scientific reports

OPEN

Automatic segmentation and landmark detection of 3D CBCT images using semi supervised learning for assisting orthognathic surgery planning

Haomin Tang[1,5], Shu Liu[2,5], Yongxin Shi[3], Jin Wei[4], Juxiang Peng[2] & Hongchao Feng[4]✉

Patients with abnormal relative position of the upper and lower jaws (the main part of the facial bones) require orthognathic surgery to improve the occlusal relationship and facial appearance. However, in addition to the retraction and protrusion of the maxillomandibular advancement, these patients may also develop asymmetry. This study aims to use a semi-supervised learning method to demonstrate the maxillary and mandible retraction, protrudation and asymmetry of patients before orthognathic surgery through automatic segmentation of 3D cone beam computed tomography (CBCT) images and landmark detection, so as to provide help for the preoperative planning of orthognathic surgery. Among them, the dice of the semi-supervised algorithm adopted in this study reached 93.41 and 96.89% in maxillary and mandibular segmentation tasks, and the average error of landmark detection tasks reached 1.908 ± 1.166 mm, both of which were superior to the full-supervised algorithm with the same data volume annotation. Therefore, we propose that the method can be applied in a clinical setting to assist surgeons in preoperative planning for orthognathic surgery.

Keywords Upper and lower jaws, Orthognathic surgery, Semi-supervised learning, Automatic segmentation, Automatic landmark detection, 3D CBCT

Nowadays, computer-assisted surgery (CAS) has become an important tool[1,2] in the preoperative planning of orthognathic surgery, which provides convenience for clinical diagnosis by automatically reconstructing surgical structures and automatically locating landmarks accurately and quickly. In Guiyang Stomatological Hospital, surgeons need to evaluate the degree of maxillary and mandibular retraction and **protrusion**[3] and asymmetry[4] of patients before orthognathic surgery, then determine the orthognathic operation area formulate the osteotomy trajectory and design the repair of bone defects after osteotomy, as well as skeleton fixation, and finally generate the preoperative plan.

With the development of artificial intelligence (AI) in recent years, deep learning has provided new ideas for some problems in medical imaging. Convolutional Neural networks (CNNs) is a class of deep learning algorithms mainly used in the field of computer vision. It is a challenging task[5–8] to use CNNs to accurately segment bone structure and detect the location of landmarks in cone beam computed tomography (CBCT) images. Manual image segmentation refers to using a mask to cover the target anatomical structure displayed by sections of multiple layers (about 20–40 layers) on the CBCT image on the annotation software. Similarly, manual image punctuation refers to marking the position of landmarks on a certain layer of the CBCT image. Therefore, the manual segmentation and punctuation of CBCT images is a time-consuming and laborious process, often requiring more than ten hours to complete a patient's CBCT data.

In view of the difficulty of medical image labeling and the need to label a large amount of image data to achieve satisfactory results, semi-supervised learning is more suitable. In recent years, semi-supervised learning has become a popular new direction in the field of deep learning. This method only requires a small number of labeled samples and a large number of unlabeled samples. Therefore, in this study, we put forward the semi-

[1]College of Medicine, Guizhou University, Guiyang 550025, China. [2]Department of Orthodontics, Guiyang Hospital of Stomatology, Guiyang 550002, China. [3]School of Stomatology, Zunyi Medical University, Guiyang 563006, China. [4]Department of Oral and Maxillofacial Surgery, Guiyang Hospital of Stomatology, Guiyang 550002, China. [5]Haomin Tang and Shu Liu contributed equally to this work. ✉email: hongchaof@126.com

supervised learning method and achieved good results with a small amount of labeled data. This study adopted mean teacher[9] as the semi-supervised learning framework, which is based on consistent regularization and evolved from the Π model and temporal ensembling[10]. The framework consists of a teacher model (TM) and a student model (SM), both of which use the same network structure (Vnet[11]). In the training process, different noise disturbances are added to the TM and SM, and training the models by minimizing the outputs difference between the two (the TM generates false labels by predicting unlabeled data as the learning target of the SM), so that the output results of the models are still consistent under different disturbances, thus effectively improving the generalization ability of the model.

The task of 3D landmark detection is to regress the position of a certain point in the space. Currently, landmark detection methods can be roughly divided into two types: The method based on direct regression[12,13] and the method based on heatmap regression[8,14,15], in which the former regression directly obtains coordinates and the latter results in a heatmap, have their own advantages and disadvantages. We consider that direct regression may lack spatial generalization (convolution weight sharing can mitigate this problem), while heatmap regression uses gaussian distribution functions as a 'soft annotation', it is more conducive to convergence in network training and easier for the network to regress to a region of the heatmap (foreground pixels) than to regress to a point, and doing so also allows the network to have spatial generalization. The Gaussian heatmap regression method has achieved good results in previous studies on medical image processing[14,15], so we also choose the heatmap regression method in this study.

Methods

Dataset

Data acquisition

In this study, we obtained 3D CBCT data from 192 patients in the radiology Department of Guiyang Stomatological Hospital. The pixel size of the obtained CBCT images are $(528\sim563)\times640\times640$, and the size of each pixel is $0.25\times0.25\times0.25\text{mm}^3$. The inclusion criteria of the data in this study are: (1) Including 50% patients with maxillary protrusion and mandibular retraction, 50% patients with maxillary retraction and mandibular retraction, aged from 18 to 40 years old. (2) Incomplete upper and lower jaw bones without visible bone defects, destruction, or resorption on CBCT. (3) Complete CBCT data, scanning range covering the top of the orbits below the head. Exclusion criteria: (1) Patients with a history of orthodontic treatment, orthognathic treatment, trauma to the upper and lower jaw bones, or bone surgery. (2) Patients with skeletal malocclusion. (3) Patients with dental malocclusion.

The dataset of this study comes from anonymous CBCTs that have been used for preoperative scanning of orthognathic surgery, most of which are in the unlabeled state, and for a small part of the data that need to be labeled, This study imported CBCT Dicom (Digital Imaging and Communications in Medicine) files into an open source software tool (3D-Slicer version 5.6.1, http://www.slicer.org/)[16] for data annotation, export the NIFTI format files marked with segmentation and TXT files marked with landmarks (recorded in IJK coordinates of 3D-Slicer) as labels for this dataset. Figure 1 shows the maxilla (green part), mandible (yellow part) and part of skull base bone (blue part, but this part was not trained as a data set) reconstructed and visualized by clinical experts in this study after manual segmentation and labeling. Based on some previous literatures[4,17,18], 18 landmarks were located at different locations (Including skull base, maxilla, mandible) in this study, as shown in (Table 1) below.

We divided the total dataset by 9:1, the number of training sets was 192 CBCT, and the number of test sets was 19 CBCT. Due to the semi-supervised learning method used in this study, we just need to label 20% of the total dataset (38 CBCT cases were labeled) in the training set, and all the data to be labeled were labeled by 3 experienced clinical experts in the field of orthognathic surgery in Guiyang Stomatological Hospital. For 3D segmentation, three clinical experts made several marks and modifications to the best, and then the three experts discussed and modified the results of their respective annotations until they reached unity. For 3D Landmark detection, Three clinical experts each repeated the labeling 3 times, and each labeling interval of 3

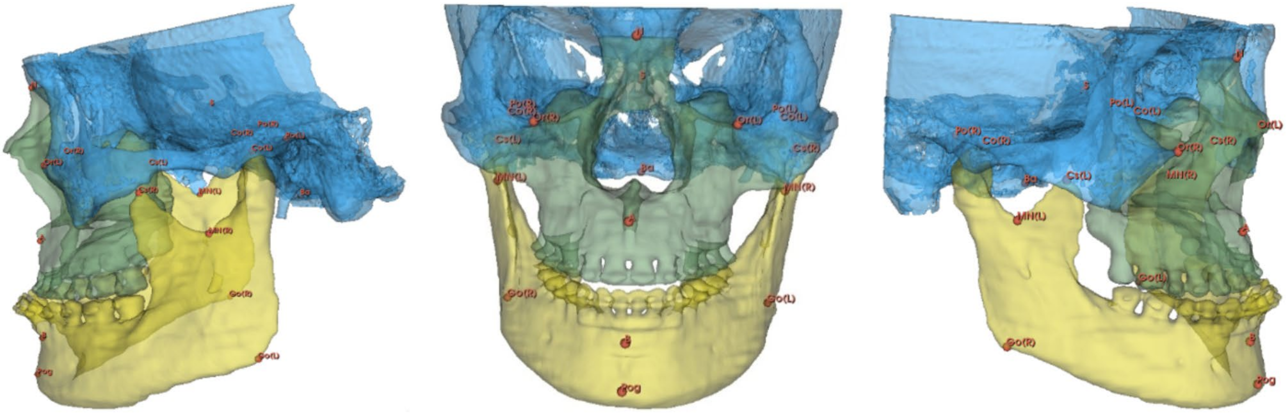


Fig. 1. The patient's 18 landmarks and segmented areas were visualized on 3D slicer.

3D landmark		Definition
Landmarks on Frankfurt horizontal plane	(L)Or	The lowest point of the left infraorbital edge
	(R)Or	The lowest point of the right infraorbital edge
	(L)Po	The most superior point of the left external auditory meatus
	(R)Po	The most superior point of the right external auditory meatus
Landmarks on Sagittal plane	Ba	The most posteroinferior point of the anterior margin of the foramen magnum in the midsagittal plane
	N	The most anterosuperior junction of the nasofrontal suture
Landmark on coronal plane	S	The most central point of sella turcica
Landmarks as reference points	B	The most concave point on the labiomental sulcus
	Pog	The most convex point of the chin profile
	A	The most concave point of the philtrum
	(L)Go	The most inferior and posterior point at the angle of the left mandible
	(R)Go	The most inferior and posterior point at the angle of the right mandible
	(R)Cs	The most superior point of the right coronoid process
	(L)Cs	The most superior point of the left coronoid process
	(L)Co	The most superior point of the left condyle
	(R)Co	The most superior point of the right condyle
	(L)Mn	The most inferior point of the left mandibular notch
	(R)Mn	The most inferior point of the right mandibular notch

Table 1. 18 Anatomical landmarks.

days. The repeatability of the three selected markers was evaluated for each clinical expert, and the intra-class correlation coefficient (ICC) of the three-dimensional coordinate values of each marker was calculated. All of them are 0.95 ~ 1.00, which means that the selection points within the clinical experts have good consistency. Based on the above markup results, we also got the mean error and variance of three times manual fixed-point markings by each of the three clinical experts: 0.62 ± 0.08(mm), 0.51 ± 0.12(mm), 0.82 ± 0.27(mm). The average of the three fixed point coordinates is taken as the doctor's artificial fixed point result, and then the ICC of the three-dimensional coordinate values of each marker is calculated using the fixed point results of three clinical experts. The ICC is greater than 0.95, it means that the manual labeling is consistent among clinical experts. At the same time, we also found that the mean error and variance of manual fixed-point labeling among the three clinical experts was 0.89 ± 0.34(mm).

In this study, the visualization of 3D CBCT image obtained by a patient with maxillary retracement and mandibular protrusion was taken as an example, as shown in (Fig. 1). The maxilla and mandible marked by manual segmentation were shown in green and yellow. To show completeness, we reconstructed part of the skull shown in blue in 3D. The 18 manually marked anatomical points (landmarks) are shown in red.

Data preprocessing

To eliminate the scale difference between features to make them are comparable and speed up the convergence of network training. We used min-max scaling to normalize the image pixel values for the input network in the dataset to a range of [0,1], as shown in Formula 1 below. x represents all the data of the image pixel value, and $\mathrm{x_i}$ is the a pixel value of the image. Max(x) and min(x) are the maximum and minimum values of the image pixel values, respectively.

$$\mathrm{x_i} = \frac{\mathrm{x_i} - \min(\mathrm{x})}{\max(\mathrm{x}) - \min(\mathrm{x})} \tag{1}$$

To keep the input CBCT image size of the model consistent and avoid the situation of insufficient computer graphics card resources, we used the Simpleitk[19] to resample the all 3D CBCT images in the dataset to reduce the image matrix size to about 1/4 of the original size (The error of the network's prediction of the landmark will be larger if the reduction ratio is too large). The resulting image pixel size is 132 × 160 × 160, and voxel spacing is 1 × 1 × 1 mm^3.

To make the network have better generalization ability and avoid overfitting problems, we called Monai framework[20] to carry out 3D data enhancement operations on the training set of this study, to increase the diversity of training data samples. For the segmentation task, we carried out enhancement processing on images and labels together, and for the landmark regression task, we only carried out enhancement processing on unlabeled data: rotate 15 to −15° with a probability of 0.2, flip along any axis of the three-dimensional coordinate system with a probability of 0.2, and translate the range of 10 mm to −10 mm with a probability of 0.2.

Maxillary and mandibular segmentation for Semi-supervised learning

Clustering hypothesis

The clustering hypothesis means that the input data points form clusters, each corresponding to an output class so that if the points are in the same cluster, they can be considered to belong to the same class. The clustering hypothesis can also be viewed as the low-density separation hypothesis, where a given decision boundary separates different clusters and is located in a low-density region.

Mean teacher
One direction of research in deep semi-supervised learning is to use unlabeled data to reinforce training models to conform to the clustering hypothesis. In this study, we adopted the Mean Teacher to realize the task of maxillary and mandibular segmentation, and the Mean Teacher is based on the above hypothesis: if two different perturbations are applied to one unlabeled data as inputs to both models, the prediction results of Mean Teacher's two model scores should not change significantly.

Mean teacher is composed of a student model (SM) and a teacher model (TM), both of which have the same segmentation model structure. For labeled data, the student model is used for supervised learning, and the summation average of cross entropy loss and dice loss is adopted as the supervised loss function. For labeled and unlabeled data, the consistency regular loss of unsupervised learning between the teacher model and the student model is made, and mean square error (MSE) is used as the consistency loss. Then, the supervised loss L_{su} and unsupervised loss L_{co} are combined into the final loss function L_f as follows (2):

$$\mathrm{L_f{=}min}\sum{}_{\mathrm{i=1}}^{\mathrm{N}}\mathrm{L_{su}}\left(\mathrm{f}\left(\mathrm{x_i;W_S}\right),\mathrm{y_i}\right)+\lambda\sum{}_{\mathrm{i=1}}^{\mathrm{N+M}}\mathrm{L_{co}}\left(\mathrm{f}\left(\mathrm{x_i;n',W_S}\right),\mathrm{f}\left(\mathrm{x_i;n,W_T}\right)\right) \tag{2}$$

Where () is a segmentation network, $\mathrm{W_S}$ and $\mathrm{W_T}$ are the learnable weights of the SM and the TM respectively. n and $\mathrm{n'}$ represent different perturbations of the SM and the TM under the same input data $\mathrm{x_i}$. N is the number of labeled data, M is the number of unlabeled data, and $\mathrm{y_i}$ is the ground-truth label. To control the balance between supervised and unsupervised consistency loss, the ramp-up weighting coefficient λ is used here:

$$\lambda\ =\ \delta\ *\ e^{-\gamma\left(1\ -\ \frac{t}{t_{\max}}\right)^{2}} \tag{3}$$

Where γ is the rate of ramp-up, δ is the consistency weight, *t* and t_{max} represent the number of current training rounds and the total training rounds. At the early stage of model training, supervision loss accounts for most of the target loss, which prevents the network from failing to obtain meaningful target prediction with unlabeled data and falling into regression.

The specific method of unsupervised learning is to add different noises to an image and input the TM and the SM respectively, and set the consistency loss constraint between the two model predictions, so that the predictions of the two models conform to the clustering hypothesis, and the data points with different labels are separated in the low-density region. The SM parameter (θSM) was optimized by AdamW, and the TM parameter (θTM) was updated by the exponential moving average (EMA). The EMA formula is as follows:

$$\boldsymbol{\theta}_{\mathrm{TM}}^{\mathrm{t}}\ =\ \boldsymbol{\alpha}\boldsymbol{\theta}_{\mathrm{TM}}^{\mathrm{t-1}}\ +(1-\boldsymbol{\alpha})\boldsymbol{\theta}_{\mathrm{SM}}^{\mathrm{t}} \tag{4}$$

where α is the ratio that controls how much weight is obtained from the SM.

Overfitting is easy in model training when there is a small amount of labeled data. By encouraging the same prediction before and after the unlabeled data disturbance, the consistency regularization method makes the decision boundary of learning located in the low-density region, which effectively alleviates the phenomenon of overfitting. The TM makes false labels for the unlabeled data and then adds the training to get a better decision boundary.

In general, training the mean teacher network is a process of seeking common ground while reserving differences, with slightly different input images and network parameters. We assume that the network is completely convergent after training, then the consistency loss will be small. At this time, the parameters of the SM and TM should be very close, and they also have good denoising ability.

Uncertainty-aware mean teacher (UA-MT)
Because the TM in Mean teacher predicts the uncertainty of false labels on the unlabeled data set, and the prediction of the TM plays a crucial role in the guidance (consistency loss) of the SM, In the previous literature[21] (Yu et al.2019), Uncertainty-Aware mean teacher solves the above problem to a certain extent. In this paper, the uncertainty of the TM is reflected by predicting the uncertainty of each pixel value of the image, and the prediction result is obtained by Q times of forward propagation for each input data, and gaussian noise is randomly added to the input data each time. Therefore, there are Q prediction results for each voxel, and prediction entropy is selected to measure the uncertainty, and the formula is expressed as:

$$\boldsymbol{\mu}_{\mathrm{c}}{=}\frac{1}{\mathrm{Q}}\sum_{\mathrm{q=1}}^{\mathrm{Q}}\mathrm{p}_{\mathrm{q}}^{\mathrm{c}}\ \mathit{and}\ \ \mathrm{u}=-\ \sum_{\mathrm{c}}\boldsymbol{\mu}_{\mathrm{c}}\mathrm{log}\boldsymbol{\mu}_{\mathrm{c}} \tag{5}$$

where is the prediction of probabilities belonging to category c in the q-th time forward propagation, expressed as p_q^c.

Under the guidance of obtaining the uncertainty of each voxel prediction, we filtered out the prediction with higher uncertainty of the teacher model and selected the prediction with more certainty as the learning goal of the student model. The final uncertainty consistency los L_{co} is as follows:

$$L_{co}(f', f) = \frac{\sum_{\nu} I(u_{\nu} < H) \, ||f'_{\nu} - f_{\nu}||^2}{\sum_{\nu} I(u_{\nu} < H)} \tag{6}$$

Where u_v is the value of the estimated uncertainty on voxel v, H is the threshold of filtering the uncertainty prediction, we also use the gaussian increasing paradigm to increase the uncertainty threshold from 3/4 Umax to Umax. Umax is the maximum voxel prediction uncertainty in the entire CBCT image (the maximum uncertainty values in the mandibular and maxillary data sets are 0.6931424 and 0.6931436, respectively). I(·) is an indicator function that returns 1 if true and 0 otherwise, used to screen out voxel samples smaller than H, and is the prediction result of the teacher and student networks at voxel v position, respectively.

In this study, we also used the method proposed by Yu et al. in 2019 to conduct maxilla and mandible segmentation for Semi-supervised learning. The semi-supervised segmentation network framework is shown in the (Fig. 2). The figure shows that the maxillary and mandibular structures in the CBCT data were separately labeled and divided into two independent data sets for training, in which the labeled CBCT images were input into the SM and the unlabeled CBCT images were input into the SM and the TM, and different noise disturbances were randomly added to each input. The SM obtains supervised loss by calculating the difference between the predicted data and the labeled data. Consistency loss is obtained by the difference between TM prediction under uncertainty filtering and SM prediction, and the supervised loss and consistency loss sum to get the final loss. The network learns to optimize the weight parameters through final loss.

Segmentation evaluation indicators

For the mandibular segmentation task, we analyzed the experimental results by considering five performance indicators[22,23] to help us understand the advantages and disadvantages of the algorithm model and the direction of improvement: Dice similarity coefficient (DSC), Positive Predictive Value (PPV), Sensitivity (SEN) and Average Surface Distance (ASD) and hausdorff distance (HD).

One of the main evaluation criteria used in the segmentation process, DSC, is a measure of ensemble similarity, which is usually used to calculate the similarity of two samples. The value ranges from 0 to 1, and the closer the segmentation result is to 1, the better, as shown in (7). Where A is the ground truth voxel region, and B is the voxel region of prediction segmentation.

$$DSC = \frac{2||A \cap B||}{||A|| + ||B||} \tag{7}$$

PPV refers to the proportion of correctly predicted voxel regions in the predicted segmentation voxel region, and its formula is shown in (8).

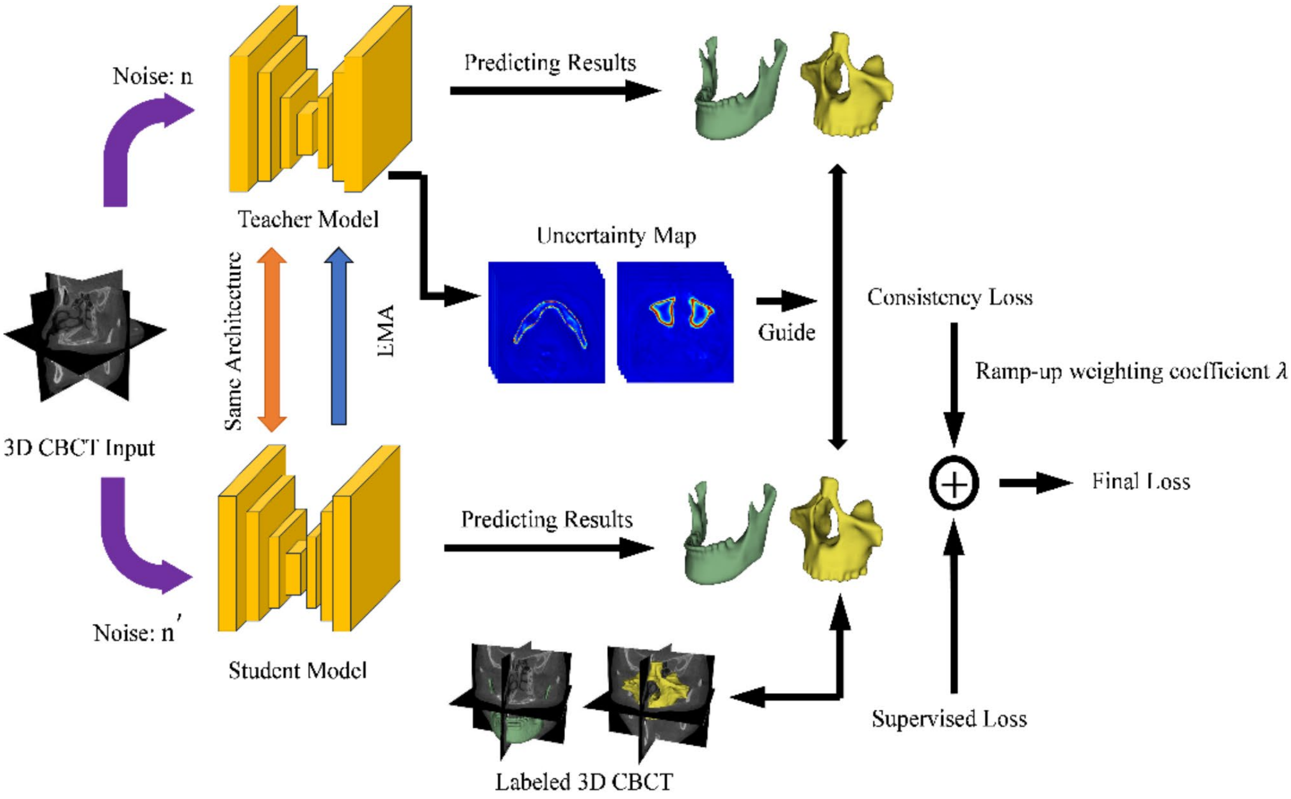


Fig. 2. A network framework for automatic semi-supervised segmentation of the maxilla and mandible.

$$\mathrm{PPV} = \frac{||\mathrm{A} \cap \mathrm{B}||}{||\mathrm{B}||} \tag{8}$$

SEN refers to the proportion of the ground truth voxel region that predicts the correct voxel region, and its formula is shown in (9).

$$\mathrm{SEN} = \frac{||\mathrm{A} \cap \mathrm{B}||}{||\mathrm{A}||} \tag{9}$$

ASD (mean surface distance) is the average distance between all points on the surface of two target regions (A and B). It can be used to calculate the error between the real label and the predicted segmentation region. The smaller the calculated value, the better. Its calculation formula (10) is shown as follows, where d(a, b) represents the minimum distance from point A on voxel a to voxel B, and table d(b, a) represents the minimum distance from point B on voxel b to voxel A.

$$\mathrm{ASD} = \frac{1}{2}\left\{\frac{\sum_{\mathrm{b}\in\mathrm{B}} min_{\mathrm{a}\in\mathrm{A}}\mathrm{d(b,a)}}{|\mathrm{B}|} + \frac{\sum_{\mathrm{a}\in\mathrm{A}} min_{\mathrm{b}\in\mathrm{B}}\mathrm{d(a,b)}}{|\mathrm{A}|}\right\} \tag{10}$$

Hausdorff distance (HD), named after Felix Hausdorff (1868–1942), is the maximum distance between one set and the nearest point in another set. In this study, Hausdorff distance is A max-min function from voxel A to voxel B. Be defined as:

$$\mathrm{D_{HD}(A,B){=}max(h(A,B),h(B,A))} \tag{11}$$

$$\mathrm{h\,(A,\ B)} = \max_{\mathrm{a}\in\mathrm{A}} \{\min_{\mathrm{b}\in\mathrm{B}} \{\mathrm{d(a,\ b)}\}\} \tag{12}$$

Landmark detection for semi-supervised learning

Mean teacher for regressing heatmaps

In previous research literature[13,24], there have been many tasks for CNNs to achieve regression landmark coordinates, among which the CNN direct regression coordinate method needs to build more layers to obtain more network parameters for better fitting. Moreover, since direct regression finally uses the full connection layer (FCN) to output the value of coordinates, it may not only lead to the lack of spatial generalization ability of the network. It may also cause the network to lose the correlation between the regression points. Heatmap regression uses the full convolutional network, which not only reduces the computational complexity of network parameters (one full connection layer is missing), but also outputs one spatial heatmap after another, making up for the problem that the above direct regression may have insufficient spatial generalization ability. Based on the above analysis, the heatmap regression method was adopted for landmark detection in this study.

Similar to semi-supervised segmentation, semi-supervised landmark detection also supports consistent regularization[25] and pseudo-label generation[26], and the landmark detection in this study adopts the heatmap regression method, and the core idea of Mean Teacher does not affect the final prediction dimension, and similar methods have been used to achieve human pose estimation in previous literature[27]. Therefore, in this study, we can transfer semi-supervised learning to the detection task.

The heatmap labeled as this image accords with the Gaussian distribution function, the formula is as follows (13): the gaussian standard deviation is set to 8, $^\wedge\mathrm{x_i}$ is any three-dimensional coordinate in the Gaussian heatmap, $\mathrm{x_i}$ is the ground truth landmark coordinate of the real label landmark in the center of the gaussian heatmap.

$$\mathrm{g_i(x;}\sigma) = \frac{\gamma}{(2\pi)^{\mathrm{d}/2}\sigma}\exp\left(-\frac{||\hat{\mathrm{x}}_{\mathrm{i}} - \mathrm{x_i}||^2}{2\sigma^2}\right) \tag{13}$$

The Adaptive Wing Loss function used in this study comes from previous literature[28]. This paper analyzes two problems of the widely used MSE loss in evaluating the difference between the ground truth heatmaps and the prediction heatmaps: 1, MSE is insensitive to small error losses, reducing the ability to accurately locate the center of the Gaussian distribution. 2. In the training process, MSE uses the same loss function and weight for all pixel values, but there are many more background pixels than foreground pixels, resulting in an imbalance of pixel categories, which results in a heatmap comparison with GT (the ground truth heatmap). The heatmaps predicted by models trained with MSE loss are fuzzy and bloated. The author wang et al. divided the pixels in heatmap regression into multiple categories, including center, foreground, difficult background and background, and proposed Adaptive Wing Loss to better complete the location of heatmap points.

$$\mathrm{A}wing(\mathrm{m},\hat{\mathrm{m}}) = \begin{cases} \omega\ln(1+|\frac{\mathrm{m}-\hat{\mathrm{m}}}{\varepsilon}|^{\alpha-\mathrm{y}}) & \mathrm{i}f\ |\mathrm{y}-\hat{\mathrm{y}}| < \theta \\ \mathrm{A}\,|\mathrm{m}-\hat{\mathrm{m}}| - \mathrm{C} & otherwise \end{cases} \tag{14}$$

$$A = w(1/(1+\left(\frac{\theta}{\epsilon}\right)^{\alpha\ - m})(\ \alpha\ - g\ )(\ \left(\frac{\theta}{\epsilon}\right)^{\alpha\ - m - 1})(\ 1/\epsilon\ ) \tag{15}$$

$$C = (\theta\ A - \omega\ ln(1+(\theta\ /\epsilon)^{\alpha\ - m})) \tag{16}$$

In the formula, A and C are set to make function Awing derivable at |m − | = θ. y is the three-dimensional heatmap pixel value of groundtruth and the predicted three-dimensional heatmap pixel value. This time, we set the parameter values of α = 2.1, ω = 14, ε = 1, θ = 0.5 with the best effect in wang et al.'s experiment.

According to the following formula (4), for N 3D volumetric heatmaps that need to be predicted for each CBCT image, the network minimizes the AWing loss between the target 3D volumetric heatmap and the regression prediction 3D volumetric heat map by training and adjusting parameters, as shown in the following formula (5). For a 3D volumetric heatmap, the argmax function is used to obtain the maximum value, which is to predict the coordinates of each landmark.

$$\min_{\mathrm{w,b}} \sum_{\mathrm{i=1}}^{\mathrm{L}} \sum_{\mathrm{x}} Awing(\hat{\mathrm{g}}_{\mathrm{i}}(\mathrm{I;w,b}),\ \mathrm{g}_{\mathrm{i}}(\mathrm{x};\sigma)) \tag{17}$$

The Gaussian distribution function was used to construct the 3D heatmap and serve as the regression target. The higher the pixel value of the 3D heatmap, the closer it is to the target pixel coordinate. As the pixel coordinate moves away, the pixel value of the 3D heatmap decreases rapidly and then the decreasing rate tends to be gentle.

The network predicts the coordinate values of x, y and z by predicting the heatmap, which can be understood as a probabilistic response map. By calculating the position coordinates of the maximum value of the heatmap, the position coordinates of the key point can be obtained in the network. As shown in the following formula, x_i is the value of each pixel in the heatmap predicted by the network, representing the scaling size of the pixel. We set it to 10 here. In order to avoid multiple large pixel peaks in the heatmap, the calculation is not accurate enough because the maximum value is not large enough, so a more accurate position coordinate can be obtained by increasing the relative maximum value and weakening the influence of other values. It contains the parameters w, b of the network after training and learning. Is the three-dimensional coordinate corresponding to the maximum pixel value of the heatmap.

$$\mathrm{h_i} = \frac{\mathrm{e}^{\beta \mathrm{x_i}}}{\sum_{\mathrm{j}} \mathrm{e}^{\beta \mathrm{x_j}}} \tag{18}$$

$$\hat{\mathrm{x}}_{\mathrm{i}} = \arg\max(\mathrm{h_i}(\mathrm{w,b})) \tag{19}$$

Uncertainty analysis of landmark detection with mean teacher

In this study, Mean teacher network was also used to carry out landmark detection tasks. Similarly, unlabeled inputs did not provide ground truth, and the predicted targets in the teacher model may be unreliable and noisy. For this problem, if we apply the method of task segmentation to the detection task, it is not realistic to estimate the uncertainty of prediction by calculating the variance of each voxel value, and the computational amount required for model training will be large. Therefore, we redesign the method of uncertainty analysis of teacher model prediction, and add the same image several times under different noises as the input of teacher model. Record the coordinates with the highest pixel value of the output heatmap of the teacher model, and then calculate the variance between the output coordinates of these several times, and take this parameter (variance) as the object of training and optimization of the network model. The specific method is as follows:

After input Gaussian noise into each input image, input it into the teacher model, and then perform T times of random forward transmission on the teacher model with dropout. Then, one of the T (T = 6) times of transmission is randomly selected to output the three-dimensional heatmap of the teacher model and argmax is used to obtain the predicted three-dimensional fixed-point coordinates. Similarly, the student model is performed once forward transmission. Then input the student model to argmax to get the three-dimensional coordinates. We use MSE to calculate the consistency loss of the predicted coordinates of the two models. MSE is shown in formula (20) below, where N is the number of landmark categories predicted on each image, 18.

$$L_{MSE} = \sum {}_{i=1}^{N} (f_i - y_i)^2 \tag{20}$$

In order to add uncertainty learning to formula (20), we use the Gaussian distribution formula for reference, bring MSE into formula (21), and perform another deformation to get formula (23), where is the weight of the variance. The almost equal contribution of variance and MSE to the loss function can be adjusted by setting the weight coefficient (the variance loss scales to about the same scale as the MSE loss), and the consistency loss function of the two models can be obtained. In this way, the magnitude of our consistency loss function is determined not only by the predicted values of the two models, but by the variance (σ) of the predicted values of the teacher model, f_a is calculated by the following formula (22), where is the average coordinates of the six teacher model predictions.

$$Gaussian\ Distribution\ Function\,(GDF) = \sum {}_{i=1}^{N} \frac{1}{\sigma_{\ i}\sqrt{2\pi}} e^{-\frac{(f_i - y_i)^2}{2\sigma_{\ i}^{\ 2}}} \tag{21}$$

$$\sigma = \sqrt{\frac{\sum_{\mathrm{t=1}}^{\mathrm{T}} (f_t - f_a)^2}{\mathrm{T} - 1}} \tag{22}$$

$$Consistency\ Loss = -ln\left(GDF\right) = \sum_{i=1}^{N} \frac{\left(f_i - y_i\right)^2}{2\sigma_i^2} + \frac{\beta}{\sqrt{2\pi}} \ln\left(\sigma_i\right) \tag{23}$$

It can be seen from formulas (21) and (23) that the larger GDF we need, the consistency loss is smaller, the variance (σ) is smaller, the better. The smaller the uncertainty of the teacher's model prediction, which means that the model can automatically learn from more certain things, thus improving the model prediction accuracy.

The semi-supervised landmark detection network proposed by us is shown in (Fig. 3). Similar to the above semi-supervised segmentation network, the input CBCT images are disturbed by random noise, wherein labeled CBCT images are input into the student model and unlabeled CBCT images are input into the student model and teacher model. Under the guidance of uncertain estimation, the teacher model generates relatively certain prediction results as false labels, and makes consistency loss with the student model prediction. Finally, the network updates the network weight of the network framework through consistency loss and supervision loss of the student model as the total loss.

Landmark detection and evaluation indicators

For the regression task of anatomical landmarks, We first calculated the inter-class correlation coefficient (ICC) to determine the confidence that the clinician marked on the CBCT images. The average error between the marked landmarks of all categories (N) in M CBCT images and the corresponding automatic positioning landmarks is calculated as the global average error (GME). And point mean squared error (PMSE) refers to the average of the errors after C tests of automatic location landmark and marked landmark for each category in the test set. To evaluate whether automatic fixation is clinically acceptable. GME, PMSE and the relevant standard deviation (SD) are defined in Eqs. (24), (25), and (26), where Δx, Δy, and Δz represent the absolute distance between the real landmark and the predicted landmark of each three-dimensional anatomical point in the x, y, and z directions, respectively. $\mathrm{R_c}$ represents the absolute distance between the real landmark and the predicted landmark of a three-dimensional anatomical point of a certain class.

$$\mathrm{Global\ mean\ error\ (GME)} = \frac{1}{\mathrm{MN}} \sum_{\mathrm{m=1}}^{\mathrm{M}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \sqrt{\boldsymbol{\Delta}\mathrm{xi}^2 + \boldsymbol{\Delta}\mathrm{yi}^2 + \boldsymbol{\Delta}\mathrm{zi}^2} \tag{24}$$

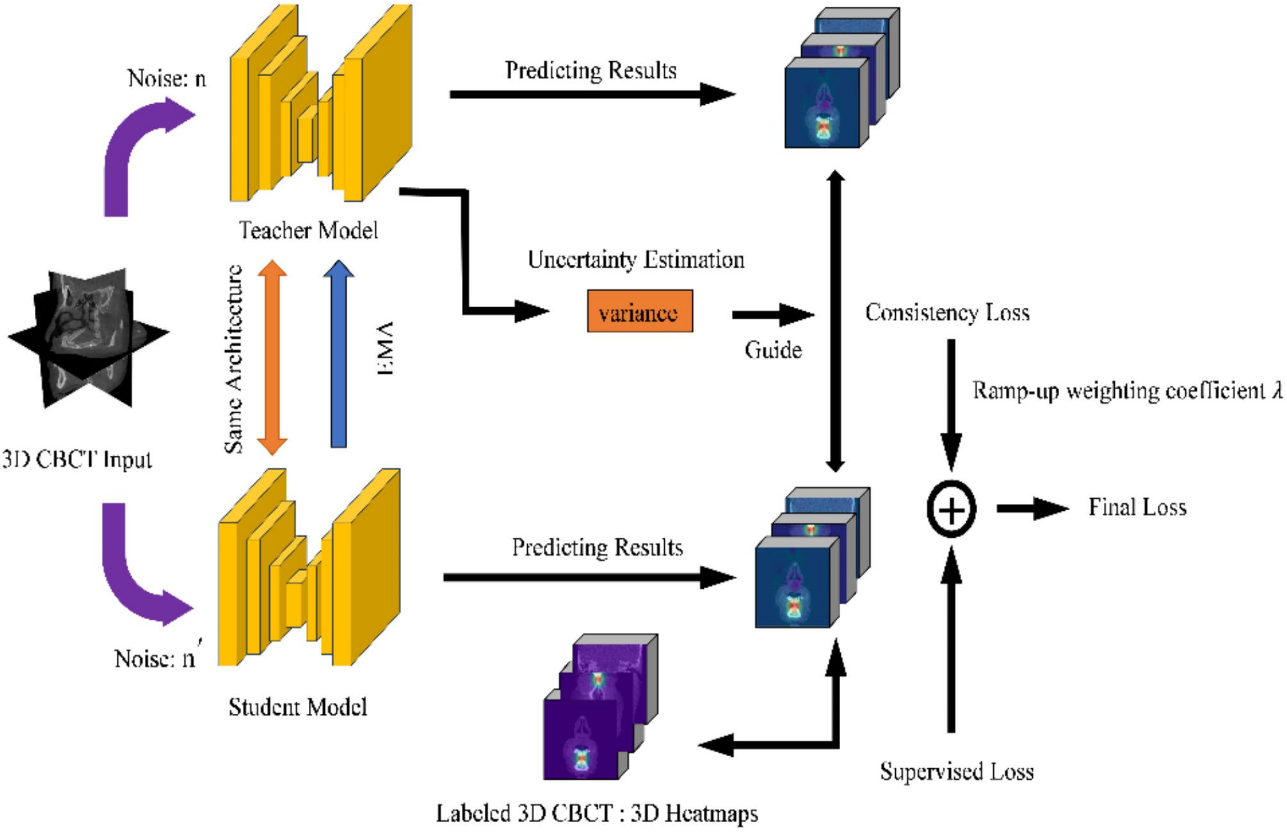


Fig. 3. A network framework for automatic semi-supervised detection of 18 landmarks.

$$\text{Point mean squared error (PMSE)} = \frac{1}{C}\sum_{c=1}^{C}\sqrt{\Delta x^2+\Delta y^2+\Delta z^2} \tag{25}$$

$$\text{Standard Deviation (SD)} = \sqrt{\frac{\sum_{c=1}^{C}(R_c - PME)^2}{C-1}} \tag{26}$$

The successful detection rate (SDR) is used to calculate whether the distance between the predicted landmark and the reference landmark is within the given range, such as 1.5, 2.5, 3, 4.0 mm. If it is within the range, the detection is considered successful. So SDR is equal to the number of successful landmark detections over the total number of detections.

Experimental setup

In this experiment, the NVIDIA RTX 3090 graphics card was used to accelerate the deep learning model to train and test the CBCT image data of this study. The operating system is Windows 10, the data analysis software used is Python3.7.11 and the deep learning framework we used is Pytorch1.10.2. The optimizer used to train the VNet or Mean teacher model selected AdamW[29], the learning rate was set to 0.0001, the weight-decay was set to 0.0001, and the beta parameters were set to 0.9, 0.999. Adjust all image specifications for training input to 132 × 160 × 160. By observing the change curve of the accuracy and loss value of the training set during the training process, the convergence information of the curve was obtained, and the training duration was set to 100 epochs.

Ethical approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. The Ethics Committee of Guiyang Hospital of Stomatology approved this retrospective study and waived the need for informed consent from patients (Approval No. GYSKLL-KY-20231222–01).

Results

Segment results

In this study, we compared the segmentation effect of semi-supervised learning method with that of supervised method, and used five different metrics (DSC, PPV, SEN, ASD, HD) to evaluate the segmentation effect of maxillary and mandible, in Table 2 below, N represents the number of labeled CBCT, and M represents the total number of trained CBCT. It is shown in the table that the semi-supervised learning method has a better overall segmentation effect on maxilla and mandible than the supervised learning method, and the DSC of the semi-supervised learning method in maxilla segmentation is 93.41, 1.62% higher than that of the supervised learning method. The DSC of mandibular segmentation was 96.89, 1.56% higher than that of supervised learning.

As shown in Fig. 4, the segmentation results of the semi-supervised learning method on the 2D CBCT slicers are masked and reconstructed with 3D segmentation in 3Dslicer (version 5.6.1, http://www.slicer.org/). It can be found that the segmentation effect of the bone edge is poor, while the segmentation effect of other bone parts is better, and the overall segmentation effect of the mandible (yellow part) is better than that of the maxilla (green part), which may be because the maxilla bone structure is more complex than that of the mandible.

In order to further test the automatic segmentation effect of our model, that is, to test the degree to which our method predicts that the 3D model deviates from the manual segmentation of the 3D model, We used python (version 3.7.9, https://www.python.org/downloads/) software to draw a color graph of the surface distance from Vnet automatic segmentation to manual segmentation under semi-supervised learning, as shown in (Fig. 5). Figure 5 is a 3D rendered color map of the distance from manual segmentation to model-predicted segmentation for the best DSC case for the maxilla and mandible in test dataset. The inaccurate areas are far away from the ground truth and appear blue, while the accurate areas are closer to the ground truth and appear green. The basket colored areas are mainly in the upper left side of the nostril of the maxilla. The most inferior and posterior region at the angle of the mandible, these regions were separately captured and shown, and were used as the focus of improvement in future research.

Landmark detection results

In this study, we compared the landmarks detection effect of semi-supervised learning (mean teacher) with supervised learning (VNet), and also compared the effect before and after adding uncertainty estimation to semi-supervised learning and without adding uncertainty estimation. Figure 6 below shows the loss curves of

Methods	*N/M* (20%)	Organ	DSC (%)	PPV (%)	SEN (%)	ASD (mm)	HD(mm)
Supervised learning method	38/173	Maxilla	91.79	93.95	93.06	0.7642	3.7623
		Mandible	95.33	94.86	93.26	0.2585	2.4460
Semi-supervised learning method		Maxilla	93.41	94.35	92.08	0.5201	2.0168
		Mandible	96.89	96.62	94.56	0.1870	1.1980

Table 2. Five metrics of supervised and semi-supervised learning methods in maxillary and mandibular segmentation results.

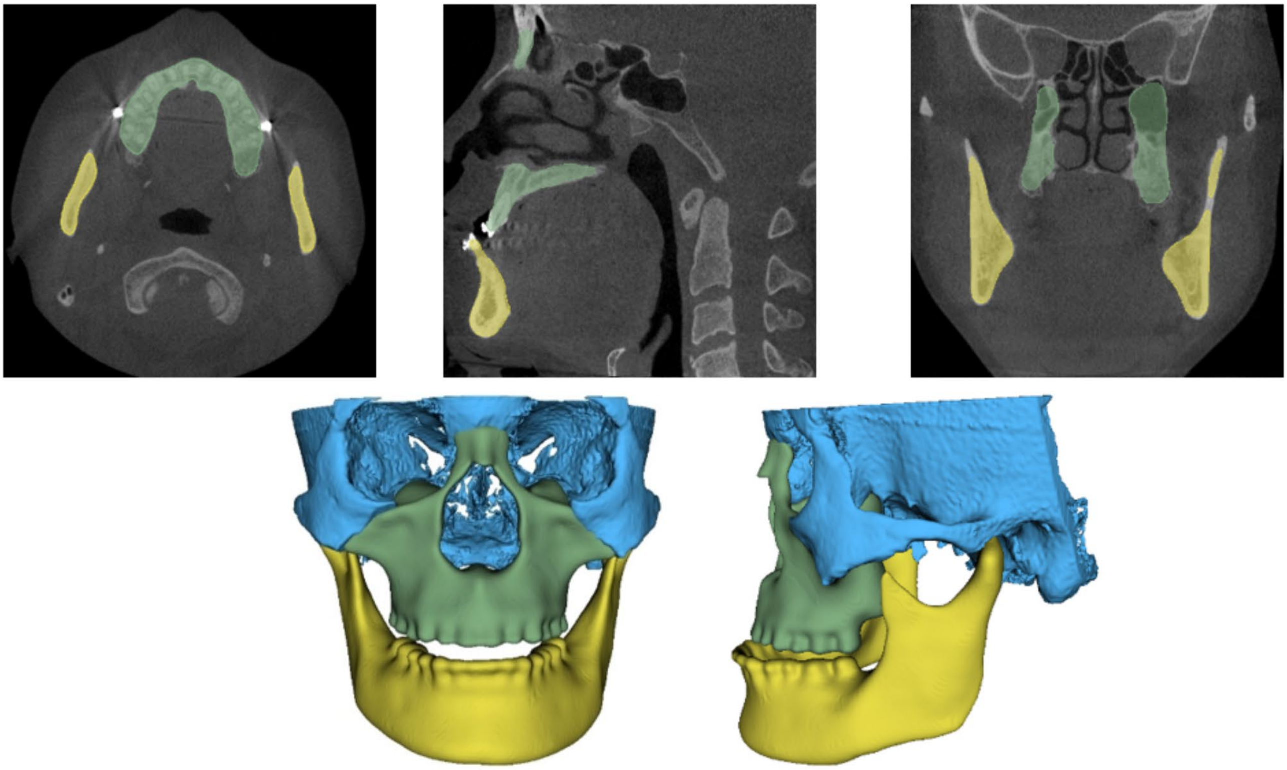


Fig. 4. Under the automatic segmentation results (CBCT slice and 3D reconstruction visualization) of our proposed method, the green bone is the upper jaw bone, the yellow bone is the lower jaw bone, and the blue bone is part of the Cranial base bone.

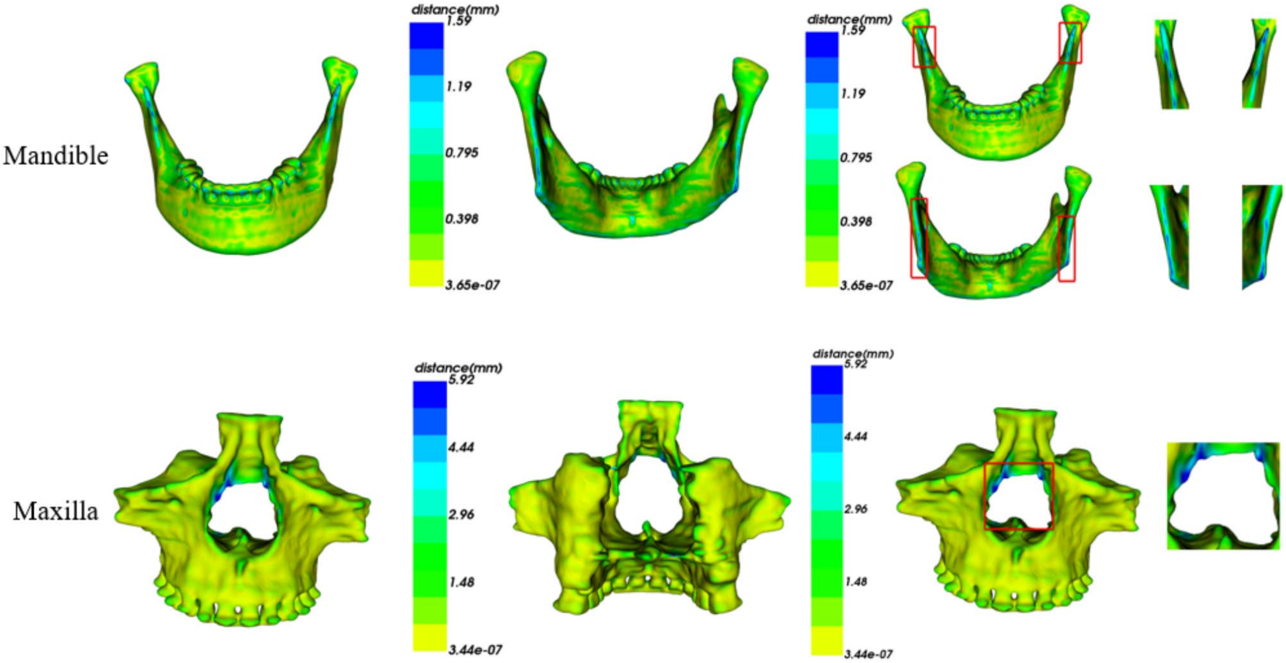
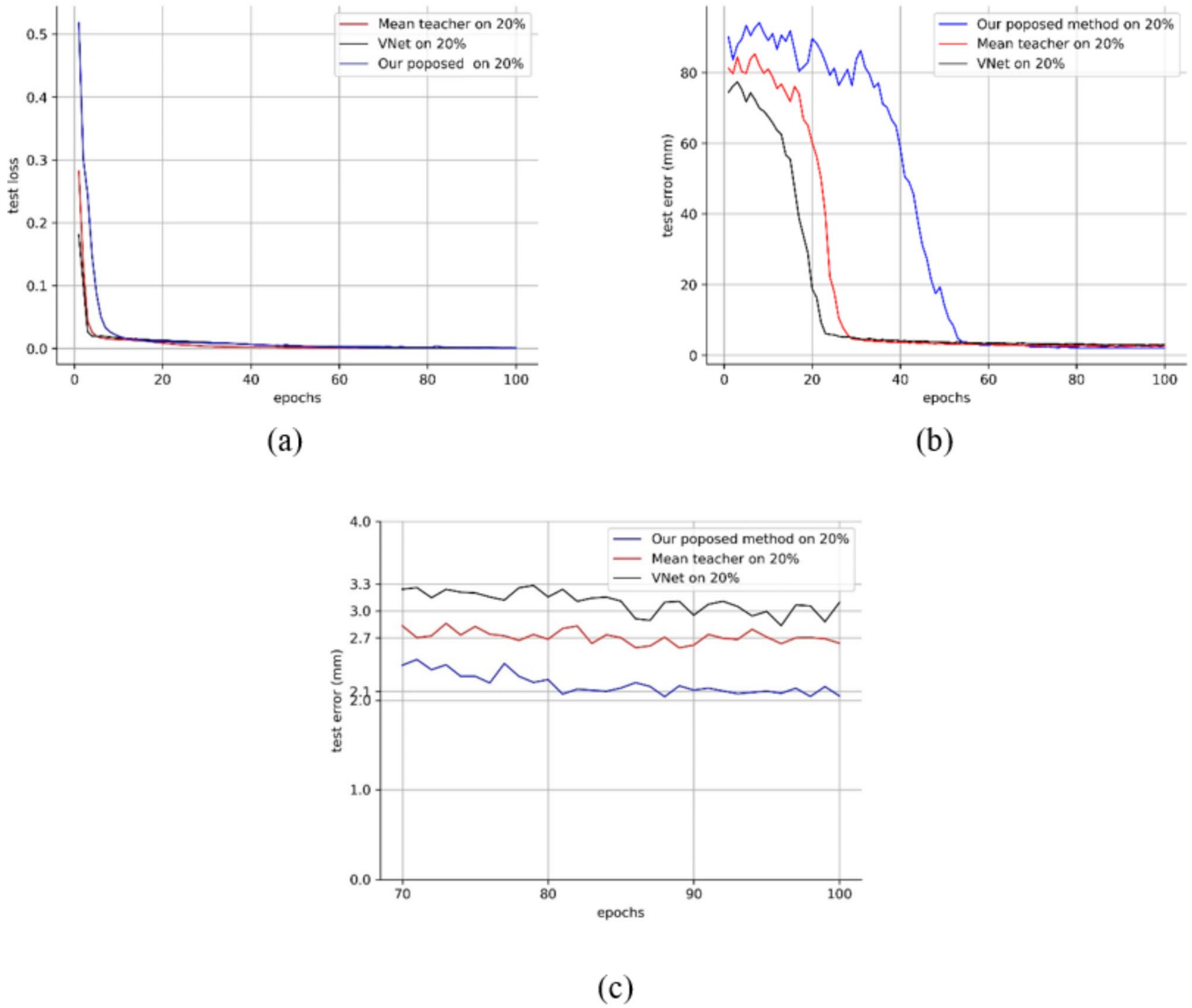


Fig. 5. In the test dataset used in this study, our proposed method outputs a color map of the surface distance between the predicted and ground truth segmentations of the mandible and maxilla, where the red boxed area appears in blue, representing the poorly segmented areas.



(a)

(b)

(c)

Fig. 6. A comparison of of VNet, mean teacher, and our poposed method (the training set is labeled 20%) in terms of convergence speed and landmarking accuracy over 100 epochs. (**a**) Test loss curves for three methods, (**b**) The average test error curves of the three methods, (**c**) Test mean error curves of 70 to 100 epoch.

the three methods trained on this data set, as well as the average tset error change curves of 18 landmarks (here we use global mean error (GME) as test error), (c) in (Fig. 6) is for the convenience of observing the convergence difference of the test error curves of the three methods. As can be seen from the figure, the loss and test error curves of the three methods converge. Although our proposed method has the slowest convergence rate, it works best of the three methods: the average test error for all landmarks is 2.11 mm, and the error of our method is reduced by 0.6 mm relative to the average teacher. At the same time, 0.9 mm lower than VNet.

In order to further demonstrate the fitting effect of the proposed semi-supervised learning model after training, we visualized the 3D heatmap results of the final output of the model. We randomly selected an image of a patient from the test set and visualized the prediction results of the CBCT image of the patient (maxillary retracted). First, the coordinates with the largest pixel value of the output heatmap were found, and then the horizontal, sagittal and coronal plane of the CBCT image were slicing with the coordinate point as the center to obtain the two-dimensional heatmaps of the three faces, as shown in the (Fig. 7) below.

The test set consists of the CBCT data of 19 patients and the real label coordinates corresponding to different positions, which are automatically positioned on 18 different location landmarks for 19 times respectively. The predicted coordinates and the error results of the real labels constitute the box plot (Fig. 8) as follows.

Table 3 shows the performance evaluation of the semi-supervised algorithm in SDR. Ten of the landmarks ((L)Po, (R)Po, N, S, Pog, A, (L)Go, (L)Co, (L)Mn and (R)Mn) were calculated (each landmark is detected 19 times) with more than 50% SDR in the 1.5 mm range, and the average SDR for all landmarks was 47.94%. Table 3 also shows the point mean square error (PMSE) detected 19 times for each landmark, among which the PMSE of most landmarks is below 2 mm and the SD is less than 1.5 mm, and the average PMSE ± SD of all landmarks is 1.908 ± 1.166 mm, and the PMSE (1.397 mm) of (L)Mn is the smallest.

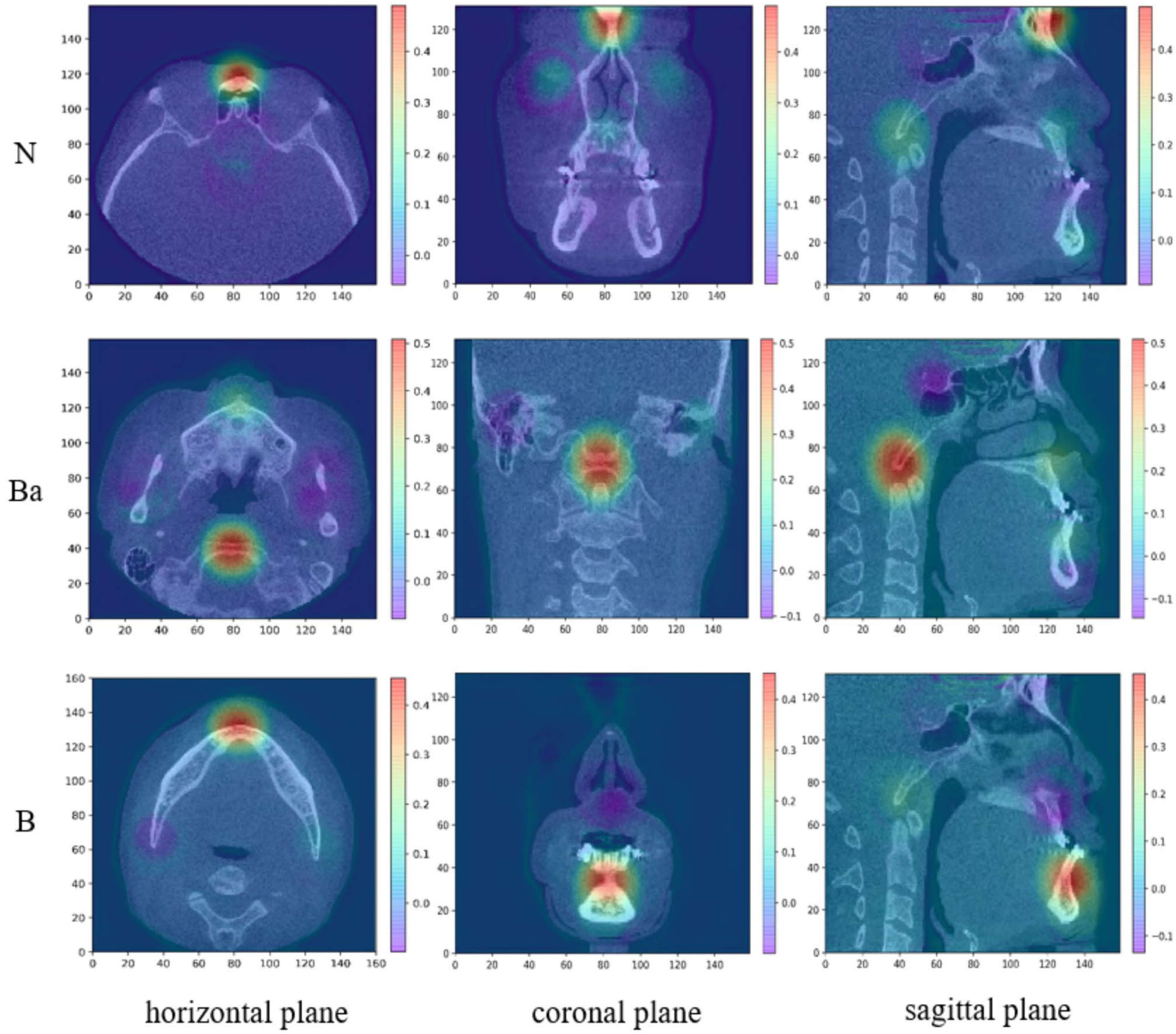


Fig. 7. The three heatmaps (B, N and Ba) in the 18 heatmaps output by the semi-supervised network were visualized in the coronal plane, sagittal plane and horizontal plane respectively.

Brief summary

The purpose of this study is to attempt to use semi-supervised deep learning method to automatically segment masks for 3D reconstruction and 3D location of landmarks for clinical planning of orthognathic surgery. Among them, the 3D segmentation mask of the maxilla and mandible is automatically generated, and the anatomic landmarks located in the skull, maxilla and mandible are automatically detected. It should be emphasized that the method of this study used a small amount of labeled training data, and the Dice of segmentation of the maxilla was 93.41% and the Dice of the mandible was 96.89%. The average PMSE of 18 landmarks was 1.908 mm, which reached the clinically acceptable accuracy standard.

Discussion

In this paper, a semi-supervised learning-based 3D CBCT method for automatic maxillary and mandible segmentation and landmark detection is proposed. This method not only provides accurate and reliable 3D reconstruction of bone structure and landmark measurement, but also reduces the workload of clinicians and provides morphometric guidance for preoperative treatment planning of **orthognathic** surgery. In addition, it also improves the drawback that deep learning methods require a large amount of manual data annotation in the early stage (serious lack of training CBCT data will lead to the problem of overfitting the model), freeing clinicians' energy and time in CBCT data annotation.

In this work, we use the consistency constraint method of semi-supervised learning ((Uncertainty-Aware Mean Teacher (UA-MT)), which aims to solve the problem of insufficient training of CBCT data by using many unlabeled segmentation and landmark data. Compared with the supervised deep learning algorithm,

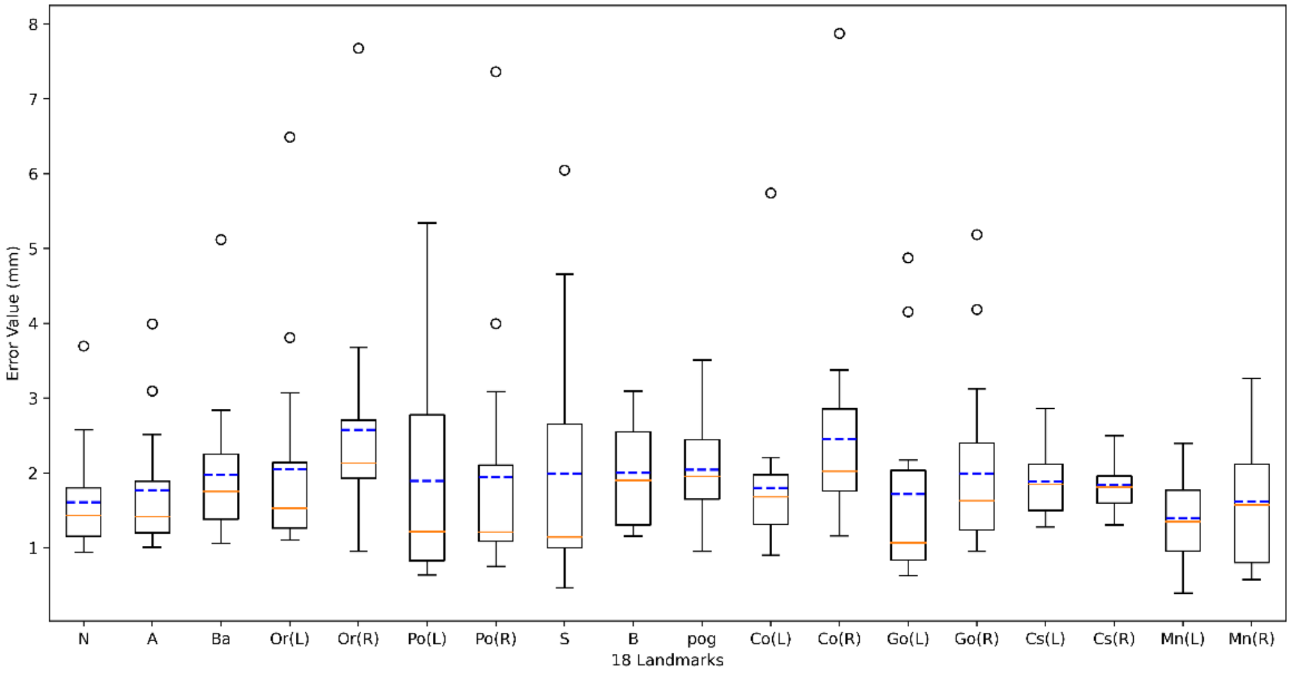


Fig. 8. The error of each landmarker is tested 19 times using our semi-supervised learning method. The orange line is the median error, and the blue dotted line is themean error. The circles are outliers.

Landmarks	SDR(%)				PMSE ± SD (mm)
	1.5 mm	2.5 mm	3.5 mm	4.5 mm	
(L)Or	42.10	84.21	89.47	94.73	2.051 ± 1.645
(R)Or	31.57	78.94	89.47	94.73	2.526 ± 1.788
(L)Po	63.15	68.42	73.68	100	1.892 ± 2.311
(R)Po	57.89	63.15	78.94	94.73	1.940 ± 2.357
Ba	42.10	73.68	94.73	94.73	1.974 ± 0.876
N	57.89	84.21	94.73	100	1.608 ± 0.451
S	52.63	63.15	84.21	100	1.993 ± 2.376
B	42.10	78.94	100	100	1.998 ± 0.462
Pog	57.89	73.68	100	100	2.040 ± 0.375
A	52.63	89.47	100	100	1.770 ± 0.656
(L)Go	57.89	78.94	89.47	94.73	1.715 ± 1.855
(R)Go	47.36	84.21	89.47	94.73	1.993 ± 1.230
(R)Cs	21.05	94.73	100	100	1.836 ± 0.119
(L)Cs	26.31	89.47	94.73	100	1.882 ± 0.233
(L)Co	52.63	84.21	89.47	94.73	1.801 ± 1.017
(R)Co	15.78	52.63	68.42	94.73	2.310 ± 2.086
(L)Mn	73.68	89.47	100	100	1.397 ± 0.360
(R)Mn	52.63	73.68	100	100	1.616 ± 0.782
Average	47.07	78.06	90.93	97.64	1.908 ± 1.166

Table 3. The success detection rate (SDR) and the detected PMSE ± SD of 18 landmarks under semi-supervised learning method.

this method can solve the problem of **inadequate** training of CBCT data. The semi-supervised algorithm significantly improved the results of segmentation and landmark detection. The dice of maxillary and mandibular segmentation tasks reached 93.41 and 96.89%, and the average error of landmarks detection tasks reached 1.908 ± 1.166 mm. In the landmark detection task, the heatmap regression method is used. In this method, we propose a calculation method for the coordinate uncertainty predicted by the TM and use the calculated value as the network learning optimization object (added to the loss function), so as to guide the network to learn in a more certain direction. Experimental results show that the proposed method (adding uncertainty learning) reduces the average error of landmark fixed point by 0.4 mm. It is believed that this article will interest a wide range of readers, including dentists, radiologists, computer scientists, and researchers in other related fields, and provide greater assistance to patient treatment.

However, the authors of this paper realize that the calculation of uncertainty added to the model processing in this paper leads to slower image processing, and the method has not been validated in a large number of clinical trials. Therefore, in the future research, we will continuously improve the calculation method of uncertainty, reduce the computational complexity of uncertainty estimation and shorten the calculation time while maintaining the prediction accuracy of the model. At the same time, we will collect more data in a clinical setting to verify the performance of the model.

Limitations of the study

The first limitation of this study is the small sample size of the data set collected. Our data came from only one medical institution, Guiyang Stomatological Hospital. Deep neural networks with high parameters are highly dependent on the amount of available training data, and their performance usually improves with the increase in the number of data points. However, in this study, our data set collection was limited due to various factors (The money and time cost of obtaining data, the difficulty of sharing data across hospitals, and the reluctance of subjects to participate in the survey). The second limitation is that the errors of some landmarks automatically located in this study did not meet clinical requirements, perhaps because we did not further improve the model (VNet). However, the proposed method can still help clinicians quickly determine the approximate location of landmarks, thus simplifying their work.

Data availability

Due to protect patient privacy, the generated and analysed datasets during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Received: 17 August 2024; Accepted: 6 March 2025
Published online: 14 March 2025

References

1. Cao, R. K., Li, L. S. & Cao, Y. J. Application of three-dimensional technology in orthognathic surgery: a narrative review. *Eur. Rev. Med. Pharmacol. Sci.* **26** (21), 7858–7865 (2022).
2. Hsu, S. S. et al. Accuracy of a computer-aided surgical simulation protocol for orthognathic surgery: a prospective multicenter study. *J. Oral Maxillofac. Surg. Off. J. Am. Assoc. Oral Maxillofac. Surg.* **71** (1), 128–142 (2013).
3. Alhammadi, M. S. et al. Orthodontic camouflage versus orthodontic-orthognathic surgical treatment in borderline class III malocclusion: a systematic review. *Clin. Oral Invest.* **26** (11), 6443–6455 (2022).
4. Cao, H. L. et al. Quantification of three-dimensional facial asymmetry for diagnosis and postoperative evaluation of orthognathic surgery. *Maxillofac. Plast. Reconstr. Surg.* **42** (1), 17 (2020).
5. Zhang, J. et al. Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization. *Med. Image. Anal.* **60**, 101621 (2020).
6. Liu, Q. et al. SkullEngine: A multi-stage CNN framework for collaborative CBCT image segmentation and landmark detection. *Mach. Learn. Med. Imaging MLMI (Workshop)* **12966**, 606–614 (2021).
7. Lian, C. et al. Multi-task dynamic transformer network for concurrent bone segmentation and large-scale landmark localization with dental CBCT. *Medical image computing and computer-assisted intervention: MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention* **12264** 807–816. (2020).
8. Zhang, R. et al. Craniomaxillofacial bone segmentation and landmark detection using semantic segmentation networks and an unbiased heatmap. *IEEE J. Biomed. Health Inf.* (2023).
9. Tarvainen, A. & Valpola, H. J. A. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *ArXiv* **30**, ArXiv (2017).
10. Laine, S. & Aila, T. J. Temporal ensembling for semi-supervised learning. *arXiv* 1610.02242. (2016).
11. Milletari, F., Navab, N. & Ahmadi, S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *fourth international conference on 3D vision (3DV)* 565–571 (IEEE, 2016).
12. Toshev, A. & Szegedy, C. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1653–1660 (2014).
13. Nishimoto, S. et al. Three-dimensional craniofacial landmark detection in series of CT slices using multi-phased regression networks. *Diagn. (Basel Switzerland)* **13** (11). (2023).
14. Payer, C., Štern, D., Bischof, H. & Urschler, M. Integrating Spatial configuration into heatmap regression based CNNs for landmark localization. *Med. Image. Anal.* **54**, 207–219 (2019).
15. Lu, G. et al. CMF-Net: craniomaxillofacial landmark localization on CBCT images using geometric constraint and transformer. *Phys. Med. Biol.* **68** (9). (2023).
16. Fedorov, A. et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imaging* **30** (9), 1323–1341 (2012).
17. Gillot, M. et al. Automatic landmark identification in cone-beam computed tomography. *Orthod. Craniofac. Res.* **26** (4), 560–567 (2023).
18. Cheng, M. et al. Prediction of orthognathic surgery plan from 3D cephalometric analysis via deep learning. *BMC Oral Health*. **23** (1), 161 (2023).
19. Beare, R., Lowekamp, B. & Yaniv, Z. Image segmentation, registration and characterization in R with simpleitk. *J. Stat. Softw.* 86. (2018).
20. Sharma, S. P. & Sampath, N. Data Augmentation for brain tumor segmentation using MONAI Framework. In *2nd International Conference on Intelligent Technologies (CONIT)* 1–8 (IEEE, 2022).
21. Yu, L., Wang, S., Li, X., Fu, C-W. & Heng, P. A. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In *Medical image computing and computer assisted intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, Proceedings, Part Ii 22* 605–613 (Springer, 2019).
22. Tong, N., Gou, S., Yang, S., Ruan, D. & Sheng, K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med. Phys.* **45** (10), 4558–4567 (2018).

23. Verhelst, P. J. et al. Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography. *J. Dent.* **114**, 103786 (2021).
24. Kim, Y. H., Lee, C., Ha, E. G., Choi, Y. J. & Han, S. S. A fully deep learning model for the automatic identification of cephalometric landmarks. *Imaging Sci. Dent.* **51** (3), 299–306 (2021).
25. Tang, P., Ramaiah, C., Wang, Y., Xu, R. & Xiong, C. Proposal learning for semi-supervised object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 2291–2301 (2021).
26. Xu, M. et al. End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF international conference on computer vision* 3060–3069 (2021).
27. Springstein, M., Schneider, S., Althaus, C. & Ewerth, R. Semi-supervised human pose Estimation in art-historical images. *arXiv* 02976 (2022).
28. Wang, X., Bo, L. & Fuxin, L. Adaptive wing loss for robust face alignment via heatmap regression. In *Proceedings of the IEEE/CVF international conference on computer vision* 6971–6981 (2019).
29. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *ArXiv* 05101 (2017).

Acknowledgements

The authors thank the staff of the Radiology Department and Oral Orthognathic Surgery Department of Guiyang Sto-matological Hospital for their support and help, as well as the Science and Technology Bureau of Guanshanhu District (2021-01) for their funding.

Author contributions

Conceptualization, H.T. and S.L. Methodology: H.T. and S.L. Investigation: H.T, Y.S, J.W and H.F. Formal analysis: H.T , J.W and J.P. Resources: Y.S and H.F. Data curation: Y.S, J.W and J.P. Writing—original draft: H.T. Writing—review and editing: H.T., S.L, Y.S, J.W, J.P and H.F. Visualization: H.T and Y.S. Supervision: J.W and H.F. Funding acquisition: H.F. All authors read and approved the final manuscript.

Funding

This research is supported by Guizhou Provincial Health Commission Science and Technology fund project (gzwkj2025-102) and the Seedling Program of Guiyang Stomatological Hospital.

Declarations

Competing interests

The authors declare no competing interests.

Ethical statement

The Ethics Committee of Guiyang Hospital of Stomatology approved this retrospective study and waived the need for informed consent from patients (Approval No. GYSKLL-KY-20240716–01).

Additional information

Correspondence and requests for materials should be addressed to H.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.