



Automatic Localization of Cephalometric Landmarks using Convolutional Neural Networks

Mogham Njikam Mohamed Nourdine^{1*}, Betül Uzbaş²

^{1*} Konya Technical University, Faculty of Engineering and Natural Science, Department of Computer Engineering, Konya, Turkey, (ORCID: 0000-0001-7068-9323), mohamednjikam25@hotmail.com

² Konya Technical University, Faculty of Engineering and Natural Science, Department of Computer Engineering, Konya, Turkey, (ORCID: 0000-0002-0255-5988), buzbas@ktun.edu.tr

(2nd International Conference on Computer, Electrical and Electronic Sciences ICCEES 2021, September 1-3, 2021)

(DOI: 10.31590/ejosat.1006517)

ATIF/REFERENCE: Mohamed Nourdine, M. N. & Uzbaş, B. (2021). Automatic Localization of Cephalometric Landmarks using Convolutional Neural Networks. *European Journal of Science and Technology*, (30), 48-52.

Abstract

Experts have brought forward interesting and effective methods to address critical medical analysis problems. One of these fields of research is cephalometric analysis. During the analysis of tooth and the skeletal relationships of the human skull, cephalometric analysis plays an important role as it facilitates the interpretation of bone, tooth, and soft tissue structures of the patient. It is used during oral, craniofacial, and maxillofacial surgery and during treatments in orthodontic and orthopedic departments. The automatic localization of cephalometric landmarks reduces possible human errors and is time saving. To performed automatic localization of cephalometric landmarks, a deep learning model has been proposed inspired by the U-Net model. 19 cephalometric landmarks that are generally manually determined by experts are automatically obtained using this model. The cephalometric X-ray image dataset created under the context of IEEE 2015 International Symposium on Biomedical Imaging (ISBI 2015) is used and data augmentation is applied to it for this experiment. A Success Detection Rate SDR of 74% was achieved in the range of 2 mm, 81.4% in the range of 2.5mm, 86.3% in the range of 3mm, and 92.2% in the range of 4mm.

Keywords: Convolutional Neural Networks, Cephalometric Landmarks Detection, Medical Image Analysis, Success Detection Rate.

Evrişimsel Sinir Ağlarını Kullanarak Sefalometrik Noktaların Otomatik Tespiti

Öz

Uzmanlar kritik tıbbi analiz problemlerinde ilginç ve etkili yöntemleri başarıyla ortaya koydular. Bu alanlardan biri de sefalometrik analizdir. Sefalometrik analiz, insan kafatasının diş ve iskelet ilişkilerinin analizi sırasında hastanın kemik, diş ve yumuşak doku yapılarının yorumlanmasını kolaylaştırdığı için önemli rol oynar. Ortodonti ve ortopedi bölümlerinde ağız, kraniyofasiyal ve çene cerrahisi ve ortodontik tedaviler sırasında kullanılmaktadır. Sefalometrik yer işaretlerinin otomatik olarak konumlandırılması, olası insan hatalarını azaltır ve zamandan tasarruf sağlar. Sefalometrik yer işaretlerinin otomatik lokalizasyonunu gerçekleştirmek için U-Net modelinden esinlenerek bir derin öğrenme modeli önerilmiştir. Genellikle uzmanlar tarafından manuel olarak belirlenen 19 sefalometrik yer işareti bu model kullanılarak otomatik olarak elde edilir. Bu araştırma için IEEE 2015 Uluslararası Biyomedikal Görüntüleme Sempozyumu (ISBI 2015) kapsamında oluşturulan sefalometrik X-ray görüntü veri seti kullanılmış ve bu verisetine veri büyüme uygulanmıştır. 2 mm aralığında 74%, 2.5 mm aralığında 81.4%, 3 mm aralığında 86.3% ve 4 mm aralığında 92.2% Başarı Tespit Oranı (SDR) elde edildi.

Anahtar Kelimeler: Evrişimsel Sinir Ağları, Sefalometrik Nokta Tespiti, Tıbbi Görüntü Analizi, Başarı Tespit Oranı.

* Corresponding Author: mohamednjikam25@hotmail.com

1. Introduction

Cephalometric analysis is an essential tool for clinical evaluation in modern craniofacial, oral and maxillofacial surgery, and results obtained provide important information during orthodontic treatment. It can also be defined as the scientific measurement of the dimensions of an x-ray lateral cephalogram which are the main resources in this analysis [1]. Cephalometric analysis is widely used in orthodontics and orthopedic sections. The results from this analysis are used to describe the relationship between the dental and skeletal parts of the human skull. This relationship plays a vital role in treatment planning, and interpretation of the bone, tooth, and soft tissue structures of the patients. The positions of cephalometric landmarks generally play a vital role during treatment planning, assessment of curative effect, and comparing different anomalies.

In general, cephalometric landmarks localization is manually computed by physicians, medical doctors, or clinicians. It takes a long time to detect these points manually. In addition, these landmarks detected by different experts can often be inconsistent. Since there is no special interpolation formula for this operation and the images differing from one to another, the results obtained after hours of calculations may not be entirely correct [2]. To solve this problem and to provide more reliable results, attempts have been made by researchers to automatically detect cephalometric landmarks to overcome these limitations in clinical practice and research environments. This operation is a difficult task mainly due to the structure of the data.

Significant progress has been made towards automatically detecting points and pattern structures in radiographic x-ray images. These approaches are done by either specifying rules based on known rules or experts' knowledge or using templates gray-scales morphological operators to regress the cephalometric points. To boost the research in this field, in 2014 and 2015, ISBI challenges on cephalometric landmark detection were launched and several researchers propose different technics to achieve this operation. One of the most prominent results validated during the 2015 symposium was developed by Lindner et al. [3]. It was based on random-forest and it achieved the SDR of 73.7% for a 2 mm precision range. Another outstanding work preserved during this symposium is by Ibragimov et al. [4] which achieved the second-highest result with a successful detection accuracy of 71.7% over a 2 mm range. They used random forest regression to apply a Haar-like feature extraction on the data and continued to make adjustments to their model using global context shape.

Many other researchers tried some other technics in the process of landmarks localization. One of these is the work proposed by Cardillo et al. [5] who developed a grayscale morphology and template matching algorithm. Using Dynamic programming, they exploited and applied the templates to extract the cranial contour on high-resolution X-Ray cephalometric images. In total, 20 landmarks were detected using this approach. Finally, they tested their work on 40 lateral X-ray images, which showed an average of 85% success detection rate. Another prominent model architecture called CephaNet was proposed in 2019 by Qian et al. [1]. In CephaNet, they designed multitasking loss to reduce the variation that existed between classes. They adopted a multi-scale training strategy to improve the success detection accuracy of the model at relatively small landmark intervals. With this method, they achieved 82.5% successful detection accuracy in the 2 mm range. In the same line of work,

Lee et al. [6] proposed a different approach that successfully regresses the correct coordinates of cephalometric landmarks directly from the image. A regression system based on CNN was trained for each cephalometric point and the coordinate location for the different variables was successfully extracted. To achieve this, 38 regression systems were respectively trained from scratch using basically the same CNN model for the corresponding locations of the coordinates and using the trained model they calculated 38 coordinate variables from the test images and extracted 19 landmarks by matching the regressed coordinate location for the coordinates. They achieved an SDR of 86.4% at 2mm.

In this study, a model for automatic cephalometric landmark detection using U-Net architecture is proposed. The dataset, U-Net architecture, and evaluation metrics are mentioned, in the second part. The experimental results are presented, in section four. The obtained results are discussed in the fifth part, and the paper is finalized in the conclusion part.

2. Material and Method

The model proposed for this experiment is built using a customized U-Net model and the results evaluated using the basic metrics of evaluations proposed during the IEEE ISBI 2015 Competition. We used the publicly available cephalometric dataset provided for this competition.

Figure (1) shows an example of a cephalogram lateral image with 19 landmarks automatically obtained using our model presented in this research paper.

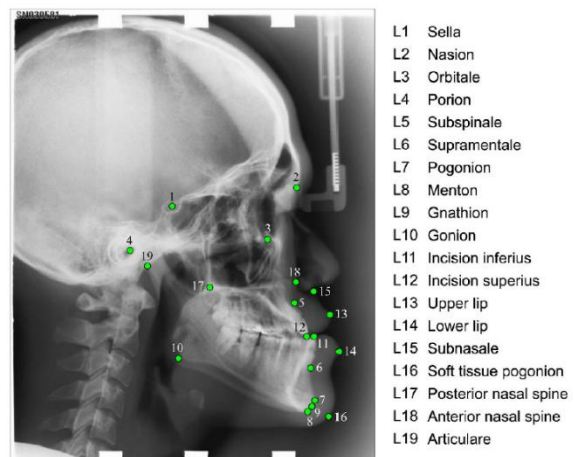


Figure. 1 A Cephalometric image showing the 19 landmarks positions with their respective annotation [7].

2.1. Dataset

The data used in the scope of this research are obtained from the publicly available dataset provided during the 2015 ISBI Grand Challenge Training Dataset. It contained 400 Lateral Cephalometric images with patients' mean age being 27 years with an age range from 7 to 76. These data were collected from 235 females and 165 males. All cephalometric images found in this dataset were generated in TIFF format using the dental X-ray Machine called CRANEXr Excel Ceph and the software Soredex SorCom (3.1.5, version 2.0) [7]. This machine is famous for its ability to give automatic preselection for panoramic imaging based on skull width. The image resolution of the images provided in this dataset is 1935 x 2400 pixels with a pixel pitch of 0.1 mm. The coordinates of all the cephalometric points shown in Figure

(1) were respectively labeled by two doctors; a junior and a senior in separated folders. We used the senior physician's annotations for this research and aimed to identify the 19 landmarks for each X-ray image. The authors of this symposium have divided the dataset into three different folders: the training set containing 150 cephalometric images, the test1 folder containing 150, and the test2 folder contains 100 images. The images used during this experiment were resized to 256x256 px sizes which are one-eight of the original size before data augmentation was applied to each image.

2.1.1. Data Augmentation

Data augmentation was performed on the training dataset which initially comprised of 150 annotated images, to increase the accuracy of our model. This operation conveniently boosted the model by permitting it to learn more efficiently. The data augmentation is performed on the dataset was shifting 1 and 3 pixels in 4 different directions (right, left, down, up) on a 256x256 cephalometric image. Every image on the training dataset is shifted using these shift metrics. Four new images are created from each image, with the shift operation. This operation is performed using the famous library for image augmentation called Imgaug [8]. After performing the data augmentation on the images, we successfully generated 1350 images used during the experiment.

2.2. U-Net Network Architecture

U-Net [9] is the CNN architecture presented in 2015 to address the issue of biomedical image processing, classification, and segmentation. The U-net model, originally used for the segmentation of neural structures, was considered as one of the best methods in the ISBI cell tracking challenge in 2015 [10]. The CNN architecture proposed for this study follows closely the U-Net model with slight modification to achieve more efficiency. The architecture of this model consists of a downsampling layer which is followed by a symmetric upsampling layer separated by a bottleneck. It is designed to learn the local characteristics of each landmark gradually in both directions. It has an input

dimension of 256x256x1. The architecture of the customized U-Net model developed for this task is shown in Figure (2).

As shown in figure (2), the proposed U-Net model is divided into three main blocks; the down-sampling block, bottleneck, and the up-sampling block. Applying down-sampling to the dimensions of input images reduces resolution and allows the network to learn more properties on the image. During this phase, the model learns the relationship that exists between the relative landmark position. It consists of 2 subsequent convolution layers with a kernel size of 3x3 which is followed by a stride of 1, a padding of 1, and a non-linear activation function layer using a rectifier linear unit (ReLU). It is then followed by a maximum pooling layer which is responsible for dividing the dimensions width and height into two.

The block size is of height 4 till the bottleneck. In this phase, the channel size of the input image is first expanded using a pair of convolutional that allow the network to model richer properties in this down-sampling phase. The maximum pooling layer halves the width and height dimensions of the property map. During each downsampling layer, the input properties passed through the double convolution increase the number of the channels by two which permit the model to learn more accurately. It is then passed through the maximum pooling layer which halves the property maps which are used during the next iteration. The filter sizes used for this operation till the bottleneck are 64, 128, 256, 512, 1024. The bottleneck layer is the base turn point for the model and consists of 1024 filters. After the bottleneck, we have the upsampling path where a transpose convolution is applied to the incoming lower channel size by doubling its dimensions. This process is repeated for each level in this path until the final layer is at the highest level where a final filter of size 1x1 is used to create the landmark heatmaps space predictions. After every convolutional layer, ReLU activation is used to speed up the training process.

Our model has a single-channel grayscale cephalometric image as the input image. All models are trained as 256x256 image size to provide faster training times and easier evaluation.

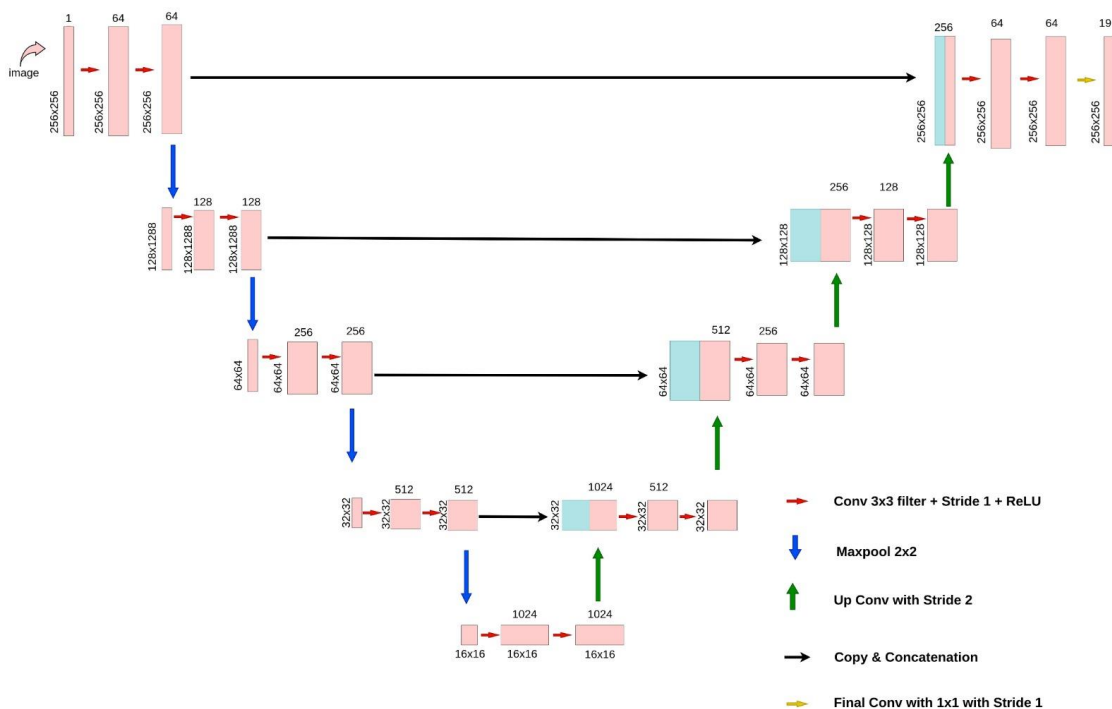


Figure 2. The Proposed U-Net Model Architecture

The application was developed using the PyTorch [11] framework on a 64-bit Intel® Core™ i7-9700F CPU @ 3.00GHz 3.00GHz 16.0GB RAM Computer. Adam Optimization [12] with a batch size of 8 was used for optimizing our training process. We have as an initial learning rate for our model 10^3 and our model is trained using 3/4 of the total image dataset. The rest of 1/3 of the images are used during the validation process during our training. We equally have a weight decay of 10^4 to avoid the model from overfitting. During our training process, we decrease our learning rate by 10 and restart the training process in case the validation loss does not improve for 10 consecutive epochs. This action stops after 10 consecutive epochs where the validation loss did not change.

2.2. Evaluation Metrics

Landmark localization model performance was evaluated using the same measurements used for the ISBI 2015 competition.

We use the Mean Square Error (MSE) as the main metric to evaluate our trained models for our heatmap regressions. The equation for this is shown in equation (1) where y values are the initial coordinates heatmaps and \hat{y} are the prediction results generated by our model and m is the batch size use during process.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y - \hat{y})^2 \quad (1)$$

To facilitate the convergence of our model, the amplitude of the coordinates that we initially labeled by the senior doctor is increased. The amplitude and the standard deviation of 1000 and 5 were used respectively, we were able to make our model focus on regressing the coordinates that have the highest amplitude.

During the process, the coordinates with lower amplitude were ignored to fasten our model's convergence.

The Mean Radial Error (MRE) and the Standard Deviation (STD) values are evaluated for performance measures. The Radial Error (R) was used to evaluate the accuracy of our prediction. It is obtained by calculating the Euclidean Distance between the predicted position y and the actual landmark position initially allocated by the senior doctor x .

$$R = \sqrt{\Delta x^2 + \Delta y^2} \quad (2)$$

In equation (2), Δx and Δy , are the distances between the actual bookmark position and the predicted position. N indicates the number of sample detections.

The model is equally evaluated based on the MRE and the STD obtained after every training epoch.

$$MRE = \frac{\sum_{i=1}^N R_i}{N} \quad (3)$$

$$STD = \sqrt{\frac{\sum_{i=1}^N (R_i - MRE)^2}{N}} \quad (4)$$

Using equation (5), we are able to determine if a coordinate is determined are in the accepted ranges of 2 mm, 2.5 mm, 3 mm and 4 mm. A coordinate is considered to be in the acceptable range if the predicted coordinates are found within these diameters.

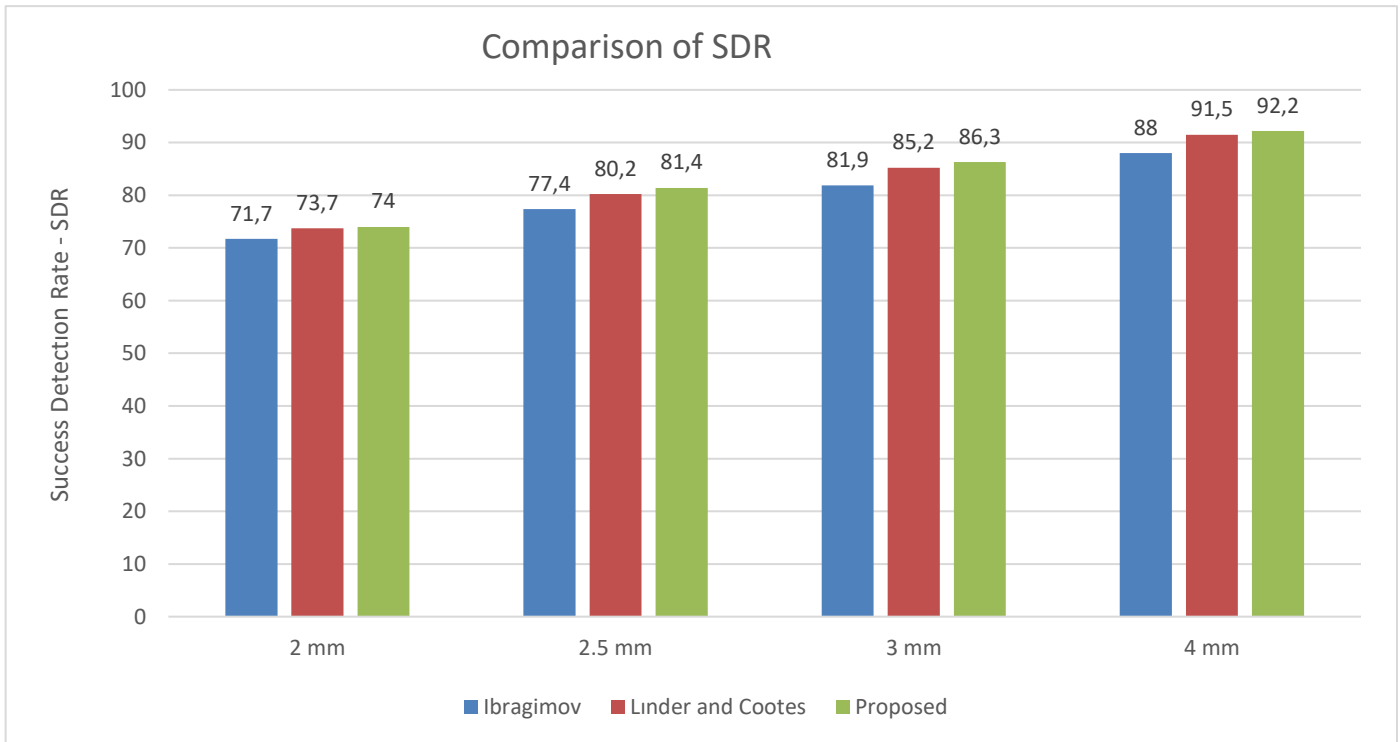


Figure. 3 SDR comparison of our model results with the results got by Ibragimov et al. [4] and Lindner et al. [3]

It is referred to as the Success Detection Rate (SDR).

$$P_z = \frac{\sum\{j:\|L_a(j)-L_b(j)\|\leq 2\}}{\sum\Omega} \times 100 \quad (5)$$

3. Results and Discussion

3.1. Result

In this study, cephalometric landmarks were automatically localized using U-Net architecture. After performing data augmentation on our training dataset, we trained our model and evaluated it using the test1 dataset. In order to virtualize the performance boost generated by data augmentation, we initially trained our model using only 150 images provided in the training dataset and subsequently trained it using the 1350 augmented data, and their success detection SDR rates were both calculated using the Test 1 dataset. Table 1 shows the obtained successful results before and after the data augmentation.

As shown in table (1); the results got after data augmentation performed better than the ones done with the original dataset in all the evaluation metrics.

Table 1. Comparisons of SDR on training data with and without data Augmentation.

Dataset	Percentage (%)			
	2mm	2.5mm	3mm	4mm
Without data Augmentation (150 images)	69.37	78.0	84.6	91.12
With data Augmentation (1350 images)	74.0	81.4	86.3	92.2

3.2. Discussion

The performance of application models in the localization role was compared to the best solution in the ISBI 2015 Challenge [13]. As shown in Figure (3), the results that we obtained using this model were compared with the results found by other experts based on the same dataset. Ibragimov et al. [4] was able to get a SDR of 71.7% on 2mm range, 77.4% on 2.5mm, 81.9% on 3mm and 88% on 4mm range. Lindner et al. [3] proposed method got 73.7% on 2mm range, 80.2% on 2.5mm, 85.2% on 3mm and 91.5% on 4mm range. Our proposed U-Net model, consisting of the base model, achieved 74% SDR in the range of 2 mm. In the range of 2.5 mm, it resulted in an SDR of 81.4%, 86.3% in the 3 mm range, and 92.2% in the 4mm range. The proposed method shows better results in all the evaluation metrics as shown in figure (3).

4. Conclusions

In this study, A modified U-net model was proposed to address the problem of the determination of cephalometric

landmarks. We applied Data Augmentation to the training dataset and successfully generated a total of 1350 images. All models were trained using images of size 256x256 to provide faster training and easier experiments. The SDR achieved through this framework shows success results as shown in figure (3). The proposed U-Net model performs better and gives better results when compared to the results accepted during the competition launched by ISBI 2015.

References

1. Qian, J., et al. *CephaNet: An Improved Faster R-CNN for Cephalometric Landmark Detection*. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019. IEEE.
2. Yue, W., et al., *Automated 2-D cephalometric analysis on X-ray images by a model-based approach*. *IEEE transactions on biomedical engineering*, 2006. **53**(8): p. 1615-1623.
3. Lindner, C. and T.F. Cootes. *Fully automatic cephalometric evaluation using random forest regression-voting*. in *IEEE International Symposium on Biomedical Imaging*. 2015. Citeseer.
4. Ibragimov, B., et al. *Computerized cephalometry by game theory with shape-and appearance-based landmark refinement*. in *Proceedings of International Symposium on Biomedical imaging (ISBI)*. 2015.
5. Cardillo, J. and M.A. Sid-Ahmed, *An image processing system for locating craniofacial landmarks*. *IEEE transactions on medical imaging*, 1994. **13**(2): p. 275-289.
6. Lee, H., M. Park, and J. Kim. *Cephalometric landmark detection in dental x-ray images using convolutional neural networks*. in *Medical Imaging 2017: Computer-Aided Diagnosis*. 2017. International Society for Optics and Photonics.
7. Lindner, C., et al., *Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms*. *Scientific reports*, 2016. **6**: p. 33581.
8. Jung, A.B., *imgaug*. Online; accessed 30-Oct-2018, 2018.
9. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
10. Goutham, E., et al. *Automatic Localization of Landmarks in Cephalometric Images via Modified U-Net*. in *2019 10th international conference on computing, Communication and Networking Technologies (ICCCNT)*. 2019. IEEE.
11. Paszke, A., et al., *Automatic differentiation in pytorch*. 2017.
12. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980, 2014.
13. Wang, C.-W., et al., *A benchmark for comparison of dental radiography analysis algorithms*. *Medical image analysis*, 2016. **31**: p. 63-76.