to the spatial continuity of the image. V-net [19] is a 3D FCN that was developed based on U-net [20]. Compared with U-net, it uses residual architecture in every convolutional stage so that the information in feature map can be utilized more efficiently. In our study, we select the V-net architecture, which can reduce information loss and learn finer structures to segment tooth region efficiently.

The probability map of tooth regions contains useful cues for the detection of the surfaces from individual tooth. However, the boundaries between a predicted tooth and the neighboring teeth may be blurred, thereby resulting in a discontinuity in the gradient of the tooth probability map. Yang *et al.* [21] proposed the segmentation of lung field in a chest X-ray by using the information of the detected lung boundary map; this method realized state-of-the-art performance. Inspired by their work, we use the network to predict not only the tooth region but also the tooth surface. With the combination of the tooth probability gradient map and the surface probability map, more information regarding the tooth boundaries is collected to realize higher performance in individual tooth segmentation.

The main objective of tooth surface prediction is to produce supplementary information to the gradient of tooth probability map. However, the tooth surface and the tooth region are strongly related. The use of a single decoder path to update the coefficients for both tooth region and tooth surface predictions would result in the learning of redundant characteristics. In addition, it would cause difficulties in gathering different contextual cues between the two targets. To effectively train the network, the feature maps are upsampled with two branches [22] in the decoder path to better learn the characteristics of the tooth region and tooth surface.

The contributions of our work mainly include the following:

1) The proposed method can automatically segment various kinds of teeth such as implanted teeth, wisdom teeth, supernumerary teeth, and the replacing teeth.
2) The method can automatically segment teeth in non-open bite positions and handle dental CBCT images with metal artifacts.

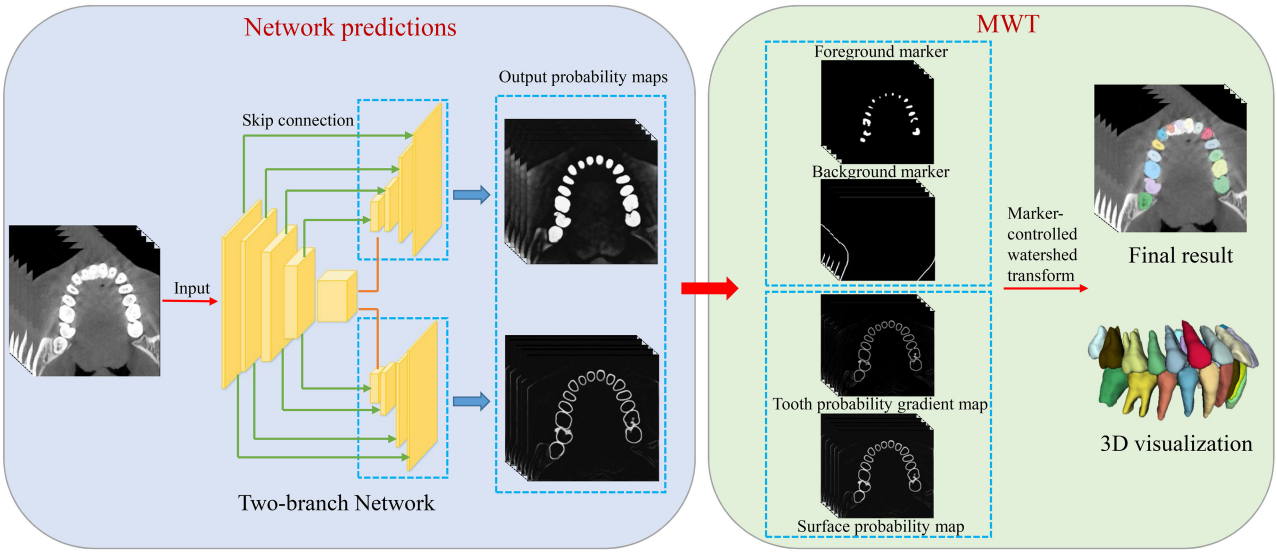


FIGURE 2. Framework of our proposed method.

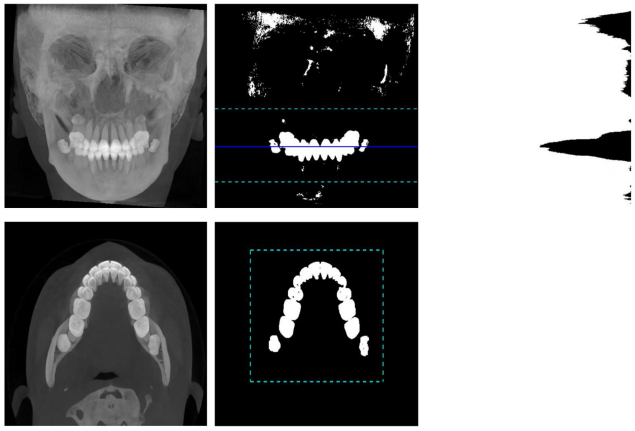


FIGURE 3. Workflow of the detection of a valid tooth region. In the first row, from left to right: a coronal MIP image of a dental CBCT image, a binary image, and a horizontal projection image. In the second row, from left to right: an axial MIP image and a tooth mask. The dashed lines indicate the valid region.

The remainder of this paper is organized as follows. The framework and details of our method are described in Section II. The experimental results are provided in Section III. Finally, we discuss our work and present our conclusions in Section IV.

II. METHOD

A. OVERVIEW

This work aims to develop an automatic method to segment individual tooth in dental CBCT images. The core of our proposed method is the effective utilization of the probability maps of both tooth region and surface. The framework of our method is illustrated in Fig. 2. First, a dental CBCT data was delimited and cropped within a valid region and then fed into a two-branch FCN for predicting tooth region and tooth surface. After the thresholding operation performed on the tooth probability map, a mask image was generated to produce a foreground marker and a background marker for MWT. From the probability map of tooth region, tooth probability gradient map can be obtained. Finally, we combined the tooth probability gradient map and the surface probability map with a specified weight, and applied MWT to yield the individual tooth segmentation.

B. PREPROCESSING

Teeth cover only a small area in dental CBCT images, which results in data imbalance. To relieve these data imbalance problems, reduce the computational load, and decrease the learning complexity, it is necessary to delimit a small and valid region from a dental CBCT image for segmentation. Yun *et al.* [23] proposed the detection of the dental arch in dental CBCT images by thresholding operations and projection analyses on the maximum intensity projection (MIP) image. Similarly, we delimited the valid region for segmentation through a MIP-based method. The workflow is illustrated in Fig. 3. Given a dental CBCT image, a coronal MIP image was generated and then binarized via the thresholding method to yield a coronal mask of teeth. Next, the coronal mask was projected horizontally. Within a tolerance, the scale of the region with the highest peak value and continuous range in the projection image was regarded as the range of axial slices that contained teeth (as seen in the cyan dashed lines of the second subgraph in Fig. 3). Based on the detected slices, we obtained an axial MIP image. Unlike Yun *et al.*'s method, we selected the threshold value for the axial MIP image according to the range that the intensity standard deviation of axial MIP image belonged to. For example, when the intensity standard deviation was lower than 750, the threshold value was set

to 2450; when the intensity standard deviation was between 750 and 810, the threshold value was set to 2600. After the binary operation, morphological opening was conducted. The pixel area which was below a specified size was regarded as the noise area and was removed. Then, we determined the tooth range in the axial plane by delimiting the range of pixels in white (tooth pixels) in the axial mask image (as seen in the cyan dashed lines of the fifth subgraph in Fig. 3). At this point, we obtained the 3D bounding box of tooth region in dental CBCT images. We denoted the region inside the bounding box as I.

The overall intensity and contrast shift often vary among dental CBCT images due to the differences in acquisition conditions and patient variabilities. This would affect the network learning and the prediction performance. Therefore, a preprocessing step for intensity normalization is required to achieve intensity consistency. The normalization of a cropped region I is expressed as $I \leftarrow (I - \mu)/\sigma$, where μ and σ are the intensity mean and standard deviation, respectively, of I.

C. PREDICTING THE TOOTH REGION AND SURFACE

1) FULLY CONVOLUTIONAL NETWORK

A typical V-net architecture contains an encoder path, a decoder path, residual blocks and skip connections. The residual block indicates that the input of each stage not only participates in nonlinear operations in convolutional layers but also is added to the output of the last convolutional layer of that stage. We utilized a modified V-net architecture to concurrently perform two prediction tasks. Different from the typical V-net, the modified V-net is composed of an encoder path that consists of four encoder layers, followed by two decoder paths, each of which consists of four decoder layers. In each encoder layer, there are volumetric convolutions with kernel size of $5 \times 5 \times 5$, where each convolution is followed by a batch normalization (BN), a rectified linear unit (ReLU) and a dropout. Considering the memory limitations of GPU, the number of channels output from the first encoder layer is set to 8. As the data proceeds through different layers along the encoder path, its resolution is reduced while the number of feature channels is doubled. This is conducted by convolution with kernel size of $2 \times 2 \times 2$ and a stride of 2. Unlike pooling layer which also reduces resolution, using a convolution operation to downscale feature maps can preserve more contextual information by controlling its stride. Each decoder layer consists of a transposed convolution with kernel size of $2 \times 2 \times 2$ and a stride of 2, a concatenation operation and several $5 \times 5 \times 5$ convolutions with BN, ReLU and dropout. The transposed convolution is utilized for upsampling and the concatenation operation is adopted for fusing the feature maps from the encoder layer into the corresponding decoder layer. At the last decoder layer in each decoder path, a convolution with kernel size of $1 \times 1 \times 1$ is employed and the probability maps are output.

In our work, a preprocessed CBCT image and the corresponding ground truth were fed into the modified V-net to learn their mapping relationships at the patch-level [24]–[26]. The efficiencies of network learning as well as feature mapping can be influenced by the loss function and the number of patches simultaneously learned by the network. The loss function is important for its role in measuring the difference between the network prediction results and the ground truth, and in updating gradients iteratively during network learning. The network simultaneously predicts several input patches; thus, the loss function learned each time is based on these several image patches.

2) LOSS FUNCTION

In this work, both tooth region segmentation and tooth surface detection are dense binary prediction tasks. The cross entropy loss (CE) is the most common loss function for binary prediction, and we used it in each prediction task to update the weights. The CE is formulated as follows:

$$CE(p) = -y \log p - (1 - y) \log (1 - p) \tag{1}$$

where $y \in \{0, 1\}$ specifies the ground-truth label and $p \in (0, 1)$ specifies the predicted probability of being the tooth region or tooth surface of each voxel. A small value of Eq. (1) corresponds to minor differences between the ground truth and the prediction. We denoted the CE of tooth surface prediction task as $CE_{surface}$ and the CE of tooth region prediction task as CE_{tooth}. Since the prediction of tooth surface was an auxiliary task in this multi-task learning, we assigned a weight λ to $CE_{surface}$. To simultaneously conduct the two tasks, the weighted $CE_{surface}$ and CE_{tooth} were summed to obtain the final loss function, which is computed as Eq. (2).

$$L(p_1, p_2) = \lambda \times CE_{surface}(p_1) + CE_{tooth}(p_2) \tag{2}$$

where p_1 and p_2 denote the prediction probabilities of the tooth surface prediction task and the tooth region prediction task, respectively.

D. SEGMENTATION OF INDIVIDUAL TOOTH

Watershed transform (WT) is a segmentation algorithm of mathematical morphology that is based on topological theory. It regards an image as topological landscape where the intensity of an image pixel corresponds to the altitude, while local minima with their affected regions correspond to ''catchment basins''. Assume that there is water rising in the ''catchment basins''. When the water level reaches the boundary point and stops spreading, the landscape is divided into several regions by watershed ridge lines, and the segmentation process is complete. To obtain the boundary information of an image, WT is often conducted in a gradient map. Due to its good response to weak edge information, WT can effectively handle segmentation targets with weak boundaries. However, it may also oversegment the image due to noise factors. To overcome this problem, MWT is developed. Here, foreground markers are seed points that represent ''catchment basins'', and the algorithm will perform segmentation surrounding these regions; in contrast, background markers represent irrelevant regions, and the algorithm will not segment

them. With markers as guidance, the noise extremal regions in images are ignored during the implementation of MWT, thereby avoiding the problem of oversegmentation. MWT can be used to deal with multi-class segmentation task.

The segmentation of individual tooth is a type of multi-class segmentation problem in essence. The interfaces between some neighboring teeth are ambiguous, thereby leading to the segmentation error-prone. With the advantages of handling weak edges, avoiding oversegmentation, and high efficiency, MWT is an effective segmentation tool for solving this multi-object segmentation task. Thus, we employed MWT for individual tooth segmentation. The results in the MWT pipeline are presented in Fig. 4. All the steps in our MWT implementation were conducted in 3D space.

Prior to the generation of markers, we transformed the tooth probability map into a binary image (Fig. 4(b)) via the thresholding method with a threshold value of 0.7. In general, there were still non-tooth structures of small volume size in the thresholded image. The non-tooth structures will influence the generation of markers. To further remove these non-tooth structures, we used a spherical structural element with radius of 2 to perform erosion operation on the thresholded image. Then, we removed the connected regions with volumes of less than 500 voxels and obtained a binary tooth image, which was denoted as $\boldsymbol{B}$.

In our study, both the foreground marker and the background marker were extracted based on $\boldsymbol{B}$. To generate the foreground marker $\boldsymbol{M}_f$, we used a $5 \times 5 \times 5$ structural element to perform opening operation on $\boldsymbol{B}$, and then we performed erosion operation using a structural element with radius of 2. The generation of the foreground marker can be typically expressed as

$$\boldsymbol{M}_f = (\boldsymbol{B} \circ \boldsymbol{C}) \Theta \boldsymbol{C} \tag{3}$$

where $\boldsymbol{C}$ denotes the structural element, $\circ$ denotes the opening operation, and Θ denotes the erosion operation. It is expected that the background marker is not in close proximity to the tooth surface. This can be realized through thinning the background by computing its skeleton. To generate the background marker $\boldsymbol{M}_b$, we firstly conducted the dilation operation on $\boldsymbol{B}$. We set the radius of the spherical structural element which was used in the dilation operation to 3. Then, we converted the dilation result (Fig. 4(c)) into a distance map (Fig. 4(d)) by Euclidean distance transformation. Finally, we performed watershed transform on the distance map, and we set the watershed ridge lines as the background marker. We used $\mathrm{E}(\cdot)$ and $\mathrm{W}(\cdot)$ to denote Euclidean distance computation function and watershed transform function, respectively. Thus, the generation of the background marker can be expressed as:

$$\boldsymbol{M}_b = \mathrm{W}(\mathrm{E}(\boldsymbol{B} \oplus \boldsymbol{C}) \tag{4}$$

where $\oplus$ denotes the dilation operation. The foreground marker and the background marker are displayed in black and white, respectively, in Fig. 4(e). We computed the gradient of the tooth probability map and combined it with the tooth surface probability map to form the input image of MWT. Guided by the foreground and background markers, MWT was used to segment individual tooth automatically.

E. EVALUATION METRICS

The automatic tooth segmentation performance was evaluated in terms of four metrics: the Jaccard similarity coefficient (Ω), the Dice similarity coefficient (DSC), the relative volume difference (RVD), and the average symmetric surface distance (ASSD).

Let us denote by D the segmentation result and by G the ground truth. The Jaccard index is computed as:

$$\Omega = \frac{|\mathrm{D} \cap \mathrm{G}|}{|\mathrm{D} \cup \mathrm{G}|} \tag{5}$$

where $|\cdot|$ is the cardinality of the set, D$\cap$G is the intersection of D and G, and D$\cup$G is the union of D and G. DSC is the overlap ratio between the ground truth G and the segmentation result D. It is formulated as Eq. (6). The larger the value of DSC is, the more accurate the segmentation result is.

$$DSC = \frac{2 \times |\mathrm{D} \cap \mathrm{G}|}{|\mathrm{D}| + |\mathrm{G}|} \tag{6}$$

RVD is used to measure the difference between the segmentation result and the ground truth. It is defined as:

$$RVD = |\mathrm{D} - \mathrm{G}| \,/\, |\mathrm{G}| \tag{7}$$

ASSD is the average distance between the estimated segmentation surface S_D and the ground truth surface S_G. Let d and g be the points on surfaces S_D and S_G, respectively. $dist(d, S_G)$ is the nearest Euclidean distance from a surface point d to surface S_G. $mean\{\cdot\}$ is the arithmetical average operator. ASSD is defined as:

$$\begin{aligned} ASSD\,(S_D, S_G) = {} & mean\,\{mean\,\{dist\,(d\ \ , S_G)\,,\ d \in S_D\}\,, \\ & \times\ mean\,\{dist\,(g,\ S_D)\ \ , g \in S_G\}\} \end{aligned} \tag{8}$$

III. RESULTS

A. DATASET AND EXPERIMENTAL SETTINGS

Clinical dental CBCT images of 25 patients were randomly collected by Stomatological Hospital of Southern Medical University, Guangzhou, China. The CBCT images were acquired by a NewTom VGi (QR s.r.l., Verona, Italy) scanner with the following imaging parameters: 110 kVp and 3−8 mA (pulse mode). The image data were cropped to different volume size, with a voxel size of 0.3 mm × 0.3 mm × 0.3 mm. The number of slices in each volume ranged from 376 to 541. In addition, the axial length ranged from 470 to 555 pixels, while the axial width ranged from 376 to 512 pixels. All the CBCT images were acquired under natural bite or closed bite conditions. The subjects included 14 males and 11 females of ages 10 to 49 years. As presented in Table 1, the dataset consists of supernumerary teeth, implanted teeth, and metal restored teeth, and the tooth location and shape vary among the samples. The dataset contained more than 770 teeth. With a large number of teeth and extensive tooth types, the 25 subjects of our study could ensure the robustness of FCN.

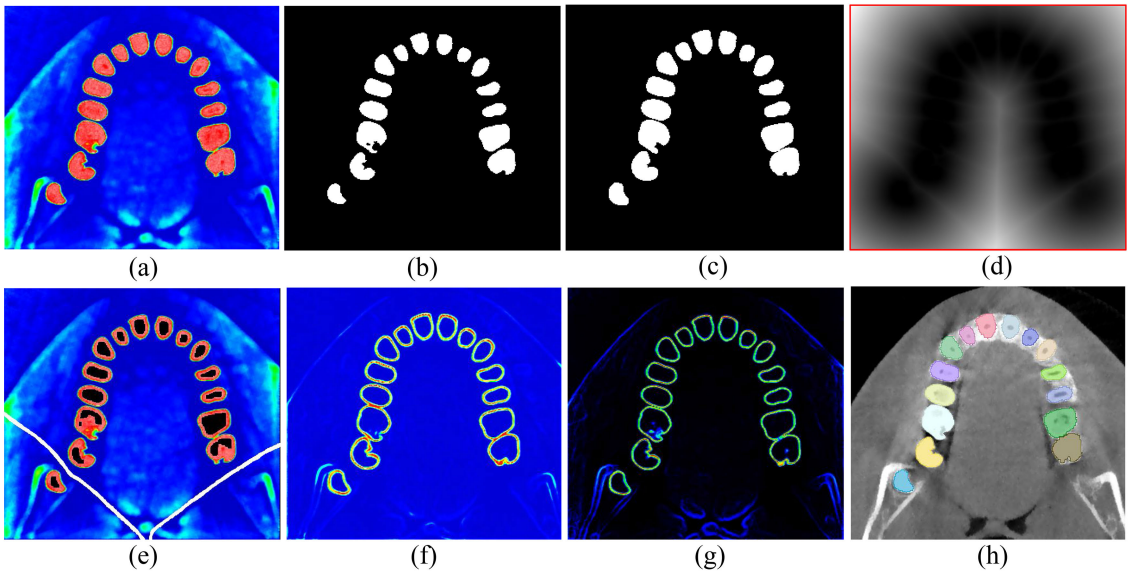


FIGURE 4. Images that are produced in the MWT pipeline. (a) a tooth probability map; (b) the thresholded image; (c) the dilation result; (d) the Euclidean distance map; (e) a visualization of the foreground marker and the background marker in black and white, respectively; (f) the tooth surface probability map; (g) the tooth probability gradient map; and (h) the final segmentation result.

TABLE 1. Description of the data conditions.

Tooth category	Tooth number	Patient number
Supernumerary teeth	7	3
Implanted teeth	3	1
Metal restored teeth	22	8
The replacing teeth	18	3
Malposed teeth	8	6
Missing teeth	/	11

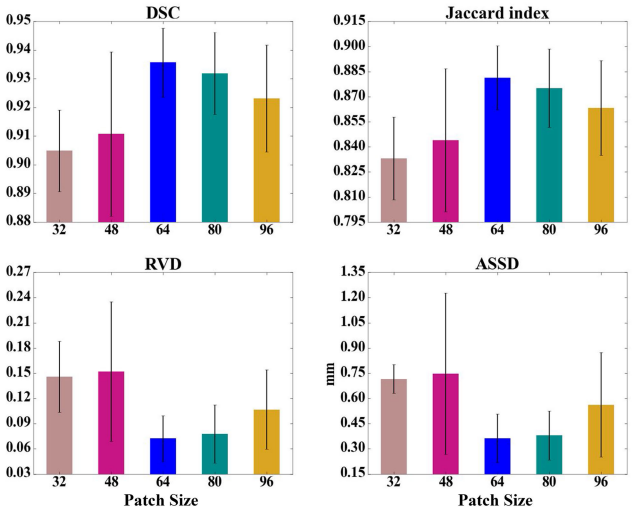


FIGURE 5. Segmentation performance using various patch sizes.

We delineated each tooth manually to form the segmentation ground truth. Moreover, we used a $3\times3\times3$ structural element to respectively perform morphological dilation and erosion operations on the manual delineations of individual tooth. After subtracting the erosion result from the dilation result, we obtained the ground truth image of the tooth surface.

The dataset was randomly divided into a training set and a test set in a ratio of 4:1. To improve the prediction performance, we augmented the training dataset on-the-fly via random rotation, zooming, and flipping to imitate the possible cases in practice. For instance, there were cases in which the head was tilted to the right or left. Therefore, we augmented the training data with random horizontal rotations in the range of [−20, 20] degrees. In addition, children may have smaller heads than adults and obese patients may have larger heads than the thin patients. Thus, we magnified the images at the scales of [0.9, 1.1]. We also randomly flipped the training data in horizontal direction to render the segmentation model robust to data with flipping. Test Time Augmentation (TTA) [27], [28] was also conducted on the test set to further improve the predictions. Our strategy for TTA consisted of the following geometric transformations: horizontal flipping, fixed angle rotation, scaling within a specified range, and a series of corresponding inverse operations. Each test case and its augmented examples were fed into the FCN model to obtain a predictive distribution, which was denoted as $\mathbf{Y}=\{y_i, y_j; i, j = 1, 2, \ldots, n\}$, where y_i is the prediction of the tooth region; y_j is the prediction of the tooth surface; and n is the serial number of predictions. We respectively computed the average of y_i and y_j to obtain the final predictions for the test case.

Considering the memory limitations of GPU, we performed patch-level learning using sliding windows in the 3D space of the images. The patches that were used to train the network were randomly cropped from the valid region especially in the tooth and the tooth surface to relieve the data

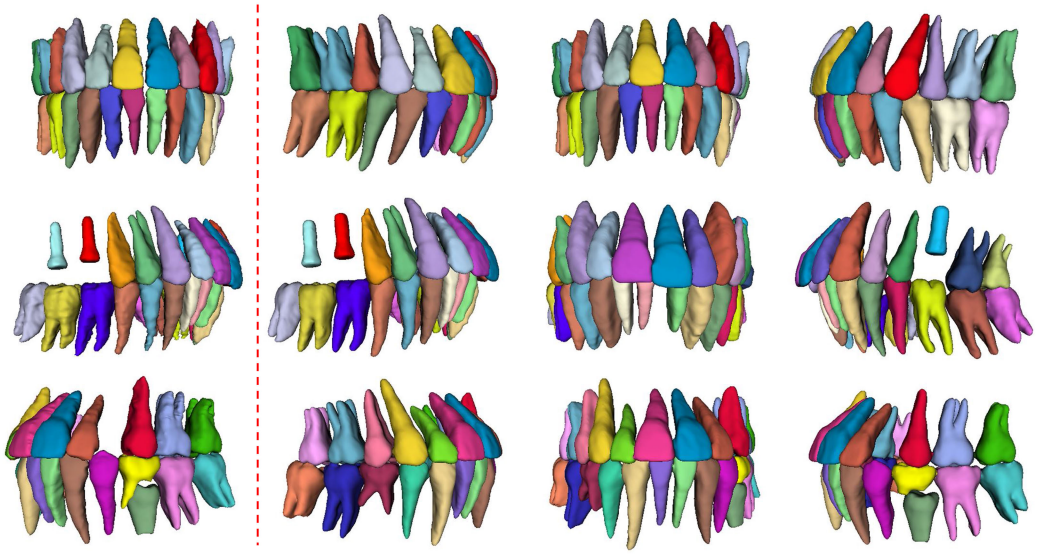


FIGURE 6. Three-dimensional visualization of individual tooth segmentation results that were obtained via the proposed method. The segmentation results from the first row to the third row correspond to three subjects. Their DSC values are 0.941 (±0.020), 0.932 (±0.020), and 0.955 (±0.014), while the ASSD values are 0.266 (±0.089) mm, 0.347 (±0.096) mm, and 0.240 (±0.143) mm, respectively.

imbalance problem. The number of image patches sampled from each image depended on the size of the CBCT data, for each patient has different head size which leads to different size of valid tooth region. All the models were trained on a Titan X GPU with 12 GB RAM. The loss function in Eq. (2) was minimized using the Adam optimizer, with an initial learning rate of 10^{-4} in the first four epochs but decreasing to 10^{-5} since the fifth epoch. The training of our proposed multi-task 3D FCN model took approximately 33 hours.

B. IMPACT OF PATCH SIZE

In this section, we investigated the effect of patch size on the segmentation performance. We conducted the experiments in which we used five input patch sizes, namely, $32 \times 32 \times 32$, $48 \times 48 \times 48$, $64 \times 64 \times 64$, $80 \times 80 \times 80$, and $96 \times 96 \times 96$, to train the multi-task network. Fig. 5 shows how varying the patch size affects the DSC, Jaccard index, RVD, and ASSD. The DSC and the Jaccard index of the model with a patch size of $64 \times 64 \times 64$ were higher than those of other models. In addition, both RVD and ASSD attained their lowest values when the model with a patch size of $64 \times 64 \times 64$ was used. The models trained with a patch size of $64 \times 64 \times 64$ could achieve better performance. This may be because that the patch with size of $64 \times 64 \times 64$ contained enough information about tooth or tooth surface for training. In addition, larger patch size caused a lower patch number to be obtained and trained in our experiment, and the patch number obtained with patch size of $64 \times 64 \times 64$ was more suitable.

With the insight that was gained through the experiment results, we set the patch size to $64 \times 64 \times 64$ throughout the following experiments. Fig. 6 displays examples of individual tooth segmentation results that were obtained via our proposed method. Compared with the ground truth, the segmentation results were visually satisfactory. The proposed method could effectively segment individual tooth, even those with unusual shapes.

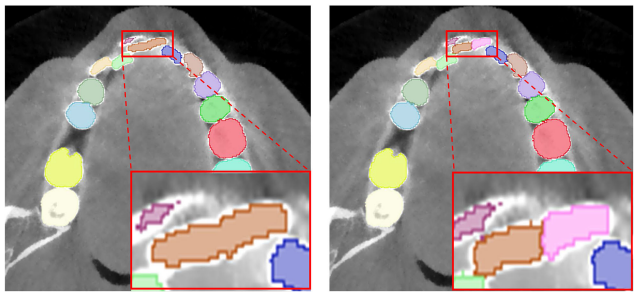


FIGURE 7. Comparison of the multi-task learning strategy and the single-task learning strategy. The bounding box in red in (a) indicates a mistake.

C. MULTI-TASK LEARNING VERSUS SINGLE-TASK LEARNING

To evaluate the necessity of using both the tooth probability gradients and the tooth surface probability to segment individual tooth, the multi-task learning strategy was compared with a single-task learning strategy. We trained a single-task network of the V-net architecture to predict only the tooth probability map. Then, MWT was conducted on the tooth probability gradient map to obtain an individual tooth segmentation. In the single-task learning strategy, voxels that belonged to a tooth might be mislabeled as neighboring tooth voxels (as shown in Fig. 7), since

TABLE 2. Performances of individual tooth segmentation via different network structures.

Network	DSC	Jaccard index	RVD	ASSD (mm)
Multi-task V-net	0.936±0.012	0.881±0.019	0.072±0.027	0.363±0.145
Single-task V-net	0.926±0.020	0.867±0.029	0.099±0.041	0.529±0.334
Modified 3D U-net	0.931±0.017	0.873±0.030	0.088±0.060	0.347±0.100
Modified 3D FC-Densenet	0.930±0.018	0.872±0.028	0.088±0.031	0.450±0.229

TABLE 3. Comparisons in the ablation study.

Method	DSC	Jaccard index	RVD	ASSD (mm)
CRE	0.727±0.016	0.604±0.021	0.714±0.122	3.603±0.429
MWT on CBCT image gradient	0.873±0.045	0.791±0.061	0.180±0.109	1.395±0.862
Simple MWT marker	0.897±0.016	0.818±0.025	0.140±0.042	0.760±0.214
Proposed	0.936±0.012	0.881±0.019	0.072±0.027	0.363±0.145

only the tooth probability gradients were used to segment the individual tooth. The mislabeled tooth voxels caused the ASSD to be larger and the Jaccard index to be lower, as seen in Table 2. Conversely, the segmentation was more accurate when using the multi-task V-net for predictions and exploiting both the tooth probability gradients and tooth surface probability for segmentation. The multi-task learning strategy achieved higher performance with DSC of 0.936 (±0.012) and ASSD of 0.363 (±0.145) mm. In addition, the Jaccard index and RVD were 0.881 (±0.019) and 0.072 (±0.027), respectively. Through combining the tooth probability gradient map and the tooth surface probability map, more information about tooth was gathered, which contributed to better performance for the multi-task learning strategy.

D. COMPARISONS WITH OTHER MODIFIED FCN

The proposed multi-task FCN was compared with a modified U-net and a modified FC-Densenet. For fair comparison, all parameters, such as the loss function, the optimizer, and the learning rate were set the same as those of our proposed method. All the modified FCNs were trained using sliding windows with input size of 64 × 64×64. After four downsamplings, their sizes were decreased to 4 × 4×4. The details of the modified 3D U-net and the modified 3D FC-Densenet are illustrated in Appendix A and Appendix B, respectively.

Table 2 shows comparisons of the quantitative evaluation results. Except for the ASSD, the modified FC-Densenet performed similar to the modified U-net. The proposed multi-task V-net achieved a higher Dice score and Jaccard index than the two compared networks. In addition, the RVD obtained by the multi-task V-net was lower than that of the modified U-net and the modified FC-Densenet. The modified U-net lacked the residual blocks for feature reuse; the modified FC-Densenet with dense blocks caused too many redundant features to be learned in our experiment. As for the modified V-net, it contained a suitable number of residual blocks, which led to more reasonable feature reuse and more accurate segmentation results.

E. ABLATION STUDY

1) MWT VERSUS CONNECTED REGION EXTRACTION

Many postprocessing methods, such as graph cut and conditional random field methods [29], [30], have strong sensitivity to the initialization or coefficient settings, and can even be of time-consuming. The segmentation method of connected region extraction (CRE) operates simply and effectively. Thus, we compared the performance of the proposed method (MWT) with that of CRE. To implement CRE, we firstly binarized the tooth probability map and the tooth surface probability map. Then we removed the surface from the tooth by subtracting the surface mask from the tooth mask. Finally, CRE was conducted, and each connected region in the binary tooth image from which the surface voxels had been removed was assigned a unique integer value that corresponded to a tooth instance.

As listed in Table 3, the DSC and ASSD of CRE were 0.727 (±0.016) and 3.603 (±0.429) mm, respectively, while our method achieved superior results in terms of the two evaluation metrics (DSC: 0.936 ± 0.012 and ASSD: 0.363 ± 0.145 mm). Fig. 8 shows an example of individual tooth segmentation results that were obtained via MWT or CRE. MWT could separate individual tooth more successfully, which was due to its effectively using of information of both the surface probability map and the tooth probability map. By contrast, the method of CRE might assign the same label to neighboring teeth erroneously since the thresholding operation performed on the surface probability map caused the surface mask to be discontinuous. The de-surface operation failed to completely remove the surface voxels from the tooth. Neighboring teeth might still be touching each other and forming a connected region. CRE assigned a single label to this connected region, thereby resulting in the failure to segment individual tooth, as seen in Fig. 8(b).

2) DENTAL CBCT IMAGE GRADIENT MAP VERSUS TOOTH PROBABILITY GRADIENT MAP

The effects of different gradient maps as the input images of MWT on the performance of individual tooth segmentation were evaluated. To maintain the normative property

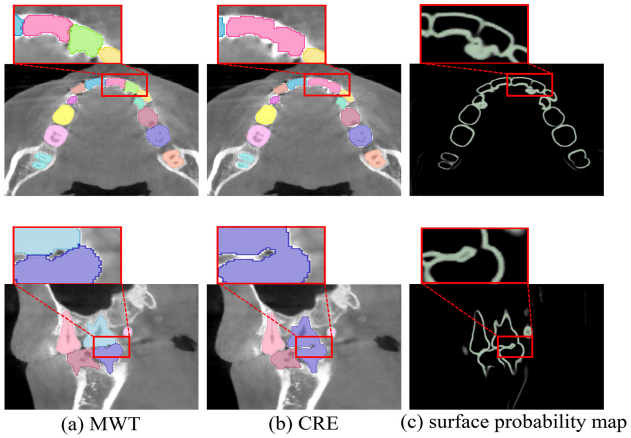


FIGURE 8. Comparison for individual tooth segmentation via different methods. (a) and (b) are the segmentation results that are obtained via MWT and CRE, respectively. (c) displays the tooth surface probability map.

of the entire procedure, only the combination way of the gradient maps was changed for comparison. The tooth surface probability map was combined with the dental CBCT image gradient map or the tooth probability gradient map to as the input image of MWT. As listed in Table 3, the DSC and the Jaccard index achieved by MWT on CBCT image gradient were 6.3% and 9.0% lower than those of our proposed method that used tooth probability gradient map as the input image, respectively; in addition, the ASSD achieved by using dental CBCT image gradient map was much higher. When the dental CBCT image gradient map was used as the input image of MWT, voxels belonging to a tooth might be mislabeled as other tooth voxels; non-tooth voxels could be incorrectly labeled as tooth voxels. In addition, the mislabeled tooth voxels only existed on the tooth surface, as shown in Fig. 9. This phenomenon may be due to the fact that, CBCT image gradients contained not only tooth gradients but also gradients of non-tooth structures; both gradients were strong, thereby confusing MWT in correctly assigning labels to tooth voxels. Instead, the gradients of non-tooth structures in the tooth probability gradient map were weaker, leading to the predominance of tooth gradients when performing individual tooth segmentation. Thus, the method using the tooth probability gradient map performed better.

3) COMPARISON OF THE EFFECTS OF DIFFERENT BACKGROUND MARKERS

We further evaluated the effectiveness of using watershed ridge lines as background marker in MWT. We performed the thresholding operation on the tooth probability map to obtain a tooth mask $\boldsymbol{B}$. Then, a spherical structural element with a radius of 3 was used to perform a dilation operation on $\boldsymbol{B}$. Finally, logical 'not' operation was conducted on the dilation result to achieve a background marker (we called it simple MWT marker).

When simple MWT marker was used as background marker for segmentation, many tooth voxels were mislabeled as background voxels, as shown in Fig. 10. The simple MWT marker was in close proximity to the tooth surface, which affected the segmentation accuracy. Accordingly, it achieved lower values in DSC and Jaccard index, and higher values in RVD and ASSD than the method that used the watershed ridge lines as the background markers, as listed in Table 3.

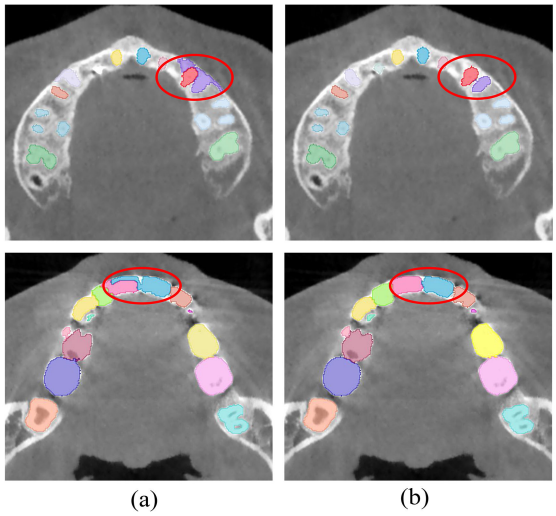


FIGURE 9. Visualization of individual tooth segmentation results that were obtained using various gradient maps as input images of MWT. The first row and the second row correspond to different subjects. (a) is the segmentation results that were produced by MWT on dental CBCT image gradient map. (b) is the segmentation results that were produced by the method using tooth probability gradient map as the input image of MWT.

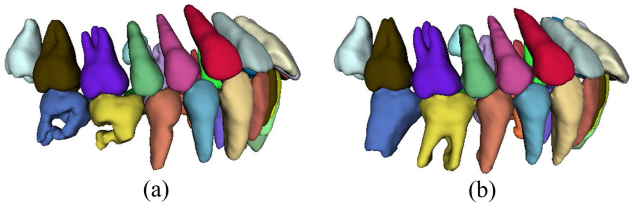


FIGURE 10. Individual tooth segmentation results produced by the methods using different background markers in MWT. (a) is produced by using simple MWT marker. (b) is produced by using the watershed ridge lines.

F. THE PROPOSED METHOD VERSUS THE LEVEL SET METHOD

To evaluate the effectiveness of our proposed method, the level set method was compared on the test set. As the segmentation procedure that mentioned in Ref. [5], we conducted the level set method for individual tooth segmentation in dental CBCT images. The details of the level set method implementation can be seen in Appendix C.

Fig. 11 illustrates the quantitative results by our proposed method and the level set method. Both RVD and ASSD achieved by the level set method were much larger than those of our method. The segmentation results achieved by the level set method were not good, as shown in Fig. 12. Some

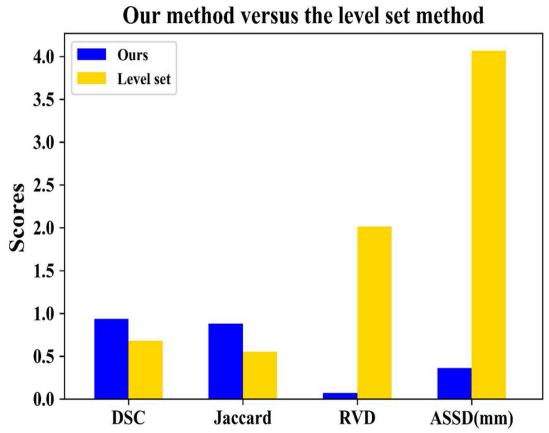


FIGURE 11. Comparisons between our proposed method and the level set method.

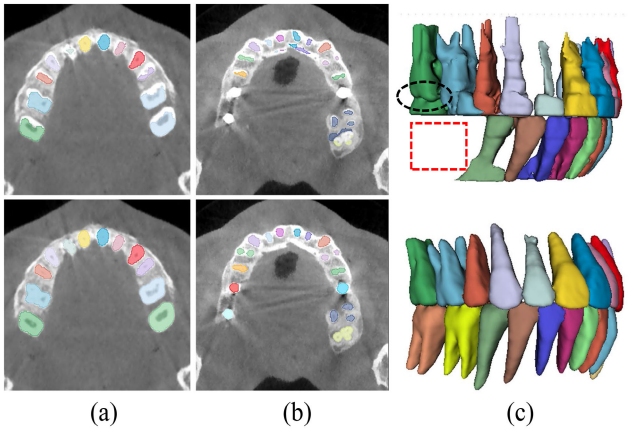


FIGURE 12. Comparisons between the proposed method and the current popular method. The first row is the segmentation results of level set method. The second row is the segmentation results of the proposed method. (a) and (b) are the segmentation performance from two different slices. (c) is the 3D visualization, and the black circle and red bounding box indicate the mistakes.

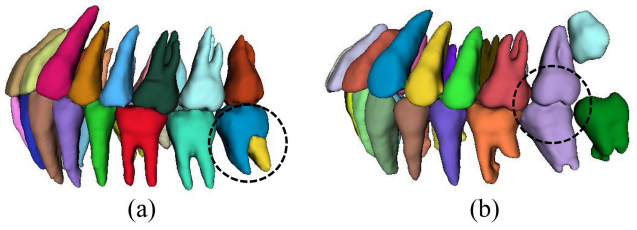


FIGURE 13. Failure cases of individual tooth segmentation of our proposed method. (a) and (b) are two examples and the dashed circles indicate the mistakes in segmentation.

teeth were missed (as the red bounding box in Fig. 12(c) indicated) because of the difficulty in selecting a suitable starting slice. In addition, when the teeth were in a natural bite position, separating the upper teeth from the lower teeth could fail (as the black circle delineated). As for dental CBCT images with metal artifacts (Fig. 12(b)), the level set method also could not effectively perform segmentation; the voxels

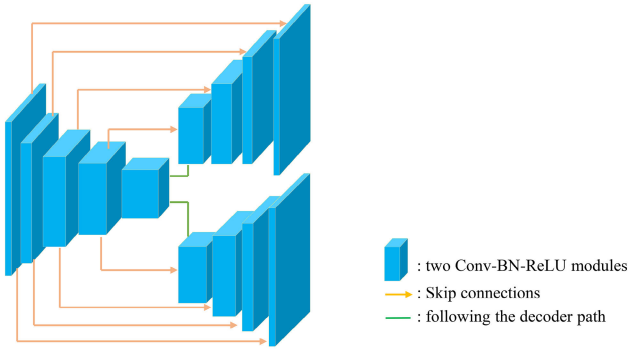


FIGURE 14. The architecture of the modified 3D U-net.

Downsampling path	
Input, m=1	
3×3×3 convolution, m=8	
DB(1 layer)+TD, m=18	
DB(2 layers)+TD, m=38	
DB(3 layers)+TD, m=68	
DB(3 layers)+TD, m=98	
DB(3 layers), m=128	
Upsampling path 1	**Upsampling path 2**
TU+DB((3 layers), m=158	TU+DB((3 layers), m=158
TU+DB((3 layers), m=128	TU+DB((3 layers), m=128
TU+DB((2 layers), m=88	TU+DB((2 layers), m=88
TU+DB((1 layer), m=48	TU+DB((1 layer), m=48
1×1×1 convolution, m=1	1×1×1 convolution, m=1

FIGURE 15. The architecture of the modified 3D FC-Densenet.

of alveolar bones might be mislabeled as tooth voxels. Due to the extensive tooth types contained in the training set and the robustness of the FCN, the proposed method can address these problems well and achieve more satisfactory segmentation results.

IV. DISCUSSION AND CONCLUSION

Our method for individual tooth segmentation employed a multi-task 3D FCN to predict the tooth region and tooth surface. Then, the predicted maps were used for individual tooth segmentation by MWT. The relatively better results achieved by our method were owing to the suitable patch size that we used, the robustness of the FCN architecture that we elaborated, and the reasonable design of the MWT implementation.

There are two main limitations in our studies. The dataset that we used contains only 25 images. Our dataset is limited since delineating teeth is a heavy workload, which renders the generation of the segmentation ground truth difficult. Additionally, we determined the foreground markers by performing morphological operations. Tooth may be broken into several fractions (several foreground markers), thereby causing the watershed transform to assign several labels to an individual tooth. Moreover, some neighboring teeth with blurred boundaries share a common foreground marker,

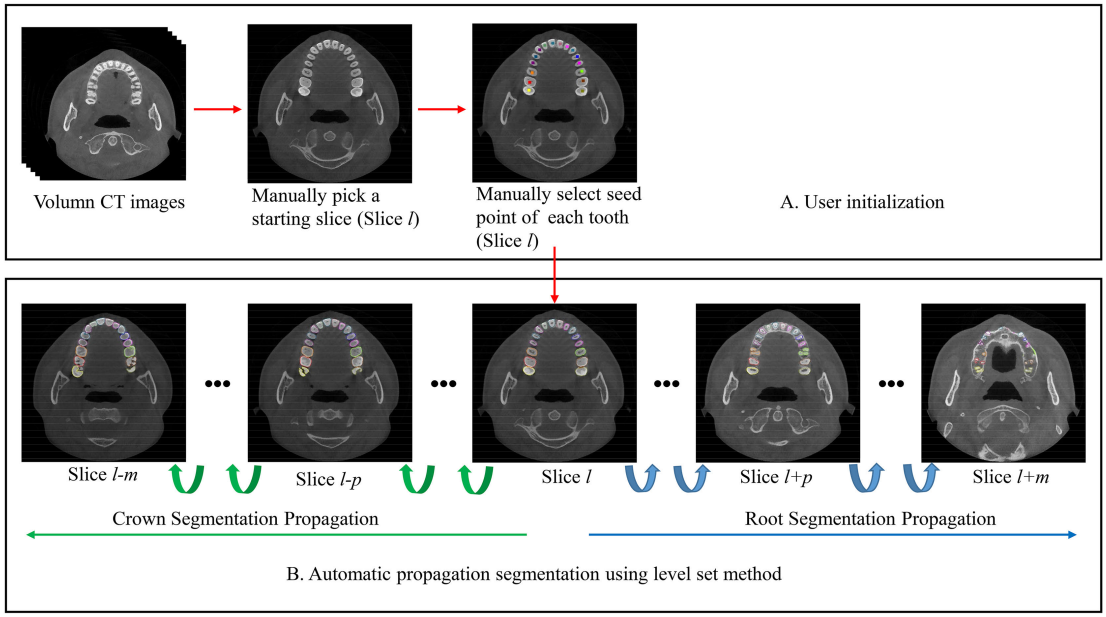


FIGURE 16. Diagram of the slice by slice segmentation procedure.

which leads to failures in the segmentation of individual tooth. Two failure cases in individual tooth segmentation are presented in Fig. 13. In fact, the foreground marker of the touching teeth in Fig. 13(b) passes through a line that separates the teeth of the upper and lower jaw. If we continue to erode this foreground marker until it does not pass through the separation line and is broken into two foreground markers, MWT will correctly assign labels to the two individual teeth. This is the task that we will realize in our following work. Generally, a large number of samples would contribute to the development of a relatively satisfactory training model. In the future, we will exploit the model that we have trained and use manual delineation to enlarge the amount of ground truth data. Then, we will retrain the model and further improve the segmentation performance.

In summary, we present an effective method for individual tooth segmentation. The experimental results demonstrate that our method performs well in individual tooth segmentation and exhibits high precision for the dental CBCT images with various teeth.

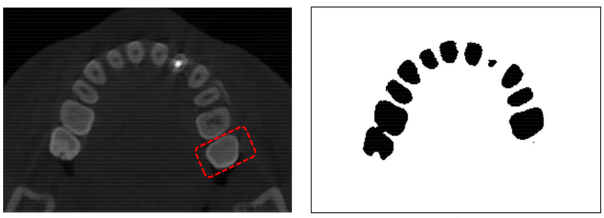


FIGURE 17. Example of a mask for restricting the level set evolution. The dashed bounding box in (a) indicates the target tooth that will be segmented by the level set. The corresponding mask is shown in (b).

V. APPENDIX

A. DETAILS OF THE MODIFIED 3D U-NET

The modified U-net contains an encoder path and two decoder paths, as shown in Fig. 14. There are four encoder layers in the encoder path. Considering the GPU memory limitation, the channel number of the feature maps that output from the first encoder layer is set to 4. Each encoder layer is composed of two Conv-BN-ReLU modules, followed by a max-pooling module for downsampling. The convolutional kernel size that we used is set to $3\times3\times3$. The two decoder paths are designed for predicting the tooth region and the tooth surface. There are four decoder layers in each decoder path. In each decoder layer, the feature maps are upsampled by a transposed convolution with kernel size of $2\times2\times2$, followed by a skip connection, and two Conv-BN-ReLU modules. In the last decoder layer of the decoder path, a convolution with kernel size of $1\times1\times1$ is used, and the prediction is output. It took about 22 hours for training this modified U-net model.

B. DETAILS OF THE MODIFIED 3D FC-DENSENET

The modified FC-Densenet is composed of a downsampling path, two upsampling paths, and some skip connections. The downsampling path consists of 4 dense blocks (DB) and 4 transition down layers (TD) while the upsampling path consists of 4 dense blocks and 4 transition up layers (TU). Dense block is an iterative concatenation of previous feature maps for feature reuse. There are some layers composed in the dense block, where the growth rate of layers is set to 10. Each layer is a BN-ReLU-Conv module, followed by a dropout. The transition down layer contains a

BN-ReLU-Conv module, a dropout of 0.2, and a $2\times2\times2$ max pooling. As for transition up layer, it is a $3\times3\times3$ transposed convolution with stride of 2 for downsampling. Fig. 15 shows the architecture of the modified 3D FC-Densenet. Training a model of the modified 3D FC-Densenet took about 28 hours.

C. DETAILS OF THE LEVEL SET METHOD

The slices of the maxillary and the mandible were segmented independently using the same segmentation procedures, as shown in Fig. 16. First, we selected an initial slice between the crown and dental root. Then the seed points were drawn manually in the initial slice for recognizing each tooth. Let $\Omega \subset \mathrm{R}^2$ denote the image domain; I: $\Omega \to \mathrm{R}$ denote the dental CBCT images; $\phi : \Omega \to \mathrm{R}$ be the level set function; and $|\mathrm{C}|$ be the length of an evolution contour. The energy function of the level set that we used for segmenting the initial slice is defined in Eq. (C.1):

$$\Gamma(\phi, f_1, f_2) = F_{\mathrm{L}}(\phi, f_1, f_2) + \nu|C| + \mu_1 P(\phi) + \mu_2 \int_\Omega g\delta_\varepsilon(\phi)|\nabla\phi|dx \tag{C.1}$$

where the first term is the local intensity energy term; the second term is the curve length penalization term; the third term is the regularization term; and the forth term is the edge detection energy term. μ_1, μ_2, and ν are empirically set to 0.8, 0.1, and 0.0008×255^2, respectively. g is the edge indicator, and δ_ε is the normalized Dirac delta function. $\mathrm{F_L}$, $|\mathrm{C}|$, and $P(\phi)$ are defined below:

$$\begin{aligned}\mathrm{F_L}(\phi, f_1, f_2) = {} & \lambda_1 \int_\Omega \left(\int_\Omega K_\sigma(x-y)|I(y)-f_1(x)|^2 \times H_\varepsilon(\phi(y))dy\right)dx \\ & + \lambda_2 \int_\Omega \left(\int_\Omega K_\sigma(x-y)|I(y)-f_2(x)|^2 \times (1-H_\varepsilon(\phi(y)))dy\right)dx \end{aligned} \tag{C.2}$$

$$|\mathrm{C}| = \int |\nabla H_\varepsilon(\phi(x))|dx \tag{C.3}$$

$$P(\phi) = \int \frac{1}{2}\left(|\nabla\phi(x)| - 1\right)^2 dx \tag{C.4}$$

Both λ_1 and λ_2 are set to 0.8; $K_\sigma(x-y)$ is the Gaussian function with a scale parameter ($\sigma = 3$). H_ε is the normalized Heaviside function. $f_1(x)$ and $f_2(x)$ are the mean intensities inside and outside the zero level set.

After segmenting the initial slice, contour propagation strategy was used to segment other slices. Based on tooth shape prior, the propagation strategy was proceeded toward the crown and the root direction, respectively. Tooth shape prior was represented by the average tooth shape of the last three previous segmented slices. In the case that there were less than three segmented slices, the shape was computed from existing segmented slices. The energy function of the level set that we used to segment the nonstarting slices is as follows:

$$\psi(\phi, f_1, f_2) = \Gamma(\phi, f_1, f_2) + \omega F_{\mathrm{G}}(\phi) + \beta \mathrm{F_{shape}}(\phi) \tag{C.5}$$

Here, the ω and β are set to 0.2 and 5, respectively. $\Gamma(\phi, f_1, f_2)$ is defined in the Eq. (C.1). $\mathrm{F_G}(\phi)$ is the global intensity prior term, while $\mathrm{F_{shape}}(\phi)$ is the shape prior constraint term. The intensity distribution of the current slice was estimated from the previous segmented slice to define a global intensity energy. Let $M = \{M_j | j = b, f\}$ be the statistical model parameter of either the foreground or the background, and ϕ_0 be the signed distance function of the prior tooth shape. $\mathrm{F_G}(\phi)$ and $\mathrm{F_{shape}}(\phi)$ are defined as follows:

$$\mathrm{F_G}(\phi) = \int \log\left(\frac{P(M_b|I(x))}{P(M_f|I(x))}\right) H_\varepsilon(\phi(x))\, dx \tag{C.6}$$

$$\mathrm{F_{shape}}(\phi) = (1 - H_\varepsilon(\phi_0(x))) \int (H_\varepsilon(\phi(x)) - H_\varepsilon(\phi_0(x)))^2 dx \tag{C.7}$$

To separate the touching teeth during the segmentation of the nonstarting slices, the level set evolution of each tooth was proceeded. A mask was used to restrict each level set evolution and to prevent from invading in neighboring teeth. The mask was assigned as 0 for points inside the contours of nontarget teeth or 1 for points outside the contours of nontarget teeth (as shown in Fig. 17). During the segmentation, all level sets were evolved simultaneously, and the mask of each level set was updated before every iteration.

REFERENCES

[1] A. G. Farman and W. C. Scarfe, "Historical perspectives on CBCT," in *Maxillofacial Cone Beam Computed Tomography*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[2] T. Evain, X. Ripoche, J. Atif, and I. Bloch, "Semi-automatic teeth segmentation in cone-beam computed tomography by graph-cut with statistical shape priors," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Melbourne, VIC, Australia, Apr. 2017, pp. 1197–1200.

[3] S. Barone, A. Paoli, and A. V. Razionale, "CT segmentation of dental shapes by anatomy-driven reformation imaging and B-spline modelling," *Int. J. Numer. Methods Biomed. Eng.*, vol. 32, no. 6, Jun. 2016, Art. no. e02747.

[4] Y. Gan, Z. Xia, J. Xiong, G. Li, and Q. Zhao, "Tooth and alveolar bone segmentation from dental computed tomography images," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 196–204, Jan. 2018.

[5] Y. Gan, Z. Xia, J. Xiong, Q. Zhao, Y. Hu, and J. Zhang, "Toward accurate tooth segmentation from computed tomography images using a hybrid level set model," *Med. Phys.*, vol. 42, no. 1, pp. 14–27, Dec. 2014.

[6] H. Gao and O. Chae, "Individual tooth segmentation from CT images using level set method with shape and intensity prior," *Pattern Recognit.*, vol. 43, no. 7, pp. 2406–2417, Jul. 2010.

[7] M. Hosntalab, R. A. Zoroofi, A. A. Tehrani-Fard, and G. Shirani, "Segmentation of teeth in CT volumetric dataset by panoramic projection and variational level set," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 3, nos. 3–4, pp. 257–265, Sep. 2008.

[8] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[9] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

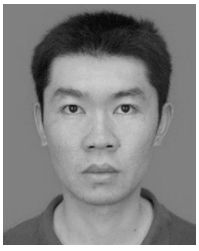
[10] Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze, "Knowledge-aided convolutional neural network for small organ segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1363–1373, Jul. 2019.

[11] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Med. Image Anal.*, vol. 53, pp. 142–155, Apr. 2019.

[12] A. Suzani, A. Rasoulian, A. Seitel, S. Fels, R. N. Rohling, and P. Abolmaesumi, "Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images," *Proc. SPIE*, vol. 9415, Mar. 2015, Art. no. 941514.
[13] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, and H. Fujita, "Classification of teeth in cone-beam CT using deep convolutional neural network," *Comput. Biol. Med.*, vol. 80, pp. 24–29, Jan. 2017.
[14] Z. Cui, C. Li, and W. Wang, "ToothNet: Automatic tooth instance segmentation and identification from cone beam CT images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, Art. no. 63686377.
[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[16] H. Huang, X. Li, and C. Chen, "Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2253–2262, Jul. 2018.
[17] L. Wang, P. Gong, and G. S. Biging, "Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 3, pp. 351–357, Mar. 2004.
[18] L. Jing, B. Hu, T. Noland, and J. Li, "An individual tree crown delineation method based on multi-scale segmentation of imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 88–98, Jun. 2012.
[19] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA Oct. 2016, pp. 565–571.
[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Germany, 2015, pp. 234–241.
[21] W. Yang, Y. Liu, L. Lin, Z. Yun, Z. Lu, Q. Feng, and W. Chen, "Lung field segmentation in chest radiographs from boundary maps by a structured edge detector," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 842–851, May 2018.
[22] Z. Deng, H. Fan, F. Xie, Y. Cui, and J. Liu, "Segmentation of dermoscopy images based on fully convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, Art. no. 17321736.
[23] Z. Yun, S. Yang, E. Huang, L. Zhao, W. Yang, and Q. Feng, "Automatic reconstruction method for high-contrast panoramic image from dental cone-beam CT data," *Comput. Methods Programs Biomed.*, vol. 175, pp. 205–214, Jul. 2019.
[24] N. Dong, L. Wang, R. Trullo, J. Li, P. Yuan, J. Xia, and D. Shen, "Segmentation of craniomaxillofacial bony structures from MRI with a 3D deep-learning based cascade framework," in *Machine Learning in Medical Imaging* (Lecture Notes in Computer Science), vol. 10541. Quebec City, QC, Canada: Springer, 2017, pp. 266–273.
[25] J. Zhang, M. Liu, L. Wang, S. Chen, P. Yuan, J. Li, S. G.-F. Shen, Z. Tang, K.-C. Chen, J. J. Xia, and D. Shen, "Joint craniomaxillofacial bone segmentation and landmark digitization by context-guided fully convolutional networks," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2017, pp. 720–728.
[26] D. Nie, L. Wang, E. Adeli, C. Lao, W. Lin, and D. Shen, "3-D fully convolutional networks for multimodal isointense infant brain image segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1123–1136, Mar. 2019.
[27] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), Granada, Spain, vol. 11384, 2018, p. 61 72.
[28] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, Apr. 2019.
[29] Y. Wang, K.-F. Loe, and J.-K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2006.
[30] F. I. Alam, J. Zhou, A. W.-C. Liew, X. Jia, J. Chanussot, and Y. Gao, "Conditional random field and deep feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1612–1628, Mar. 2019.

YANLIN CHEN received the B.S. degree in biomedical engineering from Guangdong Medical University, Dongguan, China, in 2017. She is currently pursuing the M.E. degree with the Department of Biomedical Engineering, Southern Medical University. Her researches focus on the medical image processing and radiomics.

HAIYAN DU received the B.S. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2017, where she is currently pursuing the M.S. degree in engineering with the Department of Biomedical Engineering. Her researches focus on medical image analysis and computer vision.



ZHAOQIANG YUN received the B.S. degree from the China University of Mining and Technology, Xuzhou, China, in 2007, the M.S. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2010, and the Ph.D. degree from the Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, in 2019. Since 2010, he has been a Faculty Member of the School of Biomedical Engineering, Southern Medical University. His research interests contain medical image processing, machine learning, and computer vision.

SHUO YANG received the master's degree from Southern Medical University, in 2013. He currently works with the Dental Implant Center, Stomatological Hospital, Southern Medical University. His research interest is oral implant.

ZHENHUI DAI received the B.S. degree from the Taishan Medical College, Tai'an, China, in 2007, and the M.S. degree from Southern Medical University, Guangzhou, China, in 2010. Since 2010, he has been a Radiotherapy Physicist with the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou. His research interests include medical image analysis, medical physics, and tumor tracking in radiotherapy.

LIMING ZHONG received the B.S. degree from the Department of Biomedical Engineering, South Medical University, Guangzhou, China, in 2013, and the Ph.D. degree from the Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, in 2019. She currently holds a postdoctoral position at Southern Medical University, Guangzhou. Her research interests include medical image analysis, machine learning, deep learning, computerized-aid diagnosis, and medical image reconstruction.

QIANJIN FENG received the M.S. and Ph.D. degrees in biomedical engineering from First Military Medical University, China, in 2000 and 2003, respectively. From 2003 to 2004, he was a Faculty Member of the School of Biomedical Engineering, First Military Medical University, China. Since 2004, he has been with Southern Medical University, China, where he is currently a Professor and the Dean of the School of Biomedical Engineering. His research interests are in medical image analysis, pattern recognition, and computerized-aided diagnosis.

WEI YANG received the B.Sc. degree in automation from the Wuhan University of Science and Technology, Wuhan, China, in 2001, the M.Sc. degree in control theory and control engineering from Xiamen University, Xiamen, China, in 2005, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. He is currently a Professor with the School of Biomedical Engineering, Southern Medical University, Guangzhou, China. His main research areas include medical image analysis, machine learning, and computerized-aid diagnosis.