IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 23, 2019, accepted August 27, 2019, date of publication September 10, 2019, date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940623

Locating Anatomical Landmarks on 2D Lateral Cephalograms Through Adversarial Encoder-Decoder Networks

XIUBIN DAI[1], HAO ZHAO[2], TIANLIANG LIU[2], DAN CAO[3], AND LIZHE XIE[3]

[1]School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210046, China
[2]School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[3]Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University, Nanjing 210096, China

Corresponding author: Lizhe Xie (xielizhe@njmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 31671006 and Grant 61671255, in part by the Qing Lan Project of Jiangsu Province, and in part by the Six Talent Project of Jiangsu Province under Grant JY-058.

ABSTRACT Locating anatomical landmarks in a cephalometric X-ray image is a crucial step in cephalometric analysis. Manual landmark localization suffers from inter- and intra-observer variability, which makes developing automated localization methods urgent in clinics. Most of the existing techniques follow the routine thoughts which estimate numerical values of displacements or coordinates for the target landmarks. Additionally, there are no reported applications of generative adversarial networks (GAN) in cephalometric landmark localization. Motivated by these facts, we propose a new automated cephalometric landmark localization method under the framework of GAN. The principle behind our approach is fundamentally different from the conventional ones. It trained an adversarial network under the framework of GAN to learn the mapping from features to the distance map of a specific target landmark. Namely, the output of the adversarial network in this paper is image data, instead of displacements or coordinates as the conventional approaches. Based on the trained networks, we can predict the distance maps of all target landmarks in a new cephalometric image. Subsequently, the target landmarks are detected from the predicted distance maps by an approach similar to regression voting. Experimental results validate the good performance of our method in localization of cephalometric landmarks in dental X-ray images.

INDEX TERMS Adversarial encoder-decoder networks, localization of anatomical landmarks, cephalometric analysis, prediction of distance maps.

I. INTRODUCTION

As an indispensable tool for modern orthodontics, orthognathic treatment, and maxillofacial surgery [1], the cephalometric analysis provides valuable information about patients' bony, dental, and soft tissue structures. Based on the cephalometric analysis of patients, dentists can provide with diagnosis and treatment of obstructive sleep apnea [2], assessment of mandible/lower jaw [3] and soft facial tissue [4], and so on.

In clinics, the first step of a routine cephalometric analysis is to manually mark all the anatomical landmarks after drawing craniofacial structure contours on a 2D cephalometric X-ray image. Fig. 1 displays 19 commonly used landmarks adopted in our work, and [5] describes these landmarks. After that, orthodontists calculate some cephalometric tracing clinical measurements [6], such as angles and line segments between the identified landmarks. Then, they can decide if there are anatomical abnormalities or not, according to the clinical measurements.

However, manual marking suffers from inter- and intra-observer variability, which eventually affects the accuracy of cephalometric analysis [7]. Besides, another disadvantage of manual marking is time-consuming. It will take an experienced orthodontist more than 20 minutes to locate 19 anatomical landmarks on a 2D cephalometric X-ray image [8]. Therefore, automatically locating anatomical landmarks on lateral cephalograms is in urgent need.

There have been many approaches developed for cephalometric landmark detection. The traditional methods include edge detection [9]–[12], active model [13]–[16], and template matching [8], [17]–[20]. Edge detection techniques used

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

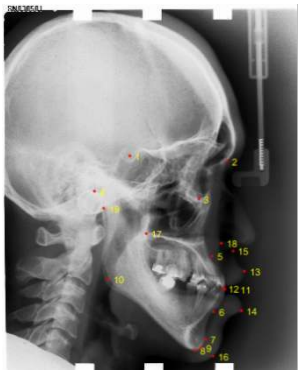


FIGURE 1. 19 anatomical landmarks used in this paper.

predefined lines, intersections, and exterior boundaries to detect landmarks. Although edge detection techniques can automatically identify the landmarks, they fail when the landmarks to be detected do not lie on the edges.

To solve these problems, some applied active models, such as the active shape model (ASM) [21] and the active appearance model (AAM) [22]. In ASM approaches, they deformed a global model of the spatial relationship between the essential structures to make an initial estimation of the location of landmarks and built a local model of image texture to determine the final positions of targets [13], [15]. Kafieh *et al.* [23] also extended ASM approaches by combing neural networks. As for AAM, using a full model combining shape and texture variability makes it different from ASM [14], [16]. For instance, Rueda incorporated mathematical morphology into AAM for landmark detection [14].

As an alternative to the active model, template matching approaches also play an important role in automated landmark localization. For examples, El-Feghi *et al.* located 20 anatomical landmarks by using a template-matching algorithm and a trained neuro-fuzzy system [8]. Kaur combined the sum of squared distance with normalized cross-correlation in template matching to obtain the location of landmarks [20]. Overall, template matching approaches can achieve higher accuracy of landmark localization than above two types of methods [20].

Due to the great success of machine learning in image processing and computer vision, researchers employed some popular techniques in cephalometric landmark localization. Early in 2003, after extracting projected principal-edge distribution (EPPD) features as input, Chakrabartty *et al.* [24] trained a support vector machine (SVM) to detect cephalometric landmarks. After that, Pouyan and Farshbaf [25] replaced EPPD with histograms of oriented gradient (HOG) descriptors as the input of SVM. Besides SVM, some articles also reported the applications of the neural network [26]–[31], game theory [32], and convex optimization [33] in locating cephalometric landmarks.

Moreover, many works that used the tree-based approach to detect landmarks on 2D lateral cephalograms have been done recently [34]–[38]. Chu *et al.* used the random forest (RF) regression trained with HOG features to obtain the initial position of the target landmarks and corrected the position iteratively via a sparse shape composition model [35]. To get accurate results of cephalometric landmark detection, Mirzaalian and Hamarneh [36] blended several hand-crafted features, e.g., local binary patterns, spatial coordinates, blobness, tubularness, and Zernike features to train the random forest models. Lindner et al. [37] developed a fully automatic landmark annotation (FALA) system in which they used their previous work about random forest regression-voting [39] in cephalometric analysis. In Vandaele's detection algorithm [38], they resorted to a variant of the random forest named as an extremely randomized tree with a multi-resolution scheme.

The general rationale behind the above machine learning-based approaches is to train a model or a network to learn the mapping from features to displacements of each voxel towards the target landmark or the relationship between features and coordinates of the target landmark directly. And, the output of the trained model or network is the numerical value of displacement or coordinate.

Fundamentally different from the previous work, the method proposed in this paper is dedicated to learning the mapping from features to the distance map of the target landmark by constructing a generative encoder-decoder network. To make the predicted results more realistic, we also build an adversarial network under the generative adversarial framework. After obtaining the displacement of each voxel towards the target landmark in a new dental X-ray image from the predicted distance map, our method can locate the target landmark in a similar way to regression voting [39]. As the distance map is an image in which the value of each pixel indicates its displacement to the target landmark, the proposed network outputs image data, instead of numerical values as the existing methods.

One can find more details about how our method works to automatically locate the cephalometric landmarks on 2D lateral cephalograms in Section 2. Section 3 provides the experimental results to evaluate the performance of the proposed method. Section 4 concludes the paper.

II. METHODS

In this paper, we propose an automated landmark localization method for in 2D lateral cephalograms. Fig. 2 shows the pipeline of the proposed method, which is composed of two stages. In this first stage, i.e., training stage, we train an encoder-decoder network as the generator to predict the distance map of input source images after preprocessing the training data. Secondly, we use a series of convolutional layers to construct the discriminator network, which is similar to the encoder part of the generator. After the training of detectors for all the target landmarks, the testing stage can produce the distance maps of the target landmarks for the newly acquired 2D lateral cephalograms. Finally, our method can locate the target landmarks through the methodology, which bears some resemblance to the regression voting.

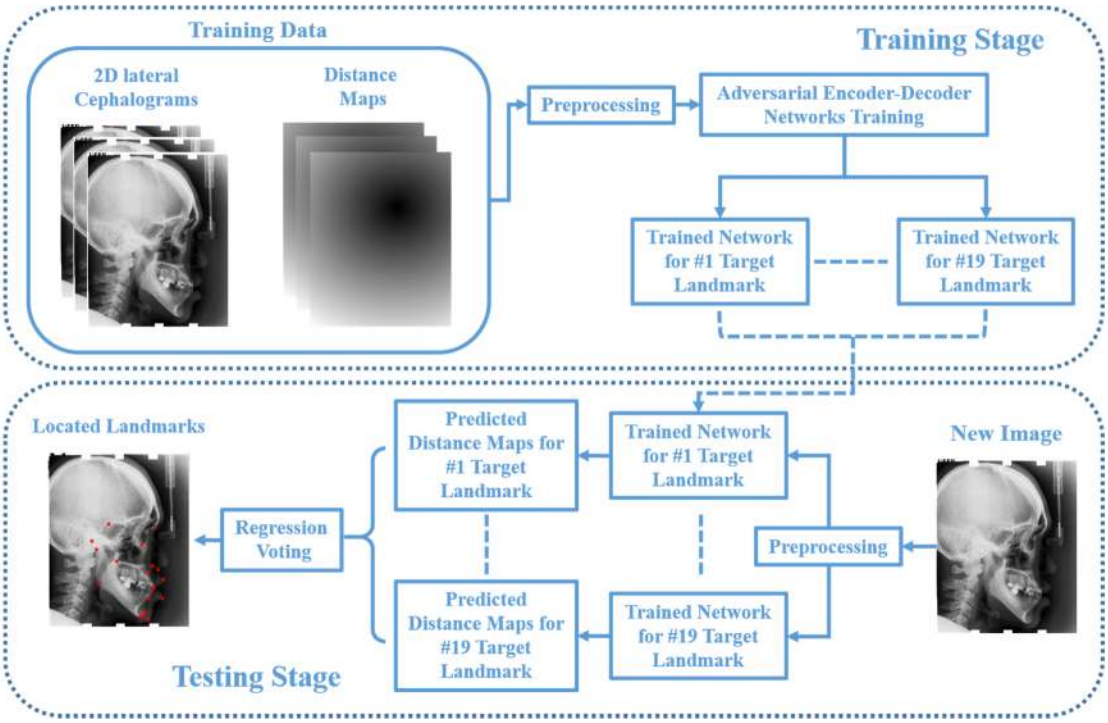


FIGURE 2. The pipeline of the proposed landmark localization method for 2D lateral cephalograms.

The next subsections will give the details of the proposed method.

A. DATA PREPROCESSING

Since the resolution and size of clinically acquired 2D lateral cephalograms are usually large, using them as training data directly is time-consuming and would have a high requirement against computing performance.

We notice that for a specific target landmark, merely the region nearby it is our concern, instead of the whole image. Inspired by this, our method crops the patches from 2D lateral cephalograms as source images before the training stage, or testing stage begins. To obtain cropped patches, we will use the template matching method in [8] to estimate the rough locations of the target landmarks in the training or testing images. Subsequently, those roughly estimated points for the target landmarks are treated as the centers of the patches cropped from the training and testing images. That will reduce the computational burden efficiently.

B. ADVERSARIAL ENCODER-DECODER NETWORK

The proposed adversarial encoder-decoder network has two components: generator and discriminator. While training the generator to estimate the distance maps of the input source images, we simultaneously train the discriminator to decide if it could consider the results produced by the generator as the real ones. We give the architecture of the adversarial encoder-decoder network in Fig. 3.

As shown in Fig. 3, the generator network has an encoder part and a decoder part. There are seven convolutional layers in the encoder part and six up-convolutional layers plus one output layer in the decoder part. And, the encoder part connects to the decoder part via four fully-connected layers. Accurately, by sequentially conducting a convolutional operation, a batch normalization (BN) and a rectified linear unit (ReLU) operation, the first convolutional layer converts the input source images into the initial feature maps. After taking similar actions as the first convolutional layer, each of the next six convolutional layers yields the feature representation from the previous layer. Unlike the other well-known networks, the encoder part, even the whole network presented in this paper, do not add any pooling layer due to its inclination to spatial resolution loss.

Next, the four fully-connected layers propagate feature representation from the seventh convolutional layer in the encoder part to the decoder part. By reshaping the feature maps of the third fully-connected layer, the fourth fully-connected layer has the same size as the seventh convolutional layer in the encoder part.

The following six layers in the decoder part are up-convolutional layers which make up-convolution operations act on the previous layer to produce new features. The difference between the first four up-convolutional layers and the last two is that as the formers use BN and ReLU operations, the latter adopts BN and Tanh operations.

While the output of ReLu operation must be more than zero, the production of Tanh operation range from -1 to 1. It indicates that Tanh operation can output a negative, which will make the features produced by Tanh operation better than the ones by ReLu operation in our network. As a proof,

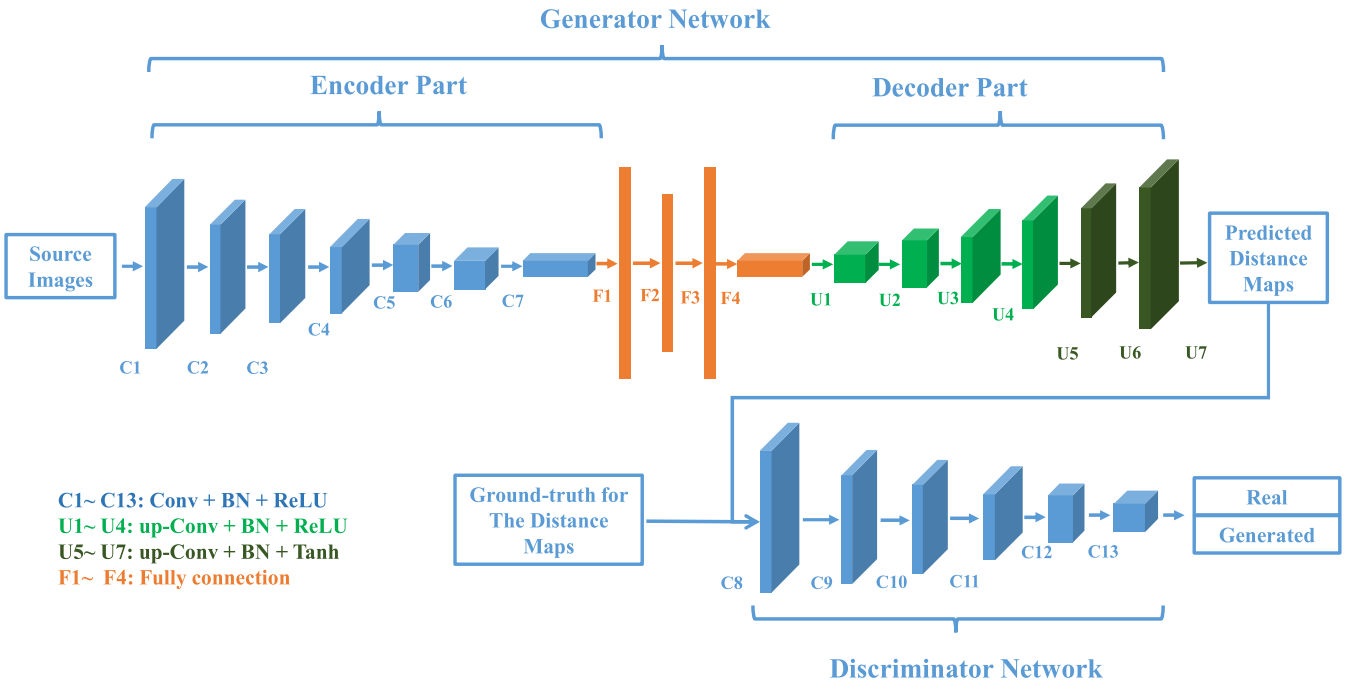
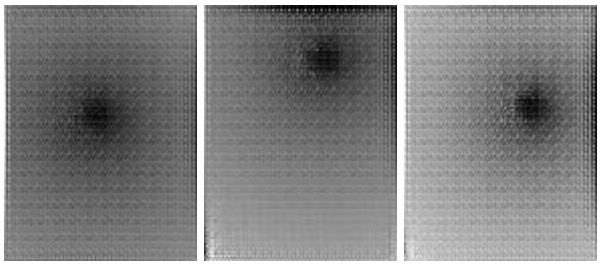
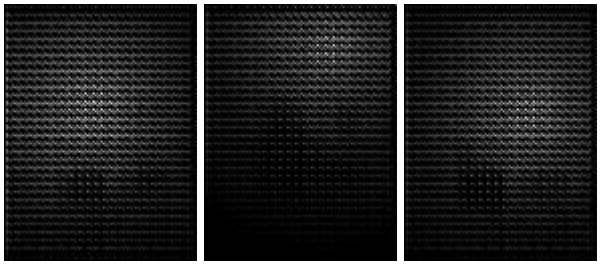


FIGURE 3. The architecture of the adversarial encoder-decoder network proposed in this paper.



(a) Some examples of the feature maps when Tanh operation is used.



(b) The corresponding feature maps when ReLu operation is used.

FIGURE 4. The comparison of some feature maps produced by Tanh and ReLu operations respectively in the fifth up-convolution layer of the decoder part. (a) is the feature maps in the fifth up-convolution layer when Tanh operation is used. (b) is the corresponding feature maps if ReLu operation substitutes for Tanh.

Fig. 4 below provides some examples of the feature maps in the fifth up-convolution layer of the decoder part which is produced through up-convolution, BN and, Tanh operations sequentially (the first row in Fig. 4). In the second row of Fig. 4, we also give the corresponding feature maps if the Tanh operation is replaced with ReLu. It is clear from Fig. 4 that the feature maps in the fifth up-convolution layer of the decoder part when using Tanh operation are closer to the real distance maps than ReLu operation.

As the final layer of the decoder, the output layer, which is built by up-convoluting the feature maps of the sixth up-convolutional layer with filters, amounts to the predicted distance maps.

Similar to the generator, the discriminator has six convolutional layers and one output layer. Having both the predicted results and their ground truth as input, we train the discriminator to provide loss gradients for the generator. The outcome of the discriminator is the probability that the predicted results are drawn from the distribution of real ones. In the convolutional layers, convolutional operation, BN, and ReLU operation are involved.

Generally, the dimensions of the input and output data are significant determinants of the network's parameters, such as the number and size of filters, stride, and padding values for each convolutional and up-convolutional operations. For the sake of simplicity, we do not make an inventory of the network's parameters here. Instead, the experiment section will give the network's settings.

C. THE TRAINING PROCEDURE

As we train the generator to predict the distance maps, the discriminator needs to be trained simultaneously to judge if it could consider the synthetic data as real ones. Only when the discriminator accepts all the generated results as real ones, will the whole training procedure stop.

The training of the discriminator D for the prediction task will utilize the favorite loss function in (1).

$$L_{\mathrm{D}} = \max_{D} \mathrm{E}_{Y}[\log(D(Y))] + \mathrm{E}_{X}[\log(1-D(G(X)))] \quad (1)$$

where X is the input source image; $G(X)$ is the generated distance map for X; and, Y is the corresponding ground-truth of $G(X)$.

As for the generator G, we first use the reconstruction loss, which measures the difference between the predicted distance map and its ground-truth. Since the reconstruction loss would lead to blur degradation of the estimated results, the loss function of the generator G also adds the term of image gradient difference (IGD) [40] as a supplement. The IGD loss tries to keep the estimated distance map with sharp edges. (2)- (4) define the loss function for the generator G.

$$L_{\mathrm{G}} = L_{\mathrm{r}} + L_{\mathrm{IGD}} \tag{2}$$

$$L_{\mathrm{r}} = \|Y - G(X)\|_2 \tag{3}$$

$$L_{\mathrm{IGD}} = \left||\nabla Y_x| - \left|\nabla [G(X)]_x\right|\right|^2 + \left||\nabla Y_y| - \left|\nabla [G(X)]_y\right|\right|^2 \tag{4}$$

with $||\bullet||_2$ being L2 distance.

In the proposed network, the training procedure for each target landmark will not stop until the loss of the generator decreases to 0.00006. At that time, the output probability of the discriminator is between 0.5 and 0.6, which means that the discriminator considers the generated results as the real ones. Besides, to avoid an infinite loop, we also set the maximum epoch in the training procedure to be 3000. In other words, even though the loss of the generator is still higher than 0.00006, our method will force the training process to stop when the epoch reaches 3000. And, we adopt the well-known Adam algorithm as the optimizer.

D. LOCALIZATION OF TARGET LANDMARKS

After the training procedure finishes, we can localize the target landmarks in the newly acquired X-ray dental images. The first step of localization is to input the new image into the trained networks to produce the distance maps of the 19 target landmarks one by one. As the intensity of the predicted distance map denotes the displacement of each pixel towards a specific landmark, we can use an approach, which is similar to regression voting method [39], to achieve the coordinates of the current target landmark from the distance map. Specifically, after deriving the displacement d of pixel $\mathbf{Z}$ towards the target landmark from the predicted distance map, every point, whose distance from $\mathbf{Z}$ is d, wins one vote from $\mathbf{Z}$. Once the voting for all the pixel in the new image finishes, the candidate point which garners the most votes is the target landmark.

III. EXPERIMENTAL RESULTS

In this section, we carry out some experiments on the database released by Automatic Cephalometric X-Ray Landmark Detection Challenge (ACXLDC), which was supported by IEEE International Symposium on Biomedical Imaging (ISBI) in 2014 [5]. The database from ACXLDC consists of 300 2D lateral cephalograms whose resolution is 1935×2400 pixels with the pixel size of 0.1mm × 0.1mm, and corresponding coordinates of the target landmarks in these cephalograms. The database from ACXLDC split all the data equally and randomly into three subsets: Training data, Test1 data, and Test2 data. In each subgroup, there are 100 2D lateral cephalograms and corresponding coordinates of the target landmarks. Note that the ACXLDC database uses 19 cephalometric landmarks which have been manually marked by two experts. Since the intensity of each pixel in a distance map is equal to the value of its displacement towards the target landmark, we artificially generate the distance maps for each target landmark in all the cephalograms by calculating the absolute Euclidean distance between each pixel and the target landmark in the cephalograms. The size of input source images cropped from the full 2D lateral cephalograms is 512×384 pixels.

The following experiments use the subset Training data and the corresponding distance maps of 19 target landmarks as training data. To keep generalization of our method, we perform k-folds cross-validation ($k = 10$). Specifically, we divide all the training images and their corresponding distance maps into ten groups randomly. For each cross-validation, our method chooses one of ten groups for validation and the other nine groups for training. And, there is no overlap between training dataset and testing dataset. Then, the subset Test1 data and the corresponding distance maps are for testing. Table 1 lists the number and size of filters, stride, and padding values for each convolutional and up-convolutional operations. The numbers of neurons used in the four fully-connected layers F1∼F4 are 12288, 400, 12288 and 12288, respectively.

We ran all the experiments on a workstation with two Intel Xeon E5-2640 v4 CPU 2.4 GHz and 64 GB memory, plus one GeForce GTX 1080 Ti GPU in the training and testing. Table 2 lists the epochs and training time for each of 19 target landmarks under this circumstance.

The batch size used in the proposed network is 100. The initial value of the learning rate is 0.001. When the loss is in the range [0.001 0.0003], we reset the learning rate as 0.0001. As the loss falls into the range [0.0003 0.0001], our method lets the learning rate is 0.00001. So long as the loss is less than 0.0001, we keep the learning rate equal to 0.000001.

Since predicting distance maps is one of the crucial steps in our method, the first a few experiments will testify the performance of our approach in prediction. Fig. 5 shows some examples of predicted patches cropped from the full distance maps, as well as the corresponding real ones. After comparing the predicted results with their ground-truth in Fig. 5, we say that all the synthetic results are visually comparable to the real ones.

Furthermore, this section also uses structural similarity index (SSIM) [41] to quantitatively assess the quality of the estimated results.

Fig. 6 gives the mean SSIM of the predicted results for 19 target landmarks. From Fig. 6, one can see that the mean SSIM values of the predicted results for all target landmarks exceed 0.93, which indicates that the proposed adversarial encoder-decoder network succeeds in preserving the structural details of the distance maps.

TABLE 1. The setting of some parameters for each convolutional and up-convolutional operation involved in the generator and discriminator. Here, C1∼C7 denote the convolutional operations used in the generator; C8∼C13 denote the convolutional operations used in the discriminator; U1∼U7 denote the up-convolutional operations used in the generator.

	Filter number	Filter size	Stride along x and y-axis (pixel)	Padding along x and y-axis (pixel)
C1	32	3×3	(2, 2)	(1, 1)
C2	32	3×3	(2, 2)	(1, 1)
C3	64	3×3	(2, 2)	(1, 1)
C4	128	3×3	(2, 2)	(1, 1)
C5	256	3×3	(2, 2)	(1, 1)
C6	512	3×3	(2, 2)	(1, 1)
C7	1024	3×3	(2, 2)	(1, 1)
C8	16	4×4	(2, 2)	(1, 1)
C9	16	4×4	(2, 2)	(1, 1)
C10	32	4×4	(2, 2)	(1, 1)
C11	64	4×4	(2, 2)	(1, 1)
C12	32	4×4	(2, 2)	(1, 1)
C13	16	4×4	(2, 2)	(1, 1)
U1	256	4×4	(2, 2)	(1, 1)
U2	128	4×4	(2, 2)	(1, 1)
U3	64	4×4	(2, 2)	(1, 1)
U4	32	4×4	(2, 2)	(1, 1)
U5	16	4×4	(2, 2)	(1, 1)
U6	16	4×4	(2, 2)	(1, 1)
U7	1	4×4	(2, 2)	(1, 1)

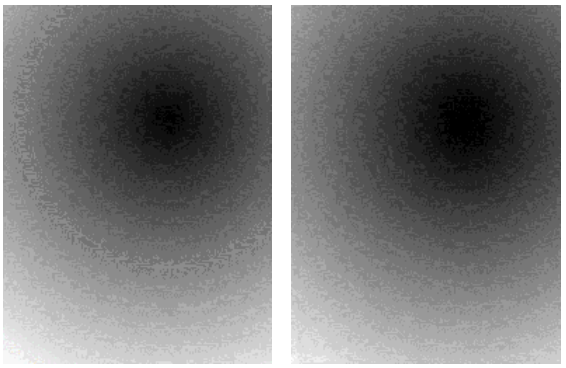
TABLE 2. The epochs and time required for the training procedures of 19 target landmarks.

No. Landmark	Epochs	Training Time (minutes)
#1	2000	440
#2	2145	472
#3	2333	513
#4	2021	445
#5	2112	465
#6	2501	550
#7	2000	440
#8	2023	446
#9	2176	479
#10	2012	443
#11	2023	446
#12	2126	468
#13	2292	504
#14	2011	442
#15	2000	440
#16	3000	660
#17	2121	467
#18	2565	564
#19	2599	572

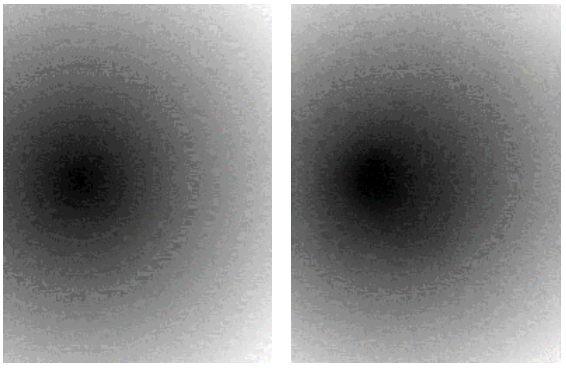
Fig. 7 provides some examples of the localized landmarks (marked in blue) and their ground-truth (marked in red) in the testing images from different patients. The results in Fig. 7 suggest that despite the morphological difference of the various patients, most of the localized landmarks seem to be close to the positions where they are supposed to be.

To quantitatively evaluate the accuracy of estimated coordinates for localized target landmarks, we adopt two frequently used measurements, namely mean radial error (MRE) in (5) and standard deviation (SD) in (6) respectively [5].

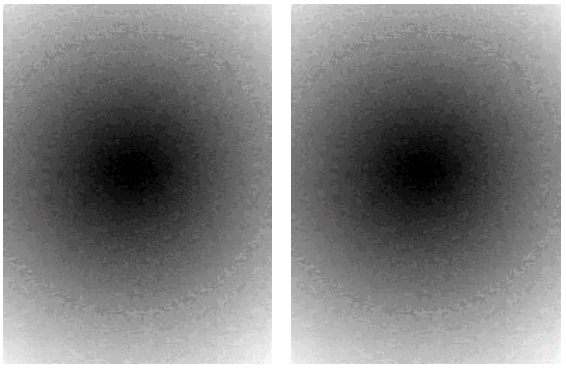
$$MRE = \frac{\sum_{i=1}^{N} R_i}{N} \tag{5}$$



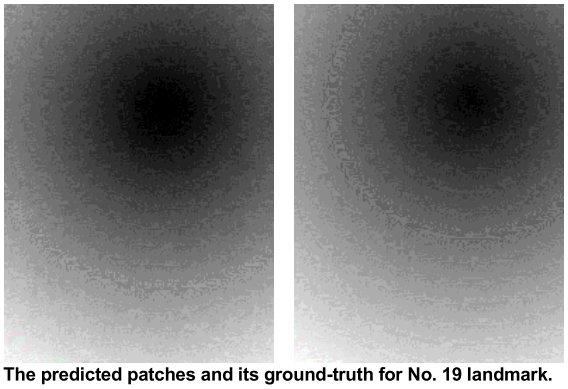
(a) The predicted patches and its ground-truth for No. 1 landmark.



(b) The predicted patches and its ground-truth for No. 7 landmark.



(c) The predicted patches and its ground-truth for No. 17 landmark.



(d) The predicted patches and its ground-truth for No. 19 landmark.

FIGURE 5. Some predicted patches cropped from the full distance maps and the corresponding real ones for No. 1, 7, 17 and 19 landmarks.

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (R_i - MRE)^2}{N-1}} \tag{6}$$

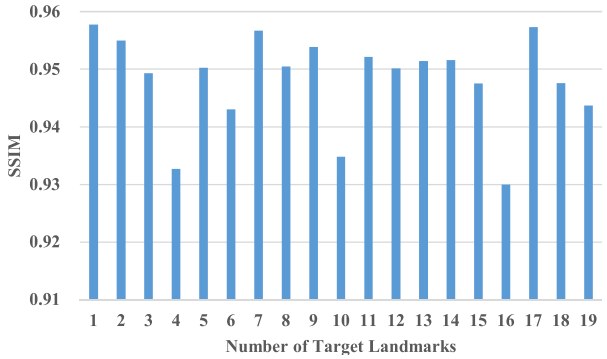


FIGURE 6. The mean SSIM of the predicted distance maps for each of the 19 target landmarks.

where R_i is the absolute Euclidean distance between the estimated coordinates of a specific target landmark and its corresponding ground-truth in i-th testing image; N is the total number of the testing images for a particular target landmark.

The evaluation also adopts success detection rate (SDR) to find out how many target landmarks can be successfully localized considering a certain margin of error. Equation (7) defines SDR [5]

$$p_b = \frac{\#\{j : \|L_{\mathrm{d}}(j) - L_{\mathrm{r}}(j)\| < b\}}{M} \times 100\% \qquad (7)$$

with L_{d} and L_{r} being the coordinates of the localized landmark and its corresponding ground-truth respectively. M is the number of all the localized landmarks; $\#\{\Omega\}$ denotes the amount of the localized landmarks which satisfy the condition Ω. As far as we know, ACXLDC first used 2mm, 2.5mm, 3mm, and 4mm as the margin of errors when calculating SDR to evaluate the submitted landmark detection methods. Since then, many localization methods, which conducted their experiments on the ACXLDC's database, also accepted the same settings for the margin of errors as ACXLDC. For the sake of comparing performance with the previous methods, we carried out the experiments on the ACXLDC's database and consequently followed the common choice of the margin of errors as well.

In Fig.8 and Fig. 9, we offer the average MRE and SD values of the localized landmarks in the testing images. It is clear that the average MRE and SD for most of the results are below about 4.0mm and 3.5mm, respectively.

And, Fig. 10 displays the average SDR values of the results when the margins of error are 2.0mm, 2.5mm, 3.0mm or 4.0mm respectively, with the comparison to the results by Chu [35] and Chen's methods [33].

Fig. 10 tells that as the margins of error rise, there is an increase of the average SDR obtained by all the methods as expected; and when the margins of error are set to 2.0mm, 2.5mm, and 3.0mm, the landmarks successfully localized by our methods are slightly less than the other two. However, the average SDR achieved by our approach is highest with the margins of error being 4.0mm. Although higher SDR at lower

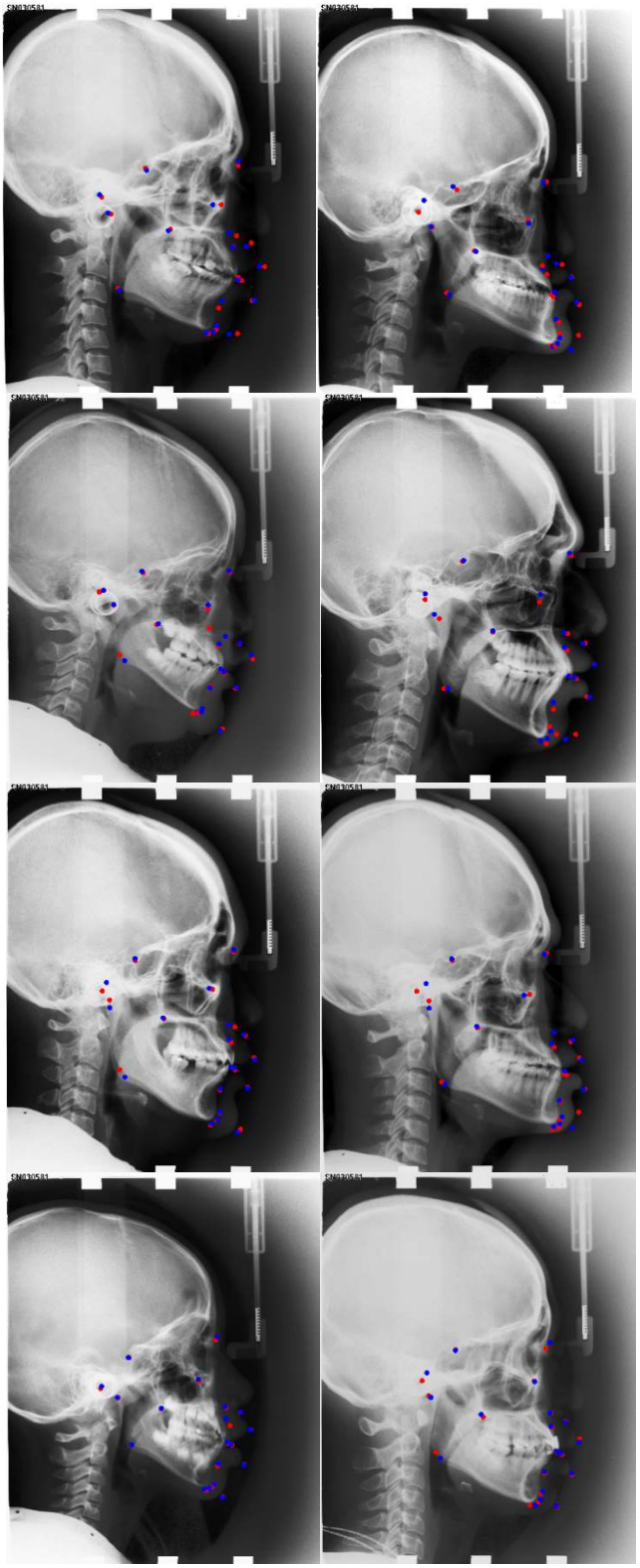


FIGURE 7. Some examples of the testing images embedded with localized landmarks and their ground-truth. We mark the localized landmarks in blue and their ground-truth in red.

margin of error, such as 2.0mm, 2.5mm, or 3.0mm, indicates the coordinates of the localized landmarks being more accurate, we think that merely considering SDR at lower

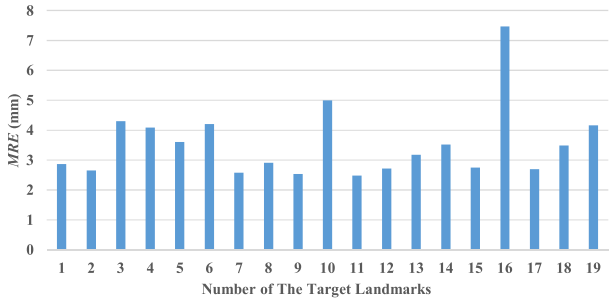


FIGURE 8. The average MRE of the localized results in the testing images.

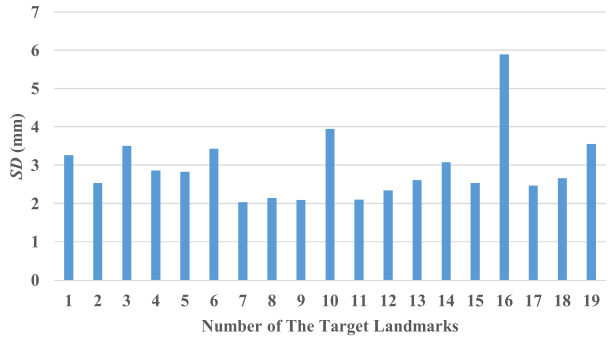


FIGURE 9. The average SD of the localized results in the testing images.

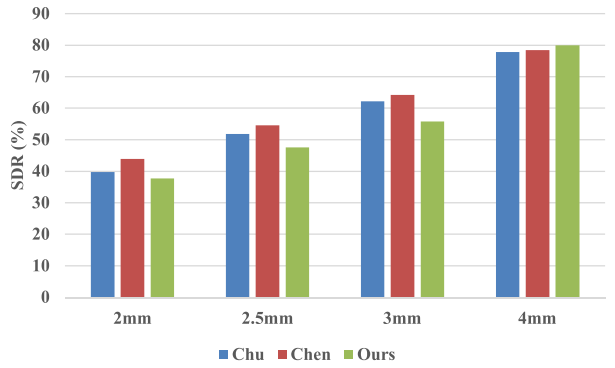


FIGURE 10. The comparison of different methods in terms of the average SDR for the localized results in the testing images when the margins of error are 2.0mm, 2.5mm, 3.0mm, or 4.0mm, respectively.

margin of error is not enough to give a comprehensive evaluation to the performance of cephalometric landmark detection algorithms. According to Yue's article [15], it considers the localized landmarks as being clinically acceptable, provided the distance between the localized results and their actual positions is less than 4.0mm. That is to say, the value of SDR with a 4.0mm margin of error suggests the proportion of the localized cephalometric landmarks which can be clinically recognized. Thus, the SDRs with the 4.0mm margin of error and lower margins of error have the same vital roles in a thorough assessment of the performance of cephalometric landmark localization algorithms.

IV. CONCLUSION AND DISCUSSION

In this paper, we propose an automated landmark localization method for 2D cephalometric X-ray images. While other existing methods estimate the coordinates of the target landmarks or the displacement vectors of each pixel towards the target landmarks directly, our approach aims to predict the distance maps of the target landmarks by constructing an adversarial encoder-decoder network. Afterward, we obtain the coordinates of the target landmarks from the predicted distance maps via an approach similar to regression voting.

The experimental results imply that our method performs well in locating most of the anatomical landmarks. However, we also observe that in Fig. 6, the mean SSIM for No. 4, No. 10 and No. 16 landmarks are the lowest three among all the landmarks. It seems that locating No. 4, No. 10, and No. 16 landmarks is more challenging than the other landmarks. In Fig.8 and Fig. 9, the relatively higher average MRE and SD values of the localized No. 10 and No. 16 landmarks confirm that it is more difficult to detect these landmarks again. Several influential review articles have also reported the difficulty of localization for No. 4, No. 10, and No. 16 landmarks [5], [6]. We will devote our future work to further investigation into this issue.

Also, since our workstation has limited computational performance, we have to crop the patches from full-size training or testing images to reduce the computational burden. However, using the cropped patches, instead of the whole images, in the training and testing step inevitably loses some context information. That would lead to degradation of performance for our method. To avoid a trade-off between performance and computational burden, we plan to incorporate a multi-resolution scheme into the proposed method in our future work.

REFERENCES

[1] R. Leonardi, D. Giordano, F. Maiorana, and C. Spampinato, "Automatic cephalometric analysis: A systematic review," *Angle Orthodontist*, vol. 78, no. 1, pp. 145–151, Jan. 2008.

[2] S. Bilici, O. Yigit, O. O. Celebi, A. G. Yasak, and A. H. Yardimci, "Relations between hyoid-related cephalometric measurements and severity of obstructive sleep apnea," *J. Craniofacial Surgery*, vol. 29, no. 5, pp. 1276–1281, Jul. 2018.

[3] R. S. Nanda and R. M. Merill, "Cephalometric assessment of sagittal relationship between maxilla and mandible," *Amer. J. Orthodontics Dentofacial Orthopedics*, vol. 105, no. 4, pp. 328–344, Apr. 1994.

[4] M. Y. Hajeer, A. F. Ayoub, and D. T. Millett, "Three-dimensional assessment of facial soft-tissue asymmetry before and after orthognathic surgery," *Brit. J. Oral Maxillofacial Surg.*, vol. 42, no. 5, pp. 396–404, Oct. 2004.

[5] C.-W. Wang, C.-T. Huang, M.-C. Hsieh, C.-H. Li, S.-W. Chang, W.-C. Li, R. Vandaele, R. Marée, S. Jodogne, P. Geurts, C. Chen, G. Zheng, C. Chu, H. Mirzaalian, G. Hamarneh, T. Vrtovec, and B. Ibragimov, "Evaluation and comparison of anatomical landmark detection methods for cephalometric X-ray images: A grand challenge," *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1890–1900, Sep. 2015.

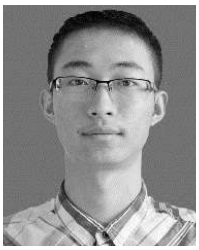
[6] C.-W. Wang, C.-T. Huang, J.-H. Lee, C.-H. Li, S.-W. Chang, M.-J. Siao, T.-M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, P. Fischer, T. F. Cootes, and C. Lindner, "A benchmark for comparison of dental radiography analysis algorithms," *Med. Image Anal.*, vol. 31, pp. 63–76, Jul. 2016.

[7] A. P. R. Durão, A. Morosolli, P. Pittayapat, N. Bolstad, A. P. Ferreira, and R. Jacobs, "Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: A comparative study," *Imag. Sci. Dentistry*, vol. 45, no. 4, pp. 213–220, Dec. 2015.

[8] I. El-Feghi, M. A. Sid-Ahmed, and M. Ahmadi, "Automatic localization of craniofacial landmarks for assisted cephalometry," *Pattern Recognit.*, vol. 37, no. 3, pp. 609–621, Mar. 2004.
[9] V. Grau, M. Alcañiz, M. C. Juan, C. Monserrat, and C. Knoll, "Automatic localization of cephalometric landmarks," *J. Biomed. Inform.*, vol. 34, no. 3, pp. 146–156, Jun. 2001.
[10] H. Mohseni and S. Kasaei, "Automatic localization of cephalometric landmarks," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2007, pp. 748–753.
[11] D. P. Jetwani, S. Kumar, and H. K. Sardana, "Cephalometric landmark identification using fuzzy wavelet edge detector," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, Bari, Italy, May 2011, pp. 349–353.
[12] T. Le-Tien and H. Pham-Chi, "An approach for efficient detection of cephalometric landmarks," in *Proc. 4th Int. Conf. Current Future Trends Inf. Commun. Technol.*, Jan. 2014, pp. 293–300.
[13] T. J. Hutton, S. Cunningham, and P. Hammond, "An evaluation of active shape models for the automatic identification of cephalometric landmarks," *Eur. J. Orthodontics*, vol. 22, no. 5, pp. 499–508, Oct. 2000.
[14] S. Rueda and M. Alcañiz, "An approach for the automatic cephalometric landmark detection using mathematical morphology and active appearance models," in *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer, Oct. 2006, pp. 159–166.
[15] W. Yue, D. Yin, C. Li, G. Wang, and T. Xu, "Automated 2-D cephalometric analysis on X-ray images by a model-based approach," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1615–1623, Aug. 2006.
[16] P. Vučinić, Z. Trpovski, and I. Šćepan, "Automatic landmarking of cephalograms using active appearance models," *Eur. J. Orthodontics*, vol. 32, no. 3, pp. 233–241, Jun. 2010.
[17] S. Sanei, P. Sanaei, and M. Zahabsaniei, "Cephalogram analysis applying template matching and fuzzy logic," *Image Vis. Comput.*, vol. 18, no. 1, pp. 39–48, Dec. 1999.
[18] W. K. Tam and H. J. Lee, "Improving point registration in dental cephalograms by two-stage rectified point translation transform," *Proc. SPIE*, San Diego, CA, USA, vol. 8314, Feb. 2012, Art. no. 83141U.
[19] I. P. Sari, R. Widayati, M. Priaminiarti, D. Danudirdjo, and T. L. Mengko, "Initial estimation of landmark location for automated cephalometric analysis using template matching method," in *Proc. 4th Int. Conf. Instrum., Commun., Inf. Technol.*, Nov. 2015, pp. 159–162.
[20] A. Kaur and C. Singh, "Automatic cephalometric landmark detection using Zernike moments and template matching," *Signal, Image Video Process.*, vol. 9, no. 1, pp. 117–132, Jan. 2015.
[21] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
[22] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[23] R. Kafieh, A. Mehri, and S. Sadri, "Automatic landmark detection in cephalometry using a modified active shape model with sub image matching," in *Proc. Int. Conf. Mach. Vis.*, Islamabad, Pakistan, Dec. 2007, pp. 73–78.
[24] S. Chakrabartty, M. Yagi, T. Shibata, and G. Cauwenberghs, "Robust cephalometric landmark identification using support vector machines," in *Proc. Int. Conf. Multimedia Expo*, Baltimore, MD, USA, Jul. 2003, pp. 825–828.
[25] A. A. Pouyan and M. Farshbaf, "Cephalometric landmarks localization based on histograms of oriented gradients," in *Proc. Int. Conf. Signal Image Process.*, Chennai, India, Dec. 2010, pp. 1–6.
[26] J.-K. Liu, Y.-T. Chen, and K.-S. Cheng, "Accuracy of computerized automatic identification of cephalometric landmarks," *Amer. J. Orthodontics Dentofacial Orthopedics*, vol. 118, no. 5, pp. 535–540, Nov. 2000.
[27] D. Giordano, R. Leonardi, F. Maiorana, G. Cristaldi, and M. L. Distefano, "Automatic landmarking of cephalograms by cellular neural networks," in *Proc. Conf. Artif. Intell. Med. Eur.*, Aberdeen, U.K., Jul. 2005, pp. 333–342.
[28] R. Leonardi, D. Giordano, and F. Maiorana, "An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images," in *Proc. BioMed Res. Int.*, Sep. 2009, Art. no. 717102.
[29] H. Lee, M. Park, and J. Kim, "Cephalometric landmark detection in dental X-ray images using convolutional neural networks," *Proc. SPIE*, Orlando, FL, USA, vol. 10134, Mar. 2017, Art. no. 101341W.
[30] J. H. Park, H.-W. Hwang, J.-H. Moon, Y. Yu, H. Kim, S.-B. Her, G. Srinivasan, M. N. A. Aljanabi, R. E. Donatelli, and S.-J. Lee, "Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD," *Angle Orthodontist*, to be published. doi: 10.2319/022019-127.1.
[31] S. Nishimoto, Y. Sotsuka, K. Kawai, H. Ishise, and M. Kakibuchi, "Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the Internet," *J. Craniofacial Surgery*, vol. 30, no. 1, pp. 91–95, Jan. 2019.
[32] B. Ibragimov, B. Likar, F. Pernus, and T. Vrtovec, "A game-theoretic framework for landmark-based image segmentation," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1761–1776, Sep. 2012.
[33] C. Chen, W. Xie, J. Franke, P. A. Grutzner, L.-P. Nolte, and G. Zheng, "Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements," *Med. Image Anal.*, vol. 18, no. 3, pp. 487–499, Apr. 2014.
[34] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, "Robust and accurate shape model matching using random forest regression-voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1862–1874, Sep. 2015.
[35] C. Chu, C. Chen, L. P. Nolte, and G. Zheng, "Fully automatic cephalometric X-ray landmark detection using random forest regression and sparse shape composition," in *Proc. Int. Symp. Biomed. Imag.*, Apr. 2014, pp. 30–38.
[36] H. Mirzaalian and G. Hamarneh, "Automatic globally-optimal pictorial structures with random decision forest based likelihoods for cephalometric X-ray landmark detection," in *Proc. Int. Symp. Biomed. Imag.*, Apr. 2014, pp. 15–27.
[37] C. Lindner, C.-W. Wang, C.-T. Huang, C.-H. Li, S.-W. Chang, and T. F. Cootes, "Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms," *Sci. Rep.*, vol. 6, Sep. 2016, Art. no. 33581.
[38] R. Vandaele, J. Aceto, M. Müller, F. Peronnet, V. Debat, C.-W. Wang, C.-T. Huang, S. Jodogne, P. Martinive, P. Geurts, and R. Maree, "Landmark detection in 2D bioimages for geometric morphometrics: A multi-resolution tree-based approach," *Sci. Rep.*, vol. 8, Jan. 2018, Art. no. 538.
[39] C. Lindner, S. Thiagarajah, J. M. Wilkinson, T. A. Consortium, G. A. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1462–1472, Aug. 2013.
[40] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.
[41] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

XIUBIN DAI was born in Zhenjiang, Jiangsu, China, in 1980. He received the B.S. degrees in biomedical engineering from the Huazhong University of Science and Technology, China, in 2002, and the M.S. and Ph.D. degrees in biomedical engineering from Southeast University, China, in 2009.

From 2008 to 2010, he held a postdoctoral position with Southeast University. Since 2010, he has been an Assistant Professor with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications. His research interests include medical image reconstruction, medical image segmentation, and pattern recognition.



HAO ZHAO was born in Anqin, Anhui, China, in 1994. He received the B.S. degrees in information engineering from the East China Jiaotong University, China, in 2017.

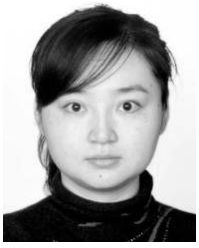
He is currently pursuing the master's degree with the Nanjing University of Posts and Telecommunications. His research interest includes medical image processing and analysis.

TIANLIANG LIU was born in Xinfeng, Jiangxi, China, in 1980. He received the B.S. degrees in telecommunication engineering from the Jiangxi Normal University, China, in 2003, and the M.S. and Ph.D. degrees in biomedical engineering from Southeast University, China, in 2010.

Since 2009, he has been an Assistant Professor with the School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interests include pattern recognition and image processing.

DAN CAO was born in Xuzhou, Jiangsu, China, in 1988. She received the B.S. and M.S. degrees in stomatology from Nanjing Medical University, China, in 2015.

Since 2015, she has been with the Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University. Her research interests include cephalometric analysis and dental X-ray image processing.



LIZHE XIE was born in Chengdu, Sichuan, China, in 1985. She received the B.S. and M.S. degrees in biomedical engineering from Sichuan University, China, in 2007, and the Ph.D. degree in biomedical engineering from Southeast University, China, in 2012.

Since 2012, she has been with the Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University. Her research interest includes medical image processing and analysis.