Automated Analysis of Three-dimensional CBCT Images Taken in Natural Head Position That Combines Facial Profile Processing and Multiple Deep-learning Models

Janghoon Ahn  
Hallym University

Thong Phi Nguyen  
Hanyang University

Yoon-Ji Kim  
University of Ulsan College of Medicine

Taeyong Kim  
Hallym University

Jonghun Yoon  
yooncsmd@gmail.com  
Hanyang University

Research Article

Keywords: Cephalometrics, Orthodontics, Mask-RCNN, Deep learning

DOI: https://doi.org/10.21203/rs.3.rs-795041/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Automated analysis of three-dimensional CBCT images taken in natural head position that combines facial profile processing and multiple deep-learning models

Janghoon Ahn\textsuperscript{1, a}, Thong Phi Nguyen\textsuperscript{2, 6, a}, Yoon-Ji Kim\textsuperscript{3}, Taeyong Kim\textsuperscript{4}, Jonghun Yoon\textsuperscript{5, 6*}

\textsuperscript{1}Department of Orthodontics, Kangnam Sacred Heart Hospital, Hallym University, Singil-ro 1 gil, Yeongdeungpo-gu, Seoul 07441, Republic of Korea
\textsuperscript{2}Department of Mechanical Design Engineering/ Major in Materials, Devices, and Equipment, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea
\textsuperscript{3}Department of Orthodontics, Asan Medical Centre, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505 Republic of Korea
\textsuperscript{4}Department of Advanced General Dentistry, Kangnam Sacred Heart Hospital, Hallym University, Singil-ro 1-gil, Yeongdeungpo-gu, Seoul 07441, Republic of Korea
\textsuperscript{5}Department of Mechanical Engineering, Hanyang University, 55, Hanyangdaehak-ro, Sangnok-gu, Ansan-si, Gyeonggi-do 15588, Republic of Korea.
\textsuperscript{6}BK21 FOUR ERICA-ACE Centre, Hanyang University, Ansan-si, Gyeonggi-do 15588, Republic of Korea

ABSTRACT

Analysing cephalometric X-rays, which is mostly performed by orthodontists or dentists, is an indispensable procedure for diagnosis and treatment planning with orthodontic patients. Artificial intelligence, especially deep-learning techniques for analysing image data, shows great potential for medical and dental image analysis and diagnosis. To explore the feasibility of automating measurement of 13 geometric parameters from three-dimensional cone beam computed tomography (CBCT) images taken in a natural head position, we here describe a smart system that combines a facial profile analysis algorithm with deep-learning models. Using multiple views extracted from the CBCT data as the dataset, our proposed method partitions and detects regions of interest by extracting the facial profile and applying Mask-RCNN, a trained decentralized convolutional neural network (CNN) that positions the key parameters. All the techniques are integrated into a software application with a graphical user interface designed for user convenience. To demonstrate the system’s ability to replace human experts, we validated the performance of the proposed method by comparing it with measurements made by two orthodontists and one advanced general dentist using a commercial
dental program. The time savings compared with the traditional approach was remarkable, reducing the processing time from about 30 minutes to about 30 seconds.

**Keywords:** Cephalometrics, Orthodontics, Mask-RCNN, Deep learning.

*Corresponding author: Tel: +82-31-400-5255; E-mail: jyoon@hanyang.ac.kr.com (Prof. Jonghun Yoon).

"The first two authors (Prof. Janghoon Ahn and Mr. Thong Phi Nguyen) contributed equally to this work."
1. Introduction

Since Broadbent [1] introduced X-ray techniques to orthodontics in 1931, they have become a common instrument for analysing the lateral or postero-anterior cephalograms on which orthodontic diagnosis and treatment planning are based. However, conventional two-dimensional X-ray images are prone to enlargement or distortion [2,3] because they generally superimpose the three-dimensional (3D) structures of the human skull onto one another [4], which provokes confusion and misinterpretation. Therefore, the cone beam computed tomography (CBCT) technique [5] was introduced to the dental profession between 2001 and 2004 with the NewTom 9000 (QR, Verona, Italy), the CB MercuRay (Hitachi Medical Systems America, Twinsburg, Ohio), and the I-CAT (Imaging Sciences, Hatfield, Pa) after Ambrose and Hounsfield [6] invented medical computed tomography (CT). With those advanced instruments, clinicians can clearly depict and digitize landmarks on 3D CBCT images from any aspect with few visual errors.

Those breakthrough visual data have inspired a lot of research to measure parameters useful for diagnosis in orthodontia. Park et al. [7] introduced a method to use the nasion true vertical plane (NTVP) and true horizontal plane (THP), based on natural head position (NHP), to analyse 3D CBCT images and show the anterior-posterior sagittal skeletal relationships and the protrusion of upper and lower anterior teeth. Also, Latif et al. [8] assessed anterior-posterior jaw relationships by analysing visualizations of the facial sagittal profile with Software Do-It Ceph (Dental Studio NX version 4.1, USA). However, the measurement tasks in that research still required a dentist to manually digitize the landmarks.

As in other fields, artificial intelligence (AI) techniques have been applied to dentistry and orthodontia. Takahashi et al. [9] proposed a method using You-Only-Look-Once version 3 (YOLOv3) to recognize dental prostheses and tooth restorations. However, that method had limited accuracy in detecting tooth-coloured prostheses.

Deep learning was used to detect periapical disease in dental radiographs in the research of Endres et al. [10]. Also, Kunz et al. [11] created an automated cephalometric x-ray analysis method using a specialized AI algorithm and compared its accuracy with the current gold standard. Kharbanda et al. [12] proposed a knowledge-based algorithm to automatically detect landmarks on 3D CBCT images.

Despite that previous research, few studies have used AI to analyse cephalometric measurements of three-dimensional CBCT images. In everyday practice, orthodontists still
spend much time and effort analysing three-dimensional CBCT images of orthodontic patients. Therefore, we here present a complete vision-based measurement system that can evaluate 13 representative parameters in 3D CBCT images to provide an integrated view of facial morphology, including sagittal and coronal analyses, occlusal cant, facial asymmetry, and anterior teeth inclination. We validated the automated measurements of those 13 parameters by comparing them with manual measurements made by two orthodontists and one advanced general dentist.

2. Materials and Methods

2.1 Landmark identification and measurements of the 13 parameters

In the sagittal view, the nasion point (N-point), which is the intersection between the nasion true vertical line (NTVL) and a true horizontal line, can be visually determined as the most concave point between the frontal bone and nasal bone. The A-point and B-point can be defined as the most concave points in the maxilla and mandible, respectively. Pog is the most prominent point on the chin, and menton (Me) is the lowest point of the mandible. Gonion (Go) is a point that intersect two lines, which go through articolare (Ar) point and Me and tangent the posterior border of a ramus and the lower border of a mandible respectively, on mandibular angle area.

In the coronal view, the N-point can be defined as the midpoint of the conjunction between the frontal bone and nasal bone. The upper dental midline (UDM) and lower dental midline (LDM) can be defined as the contact point between the upper and lower two central incisors. The canine tip (Ct, Rt and Lt) points are on the tips of the upper canines (right and left). The mesiobuccal cusp tip points (U6MB, Rt and Lt) are on the mesiobuccal cusp tips of the upper first molars (right and left). The Me-point is the midpoint of the chin profile.

The 13 parameters in this study were selected to evaluate the severity of facial asymmetry, diagnose skeletal relations, and measure anterior teeth inclination. All the parameters are given as geometrical values for length and angle, which are defined using the specific representative landmarks shown in Figure 1:

1. $d_x(A, NTVP)$: Distance from the A-point to the NTVP along the x-direction
2. $d_x(B, NTVP)$: Distance from the B-point to the NTVP along the x-direction
3. $d_x(Pog, NTVP)$: Distance from the Pog point to the NTVP along the x-direction
4. $d_y(Me, NTVP)$: Distance from the $Me$ point to the $NTVP$ along the y-direction
5. $d_y(UDM, NTVP)$: Distance from the $UDM$ to the $NTVP$ along the y-direction
6. $d_y(LDM, NTVP)$: Distance from the $LDM$ to the $NTVP$ along the y-direction
7. $d_z(Ct(Rt), THP)$: Distance from the right canine tip ($Ct(Rt)$) to the true horizontal plane ($THP$) along the z-direction
8. $d_z(Ct(Lt), THP)$: Distance from the left canine tip ($Ct(Lt)$) to the $THP$ along the z-direction
9. $d_z(U6MB (Rt), THP)$: Distance from the right upper 1st molar mesiobuccal cusp tip ($U6MB (Rt)$) to the $THP$ along the z-direction
10. $d_z(U6MB (Lt), THP)$: Distance from the left upper 1st molar mesiobuccal cusp tip ($U6MB (Lt)$) to the $THP$ along the z-direction
11. $\alpha(T_1T_2, THP)$: Angle between the vector $T_1T_2$ and the $THP$, where $T_1$ and $T_2$ are the tip of the root and the crown of the upper incisors, respectively
12. $\alpha(T_3T_4, MeGo)$: Angle between the vector $T_3T_4$ and the line $MeGo$, where $Me$ and $Go$ are the menton and gonion points of the jaw profile, respectively
13. $\alpha(MeGo, THP)$: Angle between the line $MeGo$ and the $THP$.

Figure 1. Landmarks representing the 13 parameters in the (a) sagittal view, (b) coronal view, and panoramic images.
2.2 Data preparation

The protocol of this study was reviewed and approved by the Institutional Review Board of Kangnam Sacred Heart Hospital, Hallym University (2021-02-016). The dataset of this study was constructed using 200 CBCT images taken in the NHP from 200 orthodontic patients (69 males, 131 females) who visited the Department of Orthodontics, Hallym University, Kangnam Sacred Heart Hospital, Seoul, Republic of Korea. All of CBCT images were collected with informed consent from patients. Exclusion criteria included a history of orthodontic treatment, orthognathic surgery, and prosthetic treatment of the upper or lower anterior teeth. Subjects with more than two missing teeth were also excluded.

Each set of images contains 3D scanning data with 576 slices at a resolution of 768×768 pixels and a slice thickness of 0.3 mm. Full field-of-view (23 × 17 cm) images were taken with an I-CAT CBCT scanning machine (KaVo Dental GmbH, Biberach, Germany) using the operational parameters of 120 Kv, 37.1 mA, voxel size of 0.3 mm, and scan time of 8.9 seconds. Subjects were asked to stare at their own eyes in a mirror after exercising their head up and down. A 40 × 50-cm mirror was hung on the wall 1.5 m from the subjects. Subjects’ heads were positioned without a chin support or any other type of head holder.

Of the 200 cases considered, 170 cases were used to construct the dataset, train the deep-learning models, and provide samples for the developer, and the 30 remaining cases were used as the validation set to compare the accuracy of the proposed method with manual measurements.

In the proposed method, the procedure for determining the parameters is not a straightforward process for calculating output from the loaded input. Instead, it is a flexible sequence for computing–selecting data for extraction–computing. As shown in Figure 2 (a), the input of the next stage is selected based on the results calculated in the previous stage to minimize the human factors required for operation. First, the facial profile processing algorithm, which takes the sagittal and vertical bone views extracted from the CBCT data as inputs, evaluates cephalometric landmarks as key-points for parameters 1–4 and 13. Then, the axial layer of the CBCT data, which corresponds to the mid-point between points A and B in the cephalometric landmarks, is used to define the profile of the jawbone and thereby generate a panoramic view of the tooth structure. Second, the Mask-RCNN and proposed tooth numbering method are used to detect and classify the teeth in the input panoramic view. In this step, representative key-points in the dental parameters (parameters 5–10) are also positioned. Third, the position of the incisors detected in the panoramic view is used to extract tooth sagittal views by applying
a decentralized CNN to evaluate the inclination of incisor alignment (parameters 11–12). An example of the input images used in the proposed method is shown in Figure 2 (b).

Figure 2. Multiple dental visualization modes extracted from raw CBCT images.

2.3 Cephalometric analysis algorithm

Extracting skull and mandibular profiles in the sagittal view is the first essential step before positioning cephalometric landmarks. Considering the skull profile, which is demonstrated as a red line in Figure 3, the extraction process uses the similarity between a random position, in
terms of its depth value and characteristics at the boundary of the skull, and its neighbouring points. From a starting point in the first pixel row of a sagittal depth image \( D \) that differs from zero and has the largest \( x \)-value, one pixel \( D(x_i, i) \) in pixel row \( i \) is compared with a chosen pixel in the previous row \( (i - 1) \), and it is considered to be on the facial profile if it satisfies the conditions in Eq. (1), where \( n \) is the \( y \)-size of sagittal depth image \( D \).

\[
\begin{align*}
D(x'_i, i) > 0, \forall i = 2, \ldots, n \\
|D(x'_i, i) - D(x_{i-1}, i - 1)| < d \\
|x_i - x_{i-1}| = \min|x'_i - x_{i-1}|, \forall x'_i
\end{align*}
\]  

(1)

Based on the obtained skull profile, the Me point can easily be defined as the final point of the profile, which is then considered as the starting point for extracting the mandibular profile. The procedure for this extraction is similar to the process for the facial profile, but the direction along \( y \)-axis is reversed. However, unlike the Me point, which is defined as the end point of the profile, the Ar point, which is the final point of the mandibular profile, is detected based on the change in the mean depth value of the neighbourhood region-of-interest (ROI). Specifically, a \( 5 \times 5 \) pixel ROI cropped from the sagittal depth image on the lower left side of the considered point and corresponding with a point in the jaw profile is collected, as shown in Figure 3 (b), and then the mean depth value of that ROI is calculated. The Ar point can be recognized by comparing the calculated depth values with the real distance between the cranial base and the skull’s vertical plane.

According to the skull profile thus obtained, the N-point, A-point, B-point, and Pog are determined as extreme points. In detail, the N-point is the minimum local extreme point nearest the nasal peak, and the Pog, B-point, and A-point are the 1\(^{st}\) maximum local extreme point and 1\(^{st}\) and 2\(^{nd}\) minimum local extreme points from the Me point, respectively. To detect the Go point, two tangent lines of the mandibular profile obtained from the Me and Ar points are drawn.

In the coronal depth image, the N-point in the sagittal and coronal views shares \( z_N, y_N \) with the parameter under consideration, as shown in Figure 3(c). Therefore, using the same method, the left side profiles \( T_i \) and \( T'_i \) of the nasal bone in the coronal view from \( z_N - \frac{n}{2} \) to \( z_N + \frac{n}{2} \) (\( n = 40 \) is used in this research) are extracted, and \( y_N \) is calculated by Eq. 2 as the mean value.

\[
y_N = \frac{\sum_{i} (T_i + T'_i)}{n}
\]  

(2)
Figure 3. Detection strategies for a cephalometric analysis: (a) skull (red) and mandibular (green) profile detection in the sagittal view, (b) neighbourhood ROIs for detecting Ar, (c) orbit-side profile and mandibular profile in the coronal view, (d) graph of the jaw profile’s x-values.

Meanwhile, the Me point in the coronal view is detected based on the x-values of the mandibular profile, which is also extracted using Eq. (1). The starting and end points of the scanning process are defined as intersections $S_1$ and $S_2$ between the mandibular profile and two vertical lines that go through the intersections of line $z = (z_A + z_B)/2$ and the two sides of the mandible in the coronal view, respectively. The x-values of the mandibular profile are extracted, and Me is the point with the maximum extracted x-values.

2.4 Deep-learning system for detecting tooth landmarks
The first use of the deep learning technique is to generate the dental arch curve from one CT-images of jaw as shown in Fig. 4 (a) and extract a panoramic view of the teeth. For this process, we use a landmark detection CNN, whose architecture is based on VGG-net [13], as shown in Table 1. The output of the model, which is also the label of the training dataset, is six landmarks that represent the global tooth profile, as shown in Fig. 4 (b), where $P_1$ and $P_6$ represent the head of the mandible, $P_2$ and $P_5$ are located around the third molar, and $P_3$ and $P_4$ represent canines. Three tooth-containing images from the CBCT-image set for each patient were collected with a resolution of 768×768 pixels and manually labelled. Because high-precision landmark positioning was considered unnecessary for this task, this labelling process was conducted by developers, not orthodontists. The plot of five sequential segments that connect the six detected points was smoothed using a 1D-Gaussian filter, as shown in Fig. 4 (c), before an area formed by the margins on both sides of the obtained curve was defined, as shown in Fig. 4 (d), and spread to generate a panoramic view of all teeth.

Table 1 Specifications of the layer size and shape of the landmark detection CNN used to generate the dental arch curve.

| Layer          | Filter Size/Kernel Size | Output Shape |
|----------------|-------------------------|--------------|
| Conv1          | 8 filter 9×9            | (760, 760, 8) |
| Max – pooling 1|                         | (253, 253, 8) |
| Conv2          | 8 filter 7×7            | (250, 250, 8) |
| Max – pooling 2|                         | (125, 125, 8) |
| Conv3          | 16 filter 7×7           | (119, 119, 16) |
| Max – pooling 3|                         | (59, 59, 16)  |
| Conv4          | 16 filter 5×5           | (53, 53, 16)  |
| Max – pooling 4|                         | (26, 26, 16)  |
| Conv5          | 32 filter 3×3           | (22, 22, 32)  |
| Max – pooling 4|                         | (11, 11, 32)  |
| Conv5          | 32 filter 3×3           | (9, 9, 32)    |
| Dropout        |                         | (9, 9, 32)    |
| ReLU           |                         | (9, 9, 32)    |
| Fully connected 1 | 256                    | 256          |
| Fully connected 2 | 90                     | 90           |
| Output         |                         | 12           |
Figure 4. Extraction of dental arch curve: (a) original input image, (b) six key-point detection, (c) dental arch curve after smoothing, (d) margins on both sides of the dental curve used to create the panoramic view.

After generation of the panoramic view, the next function detects the required teeth (incisors, canines, and first molars) from it. Common object detection algorithms, including RCNN, YOLO, and SSD [14], only position targets in the input images as boundary boxes. However, accurate segmentation is beneficial to auxiliary dental diagnoses and useful for future extended research topics. Therefore, we chose the Mask-RCNN method to accurately detect and segment the teeth in the panoramic view.

The designed Mask-RCNN network is shown on the left of Fig. 5. Mask-RCNN [15] adds a branch predicting the segmentation mask in each area of interest based on the Faster R-CNN [16], wherein an ROI pool is used to extract small features from each ROI. First, the ROI pool scales the ROI to the granularity of the feature graph, and then it partitions the scaled ROI to produce a summary of the characteristic values of each block of coverage area. Those calculations misplace the ROI and extracted features, though possibly without affecting the classification, which has a certain robustness to those small amplitude transformations. Nonetheless, those transformations have a large negative effect on the prediction of a precise mask at the pixel level. Mask-RCNN proposes an ROI align layer to remove the dislocation of the ROI pool and accurately align the extracted features with the input. ROI align uses bilinear
interpolation to calculate the exact value of each position and summarizes the results using the maximum or average pooling. The extracted features and input can be aligned pixel-to-pixel, which effectively maximizes the segmentation accuracy. The dataset for training the Mask-RCNN model was constructed using panoramic images as shown in Fig. 6 (a) which are generated from the original 200 CBCT images. A detection example is shown in Fig. 6 (b); Mask-RCNN successfully detected 98.75% of the 30 test cases.

![Mask-RCNN network](image)

**Figure 5.** Left: Mask-RCNN network. Right: Landmark detection procedure for teeth-of-interest.

After obtaining the detection results from the trained Mask-RCNN model, the centre points of the boundary boxes representing the positions of all teeth are calculated as shown in Fig. 6 (c). Because the teeth-of-interest for which landmarks are required (incisors, canines, and first molars) are distinguished based on their order in the horizontal direction and classification as upper/lower teeth, classifying the obtained centre points as upper and lower teeth is indispensable. To complete that task, a 2nd-degree polynomial function is generated using parameters matched to the positions of the centre points in a linear regression. Based on this function, which is represented by the green curve in Fig. 6 (d), the upper/lower category of each tooth can be conveniently determined from its relative position between the considered centre point and the representative 2nd-degree polynomial curve.
Figure 6. Example of landmark detection procedure for the panoramic view: (a) original input image, (b) Mask-RCNN detection result, (c) centre points of detected teeth, (d) upper (red scatter)/lower (yellow scatter) tooth classification, (e) numbering of upper/lower teeth, (f) teeth-of-interest (red scatter) and required landmarks (blue scatter).
Figure 7. Example dataset for training the landmark detection CNN to ROIs on teeth from the panoramic view.

Table 2 Specifications for the layer size and shape of the landmark detection CNN for the tooth crown.

| Layer        | Filter Size/Kernel Size | Output Shape |
|--------------|-------------------------|--------------|
| Conv1        | 8 filter 9x9            | (292, 142, 8)          |
| Max – pooling 1 | 3x3                    | (97, 47, 8)         |
| Conv2        | 16 filter 7x7           | (91, 41, 16)        |
| Max – pooling 2 | 2x2                    | (45, 20, 16)     |
| Conv3        | 16 filter 7x7           | (39, 14, 16)        |
| Max – pooling 3 | 2x2                    | (19, 7, 16)         |
| Conv4        | 32 filter 5x5           | (15, 3, 32)        |
| Dropout      |                         | (15, 3, 32)        |
| ReLU         |                         | (15, 3, 32)        |
| Fully connected 1 | 256                    | 256             |
| Fully connected 2 | 90                     | 90              |
| Output       |                         | 6              |

The numbering process for upper and lower teeth is implemented from the middle position of all teeth. Currently, the method can deal with patients whose teeth are in the conventional arrangement. The numbering priority is based on the horizontal distance from each centre point to the middle position, without regard to each tooth’s appearance. This numbering approach can recognize missing teeth by the abnormal distance between two teeth and the middle position, as shown in Fig. 6 (e). Then, the ROIs of the teeth-of-interest (incisors, canines, and first molars as numbers 1, 3, and 6, respectively, in Fig. 6 (e)) are resized to 300x150 pixel resolution and loaded into a landmark detection CNN with the architecture shown in Table 2.
That CNN is trained to locate three chosen points on each tooth: the middle point and left and right corners of the tooth crown on the enamel region, as shown in Fig. 7. Consequently, the characteristics in panoramic images can be defined based on the landmarks detected for the chosen teeth. In particular, UDM and LDM are the vertical lines that go through middle points of the landmarks located on the upper and lower incisors, respectively. The left and right upper 1st molar mesiobuccal cusp tips are represented by the right and left corners of the corresponding left and right first molars, and the parameters of the canines are calculated from the middle detected landmark, as shown in Fig. 6 (f).

2.5 Decentralized CNN to evaluate anterior incisor inclination

Anterior incisor inclination is evaluated by defining the tooth alignment axis, which is the line from the tip of the incisor root to the tip of the incisor crown. Therefore, landmarks that represent the tip of the root and tip of the crown in the upper and lower incisors are essential. Unlike the landmark detection already described, the accuracy of this function must be high. Therefore, we used a decentralized CNN model [17-19] previously proposed as a high-accuracy landmark detecting method for medical images. The advantage of this method is that it narrows the ROI after each order, which not only reduces the number of non-related features that can affect the results, but also increases the diversity of the training dataset. Considering the complexity of this detection, we propose a decentralized CNN model with two orders, as shown in Fig. 8: the 1st order roughly detects the region of tooth tips, and the 2nd order precisely locates the landmarks within the 1st order’s results.
The weight factors of the trained CNN models are adjusted using the difference between the output, $A^i_j$, from the deep-learning model containing the calculated position value, $j$, for the input image and the label, $Y^i_j$, created based on the actual position of the required points in the input image. The difference, $L$, is calculated using the mean square error loss function in Eq. (3), which is useful for handling features corrupted with outliers, where $n$ is the number of labels in the dataset.

$$L = \frac{\sum_{i=1}^{n}(A^i_j - Y^i_j)^2}{n}. \quad (3)$$

After each training step, the weight factor, $\theta_t$, is updated for the $(t + 1)^{th}$ iteration based on the rule shown in Eq. (4), where $M$ is the batch size ($M = 64$), $\eta$ is the learning rate (initially set as 0.001 and updated using SGDM (momentum = 0.95) [20]), and $\nabla$ is the gradient operator.

$$\theta_{t+1} = \theta_t - \eta \nabla \theta_t[L(A_m, Y_m)] \quad (4)$$

3. Results

3.1 System implementation

For user convenience in implementing the software, we developed the graphical user interface (GUI) shown in Fig 9. From the CBCT-images uploaded to the system, raw CT images in the sagittal, coronal, and axial views can be monitored in the upper left corner of the GUI. The program exports sagittal and coronal bone modes, panoramic views, and tooth modes, which are all data for the proposed system to process.

To extract dental features as described above, the GUI allows users to manipulate the data either manually or by the proposed automatic method. For manual manipulation, the dentist digitizes specific landmarks directly on each frame of the GUI for parameter calculation. This option is provided for patients with highly complicated, significantly abnormal facial profiles and tooth structures, which would cause difficulties for the developed program. The option using the proposed method, named AI extraction, automatically and sequentially locates landmarks, calculates parameters, and visualizes the results on frames without requiring any manual manipulation. The total calculation time for the AI-based option is around 12 seconds.
on our test configuration of a Core i7-7700 with 32 GB of random access memory and a Geforce RTX 2070 (NVIDIA) graphics processing unit.

In the comment box in the upper right showing the calculated parameters, the categories of sagittal cephalometric, which is related to NTVP, and vertical cephalometric, which is related to THP, are also integrated into the program to help free dentists and orthodontists from having to memorize and look up standard tables.

Figure 9. GUI designed for the developed system.

3.2 Comparison with the accuracy of manual measurements
The gold standard selected to validate the developed program is manual measurement conducted by three experienced orthodontists with more than 10 years working in this field professionally. They made their measurements using In Vivo Dental software (version 5.3; Anatomage Co., San Jose, CA, USA), which was developed especially for medical radiology diagnosis and open dental CBCT, medical CT, MRI, and other medical scans in the standard DICOM format. We chose this approach as the standard reference because it is the most popular method for measuring 3D CBCT images for orthodontic diagnosis. Furthermore, orthodontic experts deem existing commercial software with automatic measuring functions to lack reliability and demonstrate poor effectiveness in practice.

The statistical analyses of this study were performed with Excel program (Microsoft corp., Washington, USA). The deviations between the proposed method and the manual
measurements of the 13 parameters are given in Table 3, along with the mean absolute error (MAE) and correlation coefficients among the manual results.

Table 3 Deviation between the proposed method and direct measurement by three MDs.

| No. | Parameter | Mean ± 1.96×STD | MAE with MDs          | ICC  |
|-----|-----------|-----------------|----------------------|------|
| 1   | \(d_x(A, NTVP)\) (mm) | 1.14 ± 7.96 | 1st M.D.: 0.73<br>2nd M.D.: 0.88<br>3rd M.D.: 0.74 | 0.972 |
| 2   | \(d_x(B, NTVP)\) (mm) | -2.30 ± 11.6 | 1st M.D.: 0.76<br>2nd M.D.: 1.06<br>3rd M.D.: 0.71 | 0.984 |
| 3   | \(d_x(Pog, NTVP)\) (mm) | -3.00 ± 14.06 | 1st M.D.: 0.86<br>2nd M.D.: 1.09<br>3rd M.D.: 1.28 | 0.990 |
| 4   | \(d_y(Me, NTVP)\) (mm) | -2.09 ± 5.92 | 1st M.D.: 1.11<br>2nd M.D.: 1.32<br>3rd M.D.: 1.16 | 0.941 |
| 5   | \(d_y(UDM, NTVP)\) (mm) | -0.30 ± 3.29 | 1st M.D.: 1.08<br>2nd M.D.: 1.04<br>3rd M.D.: 0.96 | 0.763 |
| 6   | \(d_y(LDM, NTVP)\) (mm) | -0.5 ± 4.04 | 1st M.D.: 1.14<br>2nd M.D.: 0.97<br>3rd M.D.: 1.01 | 0.883 |
| 7   | \(d_z(Ct(Lt), THP)\) (mm) | 79.7 ± 7.57 | 1st M.D.: 0.96<br>2nd M.D.: 1.17<br>3rd M.D.: 1.00 | 0.948 |
| 8   | \(d_z(Ct(Rt), THP)\) (mm) | 80.29 ± 7.83 | 1st M.D.: 0.78<br>2nd M.D.: 1.22<br>3rd M.D.: 0.83 | 0.962 |
| 9   | \(d_z(U6MB (Lt), THP)\) (mm) | 76.11 ± 7.40 | 1st M.D.: 1.09<br>2nd M.D.: 1.35<br>3rd M.D.: 1.02 | 0.923 |
| 10  | \(d_z(U6MB (Rt), THP)\) (mm) | 77.45 ± 7.16 | 1st M.D.: 0.88<br>2nd M.D.: 1.02<br>3rd M.D.: 0.81 | 0.942 |
| 11  | \(\alpha(T_1T_2, THP)\)^\(\circ\) | 119.47 ± 14.52 | 1st M.D.: 1.89<br>2nd M.D.: 2.05<br>3rd M.D.: 1.95 | 0.969 |
| 12  | \(\alpha(T_3T_4, MeGo)\)^\(\circ\) | 92.62 ± 16.15 | 1st M.D.: 1.77<br>2nd M.D.: 1.92<br>3rd M.D.: 1.87 | 0.963 |
| 13  | \(\alpha(MeGo, THP)\)^\(\circ\) | 28.27 ± 11.71 | 1st M.D.: 0.93<br>2nd M.D.: 1.46<br>3rd M.D.: 1.36 | 0.978 |
To clearly illustrate the correlation between the developed program and measurements made by orthodontists, scatterplots of those correlations for six groups of parameters are shown in Fig. 10. Specifically, the validation data for parameters 1–3, which all consider the NTVP as the baseline for measurement along the x-axis, are visualized in Fig. 10 (a), and Fig. 10 (b) illustrates validation data for parameters 4–6, which are measured along the y-axis. The correlations between the proposed method and experts for parameters 7–10, which are measured from the THP along the z-axis, are presented in Fig. 10 (c). Because the three-angle parameters all differ in the measuring baselines they use, the validation data for parameters 11–13 are respectively shown in Fig. 10 (d), (e), and (f). Three experts participated in the validation process, and they each measured every case two times, so the values representing the manual measurements in the plots are the means of six individual measurements.
Figure 10. Scatterplots of correlations between manual measurement values and the proposed method’s results: (a) parameters 1–3, (b) parameters 4–6, (c) parameters 7–10, (d) parameter 11, (e) parameter 12, (f) parameter 13.

To further demonstrate the consistency between the measurements of the proposed method and those made by orthodontists, analysis of variance (ANOVA) tests were conducted with the validation data for the following hypotheses:

1. Null hypothesis $H_0$: all measurements are equal $\mu_1 = \mu_2 = \mu_3 = \mu_4$

2. Alternative hypothesis $H_\alpha$: At least one measurement is unequal

With $\alpha = 0.05$, the ANOVA tables for validating the parameters were analysed in Origin software (version 2018, OriginLab Corp., Northampton, Massachusetts, USA.). The F values and P values are given in Table 4 and compared with the standard Z-table. With 30 test cases, the standard $F_{dis} = 2.68$ ($\alpha = 0.05$) extracted from the table is larger than all of the collected F values, and the chosen $\alpha = 0.05$ is smaller than all the collected P values. Therefore, there is no significant evidence against the null hypothesis $H_0$, demonstrating the consistency between the measurements made using the proposed method and those of orthodontists.

Table 4 F-values and P-values from ANOVA tests.

| No. | Parameter       | F     | P-value | No. | Parameter       | F     | P-value |
|-----|----------------|-------|---------|-----|----------------|-------|---------|
| 1   | d(A, NTVP)     | 0.032 | 0.992   | 8   | d(C(Rt), THP)  | 0.018 | 0.997   |
| 2   | d(B, NTVP)     | 0.084 | 0.969   | 9   | d(U6MB(Lt), THP)| 0.073 | 0.974   |
| 3   | d(Pog, NTVP)   | 0.084 | 0.969   | 10  | d(U6MB(Rt, THP)| 0.108 | 0.955   |
3.3 Efficiency

To clarify the efficiency, we measured the time required for both the proposed method and the manual method for all 30 test cases. The time period collected for each case was from the moment the CBCT image set started loading to complete measurement of all 13 target parameters. The orthodontists needed 25–35 minutes to measure the 13 parameters for one patient, even with the support of CBCT data visualizing software. Meanwhile, the proposed program required only about 5 seconds to load the data, (or around 15 seconds for compressed CBCT-data) and about 12 seconds to measure all the parameters.

In short, compared with manual measurement, this automatic measuring program can save 30 minutes of diagnosis time per patient. In addition to the time savings, the cost of the commercial software used to assist with the manual measurement is high, and it still requires the user to perform directly.

4. Discussion

This study has introduced a hybrid system that combines a facial profile analysis with deep-learning methods (Mask-RCNN and decentralized CNN) to detect representative cephalometric and facial asymmetry landmarks. The proposed system was used to calculate 13 purposefully chosen parameters, distances and angles defined between landmarks, and we validated its results by comparing them with manual measurements made by two experienced orthodontists and one advanced general dentist as gold standard references. The small deviations obtained in the comparison results verify the practical accuracy and consistency of the proposed method. Our automatic measurement and diagnosis system has the following key innovations and functions.

1) **Advanced graphical visualization:** The interactive platform provides advanced graphic visualization by applying rendering processing algorithms to raw CBCT images so that
users can see the structure of patient skulls and the tooth arrangements in both the sagittal and coronal views, making it convenient for users to digitize landmarks and extract parameters. In this respect, the newly developed software is equivalent to common software already available commercially.

2) **Intelligent analysis:** The proposed smart dental-parameter measuring system uses a vision analysing approach that combines facial profile processing with AI-based techniques. These methods were developed using the experience of experts in this field and have demonstrated excellent consistency with manual standard references in our validation study. Therefore, this system can free laboratory staff or experts from this repetitive task.

3) **Enhancement of productivity:** The developed program offers a significant advantage in the time required for measurements, compared with the typical manual measurement approach. As the aesthetic surgery market grows, classical measuring procedures will soon fail to meet the requirements of productivity. Therefore, a computerized approach to data handling, which offers a processing speed almost 60 times faster than the manual method, is a valuable alternative option.

The implemented CBCT data allow users to observe the scans in a 3D environment instead of the usual 2D of other typical radiographs, but the quality of the images still depends on the hardware performance, which can be affected by the scanning condition and influence the operation of our proposed method. In addition, the tooth numbering method used in this study had difficulties when premolars had been extracted. That abnormal situation caused inaccurate detection of landmarks because the tooth order was different from normal cases. Therefore, our future work will give priority to improving the tooth numbering method, which is based on the classification ability of a CNN applied to an extended database. Moreover, our further work will develop an orthographic surgical plan design system that can offer options by using novel machine-learning techniques with the already extracted parameters to act as a virtual assistant for oral and maxilla-facial surgeons.

5. **Conclusion**

This study proposed a vision-based dental system that combines an analysis of facial profiles with deep-learning algorithms to measure 13 parameters in 3D CBCT images taken in NHP to automatize repetitive measuring procedures and increase productivity. The facial analysis and deep learning algorithms were applied to multiple views extracted from CBCT images to
localize representative landmarks. Furthermore, application software was designed to run the automatic extracting method proposed and allow users to visualize and digitize landmarks manually via an appropriate GUI. The geometric parameters calculated by the proposed system were compared with measurements by human experts and found to be highly consistent in terms of low deviations and significant correlations among the study population. Consequently, this vision-based dental system achieved an obvious time advantage by reducing the measuring time from about 30 minutes to about 30 seconds.

Author Contribution: T. P. N. designed the overall system, including conceptualization and the software, under the supervision of J. Y. and J. A. In addition, they wrote and revised the paper. Y. J. K. and T. K. contributed to dataset curation and annotation and helped to design the comparative analysis and experiment for validation.

Acknowledgments:
This research was financially supported by the Ministry of Trade, Industry, and Energy (MOTIE), Korea, under the “Digital Manufacturing Platform (DigiMaP)” (reference number N0002598) supervised by the Korea Institute for Advancement of Technology (KIAT).

The author Prof. Jonghun Yoon received research funding from the National Research Foundation of Korea (NRF), grant funded by the Korea Government (MSIT) (No. 2019R1A2C4070160)

Competing interests
The author(s) declare no competing interests.

Ethics declarations
This study was reviewed and approved by the Institutional Review Board of Kangnam Sacred Heart Hospital, Hallym University (2021-02-016)
All methods were performed in accordance with the relevant guidelines and regulations
All of CBCT images were collected with informed consent from patients.
References

[1] Broadbent, B. H. A new x-ray technique and its application to orthodontia. *The Angle Orthodontist, 1*(2), 45–66 (1931).

[2] Hsiao, T. H., Chang, H. P., Liu, K. M. A method of magnification correction for posteroanterior radiographic cephalometry. *The Angle Orthodontist, 67*, 137–142 (1997).

[3] Song, G. Y., Li, G., Lu, W. H., Han, B., Xu, T. M. Distortion and Magnification of Four Digital Cephalometric Units. *Nigerian Journal of Clinical Practice, 22*(12), 1644-1653 (2019).

[4] Grummons, D. C., Kappeyne van de Coppello, M. A. A frontal asymmetry analysis. *Journal of Clinical Orthodontics, 21*(7), 448-465 (1987).

[5] Arai, Y., Tammisalo, E., Iwai, K. Development of a compact computed tomographic apparatus for dental use. *Dentomaxillofacial Radiology, 28*(4), 245–248 (1999).

[6] Ambrose, J., Hounsfied, G. Computerized transverse axial tomography. *The British Journal of Radiology, 46*, 148-149 (1973).

[7] Park, Y. J., Cho, Y. S., Mah, J., Ahn, J. H. Assessment of anterior-posterior jaw relationships in Korean adults using the nasion true vertical plane in cone-beam computed tomography images. *The Korean Journal of Orthodontics, 46*(3), 163-170 (2016).

[8] Latif, V. B., Keshavarajm, Rai, R., Hedge, G., Shajahan, S. Assessment of Antero-Posterior Skeletal and Soft Tissue Relationships of Adult Indian Subjects in Natural Head Position and Centric Relation. *Journal of International Oral Health, 7*(6), 41-47 (2015).

[9] Takahashi, T., Nozaki, K., Gonda, T., Mameno, T., Ikebe, K. Deep learning-based detection of dental prostheses and restorations. *Scientific Reports, 1*, 1960-1967 (2021).

[10] Endres, M. G., et al. Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs. *Diagnostics, 10*, 430-451 (2020).

[11] Kunz, F., Stellzig-Eisenhauer, A., Zeman, F., Boldt, J. Artificial intelligence in orthodontics Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *Journal of Orofacial Orthopedics, 81*, 52–68 (2020).

[12] Gupta, A., Kharkbanda, P. O., Sardana, V., Balachandran, R., Sardana1, H. K. A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images. *International Journal of Computer Assisted Radiology and Surgery, 10*,
1737–1752 (2015).

[13] Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. in ICLR (2015).

[14] Wang, X., Chen, K., Huang, Z., Yao, C., Liu, W. Point linking network for object detection, arXiv preprint arXiv:1706.03646 (2017).

[15] He, K., Gkioxari, G., Dollar, P., Girshick, R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision, DOI 10.1109/ICCV.2017.322 (2017).

[16] He, K., Gkioxari, G., Dollar, P., Girshick, R. 2017. Mask-RCNN. arXiv preprint arXiv:1703.06870v3

[17] Nguyen, T. P., et. al. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. Computers in Biology and Medicine, 120, 103732 (2020).

[18] Chae, D., et al. Decentralized convolutional neural network for evaluating spinal deformity with spinopelvic parameters. Computer Methods and Programs in Biomedicine, 197, 105699 (2020).

[19] Nguyen, T. P., Chae, D., Park, S., Kang, K. Y., Yoon, J. Deep learning system for Meyerding classification and segmental motion measurement in diagnosis of lumbar spondylolisthesis, Biomedical Signal Processing and Control, 65, 102371 (2020).

[20] Jayalashmy, S. Sudha, G. F. Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks. Artificial Intelligence in Medicine, 103, 1018092, (2020).