2022

Automatic Cephalometric Landmark Detection on X-Ray Images Using Object Detection

Cheng-Ho King
National Chung Cheng University

Yin-Lin Wang
National Taiwan University

Chia-Ling Tsai
CUNY Queens College

AUTOMATIC CEPHALOMETRIC LANDMARK DETECTION ON X-RAY IMAGES USING OBJECT DETECTION

[1]*Cheng-Ho King,* [2]*Yin-Lin Wang,* [1]*Wei-Yang Lin,* [3]*Chia-Ling Tsai*

[1]Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan
[2]School of Dentistry, National Taiwan University, Taiwan
[3]Computer Science Department, Queens College, CUNY, New York, USA

ABSTRACT

We propose a new deep convolutional cephalometric landmark detection framework for orthodontic treatment. Our proposed method consists of two major steps: landmark detection using a deep neural network for object detection, and landmark repair to ensure one instance per landmark class. For landmark detection, we modify the loss function of the backbone network YOLOv3 to eliminate the constrains on the bounding box and incorporate attention mechanism to improve the detection accuracy. For landmark repair, a triangle mesh is generated from the average face to eliminate superfluous instances, followed by estimation of missing landmarks from the detected ones using Laplacian Mesh. Trained and evaluated on a public benchmark dataset from IEEE ISBI 2015 grant challenge, our proposed framework obtains comparable results compared to the state-of-the-art methods for cephalometric landmark detection, and demonstrates the efficacy of using a deep CNN model for accurate object detection of landmarks defined by only a single pixel location.

Index Terms— Landmark detection, Attention, Cephalometric X-ray image, repair strategy

1. INTRODUCTION

For orthodontic treatment and maxillofacial surgical planning, it is crucial that cephalometric landmarks are accurately marked out on a cephalometric X-ray image to provide angular and linear measurements of the patient's facial structures. If done manually, the operation is time-consuming and can suffer from intra- and inter-observer variability, which is further aggravated by bones overlapping. There are 19 commonly used landmarks for celphalometric analysis. To advance the state-of-the-art for automatic cephalometric landmark detection, a dataset of 400 annotated cephalometric X-ray images was made publicly available for the ISBI 2015 Grand Challenge [1]. Each image is annotated with ground truth data produced by two experienced medical doctors. Winning algorithms in this challenge are mainly machine-learning based. Ibragimov et al. [2] applied game-theoretic optimization framework and random forest to detect the landmarks. Linder et al. [3] applied Random Forest regression-voting in the Constrained Local Model framework (RFRV-CLM) to the landmark detection and achieved the best accuracy in the challenges.

In recent years, the introduction of deep convolutional neural network (CNN) has greatly improved the accuracy of cephalometric landmark detection. Two major approaches are heatmap regression and object detection. For approaches of heatmap regression, Zhong et al. [4] proposed a two-stage strategy based on the U-Net model. In stage 1, coarse positions of landmarks are obtained and small patches centered on estimated landmarks are cropped as the training data for stage 2. The positions of the 19 landmarks are refined as the final output. Oh et al. [5] proposed a Fully Convolutional Network (FCN) with the loss function enhanced by the geometric relationships among landmarks in the training stage. For object detection, Qian et al. [6] proposed a method using Faster R-CNN to detect potential landmark locations, and a repair strategy to filter out superfluous landmarks and to estimate undetected landmarks based on Laplacian transformation.

When applying object detection to solve point landmark detection, there are some common problems. (1) A bounding box is ill-defined for a landmark, which has no outline as an object to support the bounding box. (2) The number of outputs cannot be limited, unlike heatmap regression, because object detection identifies instances of object types based on the appearance. Multiple or zero non-overlapping instances of an object type can be detected.

In order to overcome these problems, we proposed an object-detection framework, CephaX, based on Darknet53, which is the backbone of YOLOv3 [7], for cephalometric landmark detection. We designed a multitask loss without bounding box constraints since bounding boxes are ill-defined for this problem. Object detection is followed by a repair strategy to handle multiple or missing instances of landmark types to ensure that each landmark type has one and only one instance detected.

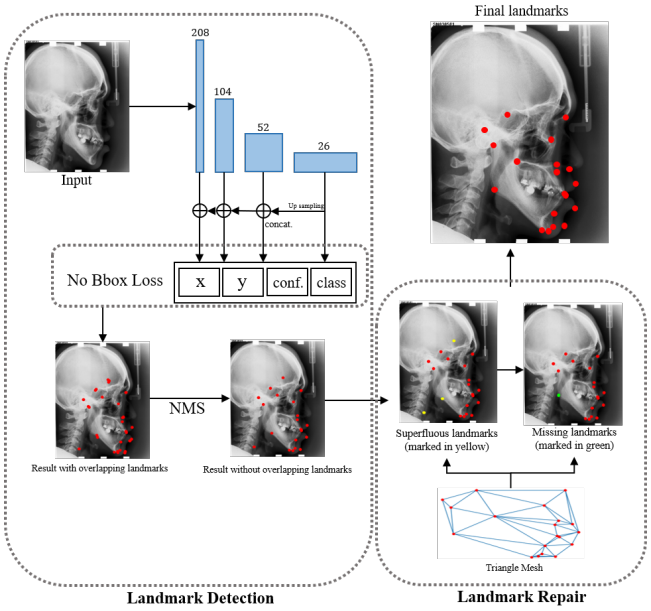

Fig. 1. An overview of the proposed method CephaX

2. METHOD

Our proposed framework consists of two stages: landmark detection and landmark repair. For landmark detection we use YOLOv3 as our modified backbone model with attention mechanism for feature selection. The multitask loss is modified to eliminate bounding box constraints so the network can focus more on predicting landmark's positions, instead of compensating for bounding boxes. For landmark repair, the repair strategy uses the relationship matrix among landmarks to remove superfluous instances not caught by Non-Maximum Suppression (NMS) at the end of landmark detection, and estimates undetected landmarks from survived landmarks using Laplacian Mesh [8]. The whole method is shown in Fig 1.

2.1. No Bbox Loss

For landmark detection, a bounding box cannot be properly estimated because a landmark specified by only its location does not have a outline to define a meaningful bounding box. For this reason, the loss of width and height of the bounding box should not be included in the total loss in our model when doing back propagation. By removing the bounding box from the loss function, the network can focus more on learning other information. We construct the coordinate loss L_{coord} using MSE (Mean Squared Error) loss, and confidence loss L_{confi} and classes loss $L_{classes}$ using BCE (Binary Cross Entropy) loss. The complete loss function of CephaX is

$$L(p, C, c) = L_{coord}(p, \hat{p}) + L_{confi}(C, \hat{C}) + L_{classes}(c, \hat{c}), \quad (1)$$

where p is the (x, y) position relative to each grid cell, and $\hat{p}$ is the ground truth. C is the confidence loss indicating whether there is an object or not, and $\hat{C}$ is the ground truth. c is the classes loss containing 19 BCE losses, one for each landmark class, and $\hat{c}$ is the ground truth.

2.2. Attention Module

In order to direct the model to focus more on the landmark coordinates, we incorporate the attention mechanism [9] in the model. We build four attention modules, one for each scale residual block in the network. Each attention module generates its own detection result (scaled up to the original resolution) and the feature maps are also propagated to other scales using Feature Pyramid Network (FPN) [10]. Detection results from all attention modules are combined and filtered using NMS, as shown in the left box of Fig 1, to generate the final output for landmark detection. The original YOLOv3 has 5 scales, down to 13×13. However, some landmarks in cephalometric X-ray image are so close that if the image is down-scaled too much, they fall in the same grid cell. This will cause the same grid cell to have multiple class labels and the model cannot determine which class to assign to the grid cell. For this reason reason, we reduce the depth of the network by taking out the 13×13 scale residual block in YOLOv3.

2.3. Landmark Repair Strategy

At the end of landmark detection, NMS is applied to identify detected instances of the same landmark class that are too close (overlapping entities), and only the one with the highest confidence in each cluster is kept. Two instances are considered too close if they overlap over 50% IoU with a 200×200 bounding box. NMS is effective for removing additional overlapping instances of the same class, but it cannot eliminate same-class instances that are far apart. To remove superfluous landmarks after NMS we exploit the geometric relationship amount landmarks by building a triangle mesh using Delaunay triangulation on the average face computed from the training data. The mesh is constructed as a 19×19 relationship matrix to represent the displacement between every two landmarks, if there is a line connecting the two landmarks in the triangle mesh. By checking the vectors in the relationship matrix, we can eliminate landmarks falling in the wrong area. After this step, each landmark should have only one or none instance left. Finally, to estimate missing landmarks, we locate the nearest face in the undeformed train data by calculating mean squared error (MSE) of the landmark relationship matrices of two faces. The missing landmarks are estimated from the corresponding landmarks of the detected ones in the nearest face using Laplacian Mesh [8].

3. EXPERIMENTS AND RESULTS

We train and validate CephaX on the public benchmark dataset from IEEE ISBI 2015 grant challenge for cephalometric landmark detection [1]. This dataset is collected from 400 patients aged 6 to 60 years and annotated by two experienced dentists manually. The ground truth is the average of the markups by both dentists. The dataset is split into three parts, 150 images for training data, 150 images for Test1 data for validation and 100 images for Test2 data for black-box testing. We adopt the Mean Radial Error (MRE), Standard Deviation (SD) of MRE, and Success Detection Rate (SDR) to evaluate the performance. MRE is defined as the distance between the predicted landmark and the ground truth. SDR is defined as the ratio of the number of accurate landmarks to the total detected landmarks. To classify a landmark as accurate, the distance between the detected landmark and the ground truth should be within a threshold of z mm (2.0 mm, 2.5 mm, 3.0 mm, 4.0mm). In clinical setting, SDR indicates the percentage of acceptable landmarks within the tolerable error range for the given medical treatment.

3.1. Implementation Details

CephaX is developed with Pytorch on Ubuntu. The optimization method is Adaptive Moment Estimation (Adam) [11]. The learning rate is set as 0.001. We train CephaX on a computer with an NVIDIA RTX 2080Ti GPU, 3.6 GHz AMD Ryzen 5 3600 CPU and a 16Gb memory.

3.2. Results and Comparison

CephaX is trained on 150 images and validated on Test1 data for model selection. The performance of the final model is evaluated on Test2 data. Table 1 shows the performance compared with our baseline model, YOLOv3, and other algorithms for cephalometric landmark detection. We compare CephaX with the methods of Lindner et al. [12], Arik et al. [13], Qian et al. [6] and Oh et al. [5]. For YOLOv3, the bounding box of a landmark is set to 200 $\times$ 200 and the instance with the highest confidence is retained for each landmark class.

CephaX achieves 86.14%, 91.72%, 94.91% and 97.96% in SDR for 2.0 mm, 2.5 mm, 3.0 mm and 4.0 mm, respective, on Test1 data and 74.58%, 81.74%, 87.26% and 94.73%, respective, on Test2 data. Compared with the method of Qian et al. [6] which is also an object detection method, our method is 3.63% higher in SDR for 2.0 mm on Test1, and 2.18% higher on Test2. CephaX is also comparable to the method of Oh et al. [5], which is based on heatmap regression, in terms of MRE and SDR on both Test1 and Test2.

Table 2 shows the ablation study of the proposed framework on Test1, which demonstrates the individual contributions of the changes made to YOLOv3. Starting with our full model, one change is removed at a time to study the impact of the change. No BBox loss and the attention mechanism for feature selection improve both the location accuracy (MRE) and the detection rate (SDR). MRE is reduced by 0.9 mm, and SDR is increased by 24.99%. The repair strategy mainly improves the location accuracy of images with facial structures that deviate more from the normal representation, as shown in Fig 2. There are only few such cases in the dataset, and the improvement is best observed in the statistically dispersion of the predicted results, reduced from 1.20 mm to 0.93 mm.

Method	150 images (Test1)			
	SDR(%)			
	2.0mm	2.5mm	3.0mm	4.0mm
YOLOv3 [7]	61.15	73.64	82.30	91.70
Lindner et al. [12]	74.95	80.28	84.56	89.68
Arik et al. [13]	75.37	80.91	84.32	88.25
Qian et al. [6]	82.51	86.25	89.31	90.62
Oh et al. [5]	**86.20**	91.20	94.40	97.70
CephaX	86.14	**91.72**	**94.91**	**97.96**
Method	100 images (Test2)			
	SDR(%)			
	2.0mm	2.5mm	3.0mm	4.0mm
YOLOv3 [7]	57.10	67.37	75.34	84.90
Lindner et al. [12]	66.11	72.00	77.63	87.42
Arik et al. [13]	67.68	74.16	79.11	84.63
Qian et al. [6]	72.40	76.15	79.65	85.90
Oh et al. [5]	**75.90**	**83.40**	**89.30**	**94.70**
CephaX	74.58	81.74	87.26	94.73

Method	Test1		Test2	
	MRE (mm)	SD	MRE (mm)	SD
YOLOv3 [7]	2.10	1.84	2.46	1.88
Linder et al. [12]	1.67	1.65	1.92	-
Arik et al. [13]	-	-	-	-
Qian et al. [6]	1.28	-	1.54	-
Oh et al. [5]	1.18	1.01	**1.46**	**0.82**
CephaX	**1.17**	**0.93**	1.50	1.00

Table 1. Experiment of CephaX and other methods on the public benchmark dataset

No bbox loss	Attention module	Repair strategy	MRE (mm)	SD	SDR(%) 2mm
✓	✓	✓	**1.17**	**0.93**	**86.14**
✓	✓		1.20	1.20	86.14
✓			1.63	1.20	73.07
			2.10	1.84	61.15

Table 2. Ablation study of CephaX performed on Test1

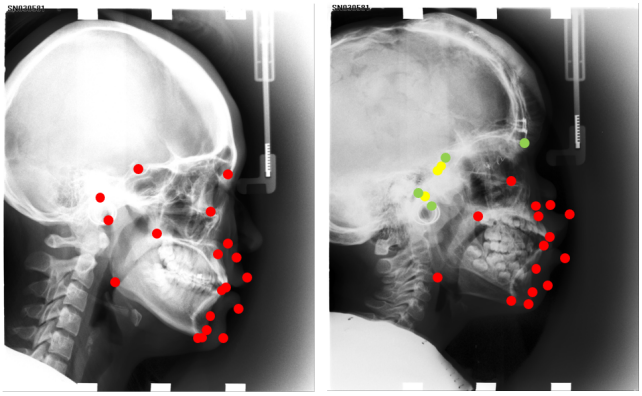

Fig. 2. Effect of repair strategy. Left: A normal facial structure not affected by the repair strategy. Right: An abnormal facial structure with landmarks before the repair marked as yellow, and after the repair marked as green. In this case, the error of landmark 2 (the right most green dot) is reduced from 68.9mm to 6.5mm.

4. CONCLUSION

In this paper, we propose a two-stage framework for Cephalometric landmark detection, combining a deep network for object detection and a repair strategy to ensure one instance per landmark class. We modify YOLOv3 to eliminate the bounding box constraint from the multi-task loss, build a triangle mesh to model the geometric relationships among landmarks to eliminate superfluous landmarks from object detection, and apply Laplacian Mesh to estimate missing landmarks from detected ones. CephaX demonstrates the efficacy of using a deep CNN model for accurate object detection on detecting extremely small landmarks defined by only one single pixel location.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by IEEE ISBI 2015 Grand Challenge. This research study does not require the institutional IRB.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Ching-Wei Wang, Cheng-Ta Huang, Meng-Che Hsieh, et al., "Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge," *IEEE transactions on medical imaging*, vol. 34, no. 9, pp. 1890–1900, 2015.

[2] Bulat Ibragimov, Boštjan Likar, F Pernus, et al., "Computerized cephalometry by game theory with shape-and appearance-based landmark refinement," in *Proceedings of International Symposium on Biomedical imaging (ISBI)*, 2015.

[3] Claudia Lindner, Ching-Wei Wang, Cheng-Ta Huang, et al., "Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[4] Zhusi Zhong, Jie Li, Zhenxi Zhang, et al., "An attention-guided deep regression model for landmark detection in cephalograms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 540–548.

[5] Kanghan Oh, IL-Seok Oh, Thang Van nhat Le, et al., "Deep anatomical context feature learning for cephalometric landmark detection," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[6] Jiahong Qian, Ming Cheng, Yubo Tao, et al., "Cephanet: An improved faster r-cnn for cephalometric landmark detection," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 868–871.

[7] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] Olga Sorkine, "Laplacian mesh processing," *Eurographics (STARs)*, vol. 29, 2005.

[9] Fei Wang, Mengqing Jiang, Chen Qian, et al., "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[10] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, et al., "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[11] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] Claudia Lindner and Tim F Cootes, "Fully automatic cephalometric evaluation using random forest regression-voting," in *IEEE International Symposium on Biomedical Imaging*. Citeseer, 2015.

[13] Sercan Ö Arik, Bulat Ibragimov, and Lei Xing, "Fully automated quantitative cephalometry using convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 1, pp. 014501, 2017.