A fully automated method for 3D individual tooth identification and segmentation in dental CBCT

Tae Jun Jang, Kang Cheol Kim, Hyun Cheol Cho, and Jin Keun Seo

Abstract—Accurate and automatic segmentation of three-dimensional (3D) individual teeth from cone-beam computerized tomography (CBCT) images is a challenging problem because of the difficulty in separating an individual tooth from adjacent teeth and its surrounding alveolar bone. Thus, this paper proposes a fully automated method of identifying and segmenting 3D individual teeth from dental CBCT images. The proposed method addresses the aforementioned difficulty by developing a deep learning-based hierarchical multi-step model. First, it automatically generates upper and lower jaws panoramic images to overcome the computational complexity caused by high-dimensional data and the curse of dimensionality associated with limited training dataset. The obtained 2D panoramic images are then used to identify 2D individual teeth and capture loose- and tight- regions of interest (ROIs) of 3D individual teeth. Finally, accurate 3D individual tooth segmentation is achieved using both loose and tight ROIs. Experimental results showed that the proposed method achieved an F1-score of 93.35% for tooth identification and a Dice similarity coefficient of 94.79% for individual 3D tooth segmentation. The results demonstrate that the proposed method provides an effective clinical and practical framework for digital dentistry.

Index Terms—Cone-beam computerized tomography, digital dentistry, tooth segmentation, tooth identification, deep learning

1 INTRODUCTION

D igital dentistry is evolving rapidly along with the rapid innovation of artificial intelligence and the development of cone-beam computerized tomography (CBCT), intra-oral and facial scanners, and dental three-dimensional (3D) printing. Digital dentistry enhances a dentist’s efficiency and improves the accuracies of orthodontic diagnoses, treatment planning, and surgical guides. A fundamental component of digital dentistry is the 3D segmentation of teeth, jaws, and skulls from CBCT images. Moreover, accurate digital models of individual tooth geometry and jaws facilitate the simulation of prosthetic evaluation, cephalometric analysis, computer-aided dental implant planning, and bite irregularity prediction.

Automatic and accurate 3D individual tooth segmentation from CBCT images is a difficult task for the following reasons: (i) similar intensities between teeth roots and their surrounding alveolar bone; (ii) attached boundary between adjacent teeth in the crown parts.

Over the last decade, there have been several attempts to develop 3D tooth segmentation methods, most of which are based on level set methods [1], [2], [3]. Unfortunately, level set-based methods have fundamental limitations in achieving fully automated segmentation. This difficulty arises from the dependence of such methods on the initialization of level set, and the automatic initialization is hindered by the complex image structure associated with adjacent teeth, the jaw, the alveolar bone, etc. Hence, user intervention through manual initialization is inevitable in this approach.

Recently, deep learning methods have been applied in 3D tooth segmentation. Lee et al. [4] and Rao et al. [5] used a fully convolutional network (FCN) [6] for whole tooth segmentation instead of individual tooth segmentation. Cui et al. [7] proposed a deep learning framework for individual tooth segmentation and identification using Mask R-CNN [8]. The limitation of these deep learning methods is a patch-based approach to handle high-dimension inputs (e.g., 800 × 800 × 400 voxels in 3D CBCT image) and a limited amount of labeled samples. It is necessary to use both local and global information to achieve accurate segmentation with individual tooth identification. Thus, the drawback of this patch-based approach is its inability to reflect contextual (global) information, since each output of a convolutional network only depends on the corresponding patch.

Automatic individual tooth identification is also a difficult task. Recently, several individual tooth identification attempts [9], [10] have been made using convolutional networks. However, these approaches suffer from misclassification errors caused by adjacent teeth similarities.

Existing 3D tooth segmentation methods may not be effective for CBCT images that are severely corrupted by metal artifacts. In a clinical dental CBCT environment (e.g., low dose radiation exposure), metal artifacts become common as the number of aged patients with metallic prosthesis increases. Hence, to be practical, it would be desirable to develop a method that works well even in images degraded by metal artifacts.

This paper aimed to address these limitations by developing a hierarchical multi-step deep learning model. The proposed method is summarized as follows. The first step is to circumvent the high-dimensionality problem associated with CT images. This step automatically generates panoramic images of the upper and lower jaws from CT images where its size is smaller than the original CT image. The panoramic images of the upper and lower jaws are separated to reduce overlaps between adjacent teeth. Notably, panoramic images generated from CBCT images are not
significantly affected by metal-related artifacts. We utilize these panoramic images to accurately perform 2D tooth detection, identification, and segmentation. The second step is to identify individual teeth by numbers according to Fédération Dentaire Internationale (FDI) dental notation. We develop a tooth detection method to localize bounding boxes that enclose each tooth and classifies them into four types (incisors, canines, premolars, and molars) according to tooth morphology. This method solve misclassification problems caused by similar adjacent teeth. The individual teeth are then identified using the results of tooth detection. Additionally, we perform 2D segmentation for individual teeth. The third step extracts loose and tight 3D tooth regions of interest (ROIs) from the detected boxes and segmented tooth regions for accurate 3D individual tooth segmentation in the final step. Tight ROIs improve the segmentation accuracy. A schematic diagram of our method is described in Fig. 1.

2 METHOD

Let $\mathbf{X}$ denote a 3D CT image with voxel grid $\Omega := \{(x, y, z) \in \mathbb{N}^3 : 1 \leq x \leq N_x, 1 \leq y \leq N_y, 1 \leq z \leq N_z\}$, where $N_x$, $N_y$, and $N_z$ are the voxel sizes in directions $x$ (sagittal axis), $y$ (frontal axis) and $z$ (longitudinal axis), respectively. The value $\mathbf{X}(x, y, z)$ at the voxel position $(x, y, z)$ is represented as the attenuation coefficient.

2.1 Step 1: Panoramic image reconstruction of the upper and lower jaws from a 3D CBCT image

This step describes the automatic reconstruction of the upper and lower jaw panoramic images from a 3D CBCT image $\mathbf{X}$. Fig. 2 illustrates the workflow.

Fig. 1: Schematic diagram of the proposed method, which consists of four steps: 1) Panoramic image reconstruction of the upper and lower jaws from a 3D CBCT image; 2) tooth identification and 2D segmentation of individual teeth in the panoramic images; 3) extraction of loose and tight 3D tooth ROIs using the detected bounding boxes and segmented tooth regions; and 4) 3D segmentation for individual teeth from the 3D tooth ROIs.

[Step 1-1] To obtain a binary bone image $\hat{\mathbf{X}}$, a 3D CT image $\mathbf{X}$ is segmented into three classes (air, soft tissues, and bones) using multi-level version of Otsu’s method [11]. The threshold values $T_0$ and $T_1$ for the histogram $h(t)$ corresponding to $\mathbf{X}$ are determined by

\[ \{T_0, T_1\} = \arg\max_{t_0, t_1} \left( \frac{\sum_{t_0 \leq t \leq t_1} i p(i)}{\sum_{0 \leq t \leq t_0} i p(i)} - \frac{\sum_{0 \leq t \leq t_0} i p(i)}{\sum_{0 \leq t \leq t_0} i p(i)} \right)^2 \]

where $p(t) = h(t)/\sum h(i)$. The binary image $\hat{\mathbf{X}}(x, y, z)$ is 1 if $\mathbf{X}(x, y, z) \geq T_1$, and 0 otherwise. As this value $T_1$ corresponds to an interface between soft tissues and bones, $\mathbf{X}$ can be viewed as a rough segmentation of upper and lower jaws.

[Step 1-2] Given the binary image $\hat{\mathbf{X}}$, the connected-component labeling (CCL) [12] is used to extract the upper jaw part ($\hat{\mathbf{X}}_{\text{upper}}$) and the lower jaw part ($\hat{\mathbf{X}}_{\text{lower}}$). The CCL method generates all the connected components in a binary image. The lower jaw part is the largest connected component and the upper jaw part is the second largest connected component.

[Step 1-3] To create a 2D image $\mathcal{M}_{\mathbf{X}_{\text{upper}}}$ displaying the upper dental arch, we apply maximum intensity projection (MIP) in the $z$ direction to the grayscale upper jaw image ($\mathbf{X}_{\text{upper}} = \mathbf{X} \odot \hat{\mathbf{X}}_{\text{upper}}$, where $\odot$ is an elementwise product), as follows:

\[ \mathcal{M}_{\mathbf{X}_{\text{upper}}}(x, y) = \max_z \mathbf{X}_{\text{upper}}(x, y, z). \]
Similarly, we obtain $X_{\text{lower}}$ and $M_{X_{\text{lower}}}$ for the lower jaw.

[Step 1-4] Next, binary dental arch regions $A_{X_{\text{upper}}}$ and $A_{X_{\text{lower}}}$ is obtained by applying the Otsu’s method [11] and the morphological closing [13] to the MIP images $M_{X_{\text{upper}}}$ and $M_{X_{\text{lower}}}$, respectively. Here, Otsu thresholding is adopted to get rough dental arch regions and the morphological closing is used to smoothen out the rough regions.

[Step 1-5] Given the upper dental arch region $A_{X_{\text{upper}}}$, we employ the morphological skeletonization [14] to extract a medial axis of the dental arch region. Cubic spline curve fitting, interpolation and extrapolation techniques are then applied to the medial axis, to obtain a smooth reference curve $C_{X_{\text{upper}}}$ passing through the dental arch region completely. The reference curve can be expressed as

$$C_{X_{\text{upper}}} = \{ r(s) = (x(s), y(s)) : s \in 1, 2, \ldots, N_s \}, \quad (3)$$

where $N_s$ is the number of curve points. Similarly, we can obtain $C_{X_{\text{lower}}}$ from the lower dental arch region $A_{X_{\text{lower}}}$.

[Step 1-6] An upper jaw panoramic image is given by

$$P_{X_{\text{upper}}}(s, z) = \int_{-\alpha}^{\alpha} X_{\text{upper}}(r(s) + n(s), z) \, dt, \quad (4)$$

where $s$ is the parameter in $[3]$, $r(s) \in C_{X_{\text{upper}}}$, and $n(s)$ is the unit normal vector at $r(s)$. Similarly, we obtain the lower panoramic image $P_{X_{\text{lower}}}$. For notational simplicity, we refer to both $P_{X_{\text{upper}}}$ and $P_{X_{\text{lower}}}$ as $P$.

### 2.2 Step 2: Bounding box detection, identification, and 2D segmentation of individual teeth in the reconstructed panoramic images

This step aims to identify and segment individual teeth in the reconstructed panoramic images. To achieve the goal, we first perform individual tooth detection. Here, the teeth are classified as incisor (class 1), canine (class 2), premolar (class 3), and molar (class 4).

[Step 2-1] To detect individual teeth in a panoramic image, we develop a deep learning method inspired by one-stage object detection [15], [16]. Given a panoramic image $P$, we obtain a detection map $f_{\text{det}}$ predicting $Y_{ij}$ for each $G_{ij}$.

with size $N_x \times N_z$ (e.g., $N_x \times N_z = 640 \times 320$), a uniform grid is created. Each grid cell $G_{ij}$ has a size of $g \times g$ (e.g., $g = 16$). Then we learn a tooth detection map $f_{\text{det}} : P \rightarrow Y$ that is given by

$$f_{\text{det}}(P) = \begin{pmatrix}
Y_{1,1} & Y_{2,1} & \cdots & Y_{N_x,1} \\
Y_{1,2} & Y_{2,2} & \cdots & Y_{N_x,2} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{1,N_z} & Y_{2,N_z} & \cdots & Y_{N_x,N_z}
\end{pmatrix}, \quad (5)$$

where $Y_{ij} = (c_{ij}, b_{ij}, p_{ij})$ predicting a confidence score $c_{ij}$, a bounding box component $b_{ij}$, and a class probability $p_{ij}$ in $G_{ij}$, as illustrated in Fig.3. Here, a confidence score $c_{ij} \in [0, 1]$ represents the existence of the tooth center in $G_{ij}$. A bounding box component [17] is denoted by

$$b_{ij} = \left( \frac{s_{ij}}{g} - (i - 1), \frac{z_{ij}}{g} - (j - 1), \log \frac{w_{ij}}{a_w}, \log \frac{h_{ij}}{a_h} \right), \quad (6)$$

where $(s_{ij}, z_{ij})$ is the center of the bounding box in $G_{ij}$, $(w_{ij}, h_{ij})$ indicates its width and height, and $(a_w, a_h)$ are the width and height of an anchor box. For a tooth in the bounding box corresponding to $b_{ij}$, we estimate a class probability

$$p_{ij} = (p_{ij,1}, p_{ij,2}, p_{ij,3}, p_{ij,4}), \quad (7)$$
where \( p_{ij,k} \) represents the probability of being tooth class \( k \).

To find exact bounding boxes among the predicted boxes for all \( G_{ij} \), we remove the boxes with scores \( c_{ij} = c_{ij} \ast (\max_k p_{ij,k}) \) less than 0.5. Several bounding boxes with high scores may appear near the center of a tooth. We adopt the non-maximum suppression (NMS) technique to filter out bounding boxes that highly overlap high-scoring boxes.

Using a labeled training dataset \( \{(p^{(n)}, Y^{(n)})\}_{n=1}^{N} \), where \( Y^{*} \) is ground-truth labeling, \( f_{det} \) is learned by minimizing the loss between the output \( Y = f_{det}(P) \) and the ground-truth \( Y^{*} \) as follows:

\[
\mathcal{L}_{det} = \sum_{n=1}^{N} \left[ \sum_{(i,j) \in \Omega_1^{(n)}} (1 - c_{ij}^{(n)})^2 + \lambda_1 \sum_{(i,j) \in \Omega_0^{(n)}} (0 - c_{ij}^{(n)})^2 \right] + \lambda_2 \left[ \sum_{(i,j) \in \Omega_1^{(n)}} |b_{ij}^{(n)} - b_{ij}^{(n)}|^2 + \sum_{(i,j) \in \Omega_0^{(n)}} CE(p_{ij}^{*}, p_{ij}^{(n)}) \right],
\]

(8)

where \( \Omega_1^{(n)} = \{(i,j)|c_{ij}^{*} = 1\} \), \( \Omega_0^{(n)} = \{(i,j)|c_{ij}^{*} = 0\} \), and CE is the cross-entropy. This multi-task loss \( \mathcal{L}_{det} \) represents the prediction errors where objects exist \( (c_{ij}^{*} = 1) \) and no objects exist \( (c_{ij}^{*} = 0) \). Since there is no object in most grid cells, the confidence score tends to be predicted as zero [15]. The constants \( \lambda_1 = 0.1 \) and \( \lambda_2 = 5 \) are used to eliminate the imbalance.

[Step 2-2] For each tooth in the detected bounding box, a number is assigned to identify the unique tooth according to the FDI system. For convenience, we first explain how the numbers are assigned to teeth in the upper jaw. As illustrated in Fig. 4, the detected bounding boxes are listed in ascending order of \( s \) coordinates of the box center. The upper right and left quadrants are divided from the middle of four sequential incisor boxes. For the two right incisors and the two left incisors, number 1 and 2 are assigned from the inside to the outside, respectively. Number 3 is assigned to the canines since there is only one in each quadrant. On each side, premolars are assigned numbers 4 and 5 from the inside to the outside. Likewise, molars are assigned numbers 6, 7 and 8 (if a wisdom tooth exists).

Fig. 4: Tooth identification process using the classification results in Step 2-1. The capital letters represent the first letters of the tooth type and the numbers are tooth codes.

[Step 2-3] The proposed 2D tooth segmentation method uses a U-shaped FCN [18] with taking advantage of bounding box knowledges obtained from Step 2-1. Let \( S \in \mathbb{R}^{N \times N^2} \) denote the segmentation image for a tooth corresponding to a detected bounding box in \( P \). We construct a training dataset \( \{I_{roi}^{(n)}, Y_{roi}^{(n)}\}_{n=1}^{N} \) for individual teeth segmentation, where \( I_{roi}^{(n)} \) and \( Y_{roi}^{(n)} \) are tooth images of \( P \) and \( S \) cropped by the bounding boxes. A segmentation map \( f_{seg} : I_{roi} \rightarrow Y_{roi} \) is learned using a U-shaped network and minimizing the following loss:

\[
\mathcal{L}_{seg} = \frac{1}{N} \sum_{n=1}^{N} \left[ -\frac{1}{M} \sum_{x} Y_{roi}^{(n)}(x) \log \left[ f_{seg} \left( I_{roi}^{(n)} \right)(x) \right] \right],
\]

(9)

where \( x \) is a pixel position and \( M \) is the number of pixels of \( Y_{roi} \).

2.3 Step 3: Extraction of loose and tight 3D tooth ROIs using the detected bounding boxes and segmented tooth regions

In this step, 3D tooth ROIs are obtained using the results of the previous steps. As described in Fig. 3 a bounding box containing one tooth is projected back into the 3D CBCT image using \( \{\vec{r}, \hat{t}\} \) and \( \{\vec{r}, \hat{t}\} \). A loose ROI domain of the target tooth is then given by

\[
D_{box} = \{(r(s) + \hat{t}(s), z) : -\alpha \leq t \leq \alpha, (s, z) \in B_{box}\},
\]

(10)

where \( B_{box} \) is the set of pixel positions in the bounding box. Similarly, a tight ROI domain \( D_{seg} \) is determined by \( B_{seg} \), which is the set of pixel positions in the 2D tooth segmented region in the box.

The loose 3D tooth ROI \( R_{box} \) is obtained by changing the voxel values outside \( D_{box} \) to 0, and extracting content in a 3D bounding box that fits closely around \( D_{box} \) as shown in Fig. 5. Similarly, we obtain the tight 3D tooth ROI \( R_{seg} \) by processing \( D_{seg} \) instead of \( D_{box} \), and using the same 3D bounding box as above.

2.4 Step 4: 3D segmentation for individual teeth from the 3D tooth ROIs

In this final step, 3D individual tooth segmentation is performed by applying the loose ROI \( (R_{box}) \) and tight ROI \( (R_{seg}) \) to the 3D version of a U-shaped FCN [18]. The tight ROI is crucial for improving the segmentation accuracy at the attached boundaries between a target tooth and its neighboring teeth.

The input of the network is \( I_{roi3} = R_{box} \oplus R_{seg} \), which represents the concatenating vector of two ROIs. Let \( Y_{roi3} \) denote a binary vector representing 3D tooth segmentation corresponding to \( I_{roi3} \). Using a training dataset \( \{I_{roi3}^{(n)}, Y_{roi3}^{(n)}\}_{n=1}^{N} \), we learn a 3D segmentation map \( f_{seg3} : I_{roi3} \rightarrow Y_{roi3} \) by minimizing the following loss:

\[
\mathcal{L}_{seg3} = \frac{1}{N} \sum_{n=1}^{N} \left[ -\frac{1}{V} \sum_{v} Y_{roi3}^{(n)}(v) \log \left[ f_{seg3} \left( I_{roi3}^{(n)} \right)(v) \right] \right],
\]

(11)

where \( v \) is a voxel position and \( V \) is the number of voxels of \( Y_{roi3} \).

3 Experiments and Results

3.1 Dataset and implementation details

Experiments were conducted on 3D CT images produced by a dental CBCT with a circular trajectory (DENTRI-X;
HDXWILL, Seoul, South Korea) using tube voltage 90kVp and tube current 10mA. All available datasets were formatted in the Digital Imaging and Communications in Medicine (DICOM) standard. The size of the CT image was \(800 \times 800 \times 400\), and its pixel spacing and slice thickness were both 0.2mm. During scanning, a bite block was used to prevent contact between the upper and lower teeth.

We received 97 dental 3D CBCT images from HDXWILL. Using these data, we generated 194 upper and lower jaws panoramic images in Step 1. We also received labeled data consisting of 97 2D individual tooth segmentation, bounding box components, and tooth codes, as well as 11 3D individual tooth segmentation. The labeling was performed by experts in HDXWILL.

Panoramic images were generated by (4) using a tomographic reconstruction software (TIGRE) [19]. The size of all panoramic images was fixed at \(640 \times 320\). The width of the panoramic images was determined by 640 reference curve points in Step 1-5. Those points were obtained by interpolating 500 points on the smooth curve and extrapolating 70 points at each end of the curve. To completely cover the teeth at both ends, we extrapolated 70 points (approximately 13.3mm) taking into account the average size of the molars. The height size 320 was determined by removing 80 CT slices that do not contain teeth from the bottom.

For the 2D detection and segmentation in Steps 2-1 and 2-3, 66 CBCT dataset were used for training and 31 dataset for testing. Because two panoramic images (upper and lower parts) were obtained from each CBCT image through Step 1, we used 132 labeled training data and 62 test data. Meanwhile, for the 3D segmentation in Step 4, 7 CBCT dataset were used for training and 4 dataset for testing. Since each patient has approximately 28 to 32 teeth, each CBCT image can provide approximately 28 to 32 training data for individual tooth segmentation. To be precise, we used 216 training data and 112 test data for the 3D segmentation in Step 4. In Steps 2-3 and 4, 2D tooth images and 3D loose and tight ROIs were resized to \(128 \times 128\) and \(128 \times 128 \times 128\), respectively.

The proposed neural networks were trained by minimizing losses in (8), (9), and (11) using Adam optimizer [20] from PyTorch [21]. We used five-fold cross validation on the training dataset. Batch sizes of 8, 32, and 4 were set for each training in Steps 2-1, 2-3, and 4, respectively. We analyzed the runtime of the proposed convolutional networks on a desktop with a GeForce RTX 2070 GPU. The processing time of the proposed three networks in Steps 2-1, 2-3, and 4 were 3.62, 5.98, and 6.82 ms/batch, respectively.

### 3.2 Evaluation and Result of the proposed method

#### 3.2.1 Bounding box detection

For a quantitative evaluation of the bounding box detection, we provide precision-recall (PR) curves [22] and their average precision (AP) [22], as shown in Fig. 6. When the intersection over union (IOU) threshold value was 0.6, according to the PR curve, the precision tends to stay high as the recall increases, and the AP was 88.11%.

#### 3.2.2 Individual tooth identification

This subsection presents the performance evaluation of tooth identification. The precision, recall, and F1-score were used to evaluate the results of the identification method. In Step 2-1, teeth are initially classified into four types, instead of directly predicting the eight tooth codes. The direct identification method can often misclassify teeth within the same tooth type. As shown in Fig. 7, the direct method confuses first premolars (code 4) and wisdom teeth (code 8) in particular. These errors hinder the performance improvement of the direct method. However, the four type-based method achieves a high accuracy by preventing misclassification due to similar tooth shape. Table 1 and Fig. 7 show that the proposed method leads to more accurate identification.

#### 3.2.3 Individual tooth segmentation

To evaluate the 2D and 3D tooth segmentation performances, we used precision, recall, Dice similarity coefficient (DSC) [23], Hausdorff distance (HD) [23], and average symmetric surface distance (ASSD) [23].
TABLE 1: Quantitative evaluation for tooth identification methods.

| Metric     | Direct method          | Proposed method          |
|------------|------------------------|--------------------------|
| Precision (%) | 93.05 ± 7.85          | 96.81 ± 1.67             |
| Recall (%)      | 87.22 ± 7.42          | 90.13 ± 5.30             |
| F1-score (%)    | 90.04 ± 7.63          | 93.35 ± 2.54             |

2D individual tooth segmentation. The proposed method proceeds in two steps consisting of the bounding box detection in Step 2-1 and individual segmentation in Step 2-3, while Mask R-CNN [8] achieves the same task in a single step. We implemented both approaches, and reported the quantitative evaluations in Table 2. The proposed method is numerically more accurate than the Mask R-CNN.

The accuracy of 2D segmentation is important because the key of precise 3D tooth segmentation is to use tight tooth ROIs obtained by 2D segmentation. Although the detection and segmentation are not performed simultaneously, two simple convolutional networks (one-stage object detection and U-shaped FCN) are designed to achieve a high accuracy.

TABLE 2: Quantitative evaluation for 2D tooth segmentation methods.

| Metric     | 2D Mask R-CNN          | Proposed method          |
|------------|------------------------|--------------------------|
| Precision (%) | 92.17 ± 5.55          | 96.18 ± 4.67             |
| Recall (%)      | 90.35 ± 5.87          | 93.17 ± 5.83             |
| DSC (%)       | 90.98 ± 3.25          | 94.41 ± 3.10             |
| HD (mm)       | 1.64 ± 0.93           | 1.34 ± 0.86              |
| ASSD (mm)     | 0.39 ± 0.12           | 0.27 ± 0.12              |

3D individual tooth segmentation. We developed a fully automated multi-step method for 3D individual segmentation. To verify the effectiveness of the proposed method, we compared it with Mask R-CNN [8], Path aggregation network (PANet) [24], Hybrid task cascade (HTC) [25], and ToothNet [7]. These methods cannot be applied to large size 3D CBCT images directly because of the computational limit. For comparison experiments, we implemented the methods in a patch-based fashion as an alternative to avoid the limitation. We also performed an ablation study of the proposed method by using either loose or tight ROI, or both. The four methods adopted for comparison show lower performances, as shown in Table 3. These methods perform individual segmentation from the original CBCT images. In contrast, the proposed method has the advantage of using loose and tight ROIs that provide considerable background region in advance. In particular, the tight ROI excludes structures on sides (e.g., adjacent teeth, jaw, etc) of the target tooth. When using only tight ROIs, the recall is the lowest because loss of tooth information may occur where the tight ROI boundary intersects the tooth boundary, as shown in Table 4. The use of only loose ROI shows that HD tends to be high because there is no information on tooth boundaries. A combination of the two ROIs enhances the segmentation performance, as the tight ROI provides detailed information on the target tooth and the loose ROI compensates for the disadvantage of the tight ROI. Wilcoxon signed-rank test [26] was used to calculate the statistical significance differences between the proposed method and other methods, as summarized in Tables 3 and 4.

3.3 Metal artifact-contaminated CBCT images

The proposed method effectively handles problems that are caused by metal-related artifacts. Fig. 8 shows that the CBCT image is severely contaminated by metal artifacts, whereas metal artifacts are significantly reduced in the corresponding panoramic image generated by the CBCT image. The panoramic images in Step 1 allow to accurately perform 2D tooth detection and segmentation. These results provide prior knowledge of each 3D tooth as loose and tight ROIs. As shown in Fig. 8 the tight ROI excludes adjacent teeth, even though the tooth boundaries are obscured by metal artifacts. Fig. 9 shows a qualitative evaluation for 3D tooth segmentation in a CBCT image with metal artifacts. As shown in Figs. 9 and 9, the segmentation results in the degraded CT image are superior to those of Figs. 9, 9, 9, 9, 9, and 9 as the tight ROIs provide robust tooth boundary information. However, Fig. 9 illustrates that using only the tight ROIs may not provide precise segmentation because it can cut out the edges of the teeth. Therefore, Fig. 9 illustrates the advantages and effectiveness of the proposed method using both loose and tight ROIs.

Fig. 7: Confusion matrix for tooth identification. (a) Result of the direct method. (b) Result of the proposed method.

Fig. 8: Image on the left is a CBCT image that is affected by metal artifacts. Image on the right is a panoramic image generated from Step 1, which is not affected by metal-related artifacts.

3.4 Missing tooth

People often lose their teeth due to factors such as cavities, periodontal disease, aging, dental trauma and orthodontic treatment. To address these cases, we applied CBCT images with missing teeth (except for wisdom teeth) to the proposed method. Four type-based classification was successful, but tooth identification was incomplete. With exception
In this paper, we developed a fully automated segmentation and identification method for individual teeth and jaws from CBCT images. Given CBCT data, the method automatically generates the maxillary and mandibular panoramic images that are projected along the reference curve representing a region-based shape feature of a dental arch. In the maxillary and mandibular panoramic images, 2D tooth segmentation and identification are performed using deep learning methods, which are vital in high-precision 3D tooth segmentation and identification. Experiments showed that the accuracy of the method is suitable for the clinical setting. Our method overcomes the limitations of existing automated methods by achieving fully automation and improved accuracy. Additionally, the method addresses the difficulty of learning high-dimensional data.

The main idea of the proposed method is the careful use of the accurate and robust 2D tooth segmentation and identification in 2D panoramic images in an indirect manner to address the difficulty of 3D segmentation from metal artifact-contaminated 3D CBCT images. In a clinical dental CBCT environment (e.g., low dose radiation exposure), metal related-artifacts are common. The proposed method utilizes the crucial observation that metal artifacts are significantly reduced in the upper and lower panoramic images generated from the CBCT images. The outcome in Step 2 serves as strong prior knowledge of 3D tooth segmentation, which plays an important role in separating teeth from 3D images, in cases where teeth are often contacted, overlapped, or connected owing to metal-related artifacts.

The automated system proposed in this study improves the efficiency of dentists by reducing the cumbersome and time-consuming manual intervention. The result provides an improved workflow for dentists to simulate preoperative orthodontic treatment and manufacture implant surgical guides. Digital occlusion analysis is potentially possible by combining our method with the intra-oral scan...
model via registration. Hence, it is expected to play an important role in digital dentistry.

ACKNOWLEDGMENTS

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI20C0127). We would like to express our deepest gratitude HDXWILL which shares dental CBCT images and ground-truth data.

REFERENCES

[1] H. Gao and O. Chae, “Individual tooth segmentation from ct images using level set method with shape and intensity prior,” Pattern Recognition, vol. 43, no. 7, pp. 2406–2417, 2010.
[2] 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,” Medical physics, vol. 42, no. 1, pp. 14–27, 2015.
[3] H.-T. Yau, T.-J. Yang, and Y.-C. Chen, “Tooth model reconstruction which shares dental CBCT images and ground-truth data. Would like to express our deepest gratitude HDXWILL which shares dental CBCT images and ground-truth data.
[4] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[5] Z. Cui, C. Li, and W. Wang, “Toothnet: automatic tooth instance segmentation and identification from cone beam ct images,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 6368–6377.
[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
[7] 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,” Computers in biology and medicine, vol. 80, pp. 24–29, 2017.
[8] D. V. Tuzloff, L. N. Tuzova, M. M. Bornstein, A. S. Krasnov, M. A. Kharchenko, S. I. Nikolenko, M. M. Sveshnikov, and G. B. Bedenko, “Tooth detection and numbering in panoramic radiographs using convolutional neural networks,” Dentomaxillofacial Radiology, vol. 48, no. 4, p. 20180051, 2019.
[9] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.
[10] H. Samet and M. Tamminen, “Efficient component labeling of images of arbitrary dimension represented by linear bintrees,” IEEE transactions on pattern analysis and machine intelligence, vol. 10, no. 4, pp. 579–586, 1988.
[11] R. M. Haralick, S. R. Sternberg, and X. Zhuang, “Image analysis using mathematical morphology,” IEEE transactions on pattern analysis and machine intelligence, no. 4, pp. 532–550, 1987.
[12] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, “Building skeleton models via 3-d medial surface axis thinning algorithms,” CVGIP: Graphical Models and Image Processing, vol. 56, no. 6, pp. 462–478, 1994.
[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in European conference on computer vision. Springer, 2016, pp. 21–37.
[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.