MASTER 2 BME-Paris

BioImaging Track

2019-2020

Contribution to an automated 3D landmark localization method on Head CT-scans using deep learning

Xiaoyang WEI

Institut de Biomécanique Humaine Georges Charpak,
Arts et Métiers ParisTech.

Laurent GAJNY, Gauthier DOT, Philippe ROUCH, Thomas SCHOUMAN.

"I attested that I have read and validated this report"

Supervisor's signature:

ECOLE NATIONALE SUPÉRIEURE D'ARTS ET
PARIS

Laurent Gajny

Summary

Craniomaxillofacial deformities present a wide variety of abnormalities of the head and the face including acquired and congenital disease [24]. Cephalometric analysis is used in clinical routine by orthodontists and maxillo-facial surgeons for diagnosis purpose. Manual landmark localization is a common practice in current cephalometric analysis on 2D images. However, 2D cephalometry only provides the visualization of lateral view of skull and is prone to projection bias so 3D cephalometry has been proposed in the literature to improve the measurement and analysis. But it is quite time-consuming to place landmarks for doctors and it often requires high level of expertise. Thanks to the development of computing facilities and machine learning research, more and more automated methods have been proposed to solve this problem [15][16][17].

In this internship, we benefited from an annotated database of 200 head CT-scans. To get the normalized data we implemented a new pre-processing strategy first including cropping and reorientation of CT image to acquire consistent position and orientation of heads for different patients based on the coordinates of annotated landmarks before feeding images to the model. On the basis of 3D U-Net architecture, we applied multi-task learning to solve this problem. Based on a limited dataset consisting of 35 patients (32 for training, 3 for testing), we localized six selected landmarks accurately with a mean Euclidean distance error of 1.69 ± 0.74 mm and segmented the upper skull and lower jaw with a pixel accuracy of 99% and a dice value of 90%. It demonstrated a promising potential to relieve working load for doctors and to help diagnosis.

Contents

Review of the Institut de Biomécanique Humaine Georges Charpak 's team

Institut de Biomécanique Humaine Georges Charpak (IBHGC) is a multidisciplinary research institute of Arts et Métiers ParisTech, which is deeply involved in sports and clinical biomechanics. It is named after Nobel prize winner Georges Charpak and is famous for EOS^{TM} low dose biplanar X-rays imaging system (EOS Imaging, Paris, France). This medical imaging modality now available in 51 countries was indeed developed in a collaborative research including IBHGC's team. In particular, researchers of the institute brought great contributions to accurate three-dimensional reconstruction of the spine.

Maxillo-facial surgery, one research topic of the institute, plays a vital role in quantification of craniomaxillofacial deformities through medical imaging. As part of this topic, 3D cephalometry have received attention from researchers of the institute. Several works about it have been explored including a systematic review [13] about current methods that have been proposed for automated landmark localization of 3D cephalometry. Moreover an intra- and inter-operator reproducibility study of manual landmarking in 3D cephalometry is currently being performed.

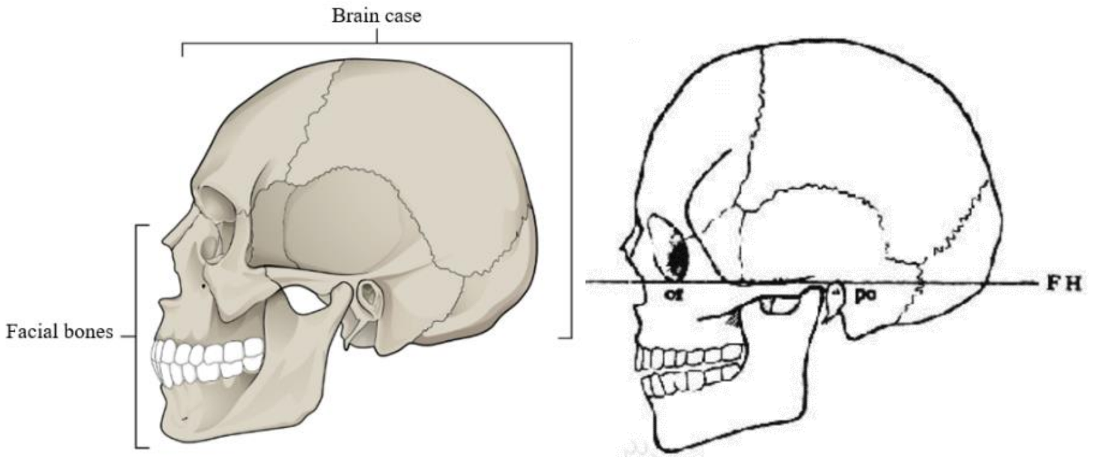
Yet, there is still no universally acceptable method that has been proved to be as reliable as manual annotation by experts. Because of the annotation cost, anatomical landmarks need to be automatically localized in volume image. Deep learning methods particularly adapted for this task are the focuses of this internship. The method has to be reliable enough for real patients with various pathologies.

1. Clinical Background

1.1 Skull anatomy and craniomaxillofacial deformities

The skull is the bony structure of the head which provides a rounded cavity to protect and house the brain (Figure 1.a). It can be subdivided into the brain case that surrounds the brain and the facial bones which not only provide support for the teeth but also form the orbits, nasal cavity and other facial structures. In adults, the lower jaw, also called mandible [1], is the biggest bone and the only bone which is movable of the skull.

Established at the World Congress on Anthropology in Frankfurt, Frankfurt horizontal plane (FH), which is also called eye-ear plane, is a plane supposed to be parallel to the ground when the head is standing in a natural position. As a result, Frankfurt line is used clinically to assist in head positioning. In craniometric study, the position of FH is determined by the coordinates of three anatomical landmarks, upper margins of both auditory meatus called Porion left and Porion rigtht (po), and the lower margin of the left orbit called Orbitale left (or).



(a) (b)

Figure1.a. The structure of the skull. Image Source [1] b. Frankfurt plane. Image Source [18]

Maxillofacial deformities are a group of various deformities of the facial bones [2] (Figure 2). Some of them are mild and some of them are severe which need surgery such as prognathism (Figure 3) or retrognathism (Figure 4). It is reported there are around 16.8 million people in America requiring surgical operations to correct the deformities [3]. Because of the complex anatomy of the skull, these operations require a detailed and precise pretreatment plan.

In addition to clinical examination, orthodontists and maxillo-facial surgeons need to quantify the spatial relationships between teeth and jaws to perform a full diagnosis. To do so, they position anatomical landmarks on medical images and use them to measure clinical indices.

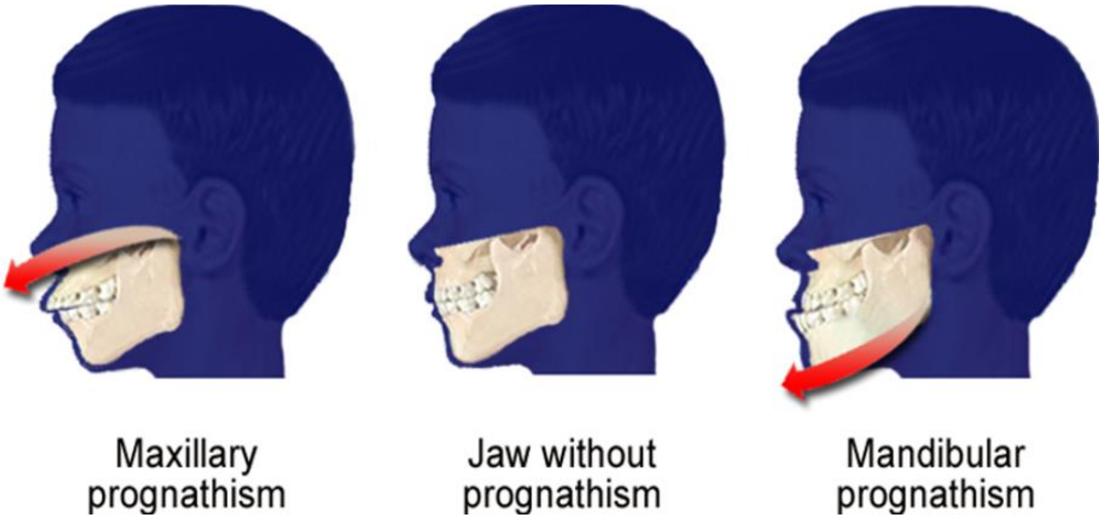


Figure2. The symptom of different craniofacial prognathism Image Source [4]

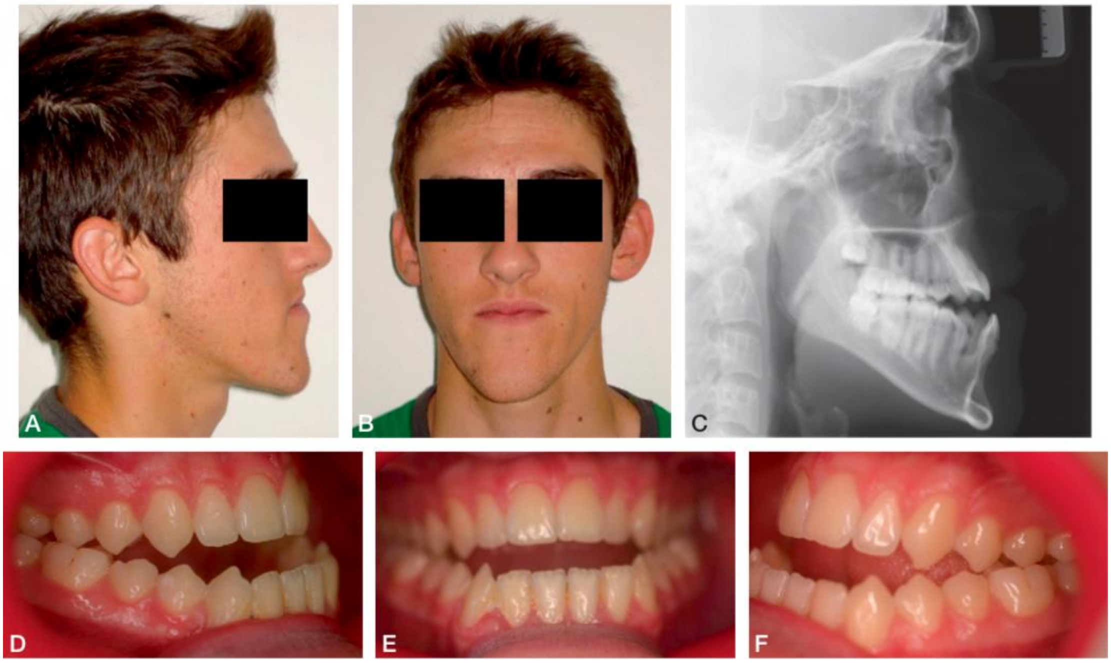


Figure3. Clinical example of mandibular prognathism. Image Source:[30].

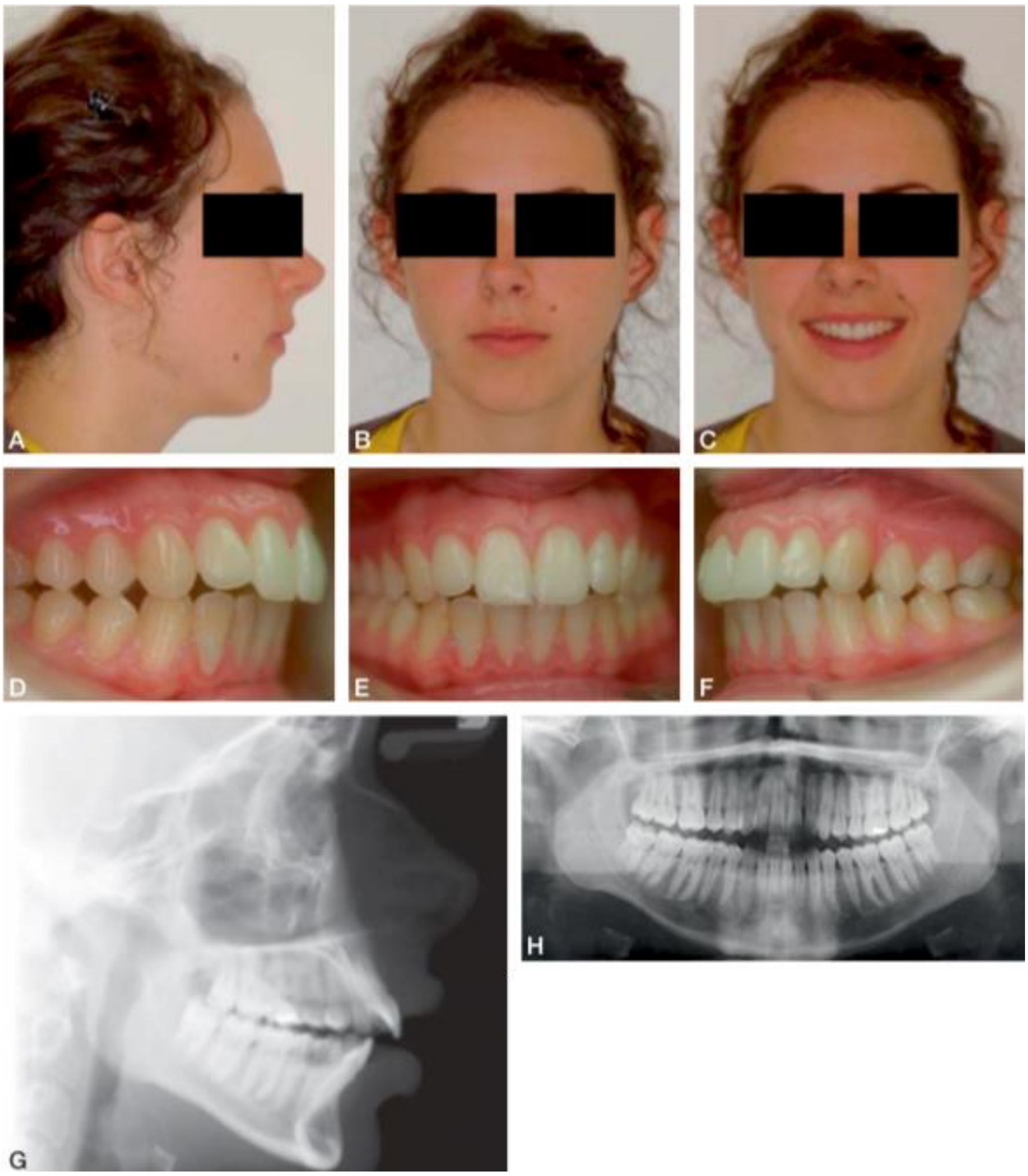
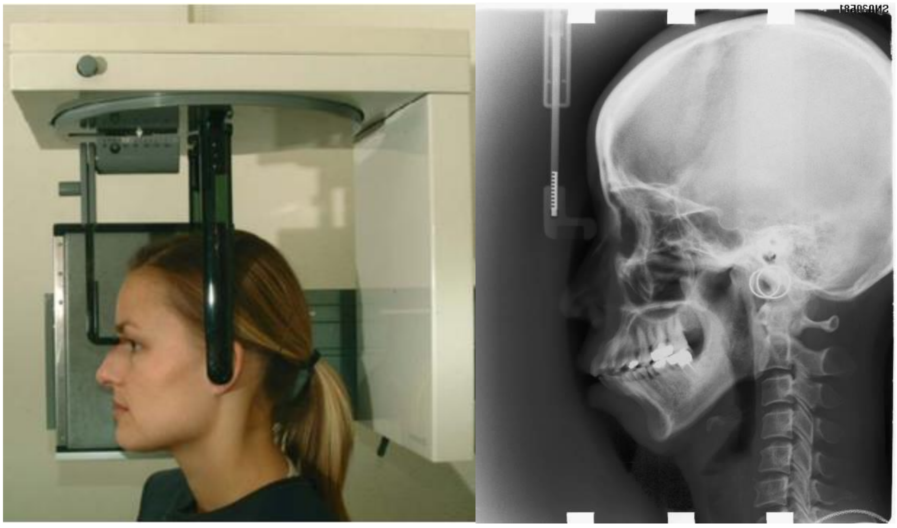


Figure4. Clinical example of mandibular retrognathism. Image Source:[30].

1.2 Imaging of the skull

1.2.1 2D imaging

A lateral cephalogram is an X-ray of the lateral view of the face with very accurate positioning. A lot of anatomical measurements can be made for clinical assessment of the deformity of the skull and the bites of patients. However, conventional radiograph is prone to the distortion caused by the nature of 2D radiography which refers to the magnification errors at different points of the images and the superimposition of bilateral structure. To acquire the lateral cephalogram, patients will be asked to stand in the correct position, and their orientation of head will be fixed by X-ray machine in two different directions orthogonal to each other (Figure 5.a). In this way by fixing the position and orientation, the impact of projection bias can be reduced to some extent. Yet, the superimposition of structures is unavoidable.



(a) (b)

Figure5. a. a patient positioned in the standard way in the X-ray machine for a lateral cephalogram. Image Source [5] b. a radiograph image of cephalometry from ISBI challenge. Image Source [6]

1.2.2 3D imaging

Computed Tomography (CT) is a 3D medical imaging technique based on computer-aided reconstruction which is often regarded as an auxiliary step for treatment planning difficult cases (Fig 4a). CT scans acquired can be exported into DICOM slices and visualized with specialized software [7]. 3D surface models can be reconstructed for these slices by performing imaging segmentation. Each pixel of slices is stored with Hounsfield units (HU), which are supposed to have similar values for similar tissues. Despite wide use in hospitals, it is blamed for relatively high radiation dose and high cost.

On the other hand, Cone Beam Computed Tomography (CBCT) has received increasingly wide use in dentistry and orthodontics (Fig 4b). During imaging acquisition, the CBCT rotates around the head of patients, acquiring hundreds of various images according to specific region of interest. CBCT is more affordable than CT-Scan, is more compact and usually delivers a lower dose of radiation to the patient [28]. However, for now there is no calibration of the grey values of CBCT voxels, which implies that imaging segmentation can be more complicated than CT-Scans [28].

(a) (b)

Figure6. a. Discovery CT750 HD. Image Source [8] b. Dental CBCT scanner. Image Source[9]

1.3 Cephalometry: state of the art

With the help of advanced medical imaging technique, we could further perform quantitative measurement studies of the skull for patients. Cephalometry is the study and measurement of the head including different anatomical landmarks. Here anatomical landmarks refer to some meaningful points of the skull to ensure the correspondence between different subjects. In clinical routine, cephalometric analysis is mainly based on 2D cephalogram and manually-annotated landmarks (Figure 7). Several angle measurements are computed and compared to normality corridors. For example, to measure mandibular retrognathism or prognathism the angle formed by points S, Na and Pog is measured: the more it increases the more it means that a mandibular prognathism is present.

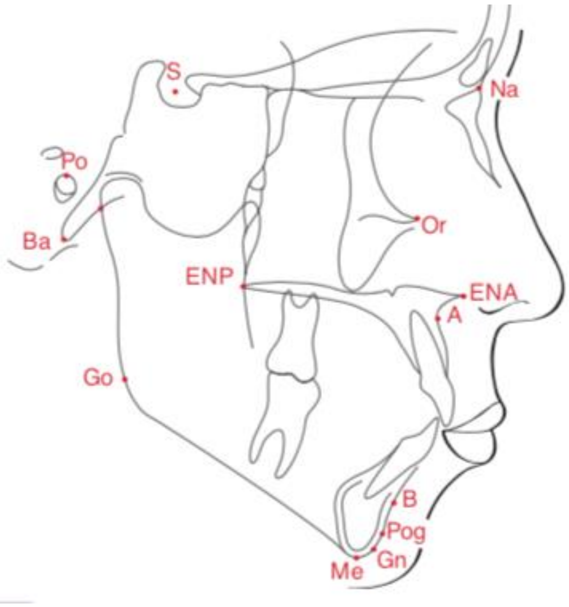


Figure7. Example of typical 2D landmarks for cephalometric analysis. Image Source:[30].

There exist manual and automatic approaches for cephalometry, with different degree of validation. In the reproducibility study of Trpkova et al[28], the authors reported an uncertainty of 0.59 mm for the x coordinate and 0.56 mm for the y coordinate. In 2019, Chen et al [12] compared the top two methods in ISBI challenge 2015 with their work using attention feature pyramid fusion. Mean radial error for all these methods was less than 2mm. Hence, automatic methods are not yet as efficient as manual operators.

As discussed before, 2D imaging does not show clearly bilateral structures and thus cannot help the diagnosis of frontal asymmetries for example. Therefore, 3D cephalometry has been proposed in the literature to improve the measurement and analysis, using CT or CBCT for image acquisition. Because of highly symmetric feature of the skull, most landmarks of the skull can be divided in two parts. The first part refers to bilateral landmarks lying on both sides of the skull symmetrically like Infraorbital Foramen Left/Right (Figure 8) which refers to two symmetric cavities below two orbits of the skull. The second part refers to landmarks exactly lying on the midsagittal plane of the skull like Point A, Point B, Anterior Nasal Spine (ANS) and Gnathion (Figure 8). Here Point A and B refer to the deepest point of the anterior silhouette of the upper and the lower alveolar arch. ANS refers to the apex of the anterior nasal spine. Gnathion refers to the most anterior-inferior mid-point of the skeletal chin.

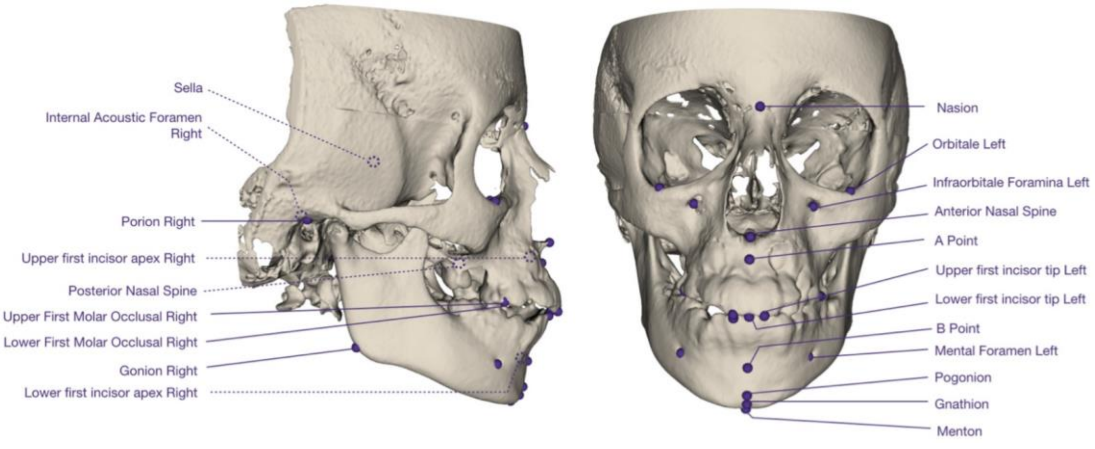

Sella
Internal Acoustic Foramen Right
Porion Right
Upper first incisor apex Right
Posterior Nasal Spine
Upper First Molar Occlusal Right
Lower First Molar Occlusal Right
Gonion Right
Lower first incisor apex Right
Nasion
Orbitale Left
Infraorbitale Foramina Left
Anterior Nasal Spine
A Point
Upper first incisor tip Left
Lower first incisor tip Left
B Point
Mental Foramen Left
Pogonion
Gnathion
Menton

Figure8. anatomical landmarks selected for the 3D cephalometry study

One major drawback of 3D landmarking is the time needed for clinicians to perform a full analysis. It is reported that it takes up to 14 minutes to annotate 22 landmarks in 3D cephalometry [25]. Thanks to the development of computing facilities and machine learning research, more and more automated methods have been proposed to solve this problem [13][14]. Among all automated methods, several deep learning-based methods received much more attention because of their high accuracy. Besides, segmentation of the skull also plays a key role in following deep learning-based approaches to provide guidance for landmark localization.

Zhang et al. [16] proposed a two-stage model with the same fully convolutional architecture composed of three down sampling and up sampling steps. The first stage is designed to generate displacement maps representing spatial context information for all landmarks. With displacement maps and raw image stacked together as input, the second stage was used to localize landmarks with and without guidance of segmentation. In their work, evaluation was based on a 5-fold cross validation. The dataset consisted of 77 CBCT from patients with maxillofacial deformities and 30 CT scans from normal control subjects. The overall mean Euclidean distance error was 1.10 ± 0.71 mm. Guidance of segmentation map helped improving the accuracy of all landmarks. But there is no additional preprocessing in this work, unnormalized position and orientation of heads could have an impact on performance.

Torosdagli et al. [17] adopted a Fully Convolutional Network (FCN) implemented using dense block for feature reuse to segment the mandible first, followed by a linear time distance transformation and geodesic learning network to generate combined geodesic map for localization of sparsely-spaced landmarks. After that, a recurrent neural network was applied to localize closely-spaced landmarks around the Menton landmark. The dataset consisted of 50 CBCT scans of heads with a high degree of maxillofacial deformities. The proposed method was evaluated using a 5-fold cross validation. Seven out of nine landmarks evaluated were localized with an average distance error less than 1 mm. For Point B and Pogonion the mean and the median errors in the volume space were less than 1.36 mm. But all methods in this pipeline were implemented with a pseudo-3D (slice-by-slice 2D) fashion which means some spatial context information in vertical direction were neglected, importantly they only focused on landmarks of the mandible.

Motivated by all these studies, we proposed a 3D U-Net based multi-task learning model which could localize landmarks and segment bones synchronously on the basis of the study of Zhang et al [16]. Besides, an additional test was included in our studies to explore the potential impact of reorientation of heads. Our method was trained on 32 CT scans and tested on 3. It will be evaluated on a bigger and already annotated dataset with 200 CT scans afterwards. The target choice could be extended to any other landmarks in cephalometry.

2. Automated landmark localization& bone segmentation

2.1 Abstract

Manual landmark localization is a common practice in current cephalometric analysis but it is a time-consuming job and requires high level expertise. Some automated deep learning-based landmarking methods have been proposed, but they could be susceptible to orientation of patients. In this project, the CT images of 200 patients were annotated manually in MIMICS by a doctor in dental surgery. The images and their segmentation maps were reoriented to a normalized direction. A 3D U-Net based model was proposed to estimate the position of each landmark using Gaussian heatmap and to segment each part of the skull. For preliminary evaluation, 32 were used for training and 3 were used for test. Results showed that the proposed model could localize six selected landmarks automatically with a Euclidean distance error of 1.69 ± 0.74 mm and have a promising potential to relieve working load for doctors and to help diagnosis.

.

2.2 Introduction

Maxillofacial deformities are a group of various deformities of the facial bones [2]. Some of them are mild and some of them are severe which need surgery. It is reported there are around 16.8 million people in America requiring surgical operations to correct the deformities [3]. Because of the complex anatomy of the skull, these operations require a detailed and precise pretreatment plan in which cephalometry plays a key role for diagnosis. Landmarking is an indispensable step in cephalometric analysis to identify the anatomical features of patients. The conventional method of landmarking is performed on 2D radiograph but it is prone to projection bias and superimposition of bilateral structure. Recently, 3D cephalometry has been proposed in the literature to solve this problem.

Manual landmarking is a gold standard in 3D cephalometry and segmentation may be needed to help operators placing the landmarks in this case. Reproducibility studies demonstrate that landmarks on midsagittal plane have greater reliability than bilateral landmarks [26]. Moreover, the inter-operator error ranges from inferior to 0.5mm to greater than 2mm [13]. Besides, manual landmarking is also a tedious job and requires high level expertise. It is reported that it takes up to 14 minutes to annotate 22 3D landmarks for a professional doctor [25]. Because of the annotation cost, there is an strong need to develop automated landmark localization method in volume image and to lighten this burden for doctors.

There are some deep learning-based approaches that have been proposed to localize landmarks automatically. Zhang et al. [16] localized 15 landmarks with and without guidance of segmentation. A mean Euclidean distance of 1.10mm was obtained with the help of segmentation but it was done without additional preprocessing, Torosdagli et al. [17] segmented the mandibles first and then localized 9 mandibular landmarks automatically. Seven out of nine landmarks were tested with a distance error less than 1mm but they used more than one models to localize landmarks and they still treated landmarking and segmentation as two separate tasks. In this context, we proposed a new preprocessing strategy including optional reorientation of heads to test whether or not it improves the results using a big dataset consisting of 200 patients. Due to the impact of the epidemic, we only present preliminary results based on a limited dataset consisting of 35 subjects of an FCN based model in this report [19].

2.3 Materials & Methods

2.3.1 Database

A database of 200 subjects with different ages and degrees of craniofacial deformity was studied in this project. The mean age was 27 years old with around 10 years standard deviation. The CT scans of most patients were acquired using Discovery CT750 HD from GE Healthcare. Each slice of images was composed of 512*512 pixels with size range from 0.32 to 0.63mm. The increment of each slice was most of the time 0.312mm and 0.5mm otherwise. The number of slices for each patient ranged from 282 to 917. 3D segmentation of different parts of the skull were performed by experts of Materialise France. All CT images were screened by Gauthier Dot, Doctor of Dental Surgery (DDS) and PhD student (IBHGC, Arts et Métiers Paris). To obtain the ground truth of landmark localization, 33 distinct landmarks were annotated by the same DDS. We selected 6 of them for a preliminary test. Midsagittal landmarks are labelled as red and bilateral landmarks are labelled as yellow in Figure 8.

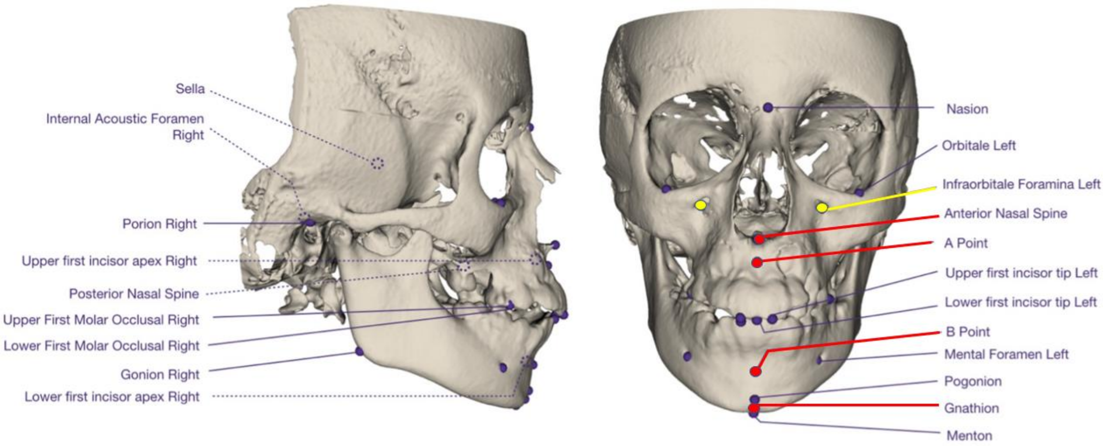

Sella
Internal Acoustic Foramen Right
Porion Right
Upper first incisor apex Right
Posterior Nasal Spine
Upper First Molar Occlusal Right
Lower First Molar Occlusal Right
Gonion Right
Lower first incisor apex Right
Nasion
Orbitale Left
Infraorbitale Foramina Left
Anterior Nasal Spine
A Point
Upper first incisor tip Left
Lower first incisor tip Left
B Point
Mental Foramen Left
Pogonion
Gnathion
Menton

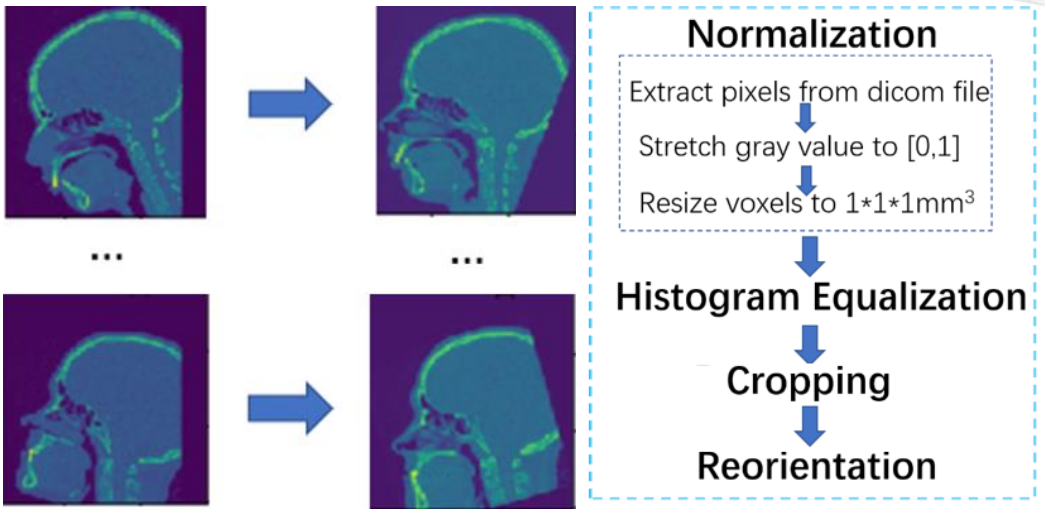
Figure8. anatomical landmarks selected for the 3D cephalometry study

2.3.2 Preprocessing

As is mentioned above, most of data were acquired using Discovery CT750 HD but some other data were acquired using different machine so the scale of Hounsfield units, image size and pixel resolution can be different for various subjects. To normalize the whole dataset adaptively we stretched the gray value of images to the range of [0,1] after extracting pixel intensity information from DICOM file. Then we stacked all slices into volume and resized voxel resolution to 1*1*1 mm^3 according to corresponding slice thickness and pixel size. By coordinate information exported using MIMICS, we cropped the images according to the minimum and the maximum of coordinates of all landmarks so that we could make sure that no landmark would be out of the field of view and there would be at least 10 pixels away from the boundary of the image for all landmarks. In this way we cropped all images to 168*152*160 size. After that we adopted adaptive histogram equalization to enhance the contrast of CT images so that we can distinguish the skull and the soft tissue more easily.

In the meantime, we generated a 3D Gaussian probability heatmap with the standard deviation of 3 mm for each landmark, and the intensity values of all voxels for heatmaps were stretched to the range of [0,1] which means that the intensity value of each voxel represents the probability that the voxel stands for the correct position of corresponding landmark. Besides, we also exported segmentation masks of upper skull and mandibula in bmp file using MIMICS with 512 x 512 size which is consistent with the size of CT scans to implement a joint model.

To acquire head CT scans, usually patients are asked to lie on supine position of CT but the orientation of their heads may vary. In order to test if reorientation of heads could have an impact on accuracy, we chose to select Frankfurt horizontal plane (FH) using anatomical landmarks and made it parallel to the horizontal plane using rotation. In craniometric study, the position of FH is determined by the coordinates of three anatomical landmarks, upper margins of both auditory meatus called Porion left and Porion rigtht, and the lower margin of the left orbit called Orbitale left. To compute the rotation matrix for correction, we created an anatomical coordinate frame first using the three landmarks mentioned above, then we calculated the rotation matrix by comparing the anatomical frame with world frame and finally we got the rotation angles in Bryant sequence. Besides, we used padding on the boundary to avoid potential distortion which could be introduced by the nearest interpolation.


...
...
Normalization
Extract pixels from dicom file
Stretch gray value to [0,1]
Resize voxels to 1*1*1mm^3
Histogram Equalization
Cropping
Reorientation

(a) (b)

Figure9. a. Example of reorientation b. Preprocessing workflow

2.3.3 3D U-Net and its extensions

U-Net was a convolutional neural network that has attracted widespread concern of researchers. It was designed for medical image segmentation at first [20] and its robustness has also been proved in other field like remote sensing [27]. The architecture of U-Net was based on a fully convolutional network [21] containing convolutional blocks only. To help grasping contextual information in vertical direction, we adopted a U-Net based network using 3D convolutional kernels to build the mapping function between original CT scans and its bone segmentation and probability heatmap corresponding to each landmark, with its architecture shown in Figure 10 below.

The proposed model was composed of a down-sampling path and an up-sampling path with two concatenations between them. The down-sampling path followed typical convolutional blocks. Each block consisted of two 3 × 3 × 3 convolutions with an activation function of rectified linear unit (ReLU), followed by a 2 × 2 × 2 max pooling layer with the stride of 2. By this way, the size of feature map was halved after getting through each block by a max pooling operation at its final step. Besides, each block in the up-sampling path consisted of a 3 × 3 × 3 deconvolution followed by two 3 × 3 × 3 convolutions using ReLU as activation function. The size of feature map was doubled using transposed convolution with stride 2 for each dimension at the top of every up-sampling block.

Therefore, after down-sampling and up-sampling the size of final feature map could be consistent with before. In this way, a new feature map was generated in up-sampling path which contained both low-level and high-level feature by concatenation. Since cephalometric landmarks were usually located on the boundary of the upper skull and the mandible, we assumed that the features for landmark localization and bone segmentation could be highly consistent, thus we adopted a hard parameter sharing architecture accompanied with a dropout regularization of 0.5 ratio in the fifth convolutional block before the output layer to implement multi-task learning and to avoid overfitting. For the final layer we adopted sigmoid function as the activation function to jointly predict both probability heatmap and binary segmentation mask. Due to the use of feature map concatenation, this fully convolutional network could grasp a big image volume with small kernel sizes while keeping high localization precision.

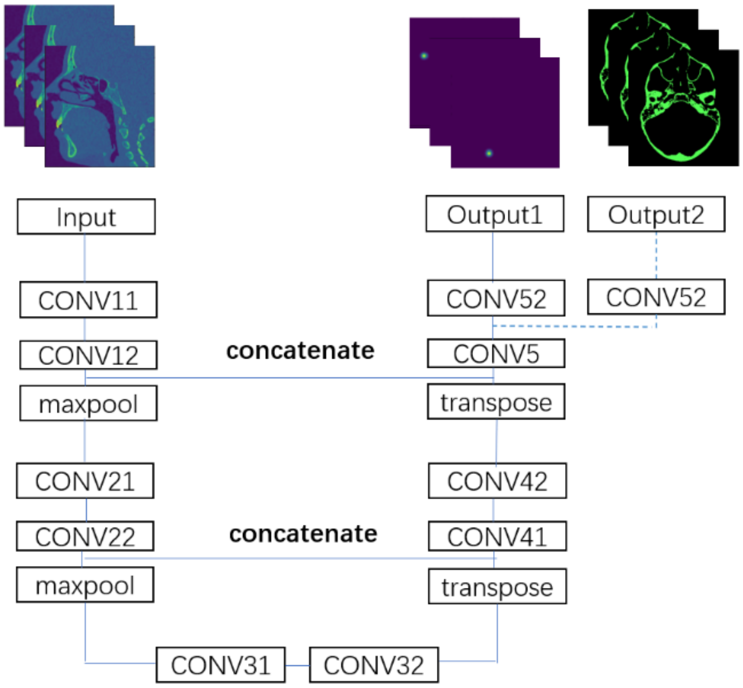


Figure10. 3D U-Net based joint model for landmark localization and bone segmentation

On the basis of proposed model, we also provided two shortcut connection-based modification as reference. The residual block was inspired by ResNet [20] proposed by Kaiming He in ImageNet challenge. The main idea was to calculate residual between convolutional layers by utilizing shortcuts to jump over several layers. Normally model degradation could be avoided using convolutional networks with residual block. Similar to ResNet, the Dense Convolutional Network (DenseNet) was implemented using convolutional blocks that concatenated every layer with all the other layers in a feed-forward mode to fully make use of extracted feature [23]. As each layer received every previous layer as input, much richer patterns could be stacked together by feature reuse. In this way, more smooth decision boundaries could be derived thanks to features of different degrees of complexity. That’s why overfitting can be resolved to some extent using dense blocks even with inadequate data. Inspired by these two models, we modified convolutional blocks in our model as following using shortcut connection. (Figure 11)

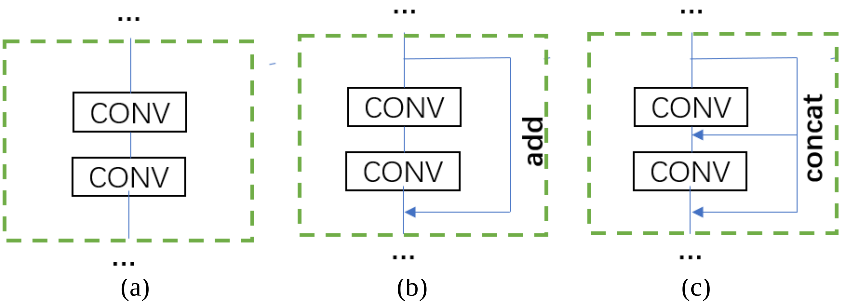


Figure11. modified blocks using shortcut connection. a. original setting b. residual block using skip connection c. dense block with multiple concatenation between each layer

2.3.4 Evaluation

On the basis of 3D U-Net and extensions, we performed experiments with and without reorientation of heads and with and without segmentation respectively to test the potential impact of these two elements. Because of limited access of our cluster, we could not fully make use of the whole dataset. In the evaluation step we used 35 patients out of 200 to perform training and testing on a cloud. The subset was further divided into 3 patients as test set and 32 patients as training set. As mentioned before, due to limited memory we selected 6 landmarks for evaluation including point A, point B, Anterior Nasal Spine and Gnathion lying on the mid sagittal plane of the head and two symmetric lateral landmarks which refer to Infraorbitale Foramen Left/Right.

In the evaluation step for the Gaussian heatmap, the coordinate of each landmark was estimated using argmax function based on probability heatmap prediction. The diagnostic accuracy for anatomical landmark localization was evaluated using the average difference and its standard deviation expressed in mm (Euclidean distance).

Since sigmoid activation was applied to the output layers for bones segmentations, the predictions of the output layer were not labeled as binary segmentation mask but with continuous values in the range of [0,1]. In this case we binarized the output of segmentation map using 0.5 as threshold. All voxels with intensity inferior to 0.5 were labelled as 0 and other voxels with value greater than 0.5 were labelled as 1. After thresholding the output of segmentation map, two various metrics were chosen to evaluate the accuracy of semantic segmentation both for the skull and mandible. The masks we used as ground truth were annotated by experts from Materialise France. The first metric is pixel accuracy whose equation is as follow (1) where N_{TP} means the number of voxels which are labelled as true positive. Similar to that, N_{TN} refers to ones that are true negative, N_{FP} for false positive and N_{FN} for false negative. The second metric is dice coefficient [24] which is expressed by Equation (2) which is often used to measure the similarity of two samples.

$$A_p(X,Y) = \frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \qquad (1)$$

$$Dice(X,Y) = \frac{2*|X \cap Y|}{|X|+|Y|} \qquad (2)$$

2.3.5 Implementation details

Deep learning models mentioned above were implemented with Keras framework [23] using Tensorflow as backend. Our fully convolutional networks were based on 3D convolutional layers. The segmentation end adopted a binary cross-entropy for loss function. The localization end used a mean squared error as loss function. The total loss function of the joint model implemented with multi-task learning is expressed by Equation (3) where the ratio for loss weight was 1:4. The standard deviation of Gaussian heatmap was 3mm. And for optimizer of our model we adopted Adam with learning rate equal to 10^{-4}. Adaptive learning rate decay was applied for training. If validation loss did not decrease for more than 20 epochs, the learning rate became one fifth as before. Adaptive decay allowed learning rate to decrease when training curve reached a plateau, which helped to fine-tune the model. Early-stopping method was also applied to avoid overfitting and to save training cost.

$$Loss=0.2*L_{heatmap} + 0.8*L_{mask} \qquad (3)$$

2.4 Experimental results

Experiments were performed on a cloud with 24 GB memory and NVIDIA Tesla P100 PCIe 16 GB graphic card with batch size equal to one. A limited dataset of 35 patients was further splitted into a training set of 32 patients and a test set of 3 patients. Limited by computation cost, no cross validation was performed. Original models without segmentation converged after training for 200 epochs (around 8 hours), joint models with segmentation converged after training for 250 epochs (around 12 hours). Besides, training cost for 3D Dense U-Net was slightly more than 3D U-Net and 3D Res U-Net.

2.4.1 landmark localization evaluation

We reported landmark localization performances for method with and without reorientation and segmentation respectively in Table 1,2 and 3. In our different tests, among the six selected landmarks, Anterior Nasal Spine (ANS) had the lowest average Euclidean distance of 1.13mm ± 0.25 mm and Infraorbital Foramen Right (IF right) had the highest average Euclidean distance 2.72 mm ± 1.43 mm. When comparing the impact of reorientation, mean Euclidean distance error increased 0.39 mm, 0.50 mm, 0.37 mm for 3D U-Net, 3D Res U-Net and 3D Dense U-Net respectively. When comparing the impact of segmentation maps, distance error increased 0.09 mm, 0.20 mm for 3D U-Net and 3D Res U-Net respectively. For 3D Dense U-Net, mean Euclidean distance error decreased 0.06 mm. Among all results we have, 3D Res U-Net without guidance of segmentation map achieved the best performance of 1.69 ± 0.74 mm when training on original images without reorientation.

Table 1: Euclidean distance: mean ± SD. (without segmentation & reorientation)

Landmarks	**3D U-Net**	**3D Res U-Net**	**3D Dense U-Net**
A	2.35 ± 0.96	1.32 ± 0.49	1.79 ± 0.62
ANS	1.13 ± 0.25	1.13 ± 0.33	1.26 ± 0.56
B	1.67± 0.73	1.70 ± 0.66	3.43 ± 1.71
Gnathion	1.72 ± 0.84	0.90 ± 0.18	1.81 ± 0.84
IF left	1.74 ± 0.92	1.95 ± 0.84	1.95 ± 0.96
IF Right	2.72 ± 1.43	3.16 ± 1.56	2.23 ± 1.03
total	1.88 ± 0.83	1.69 ± 0.74	2.07 ± 0.91

Table 2: Euclidean distance: mean ± SD. (with segmentation)

Landmarks	3D U-Net	3D Res U-Net	3D Dense U-Net
A	2.50 ± 1.04	2.18 ± 0.96	2.6 ± 1.06
ANS	1.59 ± 0.81	1.86 ± 0.75	1.92 ± 0.43
B	3.76± 1.85	4.18 ± 2.16	3.19 ± 1.54
Gnathion	0.99 ± 0.43	1.64 ± 0.56	2.25 ± 0.86
IF left	1.83 ± 0.66	1.82 ± 1.03	2.04 ± 0.79
IF Right	2.95 ± 1.37	3.28 ± 1.56	2.43 ± 0.99
total	2.27 ± 0.96	2.49 ± 1.22	2.40 ± 0.98

Table 3: Euclidean distance: mean ± SD. (with reorientation)

Landmarks	3D U-Net	3D Res U-Net	3D Dense U-Net
A	2.45 ± 0.96	1.50 ± 0.69	1.72 ± 0.60
ANS	1.23 ± 0.25	1.28 ± 0.33	1.20 ± 0.53
B	1.75± 0.73	2.05 ± 0.71	3.33 ± 1.68
Gnathion	1.80 ± 0.84	0.95 ± 0.23	1.75 ± 0.82
IF left	1.82 ± 0.92	2.00 ± 1.04	1.85 ± 0.88
IF Right	2.82 ± 1.43	3.38 ± 1.56	2.21 ± 0.92
total	1.97 ± 0.92	1.89 ± 0.84	2.01 ± 0.86

2.4.2 Bone segmentation evaluation

We obtained segmentation predictions using proposed joint model. Performances for mandible segmentation and upper skull segmentation were shown as below (Table 4,5.) For both segmentation mask, 3D Dense U-Net achieved the best performance with pixel accuracy equal to 99.5% and 99.1% and dice similar coefficients equal to 0.9033 and 0.9010, slightly better than 3D Res U-Net and 3D U-Net. Segmentation performance for mandible was slightly better than upper skull.

Table 4: Mandible segmentation: pixel accuracy & dice similarity coefficient

Metric	3D U-Net	3D Res U-Net	3D Dense U-Net
Pixel Accuracy	0.9943	0.9945	0.9946
Dice coefficient	0.9027	0.9030	0.9033

Table 5:Upper skull segmentation: pixel accuracy & dice similarity coefficient

Metric	3D U-Net	3D Res U-Net	3D Dense U-Net
Pixel Accuracy	0.9901	0.9904	0.9911
Dice coefficient	0.8998	0.9001	0.9010

2.5 Discussion

2.5.1 assessment of impact about reorientation and segmentation

We proposed a deep learning-based method for landmark localization and bones segmentation of 3D cephalometry using a fully convolutional network which was able to precisely segment upper skull and mandible in 3D CT scans. In this study, reorientation of heads did not show help to localization as expected. Interpolation during the rotation of images could be a reason why data quality is degraded. Besides, there were inevitable defects because of incomplete head on the boundary of CT scans which could cause distortion. But if we rotated images first and then crop the images, the problem could be eased to some extent but we could not implement reorientation in this way due to limited memory.

We also tried to use multi-task learning by hard parameter sharing fashion to localize landmarks and segment bones jointly. However, in terms of segmentation, the performance was satisfying with a pixel accuracy of 0.99 and a dice value of 0.90 but the performances of landmark localization were not better than original model. In fact, an inappropriate loss weight between two terms of loss function, binary cross entropy and mean squared error could be one of the reasons. Loss weights were hyperparameters that were set empirically. They could be optimized by further test.

2.5.2 Comparison of ResNet and DenseNet

In fact, a Fully Convolutional DenseNet (also called Tiramisu in short) has already been mentioned in Torosdagli et al’s work to localize cephalometric landmarks last year. In their work, a Tiramisu with 103 convolutional layers was used to segment the mandible of the skull as the first step in the pipeline. To reduce the total parameters of the model we only used 2 convolutional layers in each dense block instead of 4, with 3D convolutional kernel instead of 2D in our project. And this network was designed to predict probability heatmap instead of for preliminary segmentation only.

In terms of performance, the best result obtained among all experiments was with a mean Euclidean distance error of 1.69 ± 0.74 mm. It was acquired using 3D Res U-Net without guidance of segmentation on original data without reorientation. Actually, this result was only 0.19 mm better than the result obtained using 3D U-Net which means differences could not be considered as significative for the moment and further tests are needed.

2.5.3 Limitations and future work

A limitation of this work was the resolution of images and the size of dataset. In this project the voxel size for all CT scans was normalized to 1*1*1 mm^3. Images were fed to the model using generator in a one-by-one fashion. But voxel size could also be further normalized to 0.5*0.5*0.5 mm^3 without decreasing the field of view which would be closer to papers with state-of-the-art results [14][15].

Evaluation without decreasing the field of view on proposed model could be done by loading small patches instead of a whole image. Besides, more subjects would be included and studied using cross-validation afterwards. A mean Euclidean distance of 1.69 mm with a standard deviation of 0.74 mm proved the feasibility of proposed model for training with more data.

Training on the whole dataset consisting of 200 patients would be performed on the cluster equipped with NVIDA RTX TITAN 24 GB in next few weeks. In this case cross validation could be used to get an unbiased result. We could also try to increase batch size instead of 1, or try loading small patches instead of a whole image. Other parameters like standard deviation of the Gaussian probability heatmap could also be optimized by further test if possible.

2.6 Conclusion

A deep learning-based method for landmark localization and bone segmentation in 3D cephalometry was presented. The method should help to automate measurement and calculation of 3D cephalometry, which was an essential step in treatment planning for orthodontist and surgeons. In this context, we compared the impact of reorientation and segmentation guidance based on a dataset consisting of 35 patients. It showed no guidance for landmark digitization according to preliminary results. Our proposed method achieved 1.69 ± 0.74 mm mean Euclidean distance for landmark localization and 0.99 pixel accuracy with a dice value of 0.90 for bones segmentation. Besides, an evaluation based on larger database with 200 patients could be performed afterwards.

General Conclusion

The purpose of this internship was to replace manual annotation with automated annotation of anatomical landmarks in existing 3D cephalometry algorithms from CT scans. Methods presented were based on 3D U-Net and its extensions implemented using shortcut connection. We presented an automated landmark localization and bones segmentation method. This method was able to precisely segment bones of heads and localize landmark with a mean distance error of 1.69mm. By the way, potential impact introduced by the reorientation of heads was explored in the project and it showed no help for landmark localization. Overall, in this Master 2 internship training, we explored 3D U-Net's capability both in landmark localization and bones segmentation of 3D cephalometry. Preliminary results suggest that deep learning-based method could lead to relatively accurate results.

Personal Feedback

Under very kind guidance of supervisors, I was in charge of data normalization, image preprocessing and postprocessing and training 3D U-Net for our tasks. During this process, I improved my programming skills in Python, and developed expertise in deep learning methods. This internship experience prepares me for my future study in medical imaging analysis and machine learning.

Acknowledgements

The internship was funded by the Materialise France on subject, with the support of APHP, La Pitié-Salpétrière, Paris. And thanks a lot for the software service provided by engineer of Materialise MIMICS

References

[1] “7.2skull-anatomyandphysiology.” Accessed June 11, 2020. https://opentextbc.ca/anatomyandphysiology/chapter/7-1-the-skull/.

[2] Teichgraeber, John F., Jaime Gateno, and James J. Xia. “Congenital Craniofacial Malformations and Their Surgical Treatment.” In Encyclopedia of Otolaryngology, Head and Neck Surgery, edited by Stilianos E. Kountakis, 544–62. Berlin, Heidelberg: Springer, 2013. https://doi.org/10.1007/978-3-642-23499-6_463.

[3] W. De Vos, J. Casselman, G.R.J. Swennen, Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: A systematic review of the literature, International Journal of Oral and Maxillofacial Surgery,Volume 38, Issue 6,2009,Pages 609-625,ISSN 0901-5027, http://www-o.ntust.edu.tw/~cweiwang/celph/

[4] “Orthognathic surgery.” In Wikipedia, February 29, 2020. https://en.wikipedia.org/w/index.php?title=Orthognathic_surgery&oldid=943150052.

[5] Gwen R.J. Swennen, Filip Schutyser, Jarg-Erich Hausamen. Three-Dimensional Cephalometry A Color Atlas and Manual. Springer, Accessed June 11, 2020. https://www.springer.com/gp/book/9783540254409.

[6] Kaur, Amandeep, and Chandan Singh. “Automatic Cephalometric Landmark Detection Using Zernike Moments and Template Matching.” Signal, Image and Video Processing 9, no. 1 (January 2015): 117–32. https://doi.org/10.1007/s11760-013-0432-

[7] Hatcher DC (October 2010). "Operational principles for cone-beam computed tomography". Journal of the American Dental Association. 141 (Suppl 3): 3S–6S. doi:10.14219/jada.archive.2010.0359. PMID 20884933

[8] “Discovery CT750 HD | GE Healthcare.” Accessed June 11, 2020. https://www.gehealthcare.com/courses/discovery-ct750-hd.

[9] “Dental CBCT Scanner - All Medical Device Manufacturers - Videos.” Accessed June 11, 2020. https://www.medicalexpo.com/medical-manufacturer/dental-cbct-scanner-28330.html.

[10] Baumrind, S., and R. C. Frantz. “The Reliability of Head Film Measurements. 2. Conventional Angular and Linear Measures.” American Journal of Orthodontics 60, no. 5 (November 1971): 505–17. https://doi.org/10.1016/0002-9416(71)90116-3.

[11] Vig, K. D., and E. Ellis. “Diagnosis and Treatment Planning for the Surgical-Orthodontic Patient.” Dental Clinics of North America 34, no. 2 (April 1990): 361–84.

[12] Chen Runnan, Ma Yuexin, Chen Nenglun, Lee Daniel, and Wang Wenping. “Cephalometric Landmark Detection by AttentiveFeature Pyramid Fusion and

Regression-Voting,” August 23, 2019. https://arxiv.org/abs/1908.08841v1.

[13] G. Dot, F. Rafflenbeul, M. Arbotto, L. Gajny, P. Rouch, T. Schouman,Accuracy and reliability of automatic threedimensional cephalometric landmarking,International Journal of Oral and Maxillofacial Surgery,2020,ISSN 0901-5027,doi:10.1016/j.ijom.2020.02.015

[14] Yun HS, Jang TJ, Lee SM, Lee SH, Seo JK. Learning-based local-to-global landmark annotation for automatic 3D cephalometry. Phys Med Biol. 2020;65(8):085018. Published 2020 Apr 23. doi:10.1088/1361-6560/ab7a71

[13]Hye Sun Yun et al 2020 Phys. Med. Biol. in press https://doi.org/10.1088/1361-6560/ab7a71

[14] Jun Zhang, Mingxia Liu, Li Wang, Si Chen, Peng Yuan, Jianfu Li, Steve Guo-Fang Shen, Zhen Tang, Ken-Chung Chen, James J. Xia, Dinggang Shen, Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization,MedicalImageAnalysis,Vol.60,2020,101621,ISSN1361-8415,doi:10.1016/j.media.2019.101621.

[15] N. Torosdagli, D. K. Liberton, P. Verma, M. Sincan, J. S. Lee and U. Bagci, "Deep Geodesic Learning for Segmentation and Anatomical Landmarking," in IEEE Transactions on Medical Imaging, vol. 38, no. 4, pp. 919-931, April 2019, doi: 10.1109/TMI.2018.2875814.

[16] “Frankfurt plane baidu baike”Accessed June 11, 2020. https://baike.baidu.com/pic/%E6%B3%95%E5%85%B0%E5%85%8B%E7%A6%8F%E5%B9%B3%E9%9D%A2/664998/0/e1fe9925bc315c600845014d8db1cb13485477bb?fr=lemma&ct=single#aid=0&pic=e1fe9925bc315c600845014d8db1cb13485477bb.

[17] Özgün Çiçek and Ahmed Abdulkadir and Soeren S. Lienkamp and Thomas Brox and Olaf Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 2016

[18] Ronneberger Olaf, Fischer Philipp, and Brox Thomas. “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 18, 2015. https://arxiv.org/abs/1505.04597v1.

[19] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation,” n.d., 10.

[20] He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian. “Deep Residual Learning for Image Recognition,” December 10, 2015.

https://arxiv.org/abs/1512.03385v1.

[21] Li Xiaomeng, Chen Hao, Qi Xiaojuan, Dou Qi, Fu Chi-Wing, and Heng Pheng Ann. “H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes,” September 21, 2017. https://arxiv.org/abs/1709.07330v3.

[22] “Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index - PubMed.” Accessed June 11, 2020. https://pubmed.ncbi.nlm.nih.gov/14974593/.

[23] Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 8 November 2018).

[24] Xia JJ, Gateno J, Teichgraeber JF. New clinical protocol to evaluate craniomaxillofacial deformity and plan surgical correction. J Oral Maxillofac Surg. 2009;67(10):2093-2106. doi:10.1016/j.

[25] Hassan, Bassam, Peter Nijkamp, Hans Verheij, Jamshed Tairie, Christian Vink, Paul van der Stelt, and Herman van Beek. “Precision of Identifying Cephalometric Landmarks with Cone Beam Computed Tomography in Vivo.” European Journal of Orthodontics 35, no. 1 (February 2013): 38–44. https://doi.org/10.1093/ejo/cjr050.

[26] LISBOA, Cinthia de Oliveira, Daniele MASTERSON, Andréa Fonseca Jardim MOTTA, and Alexandre Trindade MOTTA. “Reliability and Reproducibility of Three-Dimensional Cephalometric Landmarks Using CBCT: A Systematic Review.” Journal of Applied Oral Science 23, no. 2 (2015): 112–19. https://doi.org/10.1590/1678-775720140336.

[27] Yi, Yaning, Zhijie Zhang, Wanchang Zhang, Chuanrong Zhang, Weidong Li, and Tian Zhao. “Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network.” Remote Sensing 11, no. 15 (January 2019): 1774. https://doi.org/10.3390/rs11151774.

[28] Trpkova, Biljana, Paul Major, Narasimha Prasad, and Brian Nebbe. “Cephalometric Landmarks Identification and Reproducibility: A Meta Analysis.” American Journal of Orthodontics and Dentofacial Orthopedics 112, no. 2 (August 1, 1997): 165–70. https://doi.org/10.1016/S0889-5406(97)70242-7.

[29] Sam Alycia, Currie Kris, Oh Heesoo, Flores-Mir Carlos, and Lagravére-Vich Manuel. “Reliability of different three-dimensional cephalometric landmarks in cone-beam computed tomography: A systematic review.” The Angle Orthodontist 89, no. 2 (March 1, 2019): 317–32. https://doi.org/10.2319/042018-302.1.

[30] “orthodontie de l'enfant et du jeune adulte -tome 2.” Accessed June 15, 2020. https://www.elsevier-masson.fr/orthodontie-de-lenfant-et-du-jeune-adulte-tome-2-9782294724701.html.