CEPHALOMETRIC LANDMARK 10(GONION) DETECTION Via 3-STAGE CNNs

Cheng Ta Huang[1,2,*] Tzu Wang[1] Yong-Qi Li[2]
[1] Department of Information Management, Yuan Ze University, Taoyuan, Taiwan
[2] International Bachelor Program in Informatics, Yuan Ze University, Taoyuan, Taiwan

[*]cthuang2020@saturn.yzu.edu.tw

ABSTRACT

Cephalometric analysis is a crucial application which is frequently used by dentists, orthodontists, and oral and maxillofacial surgeons as a treatment planning tool. The tracing of landmark locations on lateral cephalograms plays a critical role in the clinical analysis, but the traditional way to trace those landmarks is time-consuming and error-prone, especially the landmark 10(Gonion). To solve this problem, we propose a new 3-stage CNNs to detect the landmark 10(Gonion). The structure behind our approach is different from the conventional ones. We trained three different models to detect the features of face area and the landmark 10. Firstly, we use the categorical model to get the face area. Secondly, we process the first regression model to estimate the surrounding area of the landmark 10. Lastly, we predict the more accurate landmark coordinates with the last regression model. The experimental results show the good performance of our method in the landmark 10 detections.

Keywords: deep learning, cephalometric landmarks, convolutional neural networks, X-ray image applications

1. INTRODUCTION

Cephalometric analysis is a crucial application which is widely used by dentists, orthodontists, and oral and maxillofacial surgeons. It gives valuable information of the patients' bony, dental, and soft tissue structures from cephalograms. Based on the cephalometric analysis, dentists can provide the diagnosis and the treatment planning for the patients. In clinics, the dentist needs to draw the landmarks on a 2D cephalometric X-ray image manually, so the dentist can calculate the distance and the angle of the landmarks and discuss the treatment with the patient. But the manual method of cephalometric analysis has some problems. Owing to the low quality of X-ray cephalogram, it is such a time-consuming and error-prone work for dentists who mark those landmarks manually. Even for the senior orthodontists, it takes so much time and it is hard to do it precisely, let alone the less experienced ones. To overcome this problem, automatic landmarks detection on lateral cephalograms is in urgent need.

There have been many methods developed for cephalometric landmark detection since the last decades. The past approaches include edge detection, tree-based method, modified U-net [1], Histograms of Oriented [2], Encoder-Decoder Networks [3], CNN app-roaches and so on. Among those structures, CNN technique has achieved the greatest success in a wide range of computer vision applications, including image classification, object detection etc. Due to these big achievements, the researchers began to use this CNN framework to extract some useful features from images and now this kind of framework has gone viral. But we have noticed that those past approaches have difficulty to detect the landmark 10(Gonion) precisely; thus, this paper is devoted to resolve this problem.

In this paper, we think of cephalometric landmark detection as a multi-stage classification and regression task and we proposed a state-of-the-art 3-stage CNNs method to solve it. At the first stage, we used a classification model to detect the face area in the raw X-ray image. It helped us to narrow down the face area; therefore, we can get rid of the useless zone, which can probably affect our prediction of the landmark. Given the face area, at the second stage we used a regression model to predict the landmark 10(Gonion) preliminarily. We adopted the same concept as the first stage to get rid of the useless area and to reduce the complexity of the whole face area; otherwise, it can have a bad influence on the final result. At the last stage, we have acquired the more specific and higher resolution of the landmark 10(Gonion) area, so we can execute the last regression model to estimate the final coordinates of it.

We tested our framework's effectiveness on the IEEE 2015 ISBI Grand Challenge and we have proved that our framework is feasible and it is comparable to other methods in cephalometric landmark detection.

The following context, as below, gives more details of how we built this framework and how we tested our effectiveness. In section 2, we introduced some related works. In section 3, we explained our 3 stages thoroughly one by one. In section 4, we showed how experimental results and in the last section we made a conclusion of this research.

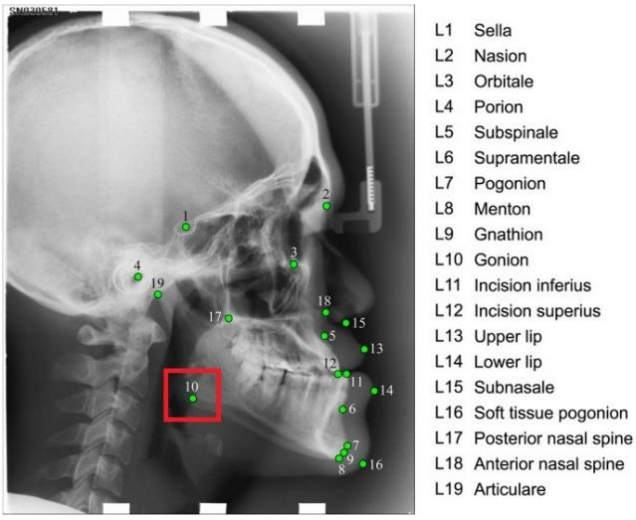


Figure 1 The example of the 19 cephalometric landmarks. We are going to find the landmark 10(Gonion).

2. Related Works

2.1 Traditional methods

In the last decades, there were so many different methods to detect the cephalometric landmarks. The conventional approaches include edge detection [8] which proposed a template matching way to use the features which was computed by image edge detection and contour segmentation operators for automated identification of cephalograms' landmarks. Although edge detection techniques can automatically mark the landmarks, they failed to detect the landmarks when the landmarks are not lying on the edges. To solve the problem like this, some paper used applied active models, such as the active shape model (ASM) [9] and the active appearance model (AAM) [10]. In ASM methods, 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 the targets.

2.1.1 Modified U-net [1]

They proposed a model which modified U-net structure [1]. In the contracting path, every convolution layer with rely is accompanied by a batch normalizer layer which would avoid internal covariance shift. The filter sizes of convolutions were modified as per the train data requirements. Max pooling was replaced by average pooling with a stride value of two. The contracting part was modified by using batch normalizer after every convolution and the remaining up sampling part is left undisturbed.

2.1.2 Encoder-Decoder Networks [3]

They proposed a new automated cephalometric landmark localization method [3] under the framework of GAN. They trained an adversarial network under the framework of GAN to learn the mapping from features to the distance map of a specific target landmark. 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, they simultaneously train the discriminator to decide if it could consider the results produced by the generator as the real ones.

2.1.3 CNN models [10], [11]

In recent years, although there is a lack of training images in the medical imaging fields, many CNN models were still widely used to solve the landmark detection work, because it has such a great performance and effectiveness compared to other traditional methods. Some people treated cephalometric landmark detection as a regression problem and proposed a single convolutional neural network structure to learn the features of all the landmarks, but it is very difficult to be optimized [10]. Some people proposed a framework which firstly used a cephalometric neural network to learn the probability whether the input images' patch center is the landmarks, and combined with a statistical shape model to refine all the landmarks' accurate positions [11].

Although those past methods can detect the cephalometric landmarks automatically, we found that they have difficulty to detect the landmark 10(Gonion) successfully and precisely. We chose CNN framework to solve this problem, because CNN models had a better result compared to other approaches.

3. Methods

3.1 Overall structure

We proposed a 3-stage structure to detect the landmark 10(Gonion). The structure is based on CNN framework. We used one classification model and two regression models to deal with this task. The first stage is designed to crop the face area which will be the input image for the second stage. The second stage is designed to detect the preliminary L10 area in the given face area. The last stage is designed to predict the final coordinate of L10. The overall structure is shown in Fig. 2.

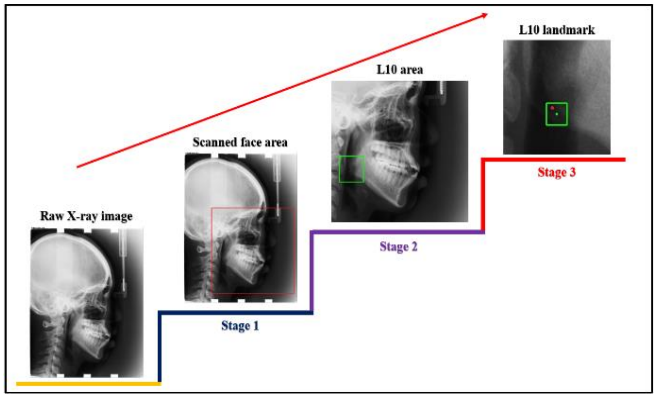


Figure 2 The 3-stage overall structure.

The following details of each stage are as below.

3.2 Stage-1

The first stage is designed to crop the face area to get rid of some useless zones, because those useless zones can affect the effectiveness of the final result; besides, cropping the face area can reduce the complexity of image, which can help our model to learn the features better.

We classify the raw image to five different areas which is the head area Fig.3, the forehead area Fig.4, the spine area Fig.5, the mouth area Fig.6 and the face area Fig.7. The reason why we classified the raw image to five different categories, it is because the face area is such a big area which is connected to other areas and can be hard to detect precisely. For instance, if our model does not include the forehead area, the prediction might detect the forehead area instead of the correct face area, similarly for other areas.

Figure 3 The head area.

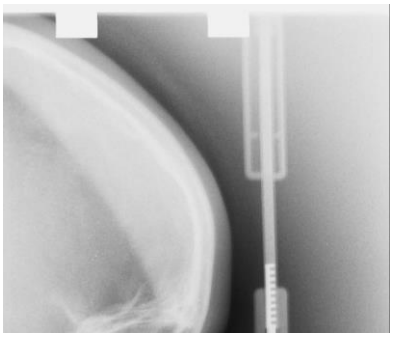


Figure 4 The forehead area.

Figure 5 The spine area.

Figure 6 The mouth area.

Figure 7 The face area.

We created a classification model for this stage and the model is using Cross Entropy as the loss function. The Cross Entropy is shown in (1).

$$CE(x) = \sum_{i=1}^{C} y_i \log f_i(x) \qquad (1)$$

Equation 1 Cross Entropy loss function

In Equation 1, x means the input, C means the number of the categories, y_i means the true label and $f_i(x)$ means the output. At this stage, we will resize our raw X-ray picture to 120*150 size and input it into the first model which is composed of three convolutional layers with 3*3 filters, followed by two fully connected layers. The detailed structure of stage-1 is displayed in Fig.8.

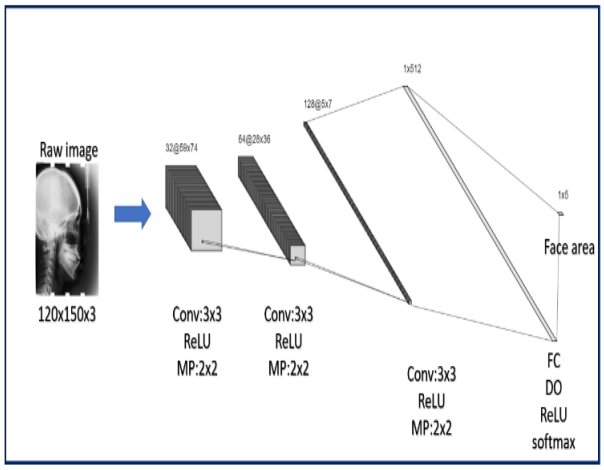


Figure 8 Stage-1 framework.

3.3 Stage-2

The second stage can be thought of as our initial prediction of the landmark 10(Gonion). In this stage, we will extract the L10 area from the face area which is given by stage-1. The structure of this stage is a regression model with Euclidean distance loss function. The Euclidean function is shown in (2).

$$EDL = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (2)$$

Equation 2 Euclidean distance loss function

In Equation 2, x_1 and y_1 mean the predicted coordinates of L10 and x_2 and y_2 mean the ground truth coordinates of L10. At this stage, we will resize the face area picture to 96*96 and input it into the regression model which is made up of three convolutional layers which filters are all 3*3, followed by two fully connected layers. The detailed structure is shown in Fig. 9.

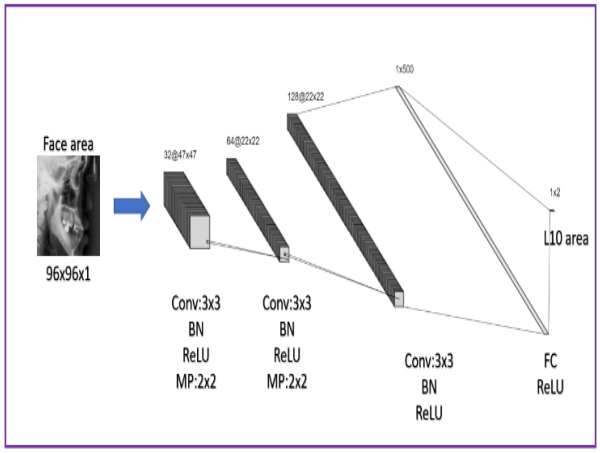


Figure 9 Stage-2 framework.

3.4 Stage-3

Due to the first two stages, we have acquired the area of L10 and we have substantially reduced the complexity of the raw X-ray image. With the previous stages, we now can predict the final coordinates of L10

precisely. The structure of this final stage is also a regression model with Euclidean distance loss function. (2).

Before starting the final stage, we will use data augmentation to get more training data to help us not to fall into the dilemma of overfitting. The way of how we increase our training data will be explained in the section 3.5. After preparing enough training data, we will resize our L10 area image to 48*48 and input it into the regression model which consists of three convolutional layers with 3*3 filters, followed by two fully connected layers. The detail of this final stage is shown in Fig.10.

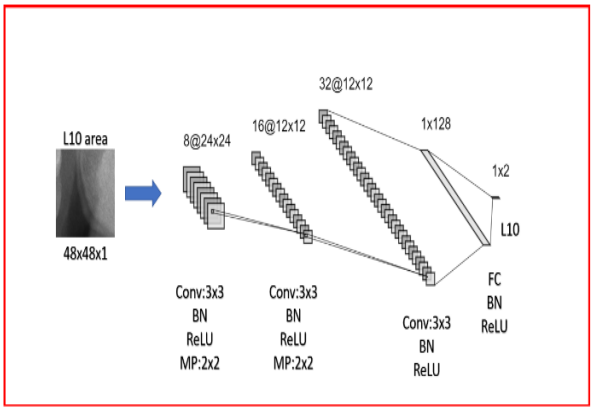


Figure 10 Stage-3 framework

3.5 Data augmentation

At stage-3, we found that there was a problem of overfitting because of the lack of training data; Thus, we decided to adopt the method of data augment to deal with this overfitting problem. We randomly tuned the area of L10 in the range of [-50,50], and we generated 30 images for each original L10 area picture, which are 4500 training images in total. The example is shown in Table 1.

4. Experiments

4.1 Evaluation Equation

In this 3-stage structure, we used the Euclidean Distance equation to estimate our Success Detection Rate (SDR) which means the detection distance is within the length. The definition of the SDR is shown in (3).

$$SDR = \frac{\#\{j: \|L_p(x) - L_g(x)\| < \beta\}}{\#\omega} \qquad (3)$$

Equation 3 Euclidean Distance

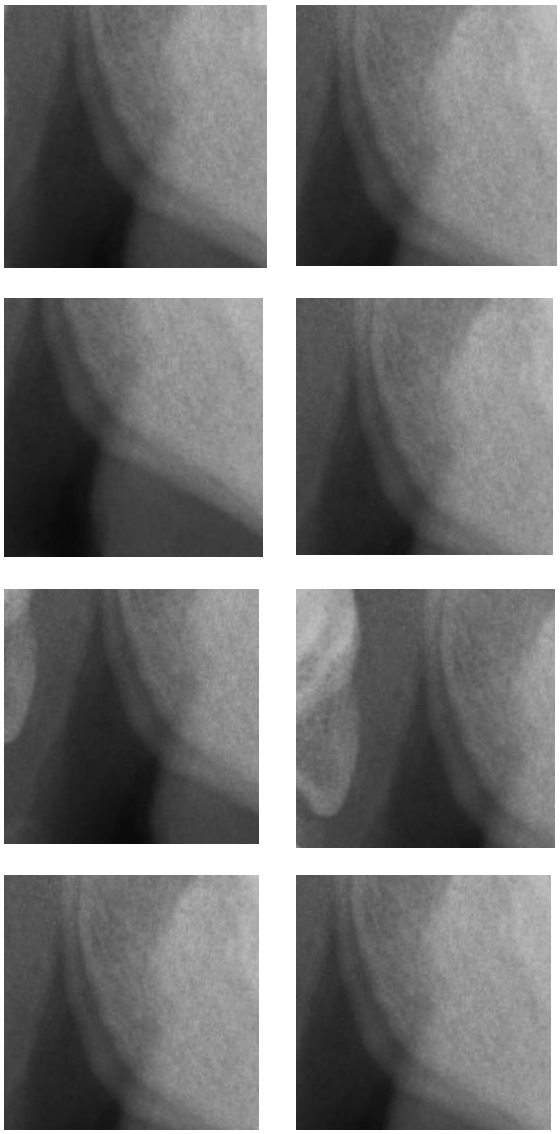
In Equation 3, L_P and L_g mean the predicted and the ground truth landmarks of the xth image. β represents the success detection distance, including 2 mm, 2.5mm, 3mm and 4mm. $\#\omega$ means the total number of testing data. With this function, we can get the SDR to evaluate our experiment results.

4.2 Dataset

In this paper, we used IEEE 2015 ISBI Grand Challenge #1: Automated Detection and Analysis for Diagnosis in Cephalometric X-ray Image. The dataset contains 400 cephalometric X-ray images, which were collected from 400 different patients and acquired by Soredex CRANEX@Excel Ceph machine (Finland) and Soredex SorCom software (3.1.5, version 2.0). The resolution of the images is 2400*1935 pixels, while the pixel spacing is 0.1mm/pixel in each dimension. For detection, 19 landmarks were marked by two senior medical doctors for each image. The ground truth landmarks are the average of those two doctors' marks.

We separated the dataset to three subsets which are training dataset (150 images), testing dataset 1 (150 images) and testing dataset 2 (100 images). We trained our model with training dataset and tested it with those two testing datasets.

Table 1 Some examples of data augmentation images



4.3 Experimental results

After training and testing the model, we have acquired our detection results. We have proved our 3-stage structure's effectiveness, which is shown in Table 1. The SDRs in the range 2mm are 49% and 45%, 2.5mm 60% and 61%, 3.0mm 74% and 67% and 4.0mm 99% and 80%, respectively in testing dataset 1 and 2. The average distances of testing dataset 1 and 2 are 3.68mm and 3.51mm. We will compare our model with other two models, Table 2 to 4. In those three tables, we can know that our 3-Stage CNN structure is better than the others. In Table 2, we compared the SDR with Modified U-net. They have only 21.3% in 2.0mm, 37.6% in 3.0mm and 51.6% in 4.0mm, but we have 49% in 2.0mm, 74% in 3.0mm and 99% in 4.0mm. In Table 3, we compared the average result with Encoder-

Decoder Networks. The average of their results is 4mm and our average of results is 3.59mm; Therefore, we can see the effectiveness of our model apparently.

Table 2 L10 detection results.

Test	SDR(%)				Average
	2.0mm	*2.5mm*	*3.0mm*	*4.0mm*	
1	49%	60%	74%	99%	3.68mm
2	45%	61%	67%	80%	3.51mm

Table 3 L10 SDR compared with Modified U-net

Model	Modified U-net			The proposed method 3-Stage CNN		
	2.0mm	*3.0mm*	*4.0mm*	*2.0mm*	*3.0mm*	*4.0mm*
Test1	21.3%	37.6%	51.6%	49%	74%	99%

Table 4 L10 average result compared to Encoder-Decoder Networks [3].

Model	Encoder-Decoder Networks	The proposed method 3-Stage CNN
Average	4mm	3.59mm

5. CONCLUSION

In this paper, we proposed this 3-stage structure for the cephalometric landmark 10(Gonion) detection. We used three different CNN models which are one classification model and two regression models to detect the landmark 10(Gonion). We wanted to overcome the error-prone problem of the detection of the landmark 10(Gonion) which happened frequently in other structures.

The experimental results display that our structure has such a great performance compared to other structures which have only 20~30% SDR in the range of 2mm and the average distance 4~5mm.

Although we have such a great performance, we still find it difficult to predict the landmark 10(Gonion) with the high SDR as other landmarks. We think the reason why we cannot reach such a high SDR like other landmarks, it is because the feature of the landmark 10(Gonion) is very different in the 400 datasets. For instance, we found that some of the landmark 10(Gonion) in the dataset are overlapping with the spine but some of them are not. We think this is the main reason why it is really difficult to detect this landmark as precisely as other landmarks. We will show the two examples of the normal L10 and the L10 overlapping with spine in Fig.11 and 12.

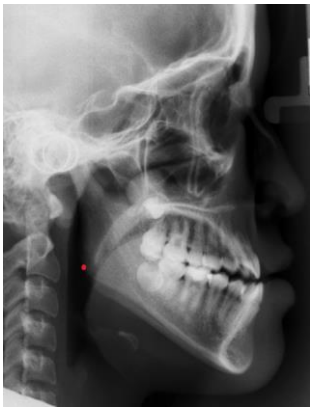


Figure 11 the example of normal landmark 10.

Figure 12 the example of the landmark 10 overlapping with the spine.

6. REFERENCE

[1] E. N. D. Goutham, S. Vasamsetti, P. V. V. Kishore and H. K. Sardana, "AUTOMATIC LOCALIZATION OF LANDMARKS IN CEPHALOMETRIC IMAGES Via MODIFIED U-Net," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944411.

[2] A. A. Pouyan and M. Farshbaf, "Cephalometric landmarks localization based on Histograms of Oriented Gradients," 2010 International Conference on Signal and Image Processing, 2010, pp. 1-6, doi: 10.1109/ICSIP.2010.5697431.

[3] X. Dai, H. Zhao, T. Liu, D. Cao and L. Xie, "Locating Anatomical Landmarks on 2D Lateral Cephalograms Through Adversarial Encoder-Decoder Networks," in IEEE Access, vol. 7, pp. 132738-132747, 2019, doi: 10.1109/ACCESS.2019.2940623.

[4] M. Farshbaf and A. A. Pouyan, "Landmark detection on cephalometric radiology images through combining classifiers," 2010 17th Iranian Conference of Biomedical Engineering (ICBME), 2010, pp. 1-4, doi: 10.1109/ICBME.2010.5704950.

[5] J. Keustermans, W. Mollemans, D. Vandermeulen and P. Suetens, "Automated Cephalometric Landmark Identification Using Shape and Local Appearance Models," 2010 20th International Conference on Pattern Recognition, 2010, pp. 2464-2467, doi: 10.1109/ICPR.2010.603.

[6] C. -W. Wang et al., "Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge," in IEEE Transactions on Medical Imaging, vol. 34, no. 9, pp. 1890-1900, Sept. 2015, doi: 10.1109/TMI.2015.2412951.

[7] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, June 2001, doi: 10.1109/34.927467.

[8] D. P. Jetwani, S. Kumar and H. K. Sardana, "Cephalometric landmark identification using fuzzy wavelet edge detector," 2011 IEEE International Symposium on Medical Measurements and Applications, 2011, pp. 349-353, doi: 10.1109/MeMeA.2011.5966733.

[9] 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.

[10] R. Kafieh, A. Mehri and S. Sadri, "Automatic landmark detection in cephalometry using a modified active shape model with sub image matching", *Proc. Int. Conf. Mach. Vis.*, pp. 73-78, Dec. 2007.

[11] Lee, H., Park, M., Kim, J., 2017. Cephalometric landmark detection in dental X-ray images using convolutional neural networks. In: Medical Imaging 2017: Computer-Aided Diagnosis, 10134. International Society for Optics and Photonics, p. 101341W.

[12] D. Giordano, R. Leonardi, F. Maiorana, G. Cristaldi and M. L. Distefano, "Automatic landmarking of cephalograms by cellular neural networks", *Proc. Conf. Artif. Intell. Med. Eur.*, pp. 333-342, Jul. 2005.

[13] R. Leonardi, D. Giordano and F. Maiorana, "An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images", *Proc. BioMed Res. Int.*, Sep. 2009.