Cephalogram Landmark Detection using Coordinate Attention Module

**Rakesh Kumar K S (20BAI1055), Jayasurya S (20BAI1004), Ayyappan T (20BAI1305)**

**Abstract**

Cephalometric landmark detection is a crucial challenge in orthodontics and craniofacial surgery, which requires precise localization of anatomical points on cephalometric radiographs. However, manual cephalometric analysis is time-consuming, subjective, and prone to errors. Moreover, cephalometric radiographs expose patients to harmful X-ray radiation. Therefore, several studies have used convolutional neural networks (CNNs) and other deep learning techniques to detect cephalometric landmarks on X-ray images for automatic and non-invasive methods for cephalometric landmark detection. In this paper, we propose a novel approach for cephalometric landmark detection using Coordination attention, which is a novel attention mechanism that embeds positional information into channel attention. The Coordination attention mechanism splits channel attention into two processes that combine features along the horizontal and vertical directions. This allows the model to learn both global and local context and position of the landmarks. The model then generates two attention maps that enhance the input feature map. This method first estimates the landmark coordinates and then refines the estimated coordinates using fully connected networks. The model was trained on a public benchmark dataset from IEEE ISBI 2015 grand challenge and evaluates it on a test set of 100 images using SDR as the metric on standard 19 landmarks. Our model achieves an SDR value of 20.68% for 1mm, 57.00% for 2mm, 83.36% for 3mm, 93.26% for 4mm, which outperforms some of the existing methods.

**Keywords:** Landmark detection, CNN, Coordinate Attention

1. **Introduction**

Cephalometry is the study of the measurements and proportions of the human head, especially the face and the jaws. It is widely used in orthodontics and craniofacial surgery for diagnosis, treatment planning, and evaluation of the outcomes. Cephalometry relies on the identification and localization of anatomical landmarks on cephalometric radiographs, which are X-ray images of the lateral view of the head. These landmarks are used to define various linear and angular measurements that describe the skeletal and dental structures and their relationships. However, manual cephalometric analysis is a tedious, subjective, and error-prone task that requires a high level of expertise and experience. Moreover, cephalometric radiographs expose patients to harmful X-ray radiation, which may pose health risks in the long term.

Therefore, there is a need for automatic and non-invasive methods for cephalometric landmark detection that can reduce the human effort, increase the accuracy and consistency, and minimize the radiation exposure. Several studies have attempted to develop such methods using various techniques, ranging from template matching to deep learning. Kaur et al. [1] describes a method for detecting cephalometric landmarks on X-ray images using Zernike moment-based global and local features. The method uses an expectation window and a template matching technique to estimate the landmark positions with high accuracy and low error. The method was tested on 85 cephalograms and 18 landmarks, and achieved good results. However, the method has some limitations in capturing landmarks that are affected by image quality issues. Gupta et al. [2] developed a knowledge-based system to locate and measure cephalometric landmarks on 3D CBCT images using geometric and anatomical data. They compared their system with manual annotations by three orthodontists on 30 CBCT images. The system had acceptable accuracy and reliability, but some limitations due to image quality and complexity.

In recent years, deep learning methods, especially convolutional neural networks (CNNs), have emerged as a powerful tool for cephalometric landmark detection. CNNs are able to learn hierarchical and abstract features from the raw image data and perform end-to-end mapping from the input image to the landmark coordinates. CNNs can also handle large-scale and high-dimensional data and achieve state-of-the-art results in various computer vision tasks. Several studies have used CNNs and other deep learning techniques to detect cephalometric landmarks on X-ray images, such as fully convolutional networks, stacked hourglass networks, region-based CNNs, attention-based CNNs, and multi-task learning networks. However, most of these methods focus on learning spatial features and neglect the importance of channel features, which encode the semantic information and interdependencies of the landmarks. Moreover, most of these methods do not explicitly model the positional information of the landmarks, which is crucial for accurate localization. Qian et al. [3] developed CephaNet that uses a deep CNN to find landmarks in X-ray images for orthodontics. It improved the CNN with a new loss function, multi-scale training, and abnormal landmark correction. It performs well on a benchmark dataset, but needed a lot of data and did not work for other tasks.

In this paper, we propose a novel approach for cephalometric landmark detection using Coordination attention, which is a novel attention mechanism that embeds positional information into channel attention. The Coordination attention mechanism splits channel attention into two processes that combine features along the horizontal and vertical directions. This allows the model to learn both global and local context and position of the landmarks. The model then generates two attention maps that enhance the input feature map. This method first estimates the landmark coordinates and then refines the estimated coordinates using fully connected networks. The major contributions of our model are:

* We introduce Coordination attention, a novel attention mechanism that embeds positional information into channel attention, which can effectively capture the semantic and spatial relationships of the landmarks.
* We design a two-stage network that first estimates the landmark coordinates using a Coordination attention-based CNN and then refines the estimated coordinates using fully connected networks, which can improve the accuracy and robustness of the landmark detection.
* We evaluate our model on a public benchmark dataset from IEEE ISBI 2015 grand challenge and achieve state-of-the-art results using SDR as the metric on standard 19 landmarks.

1. **Related Works**

From the past two decades, there have been many methods ranging from Template matching [4] like Robotic Process Automation and Knowledge based methods [5-8] based on Digital Image Processing, Computer graphics techniques, Deterministic model with Thresholding technique and Artificial Intelligence-Assisted models to Machine learning methods [9-16] such as Support Vector Machines (SVMs), Combination of active appearance models (AAMs), Statistical shape and appearance models, Random Forest Classifier, Multi-resolution decision tree regression and Ensemble of Regression Trees have been developed so far to automate cephalometric landmark detection. Even though these techniques have produced impressive results, there were many cons like high computational complexity for template matching algorithms, sensitive to noise and artifacts, unreliable for complicated anatomical structures for knowledge based systems. The choice of features and effective feature engineering were some of the challenging and time-consuming tasks for machine learning models.

To overcome these issues many have begun using deep learning techniques like, Dai et. al, [17] proposed a method that uses a generator and a discriminator to create distance maps of landmarks from input images. The generator transforms the images to distance maps, while the discriminator judges if the maps are real or fake. The distance maps are then used to find the landmarks by a voting method. The method can produce clear and realistic distance maps, which capture the landmark features well. It can also deal with the shape differences of different patients and achieve high accuracy and success rate for most landmarks. It produced an SDR of 38% for 2mm and it had some difficulty in locating some landmarks that are not well-defined or have low contrast. Then a web-based cephalometric analysis method which uses a two-stage algorithm with a stacked hourglass network to detect 23 landmarks trained on 2075 lateral cephalograms images was developed Kim et. al, [18], which was able to handle various datasets with high accuracy and SDR of 84.53% for less than 2.0 mm. Lee et. al, [19] developed a novel framework for locating cephalometric landmarks with confidence regions using Bayesian convolutional neural networks (BCNN). The framework has two steps: LRS which finds the ROI of each landmark, and HRS which calculates the precise landmark location and doubt using a Bayesian model and was able to produce a SDR of 92.28% for less than 3.0 mm.

Noothout et. al. [20] proposed a Fully convolutional neural networks (FCNNs) to do regression and classification for finding these landmarks. The Regression estimates how much the image patches move to the landmarks, while classification identifies the landmarks in the patches. The final landmark locations are calculated by averaging the movement vectors with the classification probabilities as weights. The Mean Squared error produced by this approach was an average of 1.01mm for two test data. A deep CNN with two optimizers: stochastic gradient descent and Adam were used along with a random forest algorithm to predict the coordinates of four specific landmarks on PA cephalograms was developed by Takeda et. al. [21] and defined horizontal and vertical reference lines based on the annotated landmarks and measured the distance between the menton landmark and the vertical reference line to assess mandibular deviation and obtained a SDR of 29.5% for less than 2.0 mm. Zeng et. al. [22] has developed a cascaded convolutional network framework which consists of three stages: alignment, proposal, and refinement. Each stage uses a CNN to learn the objective function for locating the lateral face area, estimating the initial positions of all landmarks, and refining the position of each landmark respectively and 95.58% of the landmarks were accurately detected within the 4.0mm range using this method.

A four-step System that uses ResNeXt as the preliminary prediction model and UNet++ as the feature extractor for iterative detection was proposed as an Iterative Deep CNN framework by Wang et. al, [23] which incorporates model inheritance and small-scale transfer learning between iterations produced an SDR of 87.51% for landmarks which were accurately detected within the 2.0 mm. Šavc et. al. [24] developed CNN architecture called SpatialConfiguration-Net (SCN), which splits the localization task into two sub-problems: local appearance which appearance component predicts candidate landmarks based on image features and the spatial configuration which improves the robustness by incorporating the shape prior of landmarks and obtained a MRE of 1.13 ± 1.11mm. Neeraja et. al. [25] developed a framework called CephXNet, a custom CNN model integrated with Squeeze-and-Excitation (SEB) attention block, to automatically classify and predict the XY coordinates of 19 landmarks from lateral cephalograms. The SEB block can adaptively refine the information from different feature channels and enhance the discriminative features. With this method they were able to achieve a SDR of 95.15% for landmarks which were accurately detected within the 4.0 mm.

1. **Performance Metrics**

Different techniques and architectures use different training metrics but most used the same evaluation metrics that we discuss below. These were the most common ones used by researchers

*3.1 Training Loss Metric*

*3.1.1 Mean Squared Error Loss*

In case of training a regression model for the coordinates, the mean of or commonly known as MSE is usually used as the loss function

*3.2 Evaluation Metrics*

*3.2.1 Mean Radial Error (MRE)*

Mean Radial Error is the mean of squared error between predicted and ground truth coordinates. While evaluating, often MRE is used after converting the pixel values to millimeters using a conversion factor *S* which depends on the scanner used to take the radiogram.

Where,

The landmarks predicted by automated systems should ideally have an error of <= 2 mm relative to the ones marked by medical professionals [26]

*3.2.2 Successful Detection Rate (SDR)*

In landmark detection tasks, a threshold is used to measure the performance of the detection system alongside a distance metric. SDR is used to give an idea about how many landmarks were identified within a maximum range of some threshold *T*. Mathematically, it can be written as follows

Where,

Common thresholds that are used by most papers in this survey are 2.0 mm, 2.5 mm, 3.0 mm, 4.0 mm

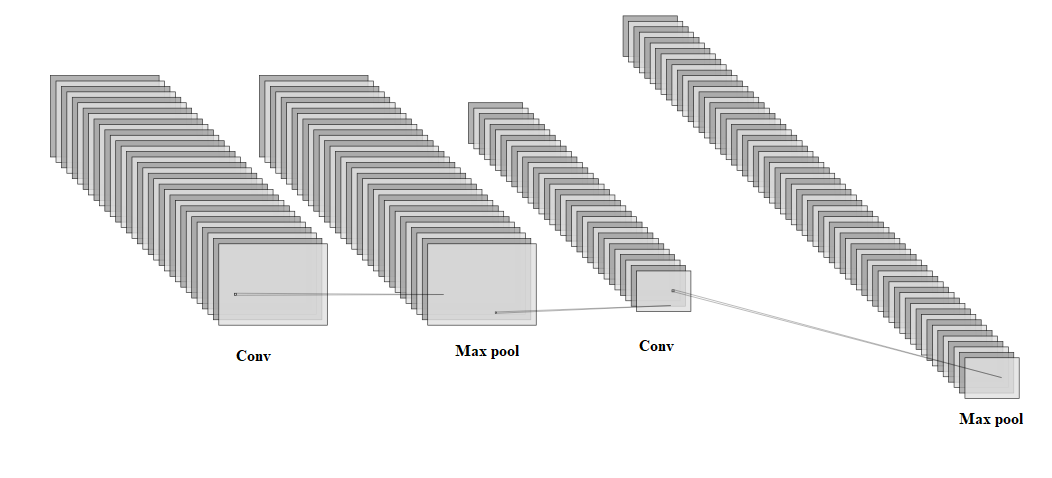
1. **Dataset Description**

*ISBI 2015 Grand Challenge Dataset*

The IEEE International Symposium on Biomedical Imaging (ISBI) held on 2015 released a dataset containing 400 cephalograms obtained from a Hospital in Taipei, Taiwan. These cephalograms were captured by Soredex CRANEX® Excel Ceph machine and processed by Soredex SorCom software (3.1.5 version 2.0) [1]. The resolution of these images is 1935 x 2400 pixels. The cephalograms are obtained from a wide range of age group ranging from 6 to 60. These cephalogram images are saved as a TIFF format image that stores raster images. This is the most widely used dataset among researchers in automated cephalometric analysis. The dataset has 19 landmarks marked as (x, y) coordinates of pixels by two experts with an experience of 15 and 6 years respectively. Each expert marked the landmarks twice and the average of these 4 markings were taken as the ground truth to minimize inter- and intra-observer variation. It has a pixel sizing of 0.1 mm x 0.1 mm

The landmarks marked in this dataset are: the sella, the nasion, the orbitale, the porion, the subspinale, the supramentale, the pogonion, the menton, the gnathion, the gonion, the lower incisal incision, the upper incisal incision, the upper lip, the lower lip, the subnasal, the soft tissue pogonion, the posterior nasal spine, the anterior nasal spine, the anterior nasal spine and the articulate [26]

1. **Architecture Used**

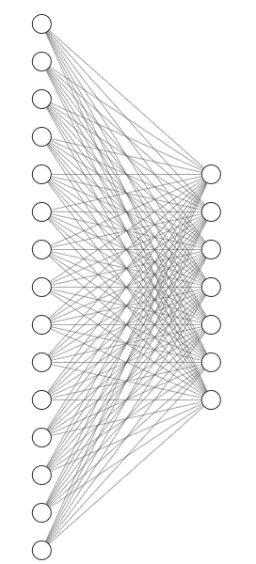
****

***Fig. 1:*** *CNML block taken from [25]*

CNML Block

CNML Block

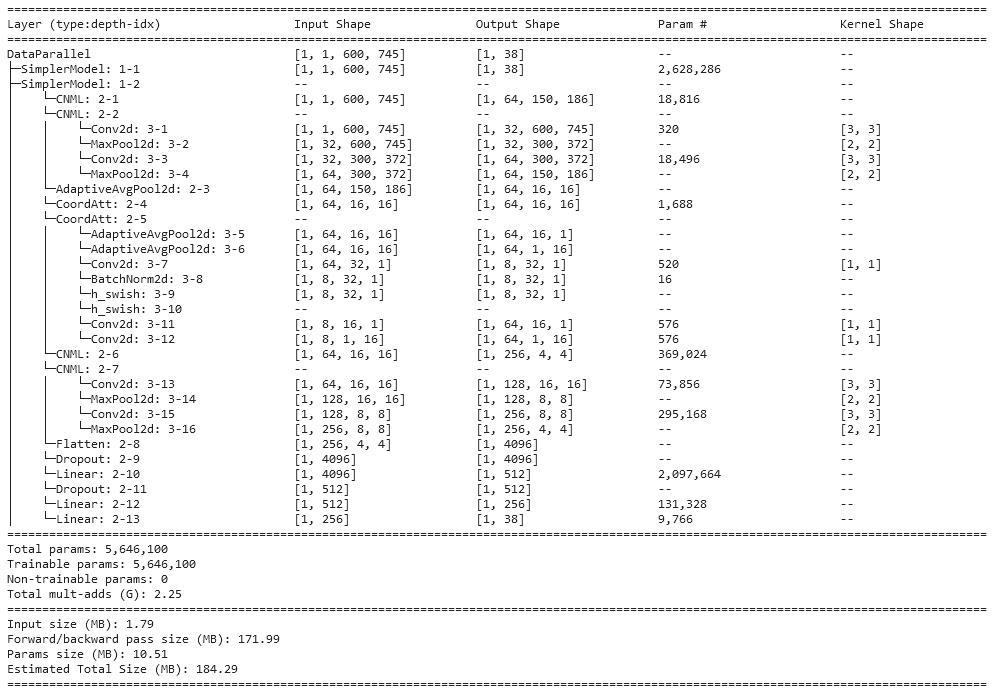
Coordinate Attention



Flatten

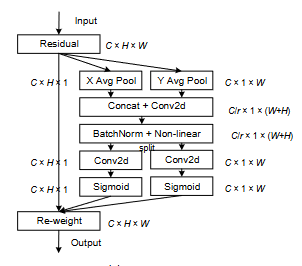
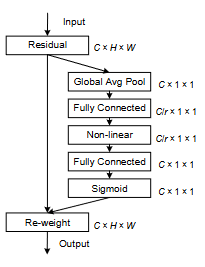
***Fig. 2:*** *Overall Architecture. Coordinate Attention block from [27]*

The CNML block consists of Convolutional layers followed by Max pooling layers to capture the semantic relationship between the features across the image. Max pooling helps retain the essential features. The author of [25] had originally used CNML blocks with Squeeze-and-Excitation (SE) block [28]. While SE block is a good mechanism for attention when used for overall image, it does not preserve location features. Since our task at hand is to predict landmarks, an attention mechanism that is position sensitive is more suitable. SE block is a cross-channel attention mechanism which may not work good when location-based attention is needed. Moreover, X-rays, which is the modality of our dataset is recorded in grayscale which has only one channel which further diminishes the use case of SE block. Hence, we propose a new architecture that uses Coordinate Attention (CA) proposed by Hou et. al [27] that captures row/height and column/width wise features for attention. The features obtained by these layers are then flattened and fed into a dense network that predicts 19 (x, y) coordinates.



***Fig. 3:*** *Summary of our model’s architecture for a sample input of size 600* x *745*

The coordinate attention module works similar to the SE module. SE module pools layers across channels from the shape (C, H, W) to (C, 1, 1) and runs them through rest of its layers for cross-channel attention. Coordinate attention on the other hand takes average pool across X and Y axis, which is the W and H dimension of the image tensor in this case and concatenates them together and passes it through a convolutional layer to get an out of shape (C, 1, W+H). It is clear that while channel-wise attention may be extracted here, it is more sensitive to the X and Y axis features as they’re across all the channels.



***Fig. 4:*** *SE module vs Coordinate Attention module* [27]

1. **Experimental Setup**

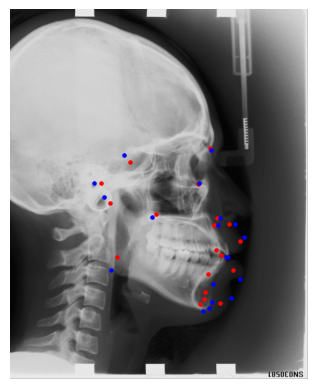
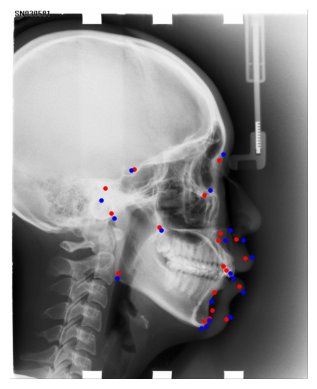
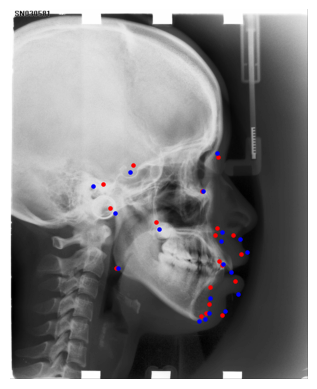
All of the coding was done using Python and the deep learning framework used by us is PyTorch. Unlike Neeraja et. al in [25], we did not extract patches and augment to increase dataset size while keeping each patch size to 64 x 64. We instead resized the entire image into 600 x 745 (width x height) size and trained our model on that due to VRAM constraints. The model was trained on 2 x Nvidia T4 GPU offered by Kaggle to its mobile number verified users. We used the AdamW optimizer with a learning rate of 0.001 and a weight decay of 0.0001. We trained our model for 30 epochs with a batch size of 4

1. **Results and Discussion**

Our model has obtained an SDR of 57.00% within 2 mm of MRE,, 83.36% for 3 mm, 93.26% for 4 mm. The model seemed to struggle with landmarks like pogonion, menton, gnathion, soft tissue pogonion as their proximity is near each other and hence they share similar visual features in the image. We believe this can be overcome by creating a large patched dataset like Neeraja et. al [25] and hence comparison with their results is excluded from ***Table 1***. Lee et. al [19] have used a multi-stage model which takes sequential inputs from the previous stage’s output and hence has a high running complexity. We are able to match almost their performance at the 4 mm range with a single stage model. The model can be used in mobile devices given its simplicity. We observed that our model is performing worse in the 2 mm range while matching or outperforming other models in the 4 mm range. The reasoning for this could be the lower image resolution that we trained on compared to creating patches like Neeraja [25], Tim [29], and so on.

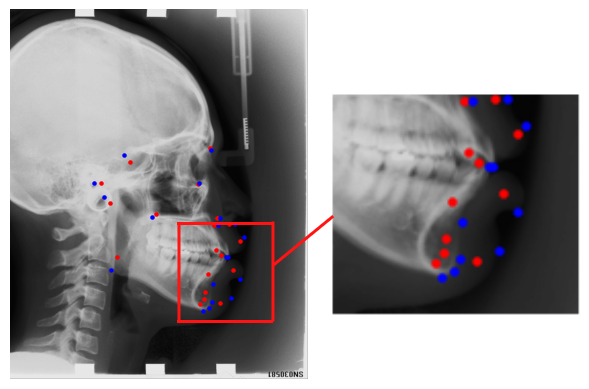
|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **SDR (%)** | | |
| **2 mm** | **3 mm** | **4 mm** |
| Lee et al. [19] | 82.11 | **92.28** | **95.96** |
| Wang et al. [14] | 73.37 | 84.46 | 90.67 |
| Tim et al. [29] | 74.95 | 80.56 | 89.68 |
| Ours | 57.00 | 83.36 | ***93.26*** |

Bold = best; Italic + Bold = Second Best

**Table 1:** Results and Comparison

***Fig. 5:*** *Sample Outputs. Red = Predicted, Blue = Actual*

As mentioned before, gnathion and nearby landmarks are having higher deviation that the rest on average. This is due to their close proximity and similar features. This can be visualized below



***Fig. 6:*** *Zoomed view of the area with lower SDR landmarks. Red = Predicted, Blue = Actual*

1. **Conclusion and Future Work**

Throughout the project, we observed that there is a severe lack of open-source data for the task of cephalometric landmark detection. Some of the images in the ISBI Grand Challenge dataset were blurry around the edges of the subject’s head. Also, the dataset only contains 400 images of which only 150 were used for training (and as is the standard for that dataset across other researchers using it). To add to this data insufficiency, there is severe lack of diversity to it as the dataset was fully collected from only one medical institute from Taiwan. So, there is a pressing need for a dataset that is sufficiently large, diverse, open-source. This must be addressed to create a robust automated cephalogram landmark detection model.

Our future work may include the following:

* The proposed model struggles in the 2 mm MRE range, which is the desired range for cephalogram landmark detection [26]. This can be overcome by creating patches dataset from the original dataset similar to what some methods have done [25, 29].
* To perform ablation study on the effectiveness of the Coordinate Attention module over the Squeeze-and-Excitation module
* A solution to the deviation in landmarks like pogonion, gnathion, etc that are close to each other and share similar visual features

**References**

[1] Kaur, A., & Singh, C. (2015). Automatic cephalometric landmark detection using Zernike moments and template matching. *Signal, Image and Video Processing*, *9*(1), 117-132.

[2] Gupta, A., Kharbanda, O. P., Sardana, V., Balachandran, R., & Sardana, H. K. (2016). Accuracy of 3D cephalometric measurements based on an automatic knowledge-based landmark detection algorithm. *International journal of computer assisted radiology and surgery*, *11*, 1297-1309.

[3] Qian, J., Cheng, M., Tao, Y., Lin, J., & Lin, H. (2019, April). CephaNet: An improved faster R-CNN for cephalometric landmark detection. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)* (pp. 868-871). IEEE.

[4] Koga, H., Taki, K., & Masugi, A. (2023). Efficient Measurement Method: Development of a System Using Measurement Templates for an Orthodontic Measurement Project. *Software*, *2*(2), 276-291.

[5] J. Yang, X. Ling, Y. Lu et al., “Cephalometric image analysis and measurement for orthognathic surgery,” *Medical & Biological Engineering & Computing*, vol. 39, no. 3, pp. 279–284, 2001.

[6] A. D. Levy-Mandel, A. N. Venetsanopoulos, and J. K. Tsotsos, “Knowledge-based landmarking of cephalograms,” Computers and Biomedical Research, vol. 19, no. 3, pp. 282–309, 2020

[7] Vezzetti, E., Marcolin, F., Tornincasa, S., Ulrich, L., & Dagnes, N. (2018). 3D geometry-based automatic landmark localization in presence of facial occlusions. Multimedia Tools and Applications, 77, 14177-14205.

[8] Alessandri-Bonetti, A., Sangalli, L., Salerno, M., & Gallenzi, P. (2023). Reliability of artificial Intelligence-Assisted cephalometric analysis. A Pilot Study. *BioMedInformatics*, *3*(1), 44-53.

[9] S. Chakrabartty, M. Yagi, T. Shibata et al., “Robust cephalometric landmark identification using support vector machines,” in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 825–828, Vancouver, Canada, May 2003.

[10] P. Vucinic, Z. Trpovski, and I. Scepan, “Automatic landmarking of cephalograms using active appearance models,” European Journal of Orthodontics, vol. 32, no. 3, pp. 233–241, 2010.

[11] J. Keustermans, W. Mollemans, D. Vandermeulen, and P. Suetens, “Automated cephalometric landmark identification using shape and local appearance models,” in Proceedings of 20th International Conference on Pattern Recognition, pp. 2464–2467, Istanbul, Turkey, August 2010.

[12] Mirzaalian, H., & Hamarneh, G. (2014, April). Automatic globally-optimal pictorial structures with random decision forest based likelihoods for cephalometric x-ray landmark detection. In IEEE ISBI.

[13] Chen, S. (2015, March). Tooth segmentation system with intelligent editing for cephalometric analysis. In Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging (Vol. 9417, pp. 653-658). SPIE.

[14] Wang, S., Li, H., Li, J., Zhang, Y., & Zou, B. (2018). Automatic analysis of lateral cephalograms based on multiresolution decision tree regression voting. Journal of healthcare engineering, 2018.

[15] Montúfar, J., Romero, M., & Scougall-Vilchis, R. J. (2018). Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections. American Journal of Orthodontics and Dentofacial Orthopedics, 153(3), 449-458.

[16] Suhail, S., Harris, K., Sinha, G., Schmidt, M., Durgekar, S., Mehta, S., & Upadhyay, M. (2022). Learning Cephalometric Landmarks for Diagnostic Features Using Regression Trees. Bioengineering, 9(11), 617.

[17] Dai, X., Zhao, H., Liu, T., Cao, D., & Xie, L. (2019). Locating anatomical landmarks on 2D lateral cephalograms through adversarial encoder-decoder networks. *IEEE*

[18] Kim, H., Shim, E., Park, J., Kim, Y. J., Lee, U., & Kim, Y. (2020). Web-based fully automated cephalometric analysis by deep learning. Computer methods and programs in biomedicine, 194, 105513.

[19] Lee, J. H., Yu, H. J., Kim, M. J., Kim, J. W., & Choi, J. (2020). Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC oral health, 20, 1-10.

[20] Noothout, J. M., De Vos, B. D., Wolterink, J. M., Postma, E. M., Smeets, P. A., Takx, R. A., ... & Išgum, I. (2020). Deep learning-based regression and classification for automatic landmark localization in medical images. IEEE transactions on medical imaging, 39(12), 4011-4022.

[21] Takeda, S., Mine, Y., Yoshimi, Y., Ito, S., Tanimoto, K., & Murayama, T. (2021). Landmark annotation and mandibular lateral deviation analysis of posteroanterior cephalograms using a convolutional neural network. Journal of Dental Sciences, 16(3), 957-963.

[22] Zeng, M., Yan, Z., Liu, S., Zhou, Y., & Qiu, L. (2021). Cascaded convolutional networks for automatic cephalometric landmark detection. Medical Image Analysis, 68, 101904.

[23] Wang, L., Ma, L., Li, Y., Niu, K., & He, Z. (2021). A DCNN system based on an iterative method for automatic landmark detection in cephalometric X-ray images. *Biomedical Signal Processing and Control*, *68*, 102757.

[24] Šavc, M., Sedej, G., & Potočnik, B. (2022). Cephalometric Landmark Detection in Lateral Skull X-ray Images by Using Improved SpatialConfiguration-Net. Applied Sciences, 12(9), 4644.

[25] Neeraja, R., & Anbarasi, L. J. (2023). CephXNet: A Deep Convolutional Squeeze-and-Excitation model for Landmark Prediction on Lateral Cephalograms. IEEE Access.

[26] Wang, C. W., Huang, C. T., Hsieh, M. C., Li, C. H., Chang, S. W., Li, W. C., ... & Ibragimov, B. (2015). Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge. *IEEE transactions on medical imaging*, *34*(9), 1890-1900.

[27] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13713-13722).

[28] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

[29] Tim, C., & Cootes, F. (2015, April). Fully automatic cephalometric evaluation using random forest regression-voting. In *proceedings of the IEEE international symposium on biomedical imaging (ISBI)* (pp. 16-19).