Heliyon 10 (2024) e34583

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Craniomaxillofacial landmarks detection in CT scans with limited labeled data via semi-supervised learning

Leran Tao [a,b,c,1], Xu Zhang [d,1], Yang Yang [e,f], Mengjia Cheng [a,b,c], Rongbin Zhang [g], Hongjun Qian [h], Yaofeng Wen [e,**], Hongbo Yu [a,b,c,*]

[a] Department of Oral and Cranio-Maxillofacial Surgery, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200011, China
[b] National Center for Stomatology & National Clinical Research Center for Oral Diseases, Shanghai, 200011, China
[c] Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology, Shanghai, 200011, China
[d] Mechanical College, Shanghai Dianji University, Shanghai, 201306, China
[e] Shanghai Lanhui Medical Technology Co., Ltd, Shanghai, 200333, China
[f] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200030, China
[g] College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200125, China
[h] Appleby College, ON, L6K 3P1, Canada

ARTICLE INFO

Keywords:
Semi-supervised learning
Landmarks detection
3D cephalometric analysis
Computer-assisted surgery design
Dentomaxillofacial deformities

ABSTRACT

Background: Three-dimensional cephalometric analysis is crucial in craniomaxillofacial assessment, with landmarks detection in craniomaxillofacial (CMF) CT scans being a key component. However, creating robust deep learning models for this task typically requires extensive CMF CT datasets annotated by experienced medical professionals, a process that is time-consuming and labor-intensive. Conversely, acquiring large volume of unlabeled CMF CT data is relatively straightforward. Thus, semi-supervised learning (SSL), leveraging limited labeled data supplemented by sufficient unlabeled dataset, could be a viable solution to this challenge.

Method: We developed an SSL model, named CephaloMatch, based on a strong-weak perturbation consistency framework. The proposed SSL model incorporates a head position rectification technique through coarse detection to enhance consistency between labeled and unlabeled datasets and a multilayers perturbation method which is employed to expand the perturbation space. The proposed SSL model was assessed using 362 CMF CT scans, divided into a training set (60 scans), a validation set (14 scans), and an unlabeled set (288 scans).

Result: The proposed SSL model attained a detection error of 1.60 ± 0.87 mm, significantly surpassing the performance of conventional fully supervised learning model (1.94 ± 1.12 mm). Notably, the proposed SSL model achieved equivalent detection accuracy (1.91 ± 1.00 mm) with only half the labeled dataset, compared to the fully supervised learning model.

Conclusions: The proposed SSL model demonstrated exceptional performance in landmarks detection using a limited labeled CMF CT dataset, significantly reducing the workload of medical professionals and enhances the accuracy of 3D cephalometric analysis.

* Corresponding author. Department of Oral and Cranio-Maxillofacial Surgery, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200011, China.
** Corresponding author.
E-mail addresses: ericwene@sjtu.edu.cn (Y. Wen), yhb3508@163.com (H. Yu).
[1] Leran Tao and Xu Zhang have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e34583
Received 16 February 2024; Received in revised form 21 May 2024; Accepted 11 July 2024
Available online 16 July 2024

1. Introduction

Cephalometric analysis [1], primarily employed in orthognathic and orthodontic treatments, traditionally relies on cephalometric lateral radiographs, which overlook stereoscopic and coronal details. Advancements in computing and nuclear medicine have made three-dimensional cephalometric analysis [2] using craniomaxillofacial (CMF) CT/CBCT an increasingly viable method for delivering precise, comprehensive insights into patient deformities. However, digitizing landmarks in this 3D context is challenging, impeding its broader clinical adoption.

In recent years, the emergence of deep supervised learning provided us chances to solve this problem [3–8]. Nevertheless, the development of robust deep supervised learning models for CMF CT landmarks detection demands extensive, high-quality labeled data and incurs substantial costs, particularly in the medical field. In contrast, amassing a large volume of unlabeled CMF CT images is relatively cost-effective. Thus, semi-supervised learning (SSL) [9], which combines limited labeled data with ample unlabeled data, becomes an available technique for reducing training expenses and efficient landmarks detection.

Contemporary SSL methodologies predominantly concentrate on consistency regularization [10–12] and proxy-label [13,14] techniques. By analyzing connections among unlabeled data, these methods can mitigate the limitation in the volume of labeled data, aiding in the development and refinement of deep learning models. However, variations in patient head positioning during CT scans and the consequent discrepancies between labeled and unlabeled data have hindered SSL from achieving optimal results in CMF CT landmark detection.

The state-of-the-art SSL model in nature image processing was proposed by Yang et al. [15], named Unimatch, where a unified dual-stream perturbation method that guides the outputs of strongly perturbed images with their weak counterparts was developed and dropout processes between the encoder and decoder in the baseline model (DeepLabv3 [16]) was introduced to expand the perturbation space. In the field of CMF CT landmarks detection, complex models like DeepLabv3 are less effective given that the data is always scarce and frameworks like U-Net [17] and its variants [18–20] are widely favored due to their simple structure and efficient performance with limited data. However, the multilayers skip connections in U-Net structure mean that dropout processes, as proposed in UniMatch [15] and limited only between the encoder and decoder, are insufficient to generate an adequate perturbation space.

To address the issues mentioned above, we therefore proposed an SSL model named CephaloMatch for efficient CMF CT landmarks detection with limited labeled data. The proposed SSL model introduced a head position rectification method through Frankfurt horizontal plane (FHP) and midsagittal plane (MSP) adjustments based on coarse landmarks detection, ensuring uniformity in head positions across labeled and unlabeled CMF CT datasets. Moreover, a multilayers perturbation pattern tailored to the U-Net

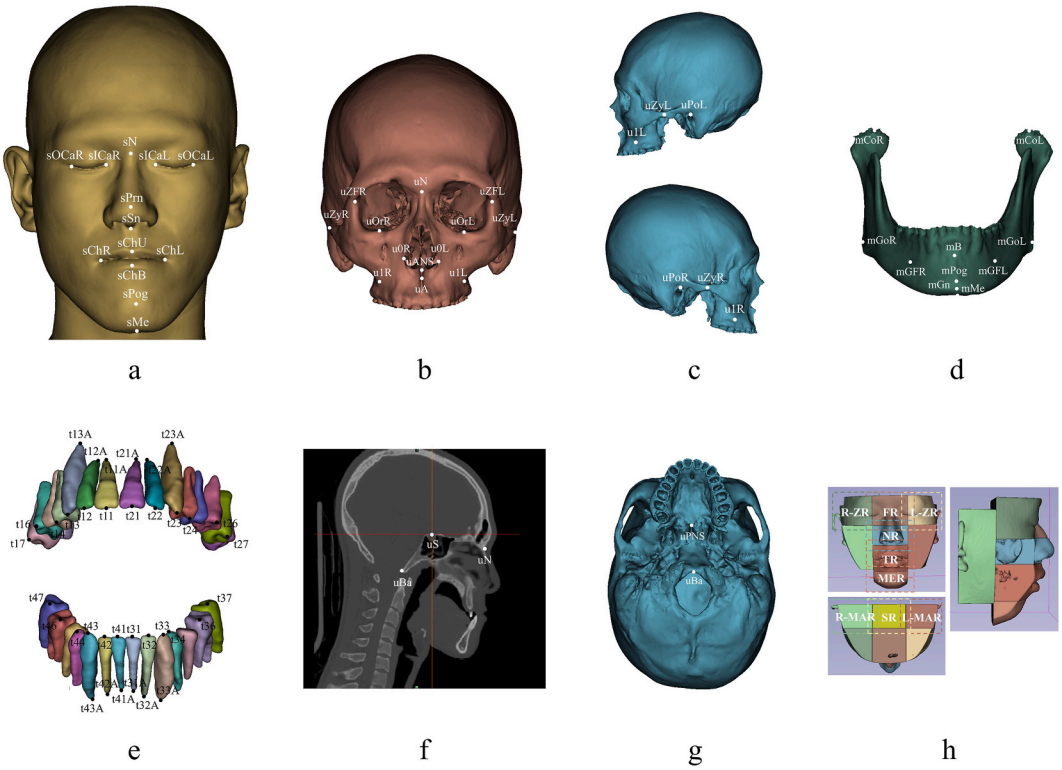

Fig. 1. 77 CMF landmarks and 9 regions. a. 13 facial soft tissue landmarks. b-f. 28 skeletal landmarks. g. 36 dental landmarks. h. 9 divided regions (Abbreviation: R-ZR: Right zygomatic region; L-ZR: Left zygomatic region; FR: Frontal region; NR: Nasal region; TR: Teeth region; MER: Mental region; R-MAR: Right mandibular region; L-MAR: Left mandibular region; SR: Sphenoid region).

Table 1
Baseline characteristics of patients in three divided datasets.

Characteristics	Labeled set (74)		Unlabeled set (288)
	Training set (60)	Validation set (14)	
Age			
In year	22.1 ± 3.2	22.3 ± 4.1	23.6 ± 4.7
Gender – No. (%)			
Male	27 (45 %)	6 (42.9 %)	102 (35.4 %)
Female	33 (55 %)	8 (57.1 %),	186 (64.6 %)
Skeletal Classification – No. (%)			
I	15 (25 %)	4 (28.6 %)	58 (20.1 %)
II	7 (11.7 %)	1 (7.1 %)	54 (18.8 %)
III	38 (63.3 %)	9 (64.3 %)	176 (61.1 %)

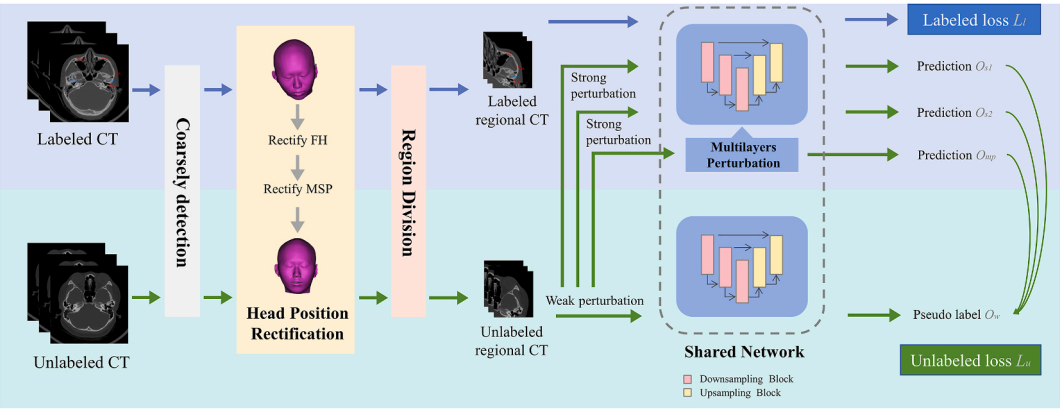

Fig. 2. The overall architecture of the proposed semi-supervised learning model.

architecture was designed to implement dropout processes not only between the encoder and decoder but also after each skip connection across multiple layers, enhancing the perturbation space in strong-weak perturbation consistency.

2. Materials and methods

2.1. Data collection and preprocessing

For our investigation, 362 craniomaxillofacial (CMF) CT scans of patients with dentomaxillofacial deformities were retrospectively collected from Shanghai Ninth People's Hospital, Shanghai, China, from 2015 to 2022. The inclusion criteria were: 1) patients diagnosed with dentomaxillofacial deformities and orthognathic-orthodontic joint treatment were required; 2) CMF CT scanned before treatment. The exclusion criteria were: 1) patients diagnosed with congenital dentofacial deformities; 2) have a history of orthognathic treatment.

Each scan featured a pixel size of 0.45 mm × 0.45 mm, a slice interval of 1 mm, and a resolution of 512 × 512 × 231. To optimize the computational process and minimize the graphics memory usage, CT images were resampled. In resizing, the pixel dimensions were adjusted to 1 mm × 1 mm, resulting in a revised resolution of 229 × 229 × 231 for each CT scan.

77 landmarks were designated for detection, as demonstrated in Fig. 1a–g, which is based on clinical requirements and researches with regard to 3D cephalometric analysis [21–23]. A subset of 74 CMF CT sets underwent manual digitization by two junior CMF surgeons and subsequent review by a senior CMF surgeon. These scans constituted our training and validation sets. The remaining 288 CMF CT scans, which were not labeled, were allocated for use in the semi-supervised learning (SSL) training process, forming our unlabeled dataset. Baseline characteristics of these three datasets is displayed in Table 1, including age, gender, and skeletal classification. To validate the reliability of manually digitization, we invited two senior surgeons manually digitize all landmarks in 10 same CMF CT scans and one senior surgeon repeat digitized these 10 CMF CT scans after one week. The inter-observer variation was 1.27 ± 0.70 mm and intra-observer variation was 1.01 ± 0.74 mm. The intraclass correlation coefficient of two observers was greater than 0.99 [24].

2.2. Model architecture

The labeled and unlabeled CMF CT scans formed our overall dataset. Initially, a coarse model based on 3D U-Net was pre-trained to

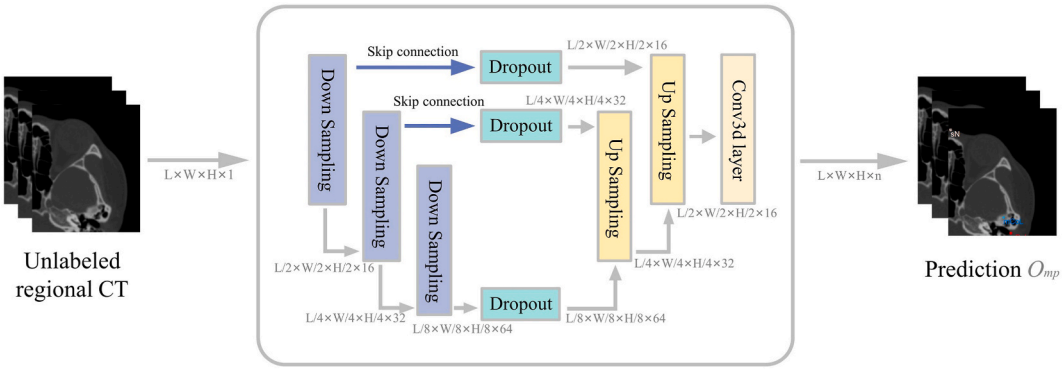

Fig. 3. The illustration of multilayers perturbation based on 3D U-Net.

approximately detect coarse landmarks based on downscaled CMF CT scans with a resolution of 96 × 96 × 96. Following this, a head position rectification method was introduced to minimize disparities between the labeled and unlabeled data using the coarse landmarks identified in the earlier stage.

To address the high computational demands associated with the original high-resolution CMF CT scans, a region division pattern was implemented to divide the CMF CT scans into nine distinct regions [24]. (Fig. 1h) Subsequently, both the labeled and unlabeled regional CT scans were processed through a shared network to complete the multilayers perturbation-based semi-supervised training phase (Fig. 2).

There are three key components of this SSL model: head position rectification, supervised learning section and multilayers perturbation pattern.

2.2.1. Head position rectification for data consistency

In this period, a head position rectification approach was proposed to standardize head positions in CT scans of different patients, which enhances the consistency of overall data.

Firstly, the Frankfort horizontal plane (FHP), defined by points uPoR (right ear point), uPoL (left ear point), and the midpoint of uOrR (right infraorbital point) and uOrL (left infraorbital point), was aligned parallel to the horizontal xOy plane in the three-dimensional coordinate system.

Subsequently, the midsagittal plane (MSP), defined by points uN (nasal root point), uS (sella turcica center point), and uBa (anterior skull base point), was aligned parallel to the yOz plane. Due to the limitations of geometric relation, only two points (uS and uBa) were adopted and the line connecting uBa and uS was aligned parallel to the sagittal plane.

To achieve these alignments, a rotation matrix technique was employed. The normal vector for the FHP was calculated using the coordinates of designated landmarks. Then, the angle between this vector and the z-axis was determined, defining the axis of rotation. Utilizing the Rodriguez formula, a rotation matrix (R_FHP) was derived to align the FHP with the horizontal plane. Similarly, a rotation matrix (R_MSP) was calculated to align the MSP with the sagittal plane. Finally, the composite rotation matrix was obtained by combining R_FHP and R_MSP. Applying this matrix to the original CT dataset yields the corrected sample.

2.2.2. Supervised learning section with labeled CMF CT

Based on our previous research [24], two different landmarks detection training strategies were adopted for constructing the supervised learning section. For regions with dense landmarks, like teeth region (TR), a gaussian heatmap-based strategy [25] was utilized to avoid confusion over proximity landmarks. For regions with sparse landmarks, a random mask-based strategy [5] was employed to enhance linkages between remote landmarks.

2.2.3. Multilayers perturbation pattern in strong-weak perturbation consistency (Fig. 3)

In the original UniMatch model [15], dropout processing is strategically situated between the encoder and decoder to broaden the perturbation space, which is crucial for maintaining strong-weak perturbation consistency. This study fused this technique with U-Net, a prevalent model in CMF CT landmarks detection, known for its skip connection architecture designed to prevent information loss and gradient vanishing. By integrating dropout processing subsequent to skip connections in various U-Net layers, a larger perturbation space was generated, thereby augmenting the robustness of the SSL model.

During this process, an unlabeled regional CT scan was input into a Down Sampling Block. Then, two streams of output data were generated. One was directed to an Up Sampling Block via a Skip Connection and augmented with Dropout processing; another was fed into the Down Sampling. Following the last Down Sampling Block, inclusive of Dropout Processing, the data was transitioned to an Up Sampling Block. It was then concatenated with direct outputs from the corresponding Down Sampling Blocks.

Table 2
Comparing with conventional supervised learning model by metrics of error(mm), standard deviation(mm), detection failure rate (%) in nine divided regions.

Method	R-ZR (4)		L-ZR (4)		FR (4)		NR (8)		TR (40)		MER (8)		R-MAR (3)		L-MAR (3)		**SR (3)**		**Overall (77)**	
	Error (std)	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**
Supervised learning	1.65 (0.64)	1.786	1.90 (0.96)	1.786	1.71 (0.58)	5.357	1.90 (1.20)	0	2.06 (1.34)	2.857	2.04 (1.09)	8.929	1.96 (0.68)	2.381	**1.63 (0.66)**	4.762	1.22 (0.60)	0	1.94 (1.12)	3.154
CephaloMatch (ours)	**1.46 (0.64)**	3.571	**1.59 (0.74)**	1.786	**1.53 (0.54)**	3.571	**1.62 (1.15)**	0	**1.60 (0.91)**	0.536	**1.79 (1.00)**	7.143	**1.70 (0.70)**	2.381	1.67 (0.71)	4.762	**1.14 (0.47)**	0	**1.60 (0.87)**	1.762

2.3. Evaluation metrics

To assess the efficacy of the SSL model in landmarks detection, three key metrics were employed: mean prediction error (Eq. (1)), mean prediction standard deviation (Eq. (2)), and detection failure rate (Eq. (3)). This tripartite evaluative approach offered a comprehensive analysis of the model's precision and reliability in identifying landmarks, with the mean prediction error set as the principal metric.

Mean Prediction Error (l_e): This metric calculates the average Euclidean distance between the predicted coordinates (x_{pre}, y_{pre}, z_{pre}) and the ground truth coordinates (x_{gt}, y_{gt}, z_{gt}). It is defined as:

$$l_e = \frac{1}{n}\sum_{i=1}^{n}\left[\left(x_{pre}^i - x_{gt}^i\right)^2 + \left(y_{pre}^i - y_{gt}^i\right)^2 + \left(z_{pre}^i - z_{gt}^i\right)^2\right]^{\frac{1}{2}} \quad \text{(Eq. 1)}$$

Mean Prediction Standard Deviation (s_e): This metric measures the variability of the prediction error across all samples. It is defined as:

$$s_e = \left[\frac{1}{n}\sum_{i=1}^{n}(l_i - l_e)^2\right]^{\frac{1}{2}} \quad \text{(Eq. 2)}$$

where l_i represents the prediction error for a given sample.

Detection Failure Rate (f_d): This metric calculates the percentage of samples where the prediction error exceeds a predefined threshold (8 mm in this study) or where the landmarks were not detected. It is defined as:

$$f_d = (n - x)/n \times 100\% \quad \text{(Eq. 3)}$$

where x is the number of detection failure samples, calculated as:

$$x = \sum_{i=1}^{n} 1 \; if \; \left(l_i > 8\ mm \quad or \quad \left(x_{pre}, y_{pre}, z_{pre}\right) \; was \; missing\right) \quad (4)$$

Failure prediction landmarks were not calculated in l_e and s_e.

2.4. Implement detail

The labeled dataset was divided into training dataset (60 samples) and validation set (14 samples), and the unlabeled dataset included 288 samples. All samples were downsampled to make sure each voxel was 1mm × 1mm × 1 mm and divided into 9 regions where the 9 corresponding detection models were developed separately. Each model was validated every 10 epochs, and the best validation result would be taken as final result. The training detail was set as follows: Backbone: 3D U-Net, optimizer: Adam, loss function: focal loss (labeled loss computation) and l_1 loss (unlabeled loss computation), learning rate: 0.0001 (teeth region: 0.001), epochs: 500, data augmentation: randomly shift [−10, 10] in three dimensions, batch size of training dataset: 2, batch size of unlabeled dataset: 1, weak perturbation: randomly shift [−10, 10] in three dimensions, strong perturbation: CutMix [26], dropout level: 0.5. All models were implemented in PyTorch and trained on an NVIDIA Tesla A100.

To compare the landmarks detection performance and labeled data dependency of proposed SSL model with conventional supervised learning model, two comparative experiments were conducted, where one was under the same amount of labeled data and another was under different amount. Furthermore, in order to demonstrate novelty of the proposed SSL model, state-of-the-art SSL model from natural image processing [11,14,15] were adapted to this task and compared to our proposed method. Ablation experiment on the proposed head position rectification method and multilayers perturbation pattern was also conducted to evaluate their effectiveness.

This research was approved by the Research Ethics Committee in Shanghai Ninth People's Hospital (IRB No. SH9H-2022-TK12-1) on February 8, 2022. All methods were carried out in accordance with relevant institutional guidelines and regulations. All patients are aware and have given their written informed consent.

3. Results

3.1. Comparing with conventional supervised learning model

3.1.1. Conventional supervised learning

Using 60 samples as training set and 14 samples as validation set, the supervised learning method had a mean error of 1.94 ± 1.12 mm and a failure rate of 3.154 % (Table 2).

3.1.2. The proposed SSL model (CephaloMatch)

Using 60 samples as training set, 14 samples as validation set and 288 samples as unlabeled set, the proposed SSL model (CephaloMatch) had a mean error of 1.60 ± 0.87 mm and a failure rate of 1.762 % (Table 2).

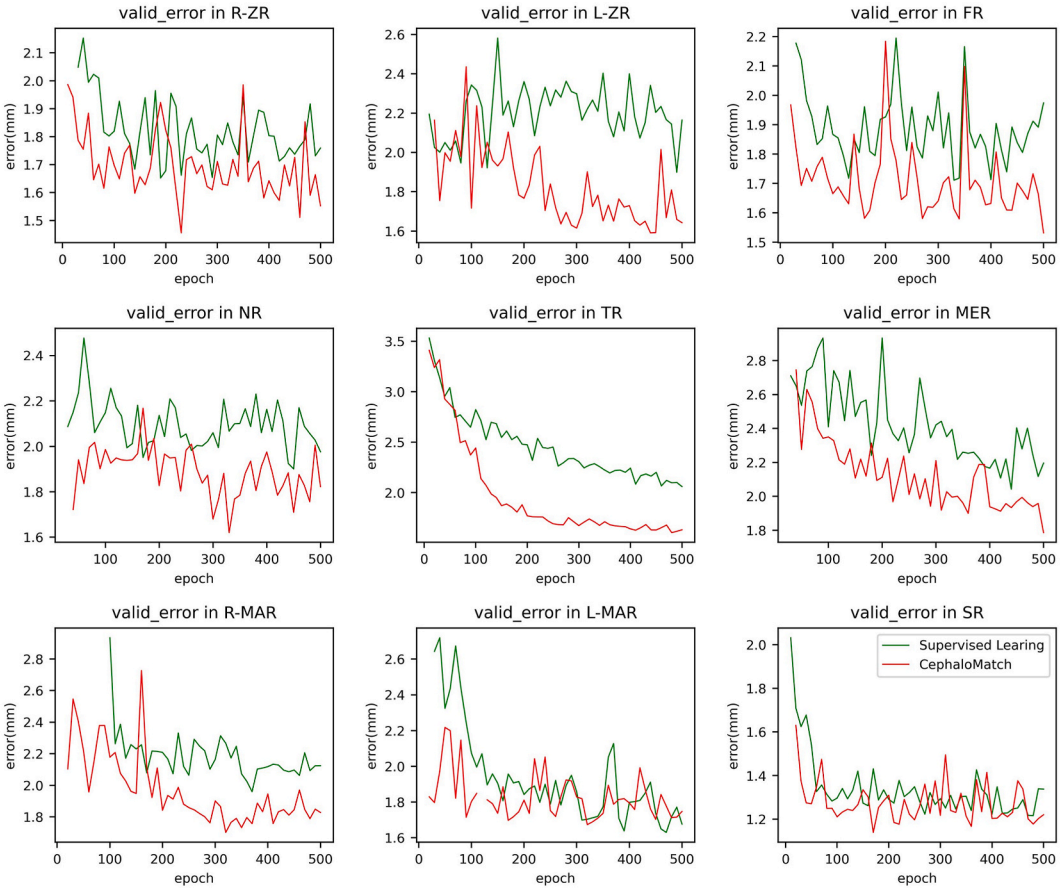

Fig. 4. Validation error curve when comparing the proposed SSL model with conventional supervised learning model in nine divided regions (When detection failure rate was more than 50 %, it was not demonstrated in error curve).

Validation error curve was plotted to demonstrate the training process of this comparative experiment in nine divided regions (Fig. 4).

3.2. *Training with fewer labeled dataset*

Compared to the conventional supervised learning, which utilized a training set of 60 labeled samples and achieved a mean error of 1.94 ± 1.12 mm, the proposed SSL model demonstrated improved performance with varying proportions of the labeled dataset. Specifically, the SSL model attained a mean error of 1.60 ± 0.87 mm using the full set of 60 labeled samples (100 %), 1.66 ± 0.93 mm with a reduced set of 48 labeled samples (80 %), 1.79 ± 0.97 mm with 36 labeled samples (60 %), and 1.91 ± 1.00 mm with 30 labeled samples (50 %) (Table 3).

3.3. *Comparing with state-of-the-art SSL models from nature image processing*

3.3.1. *Mean Teachers (2017) [11]*

Using 60 samples as training set, 14 samples as validation set and 288 samples as unlabeled set, Mean Teachers had a mean error of 1.71 ± 0.91 mm and a failure rate of 3.618 % (Table 4).

3.3.2. *FixMatch (2021) [14]*

Using 60 samples as training set, 14 samples as validation set and 288 samples as unlabeled set, FixMatch had a mean error of 1.69 ± 0.92 mm and a failure rate of 1.670 % (Table 4).

3.3.3. *UniMatch (2023) [15]*

Using 60 samples as training set, 14 samples as validation set and 288 samples as unlabeled set, UniMatch had a mean error of 1.67 ± 0.95 mm and a failure rate of 1.948 % (Table 4).

Validation error curve was plotted to demonstrate the training process of this comparative experiment in nine divided regions

Table 3
Comparing with conventional supervised learning method using fewer labeled dataset in nine divided regions.

Method	Training set	R-ZR (4)		L-ZR (4)		FR (4)		NR (8)		TR (40)		MER (8)		R-MAR (3)		L-MAR (3)		SR (3)		Overall (77)	
		Error (std)	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**
Supervised learning	60 (100 %)	1.65 (0.64)	1.786	1.90 (0.96)	1.786	1.71 (0.58)	5.357	1.90 (1.20)	0	2.06 (1.34)	2.857	2.04 (1.09)	8.929	1.96 (0.68)	2.381	1.63 (0.66)	4.762	1.22 (0.60)	0	**1.94 (1.12)**	3.154
CephaloMatch	48 (80 %)	1.50 (0.73)	3.571	1.85 (0.89)	1.786	1.65 (0.64)	3.571	1.74 (1.18)	1.786	1.60 (0.98)	0.536	1.92 (1.07)	7.143	1.79 (0.75)	2.381	1.81 (0.53)	4.762	1.13 (0.47)	0	1.66 (0.93)	1.948
	36 (60 %)	1.60 (0.63)	3.571	1.89 (0.80)	1.786	1.70 (0.61)	3.571	2.06 (1.61)	2.679	1.76 (1.01)	0.893	1.92 (0.94)	9.821	1.83 (0.69)	4.762	1.85 (0.74)	4.762	1.18 (0.46)	0	1.79 (0.97)	2.597
	30 (50 %)	1.62 (0.70)	3.571	1.99 (0.96)	1.786	1.82 (0.61)	3.571	2.04 (1.41)	3.571	1.93 (1.07)	2.679	2.01 (1.07)	8.036	2.02 (0.67)	4.762	2.02 (0.58)	4.762	1.15 (0.42)	2.381	**1.91 (1.00)**	3.525

Table 4
Comparison with state-of-the-art SSL models from nature image processing by metrics of error(mm), standard deviation(mm), detection failure rate (%) in nine divided regions.

Method	R-ZR (4)		L-ZR (4)		FR (4)		NR (8)		TR (40)		MER (8)		R-MAR (3)		L-MAR (3)		SR (3)		Overall (77)	
	Error (std)	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**
Mean Teachers	1.56 (0.70)	1.786	1.83 (0.94)	1.786	1.75 (0.64)	3.571	1.74 (1.14)	0.893	1.74 (0.99)	4.107	**1.73 (0.84)**	5.357	1.83 (0.64)	2.381	1.71 (0.52)	7.143	1.09 (0.56)	2.381	1.71 (0.91)	3.618
FixMatch	1.60 (0.67)	1.786	1.75 (1.00)	1.786	1.73 (0.78)	5.357	1.72 (1.19)	0	1.69 (0.96)	0.536	1.82 (0.96)	5.357	1.89 (0.74)	2.381	1.74 (0.59)	7.143	1.05 (0.42)	0	1.69 (0.92)	1.670
UniMatch	1.54 (0.69)	5.357	1.75 (0.78)	3.571	1.66 (0.63)	5.357	1.79 (1.30)	0	1.67 (1.01)	0.893	1.76 (1.11)	3.571	1.88 (0.64)	4.762	**1.63 (0.53)**	4.762	**1.00 (0.46)**	0	1.67 (0.95)	1.948
CephaloMatch (ours)	**1.46 (0.64)**	3.571	**1.59 (0.74)**	1.786	**1.53 (0.54)**	3.571	**1.62 (1.15)**	0	**1.60 (0.91)**	0.536	1.79 (1.00)	7.143	**1.70 (0.70)**	2.381	1.67 (0.71)	4.762	1.14 (0.47)	0	**1.60 (0.87)**	1.762

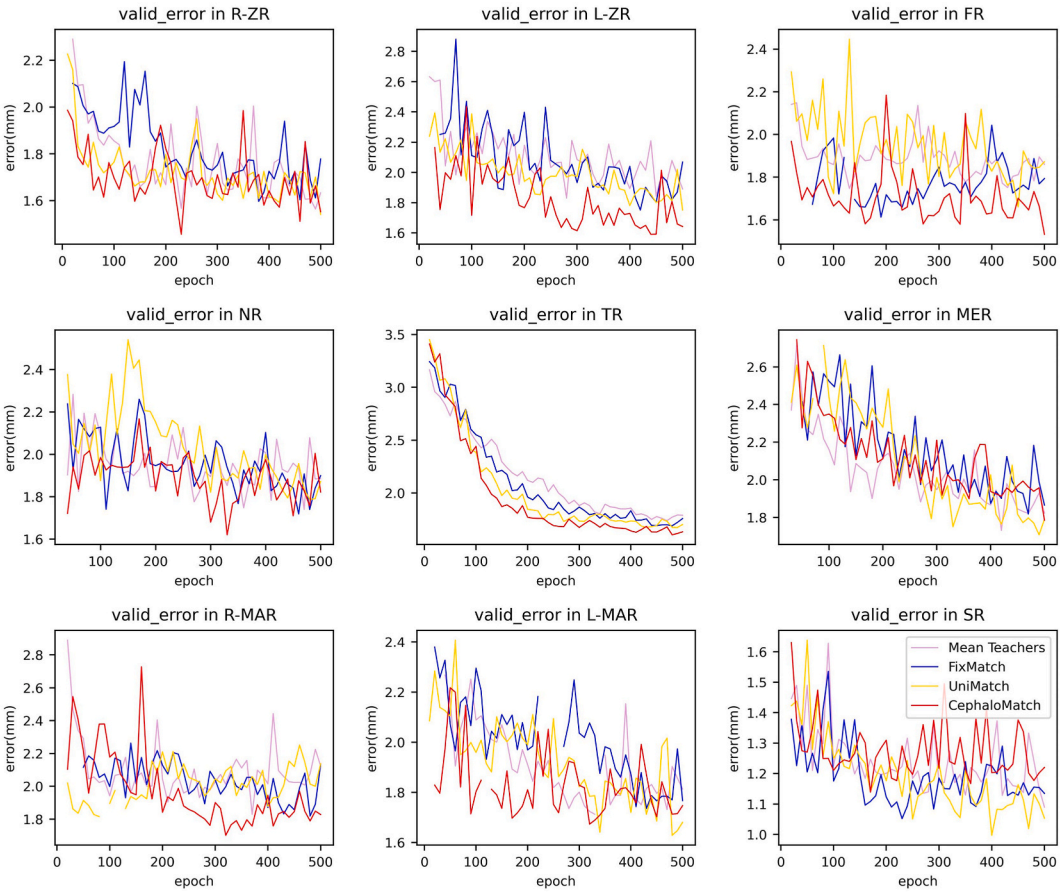

Fig. 5. Validation error curve when comparing the proposed SSL model with state-of-the-art SSL model from nature image processing in nine divided regions (When detection failure rate was more than 50 %, it was not demonstrated in error curve).

(Fig. 5).

3.4. Ablation experiment on head position rectification and multilayers perturbation

3.4.1. Head position rectification

Regarding UniMatch as baseline model, the SSL model developing with head position rectification had a mean error of 1.64 ± 0.87 mm, outperforming the baseline model (1.67 ± 0.95 mm) (Table 5).

3.4.2. Multilayers perturbation

Regarding UniMatch as baseline model, the SSL model developing with multilayers perturbation had a mean error of 1.65 ± 0.93 mm, outperforming the baseline model (1.67 ± 0.95 mm) (Table 5).

3.4.3. Jointly using rectification of head position and multilayers perturbation

Regarding UniMatch as baseline model, the SSL model developing with head position rectification and multilayers perturbation had a mean error of 1.60 ± 0.87 mm, outperforming the baseline model (1.67 ± 0.95 mm) and SSL model in 3.4.1&3.4.2 (Table 5).

4. Discussion and conclusion

This study presented an innovative application of SSL in the field of CMF CT landmarks detection, which had remarkable efficacy in achieving high levels of accuracy with a constrained labeled dataset comparing with conventional supervised learning. With the same quantity of labeled data, the proposed SSL model achieved a lower detection error compared to conventional supervised learning (Table 2). Particularly in cases where conventional supervised learning struggles to achieve accurate detection results, such as landmarks situated on smooth surfaces, the proposed SSL model demonstrated superior detection accuracy by leveraging structural features learned from an unlabeled dataset, which was several times larger than the labeled dataset (Fig. 6). In our workflow of manually landmarks digitization, it often takes an experienced surgeon 15–25 min to label a CMF CT with locations of all 77 landmarks

Table 5
Ablation experiment on head position rectification and multilayers perturbation in nine divided regions.

Head position rectification	Multilayers perturbation	R-ZR (4)		L-ZR (4)		FR (4)		NR (8)		TR (40)		MER (8)		R-MAR (3)		L-MAR (3)		SR (3)		Overall (77)	
		Error (std)	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**	**Error (std)**	**Fail**
		1.54 (0.69)	5.357	1.75 (0.78)	3.571	1.66 (0.63)	5.357	1.79 (1.30)	0	1.67 (1.01)	0.893	1.76 (1.11)	3.571	1.88 (0.64)	4.762	**1.63 (0.53)**	4.762	**1.00 (0.46)**	0	1.67 (0.95)	1.948
	✓	1.56 (0.71)	1.786	1.65 (0.93)	1.786	1.62 (0.63)	5.357	1.74 (1.13)	0	1.63 (0.90)	0.536	**1.76 (0.93)**	6.250	1.78 (0.65)	2.381	1.71 (0.56)	4.762	1.11 (0.52)	0	1.64 (0.87)	1.670
✓		1.56 (0.67)	3.571	1.73 (0.77)	1.786	**1.53 (0.47)**	3.571	1.69 (1.35)	0.893	1.64 (0.98)	0.893	1.82 (1.01)	7.143	1.78 (0.67)	4.762	1.70 (0.75)	4.762	1.18 (0.59)	0	1.65 (0.93)	2.134
✓	✓	**1.46 (0.64)**	3.571	**1.59 (0.74)**	1.786	1.53 (0.54)	3.571	**1.62 (1.15)**	0	**1.60 (0.91)**	0.536	1.79 (1.00)	7.143	**1.70 (0.70)**	2.381	1.67 (0.71)	4.762	1.14 (0.47)	0	**1.60 (0.87)**	1.762

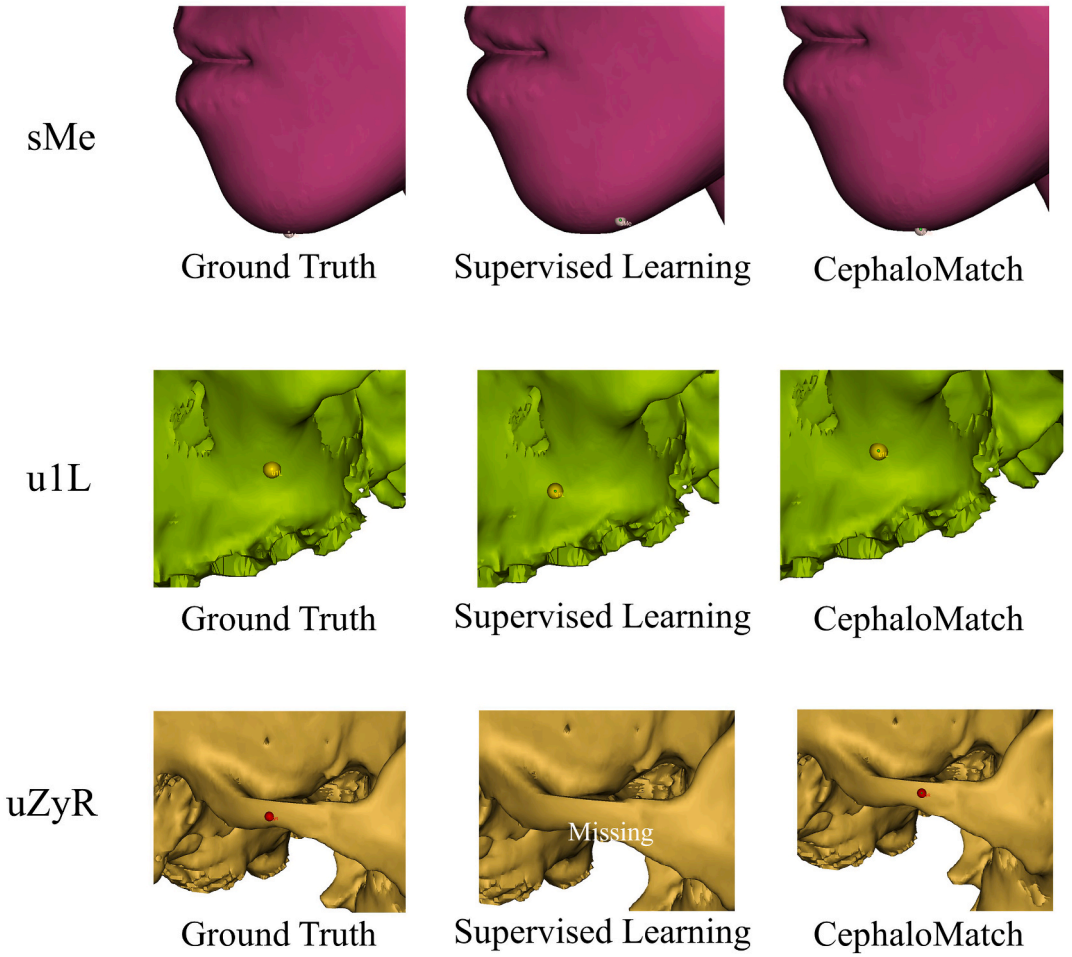

Fig. 6. An illustration comparing the proposed SSL model with conventional supervised learning with the same amount of labeled data (Abbreviation: sMe: the most inferior midpoint of the chin on the outline of the soft tissue; u1: the midpoint of zygomatic alveolar ridge; uZy: the most lateral point of zygomatic arch).

and a beginner more than 30 min. This significant time investment hinders the availability of labeled data and prompted us to explore whether a large volume of unlabeled data could compensate for the shortage of labeled data. Under this hypothesis, we tried to reduce the amount of labeled data used in the proposed SSL model and compare its landmarks detection performance with conventional supervised learning using the entire labeled data. The experimental result shows that the proposed SSL model based on 30 labeled data and 288 unlabeled data achieve equivalent detection error with conventional supervised learning based on 60 labeled data, initially validate its performance with limited labeled data (Table 3). This substantial reduction in the reliance on extensive labeled datasets means that it not only can relieve the heavy workload of the medical professionals but also can expedite the integration of deep learning models into clinical applications.

This study introduced two innovative methodologies within the proposed SSL model framework: a head position rectification method and a multilayers perturbation pattern. The ablation experiment of these two techniques was demonstrated in Table 5. The diversity in patient head positioning within CT scans potentially undermines the efficacy of regularization consistency in SSL. The introduction of head position rectification within the proposed SSL model standardized head positioning, aligning it with a uniform paradigm. This standardization effectively mitigated landmark detection errors. When adapting the state-of-the-art SSL methods to this task, UniMatch exhibited the lowest detection error, attributable to its novel integration of a dropout mechanism between the encoder and decoder within a weak-strong perturbation consistency framework. The deliberate information attenuation in the strong perturbation phase broadened the perturbation spectrum, thereby enhancing the SSL model's adaptability. Inspired by this principle and the unique architecture of U-Net, a multilayers perturbation pattern was devised to further diversify the perturbation landscape for this specific task. During the skip connection phase, a dropout process was employed to intentionally restrict information transmission, thereby inducing a more robust perturbation effect. An ablation study confirmed that the multilayers perturbation effectively augmented landmark detection accuracy. Ultimately, the integration of these two novel strategies within an SSL framework named CephaloMatch achieved substantial performance enhancements, surpassing existing state-of-the-art SSL methodologies, as evidenced

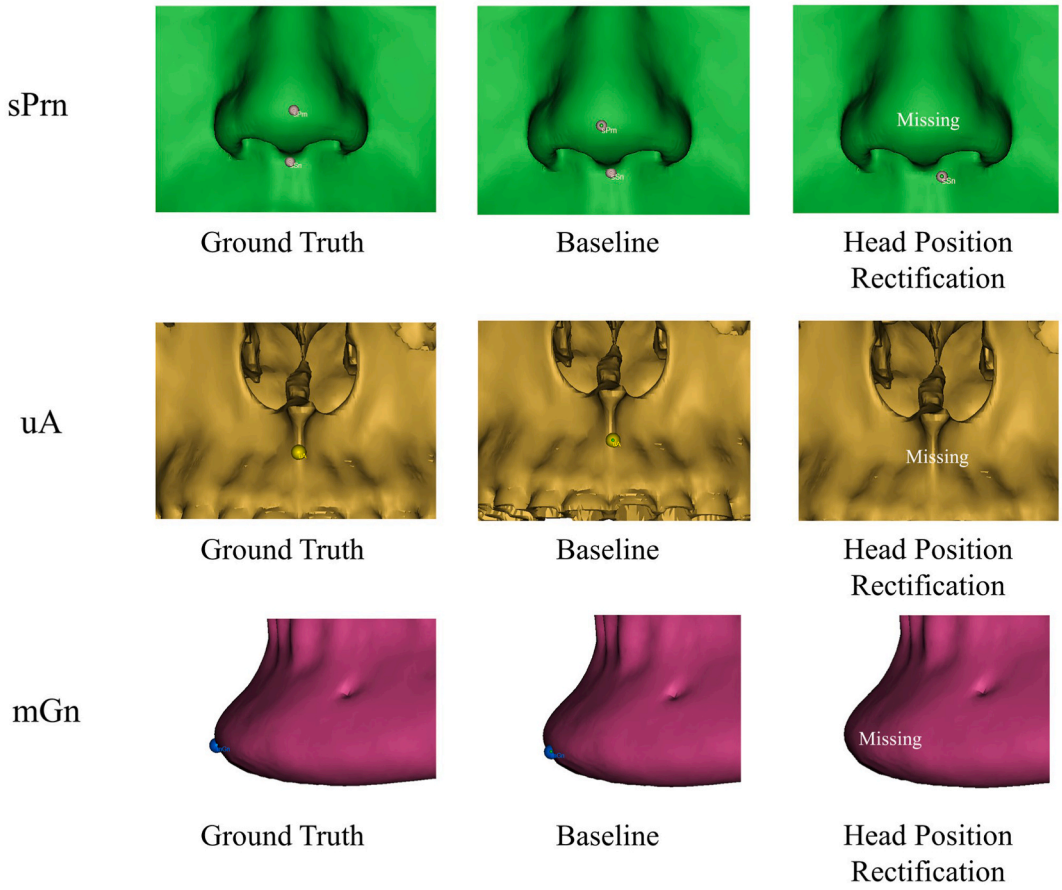

Fig. 7. Several landmarks in NR and MER not detected after head position rectification (Abbreviation: sPrn: the most anterior point at the nasal tip; uA: the most posterior midline point on the premaxilla between the anterior nasal spine and prosthion; mGn: the midpoint of mPog and mMe on the mandible in the midline.).

by the comparative data (Table 4, Table 5).

Despite the noteworthy advancements presented in our research, certain limitations were observed. Firstly, due to the difficult access to labeled data, validation set in this study is relatively small, making it difficult for statistical tests to show significant differences. Nonetheless, the experimental results serve as a foundational step for subsequent large-scale clinical trials. Additionally, while head position rectification contributed to a reduction in landmark detection error, it also resulted in a higher detection failure rate compared to the baseline. Upon comparing predicted landmarks with ground truth, instances were identified where landmarks located at NR and MER were undetected following head position rectification (Fig. 7). Lastly, 3D U-Net was set as backbone of all above models, including conventional supervised learning, CephaloMatch, UniMatch, etc., and extensive experiments using other U-Net structured network as backbone were not conducted. In future research, we believe that with a larger multicenter labeled/unlabeled dataset and extensive experiments on more different algorithms including self-supervised learning and reinforcement learning, a more comprehensive experiment result will be demonstrated with regard to accurate and cost-effective CMF CT landmarks detection.

In conclusion, our study underscores the transformative potential of semi-supervised learning (SSL) in overcoming the challenges associated with craniomaxillofacial (CMF) CT landmarks detection in scenarios with limited labeled data. The proposed SSL model, CephaloMatch, exhibited superior landmarks detection performance compared to conventional supervised learning when faced with limited labeled data, and achieved equivalent performance with fewer labeled data. With continued refinement and clinical validation, we anticipate that the proposed model will play a pivotal role in clinical settings, offering a cost-effective and efficient solution for diagnostic and treatment planning in dentomaxillofacial deformities.

Data availability statement

The data presented in this study are available on request from Yang Yang (17732239091@163.com). The data are not publicly available due to plans for further research, requirements from the hospital and privacy restrictions.

CRediT authorship contribution statement

Leran Tao: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xu Zhang:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation. **Yang Yang:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition. **Mengjia Cheng:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation. **Rongbin Zhang:** Writing – review & editing, Data curation. **Hongjun Qian:** Writing – review & editing, Data curation. **Yaofeng Wen:** Project administration, Data curation. **Hongbo Yu:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgments

This work was supported by National Natural Science Foundation of China (81571022), Multi-center clinical research project of Shanghai Jiao Tong University School of Medicine (DLY201808), Shanghai Natural Science Foundation (23ZR1438100), and National College Student Innovation and Entrepreneurship Training Program (202310248128).

References

[1] B. Broadbent, Y. Yang, A new X- ray Technique and its application to orthodontia: the Introduction of Cephalometric Radiography, Angle Orthod. 51 (2) (1981) 93–114.
[2] H. Pinsky, S. Dyda, R. Pinsky, K. Misch, D. Sarment, Accuracy of three-dimensional measurements using cone-beam CT, Dentomaxillofacial Radiol. 35 (6) (2006) 410–416.
[3] R. Chen, Y. Ma, N. Chen, L. Liu, Z. Cui, Y. Lin, W. Wang, Structure-aware long short-term memory network for 3D cephalometric landmark detection, IEEE Trans. Med. Imag. 41 (7) (2022) 1791–1801.
[4] G. Dot, T. Schouman, S. Chang, F. Rafflenbeul, A. Kerbrat, P. Rouch, L. Gajny, Automatic 3-dimensional cephalometric landmarking via deep learning, J. Dent. Res. 101 (11) (2022) 1380–1387.
[5] Q. Liu, H. Deng, C.F. Lian, X. Chen, D. Xiao, L. Ma, X. Chen, T. Kuang, J. Gateno, P. Yap, J. Xia, SkullEngine: a multi-stage CNN framework for collaborative CBCT image segmentation and landmark detection, Machine learning in medical imaging MLMI (Workshop) 12966 (2021) 606–614.
[6] C. Payer, D. Stern, H. Bischof, M. Urschler, Integrating spatial configuration into heatmap regression based CNNs for landmark localization, Med. Image Anal. 54 (2019) 207–219.
[7] Y. Lang, C. Lian, D. Xiao, H. Deng, P. Yuan, J. Gateno, S. Shen, D. Alfi, P. Yap, J. Xia, Automatic localization of landmarks in craniomaxillofacial CBCT images using a local attention-based graph convolution network. Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2020, pp. 817–826.
[8] Y. Lang, C. Lian, D. Xiao, H. Deng, K. Thung, P. Yuan, J. Gateno, T. Kuang, M. Alfi D, L. Wang, D. Shen, J. Xia, P. Yap, Localization of craniomaxillofacial landmarks on CBCT images using 3D mask R-CNN and local dependency learning, IEEE Trans. Med. Imag. 41 (10) (2022) 2856–2866.
[9] Y. Ouali, C. Hudelot, M. Tami, An overview of deep semi-supervised learning, arXiv preprint arXiv:200605278 (2020), https://doi.org/10.48550/arXiv.2006.05278.
[10] L. Samuli, A. Timo, Temporal ensembling for semi-supervised learning, International Conference on Learning Representations (ICLR). 4 (5) (2017) 6.
[11] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Process. Syst. 30 (2017).
[12] Q. Xie, Z. Dai, E. Hovy, T. Luong, Q. Le, Unsupervised data augmentation for consistency training, Adv. Neural Inf. Process. Syst. 33 (2020) 6256–6268.
[13] E. Arazo, D. Ortego, P. Albert, N.E. O'Connor, K. McGuinness, Pseudo-labeling and confirmation bias in deep semi-supervised learning, International Joint Conference on Neural Networks (IJCNN) (2020) 1–8.
[14] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C.A. Raffel, E.D. Cubuk, A. Kurakin, C.L. Li, Fixmatch: simplifying semi-supervised learning with consistency and confidence, Adv. Neural Inf. Process. Syst. 33 (2020) 596–608.
[15] L. Yang, L. Qi, L. Feng, W. Zhang, Y. Shi, Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7236–7246.
[16] L. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:170605587, https://doi.org/10.48550/arXiv.1706.05587, 2017.
[17] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.
[18] F. Milletari, N. Navab, S. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation. 4th IEEE International Conference on 3D Vision (3DV), 2016, pp. 565–571.
[19] Z. Zhou, M. Siddiquee, N. Tajbakhsh, J. Liang, UNet plus plus: a nested U-net architecture for medical image segmentation. 4th International Workshop on Deep Learning in Medical Image Analysis (DLMIA)/8th International Workshop on Multimodal Learning for Clinical Decision Support (ML-CDS), 2018, pp. 3–11.
[20] H. Lee, S. Bao, Y.A. Huo, B. Landman, 3D UX-net: a large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation, arXiv (2022), https://doi.org/10.48550/arXiv.2209.15076.
[21] C.K. Liang, S.H. Liu, Q. Liu, B. Zhang, Z.J. Li, Norms of McNamara's cephalometric analysis on lateral view of 3D CT imaging in adults from northeast China, J. Hard Tissue Biol. 23 (2) (2014) 249–254.
[22] L.K. Cheung, Y.M. Chan, Y.S.N. Jayaratne, J. Lo, Three-dimensional cephalometric norms of Chinese adults in Hong Kong with balanced facial profile, Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology 112 (2) (2011) E56–E73.

[23] C.T. Ho, R. Denadai, H.C. Lai, L.J. Lo, H.H. Lin, Computer-aided planning in orthognathic surgery: a comparative study with the establishment of burstone analysis-derived 3D norms, J. Clin. Med. 8 (12) (2019).
[24] L. Tao, M. Li, X. Zhang, M. Cheng, Y. Yang, Y. Fu, R. Zhang, D. Qian, H. Yu, Automatic craniomaxillofacial landmarks detection in CT images of individuals with dentomaxillofacial deformities by a two-stage deep learning model, BMC Oral Health 23 (1) (2023) 876.
[25] Y. Gao, D. Shen, Collaborative regression-based anatomical landmark detection, Phys. Med. Biol. 60 (24) (2015) 9377–9401.
[26] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: regularization strategy to train strong classifiers with localizable features. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6023–6032.