



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Cephalometric Landmark Detection across Ages with Prototypical Network

Han Wu^{1*}, Chong Wang^{2*}, Lanzhuju Mei¹, Tong Yang⁶, Min Zhu⁵,
Dinggang Shen^{1,3,4}, and Zhiming Cui^{1(✉)}

¹ School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China

cuizhm@shanghaitech.edu.cn

² Australian Institute for Machine Learning, The University of Adelaide

³ Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China

⁴ Shanghai Clinical Research and Trial Center, Shanghai, China

⁵ Shanghai Ninth People's Hospital, Shanghai Jiao Tong University, Shanghai, China

⁶ Shanghai Linkedcare Information Technology Co., Ltd., Shanghai, China

Abstract. Automated cephalometric landmark detection is crucial in real-world orthodontic diagnosis. Current studies mainly focus on only adult subjects, neglecting the clinically crucial scenario presented by adolescents whose landmarks often exhibit significantly different appearances compared to adults. Hence, an open question arises about how to develop a unified and effective detection algorithm across various age groups, including adolescents and adults. In this paper, we propose CeLDA, the first work for **C**ephalometric **L**andmark **D**etection across **A**ges. Our method leverages a prototypical network for landmark detection by comparing image features with landmark prototypes. To tackle the appearance discrepancy of landmarks between age groups, we design new strategies for CeLDA to improve prototype alignment and obtain a holistic estimation of landmark prototypes from a large set of training images. Moreover, a novel prototype relation mining paradigm is introduced to exploit the anatomical relations between the landmark prototypes. Extensive experiments validate the superiority of CeLDA in detecting cephalometric landmarks on both adult and adolescent subjects. To our knowledge, this is the first effort toward developing a unified solution and dataset for cephalometric landmark detection across age groups. Our code and dataset will be made public on Github.

Keywords: Cephalometric Landmark · Prototypical Network · Landmark Prototypes · Relation Mining · Prototype Alignment

1 Introduction

Automatic and accurate detection of cephalometric landmarks holds significant importance in clinical practice, particularly for orthodontic diagnosis and therapy planning [14]. With the remarkable achievements of deep learning [13,21],

* indicates equal contribution.

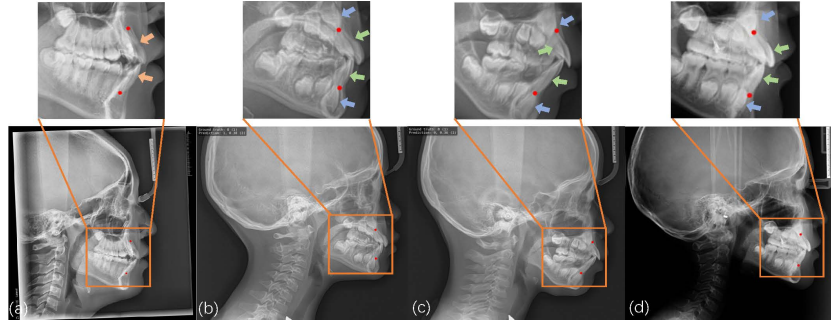


Fig. 1: (a) An adult case, with regular anatomical structures and permanent teeth (orange arrow); (b,c,d) adolescent cases, with complicated anatomical changes due to unerupted teeth (blue arrow) and baby teeth (green arrow). These changes on adolescent cases cause significant landmark shifts. Here we show only two landmarks (red points) out of ten for better visualization.

there are many learning-based efforts made for detecting cephalometric landmarks, i.e., regressing landmarks with deep convolutional neural networks [7,10], improving detection performance with two-stage networks [6,8,16,26], and modelling landmark relationships with anatomical prior information [2,9].

Regardless of their encouraging performances, these existing approaches are mostly dedicated to detecting cephalometric landmarks on adult subjects, which has clear skull bone and regular tooth arrangement shown in Fig. 1(a), ignoring more challenging adolescent subjects that often have complicated morphological changes in anatomy due to the presence of unerupted and permanent teeth as in Fig. 1(b,c,d). Such changes are prone to cause significant shifts of the cephalometric landmarks [17]. Very recently, Ceph-Net [23] targets the cephalometric landmark detection on adolescent cases and utilizes an attention-based stacked regression network to progressively refine detection results. However, Ceph-Net considers only adolescent cases but the common adult cases are not included. To date, it remains unexplored and needs to be addressed to develop a unified and effective cephalometric landmark detection algorithm across different age groups, including both adolescent and adult cases. Generally, the main obstacle in approaching such an algorithm comes from the landmark shifts across age groups, necessitating robust learning capabilities of the algorithm.

In this paper, we propose CeLDA for age-inclusive cephalometric landmark detection. Specifically, our CeLDA relies on a prototypical network to realize landmark detection by comparing image features with landmark prototypes. To ensure robust prototypes against the landmark shifts from different age groups, we present new strategies for CeLDA to promote prototype alignment and obtain a holistic estimation of landmark prototypes from a large set of training samples. Furthermore, a novel prototype relation mining paradigm is introduced to leverage anatomical relations among landmarks. Extensive experimental results illustrate that our CeLDA outperforms existing state-of-the-art (SOTA)

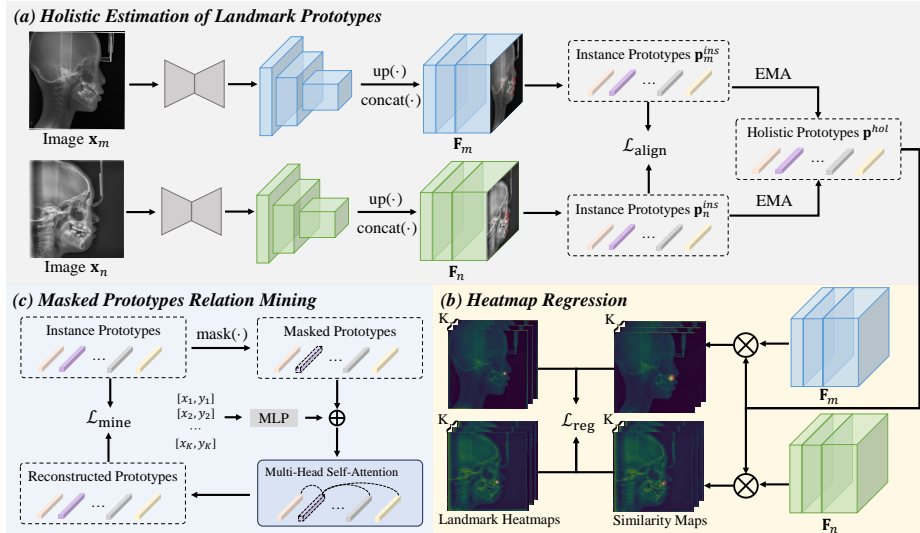


Fig. 2: An overview of the proposed CeLDA method for cephalometric landmark detection, based on a set of holistic landmark prototypes.

approaches in detecting cephalometric landmarks on adolescent subjects, adult subjects, and both. To summarise, **our major contributions are:** 1) the first prototype-based approach for age-inclusive cephalometric landmark detection, where the holistic prototypes are obtained to improve the learning robustness and predictive performance; 2) a novel prototype relation mining paradigm to take advantage of crucial anatomical relationships between landmarks; 3) a new comprehensive benchmark dataset for landmark detection that consists of cephalometric images from both adolescent and adult subjects.

2 Methodology

Our dataset \mathcal{D} comprises image-label pairs represented by (\mathbf{x}, \mathbf{y}) , where $\mathbf{x} \in \mathbb{R}^{H \times W}$ denotes a cephalometric image of size $H \times W$, and $\mathbf{y} \in \{0, 1\}^{K \times H \times W}$ represents K binary ground-truth landmark maps. Each of the K landmark maps only has one single annotated landmark point, i.e., $\sum \mathbf{y}_{k,:} = 1$. Following existing approaches [2,5,25,27], we transform the sparsely-distributed landmark maps into K landmark heatmaps $\mathbf{H}_k = \text{Gaussian}(\mathbf{y}_k) \in \mathbb{R}^{H \times W}$ for model training, using a Gaussian smoothing strategy as in [2,29].

2.1 Overview

An overview of our proposed method is shown in Fig. 2. For an input cephalometric image \mathbf{x} , we employ a network backbone f_θ , i.e., U-Net [12], to extract multi-level high-resolution feature maps $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\}$, where $\mathbf{F}_1 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times D_1}$,

$\mathbf{F}_2 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times D_2}$, $\mathbf{F}_3 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times D_3}$. In order to enable an accurate detection of the sparsely-distributed landmark, these feature maps are up-sampled to the original resolution of the input image and then concatenated into a composite feature map $\mathbf{F} = \text{concat}(\text{up}(\mathbf{F}_1), \text{up}(\mathbf{F}_2), \text{up}(\mathbf{F}_3)) \in \mathbb{R}^{H \times W \times D}$, where $D = (D_1 + D_2 + D_3)$ and $\text{up}(\cdot)$ denotes an up-sampling operation.

In Fig. 2(a), our CeLDA method leverages K holistic prototypes $\mathcal{P}_{hol} = \{\mathbf{p}_k^{hol}\}_{k=1}^K$ that are estimated from a large set of training samples, as introduced in Section 2.2. In \mathcal{P}_{hol} , each prototype $\mathbf{p}_k^{hol} \in \mathbb{R}^{1 \times 1 \times D}$ corresponds to one landmark and captures robust landmark-representative features. After that, we derive K similarity maps, see Fig. 2(b), by calculating the dot-product between the feature maps \mathbf{F} and each prototype \mathbf{p}_k^{hol} , which is formulated as:

$$\mathbf{S}_k = \mathbf{p}_k^{hol} \cdot \mathbf{F} \in \mathbb{R}^{H \times W}. \quad (1)$$

Finally, the detection prediction for the k -th landmark is obtained by selecting the location in \mathbf{S}_k that has the highest similarity. For model training, the standard regression loss is utilized to supervise our CeLDA:

$$\mathcal{L}_{\text{reg}} = \frac{1}{K} \sum_k \|\mathbf{S}_k - \mathbf{H}_k\|_2^2 \quad (2)$$

where \mathbf{H}_k is the k -th ground-truth landmark heatmap.

In the following section, we elaborate how to estimate and obtain the prototypes \mathcal{P}_{hol} for robust cephalometric landmark detection across age groups.

2.2 Holistic Estimation of Landmark Prototypes

Prototypes have been studied for classification [15] and segmentation [28] for a long time [20], where their essence is representing classes by prototypes to encode class-representative features. Making an analogy to our landmark detection task, it is natural to define prototypes to represent landmarks, i.e., capturing landmark-representative features. To achieve this, we propose to first create instance-level landmark prototypes $\mathcal{P}_{ins} = \{\mathbf{p}_k^{ins}\}_{k=1}^K$ for each individual training image \mathbf{x} , where $\mathbf{p}_k^{ins} \in \mathbb{R}^{1 \times 1 \times D}$ and each of them is calculated as:

$$\mathbf{p}_k^{ins} = \frac{\sum_{i,j} \mathbf{H}_k(i,j) \cdot \mathbf{F}(i,j)}{\sum_{i,j} \mathbf{H}_k(i,j)} \quad (3)$$

where $i \in \{1, \dots, H\}$ and $j \in \{1, \dots, W\}$ are spatial indexes. From Eq. (3), the instance prototypes \mathcal{P}_{ins} are generated by averaging the local contextual features around the landmark point. Although straightforward and easy to implement, one noticeable shortcoming of the instance-level prototypes is that they consider only individual-image information, which is insufficient to encapsulate the drastic appearance variations of the cephalometric landmarks, particularly for different age groups. To overcome this problem, we propose a new strategy to achieve a holistic estimation of the landmark prototypes. Specifically, inspired by the

well-established exponential moving averaging (EMA) technique, we obtain the holistic prototypes \mathcal{P}_{hol} in an on-the-fly fashion by exploiting a large set of training samples, as formulated below:

$$\mathcal{P}_{hol}^{(t+1)} = \alpha \cdot \mathcal{P}_{hol}^{(t)} + (1 - \alpha) \cdot \frac{1}{|\mathcal{B}|} \sum_{b=1}^{|\mathcal{B}|} \mathcal{P}_{ins}^{(t),b}, \quad (4)$$

where \mathcal{B} denotes a training mini-batch with size $|\mathcal{B}|$, α is a momentum update coefficient, t indicates the training iteration, and \mathcal{P}_{hol} is our holistic prototypes used for reliable detection of the landmarks, as described in Section 2.1. It is worth noting in Eq. (4) that during training our holistic prototypes are slowly progressing to take advantage of information from not only the current mini-batch but also historical prototypes. Therefore, they will gradually gain a global picture of the whole training set, allowing a robust landmark detection from cephalometric images across ages, such as adolescent and adult stages.

2.3 Cross-image Prototype Alignment

According to Eq. (4), our holistic prototypes are obtained by accumulating instance-level prototypes during training, to increase the prototype robustness we also propose to encourage prototype alignment across individual images:

$$\mathcal{L}_{align} = \frac{1}{K} \sum_{k=1}^K \|\mathbf{p}_{m,k}^{ins} - \mathbf{p}_{n,k}^{ins}\|_2^2, \quad (5)$$

where $\mathbf{p}_{m,k}^{ins}$ and $\mathbf{p}_{n,k}^{ins}$ denote the k -th instance-level prototypes for the image \mathbf{x}_m and \mathbf{x}_n within the mini-batch \mathcal{B} , respectively. Notice that Eq. (5) is able to enforce prototype consistency for training samples from not only within the same age group but also across different age groups.

2.4 Masked Prototype Relation Mining

As illustrated in Fig. 1, landmarks naturally have crucial anatomical relations within a cephalometric image [24]. Given that our CeLDA harnesses prototypes to represent landmarks, we further present a novel prototype relation mining paradigm to exploit the anatomical dependency between landmarks.

Motivated by the great success of masked modeling in language [3] and vision [4] applications, in this paper, we propose to mask the instance-level landmark prototypes. As demonstrated in Fig. 2(c), after obtaining K instance prototypes \mathcal{P}_{ins} from a training image, we randomly mask out a proportion of prototypes in \mathcal{P}_{ins} and replace them with zero, where the landmark positional embeddings are introduced as a location indicator. The combination of masked prototypes and positional embeddings is processed by a multi-head self-attention (MSA) layer, which reconstructs the masked prototypes as follows:

$$\hat{\mathcal{P}}_{ins} = \text{MSA}(\text{mask}(\mathcal{P}_{ins}) \oplus \mathbf{E}_{pos}), \quad \mathbf{E}_{pos} = \text{MLP}(\bar{\mathbf{y}}), \quad (6)$$

where $\hat{\mathcal{P}}_{ins}$ is the reconstructed prototypes, $\text{mask}(\cdot)$ is a mask-out operation to randomly exclude a ratio (denoted by R) of prototypes from \mathcal{P}_{ins} , \oplus represents the element-wise summation, and \mathbf{E}_{pos} denotes the landmark positional embeddings, encoding the ground-truth landmark coordinates $\bar{\mathbf{y}} \in \mathbb{R}^{K \times 2}$ using a multi-layer perceptron (MLP), where the landmark coordinates $\bar{\mathbf{y}}$ can be easily derived from the ground-truth landmark maps \mathbf{y} . The reconstructed prototypes are supervised by the original prototypes in \mathcal{P}_{ins} :

$$\mathcal{L}_{\text{mine}} = \sum_{k=1}^K \|\hat{\mathbf{p}}_k^{ins} - \mathbf{p}_k^{ins}\|_2^2, \quad (7)$$

where $\hat{\mathbf{p}}_k^{ins}$ denotes a reconstructed prototype in $\hat{\mathcal{P}}_{ins}$, and \mathbf{p}_k^{ins} is the corresponding raw prototype in \mathcal{P}_{ins} . Relying on Eq. (7), our CeLDA can make full use of the structural information regarding landmark relations during the process of learning the instance-level prototypes for each training sample, benefiting its understanding of the anatomical landmark dependency.

2.5 Overall Training Objective

The overall optimization objective of CeLDA is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{reg}} + \lambda_1 \mathcal{L}_{\text{align}} + \lambda_2 \mathcal{L}_{\text{mine}}, \quad (8)$$

where λ_1 and λ_2 are hyper-parameters to control the weight of the loss terms.

3 Experiments

3.1 Dataset and Evaluation Metric

Dataset: For the task of cephalometric landmark detection across age groups, we collected a new benchmark dataset, named CephAdoAdu, with both adolescent and adult cases, distinguishing it from existing datasets that solely consist of either adolescent or adult cases. CephAdoAdu has a total of 1000 (500 adult cases, 500 adolescent cases) cephalometric X-ray images, acquired from eight clinical centers. Every cephalometric image underwent manual annotations to mark 10 typical landmarks, by an experienced dental radiologist with over ten years of expertise. Our new dataset has two advantages over existing ones: 1) a more clinically practical coverage of subjects across different age groups; 2) a larger number of annotated images, ensuring a comprehensive and faithful model evaluation. The whole dataset is randomly divided into training set (400 images), validation set (300 images), and testing set (300 images). Notice that our data split is evenly performed in terms of the adult and adolescent cases.

Table 1: Cephalometric landmark detection results with both adult and adolescent cases, only adult cases, and only adolescent cases, respectively.

Methods	Adult + Adolescent					Adult					Adolescent				
	MRE ↓ (mm, std.)	SDR (%) ↑				MRE ↓ (mm, std.)	SDR (%) ↑				MRE ↓ (mm, std.)	SDR (%) ↑			
		2mm	2.5mm	3mm	4mm		2mm	2.5mm	3mm	4mm		2mm	2.5mm	3mm	4mm
Cascade RCNN [1]	2.31 (0.94)	61.47	73.20	81.13	90.77	2.19 (0.97)	59.93	72.13	80.47	90.80	2.43 (0.94)	63.00	74.27	81.80	90.73
SCN [11]	1.73 (1.06)	82.97	90.40	93.37	96.57	1.40 (0.48)	82.07	91.20	94.33	97.33	2.05 (1.70)	83.87	89.60	92.40	95.80
GU2Net [29]	1.69 (0.91)	80.33	88.13	91.47	95.57	1.46 (0.50)	80.27	88.80	92.07	96.33	1.93 (1.35)	80.40	87.47	90.87	94.80
Wu et al. [22]	1.34 (1.24)	87.17	91.93	95.57	97.10	1.13 (0.66)	86.60	92.13	95.00	97.80	1.55 (1.87)	87.73	91.73	94.13	96.40
CeLDA	1.05 (0.33)	89.13	93.60	96.17	98.67	1.10 (0.37)	88.33	92.93	96.20	98.80	1.00 (0.32)	89.93	94.27	96.13	98.53

Evaluation Metric: Following previous studies [18,19], we evaluate the model performance with the two commonly-used metrics: 1) Mean Radial Error (MRE) computes the average Euclidean distance between the predicted and ground-truth landmarks; 2) Successful Detection Rate (SDR) is defined as the percentage of landmarks that are accurately detected within a range of 2.0 mm, 2.5 mm, 3.0 mm, and 4 mm from the ground-truth landmarks.

3.2 Implementation Details

Our CeLDA employs U-Net [12] as the network backbone f_θ . In Eq. (4), $\alpha = 0.99$ and the mini-batch size $|\mathcal{B}| = 8$. All images are resized to 512×512 as model input. Training images are augmented to introduce random changes in brightness, contrast, and Gaussian noise. Our CeLDA is optimized for a total of 150 training epochs, using SGD optimizer with a learning rate of 0.005 which is decreased by a factor of 0.1 per 50 epochs. λ_1 and λ_2 in Eq. (8) are set to 1.0 and 3.0 respectively. We have the mask ratio $R = 0.7$ for the prototype relation mining. All methods were implemented in the PyTorch framework and trained on an NVIDIA Tesla A100 GPU with 40GB memory.

3.3 Comparison with SOTA Approaches

We compare our CeLDA with the following typical landmark detection models. Cascade RCNN [1] detects cephalometric landmarks with a multi-stage object detection architecture to progressively eliminate noisy predictions. SCN [11] regresses landmark heatmaps with a fully convolutional network that considers the spatial configuration of landmarks. GU2Net [29] is a universal landmark detection method that solves multiple detection tasks with end-to-end training on mixed datasets. We also compare with the recent champion method, proposed by Wu et al. [22], in the MICCAI CL-Detection2023 leaderboard. It is worth noting that all the above approaches are designed for detecting landmarks from only adult images. To achieve comparison fairness, all these competing approaches employ the same image augmentation strategies mentioned in Section 3.2.

Table 1 provides landmark detection results on CephAdoAdu test set. As evident, our CeLDA consistently outperforms other competing approaches on both adult and adolescent cases, only adult cases, and only adolescent cases. In particular, our CeLDA exhibits more improvements (in both MRE and SDR metrics)

Table 2: Ablation analysis for our proposed CeLDA method.

\mathcal{L}_{reg}	\mathcal{L}_{align}	\mathcal{L}_{mine}	Adult + Adolescent				Adult				Adolescent						
			MRE ↓ (mm, std.)	SDR (%) ↑				MRE ↓ (mm, std.)	SDR (%) ↑				MRE ↓ (mm, std.)	SDR (%) ↑			
				2mm	2.5mm	3mm	4mm		2mm	2.5mm	3mm	4mm		2mm	2.5mm	3mm	4mm
✓			1.16 (0.38)	86.77	92.00	95.30	98.10	1.16 (0.36)	85.93	91.73	95.60	98.53	1.17 (0.42)	87.60	92.27	95.00	97.67
✓	✓		1.13 (0.34)	87.83	92.97	95.70	98.30	1.14 (0.36)	86.20	92.47	96.07	98.87	1.12 (0.44)	89.47	93.47	95.33	97.73
✓	✓	✓	1.05 (0.33)	89.13	93.60	96.17	98.67	1.10 (0.37)	88.33	92.93	96.20	98.80	1.00 (0.34)	89.93	94.27	96.13	98.53

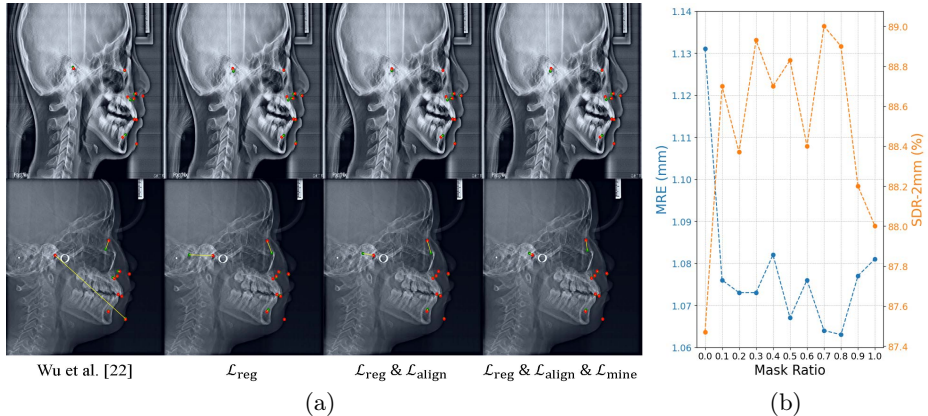


Fig. 3: (a) Visual comparison of different landmark detection methods, where the ground truth and predicted landmarks are indicated in green and red dots respectively, and each pair of them is linked with a yellow line. (b) Effect of the mask ratio R for mining landmark prototype relations.

for only adolescent cases compared with only adult cases, showing its strength in detecting more challenging adolescent cephalometric landmarks. Moreover, CeLDA also greatly surpasses other approaches in both adult and adolescent cases, verifying its effectiveness in detecting landmarks across age groups.

3.4 Analytical Ablation Studies

We perform ablation experiments to study the effectiveness of our prototype alignment \mathcal{L}_{align} and prototype relation mining \mathcal{L}_{mine} , with results given in Table 2. We observe that the baseline (using only \mathcal{L}_{reg}) achieves an MRE of 1.16 mm on adult and adolescent cases, which reduces to 1.13 mm, upon the use of prototype alignment loss. Remarkable performance improvements can be observed when further utilizing the prototype relation mining strategy, showing the advantage of mining prototype relations to harness the anatomical dependency between landmarks. We show typical visual landmark detection results in Fig. 3 (a), where we observe progressive improvements with the incorporation of each key component in our method.

In Fig. 3 (b), we explore the sensitivity of our CeLDA to the mask ratio R used for the landmark prototype relation mining. As evident, a small mask ratio

is inadequate to mine the prototype relations, resulting in sub-optimal results. Conversely, a large mask ratio may lead CeLDA to reconstruct wrong landmark prototypes, causing a decline in predictive performance. According to Fig. 3 (b), we set the mask ratio at 0.7 in all other experiments.

4 Conclusion

In this work, we presented the CeLDA method to address cephalometric landmark detection across different age groups with the prototypical network. Our CeLDA detects cephalometric landmarks by comparing image features with a set of holistic landmark prototypes, where their anatomical relations are exploited with a masking-based mining strategy. Our CeLDA shows great superiority over existing approaches on adolescent and adult cases. We established and released the first cephalometric benchmark dataset covering a large number of both adult and adolescent cases, with the hope that it will provide a more comprehensive evaluation for the landmark detection community.

Acknowledgments. This work was supported in part by NSFC grants (No. 6230012077) and Shanghai Municipal Central Guided Local Science and Technology Development Fund Project (No: YDZX20233100001001).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article

References

1. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 6154–6162 (2018)
2. Chen, R., Ma, Y., Liu, L., Chen, N., Cui, Z., Wei, G., Wang, W.: Semi-supervised anatomical landmark detection via shape-regulated self-training. *Neurocomputing* **471**, 335–345 (2022)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
4. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16000–16009 (2022)
5. He, T., Yao, J., Tian, W., Yi, Z., Tang, W., Guo, J.: Cephalometric landmark detection by considering translational invariance in the two-stage framework. *Neurocomputing* **464**, 15–26 (2021)
6. Jiang, Y., Li, Y., Wang, X., Tao, Y., Lin, J., Lin, H.: Cephalformer: Incorporating global structure constraint into visual features for general cephalometric landmark detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 227–237. Springer (2022)

7. Lee, H., Park, M., Kim, J.: Cephalometric landmark detection in dental x-ray images using convolutional neural networks. In: *Medical imaging 2017: Computer-aided diagnosis*. vol. 10134, pp. 494–499. SPIE (2017)
8. Lee, J.H., Yu, H.J., Kim, M.j., Kim, J.W., Choi, J.: Automated cephalometric landmark detection with confidence regions using bayesian convolutional neural networks. *BMC Oral Health* **20**, 1–10 (2020)
9. Li, W., Lu, Y., Zheng, K., Liao, H., Lin, C., Luo, J., Cheng, C.T., Xiao, J., Lu, L., Kuo, C.F., et al.: Structured landmark detection via topology-adapting deep graph learning. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX* 16. pp. 266–283. Springer (2020)
10. Oh, K., Oh, I.S., Lee, D.W., et al.: Deep anatomical context feature learning for cephalometric landmark detection. *IEEE Journal of Biomedical and Health Informatics* **25**(3), 806–817 (2020)
11. Payer, C., Štern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based cnns for landmark localization. *Medical Image Analysis* **54**, 207–219 (2019)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. pp. 234–241. Springer (2015)
13. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural Networks* **61**, 85–117 (2015)
14. Schwendicke, F., Chaurasia, A., Arsiwala, L., Lee, J.H., Elhennawy, K., Jost-Brinkmann, P.G., Demarco, F., Krois, J.: Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clinical Oral Investigations* **25**(7), 4299–4309 (2021)
15. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems* **30** (2017)
16. Song, Y., Qiao, X., Iwamoto, Y., Chen, Y.w.: Automatic cephalometric landmark detection on x-ray images using a deep-learning method. *Applied Sciences* **10**(7), 2547 (2020)
17. Tanikawa, C., Yamamoto, T., Yagi, M., Takada, K.: Automatic recognition of anatomic features on cephalograms of preadolescent children. *The Angle Orthodontist* **80**(5), 812–820 (2010)
18. Wang, C.W., Huang, C.T., Hsieh, M.C., Li, C.H., Chang, S.W., Li, W.C., Vandaele, R., Marée, R., Jodogne, S., Geurts, P., et al.: 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 (2015)
19. Wang, C.W., Huang, C.T., Lee, J.H., Li, C.H., Chang, S.W., Siao, M.J., Lai, T.M., Ibragimov, B., Vrtovec, T., Ronneberger, O., et al.: A benchmark for comparison of dental radiography analysis algorithms. *Medical Image Analysis* **31**, 63–76 (2016)
20. Wang, C., Chen, Y., Liu, F., Elliott, M., Kwok, C.F., Peña-Solorzano, C., Frazer, H., McCarthy, D.J., Carneiro, G.: An interpretable and accurate deep-learning diagnosis framework modelled with fully and semi-supervised reciprocal learning. *IEEE Transactions on Medical Imaging* (2023)
21. Wang, C., Cui, Z., Yang, J., Han, M., Carneiro, G., Shen, D.: Bowelnet: Joint semantic-geometric ensemble learning for bowel segmentation from both partially and fully labeled ct images. *IEEE Transactions on Medical Imaging* **42**(4), 1225–1236 (2022)

22. Wu, Q., Yeo, S.Y., Chen, Y., Liu, J.: Revisiting cephalometric landmark detection from the view of human pose estimation with lightweight super-resolution head. arXiv preprint arXiv:2309.17143 (2023)
23. Yang, S., Song, E.S., Lee, E.S., Kang, S.R., Yi, W.J., Lee, S.P.: Ceph-net: automatic detection of cephalometric landmarks on scanned lateral cephalograms from children and adolescents using an attention-based stacked regression network. *BMC Oral Health* **23**(1), 803 (2023)
24. Yao, Q., Quan, Q., Xiao, L., Kevin Zhou, S.: One-shot medical landmark detection. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II* 24. pp. 177–188. Springer (2021)
25. Yueyuan, A., Hong, W.: Swin transformer combined with convolutional encoder for cephalometric landmarks detection. In: *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (IC-CWAMTIP)*. pp. 184–187. IEEE (2021)
26. Zeng, M., Yan, Z., Liu, S., Zhou, Y., Qiu, L.: Cascaded convolutional networks for automatic cephalometric landmark detection. *Medical Image Analysis* **68**, 101904 (2021)
27. Zhong, Z., Li, J., Zhang, Z., Jiao, Z., Gao, X.: An attention-guided deep regression model for landmark detection in cephalograms. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI* 22. pp. 540–548. Springer (2019)
28. Zhou, T., Wang, W., Konukoglu, E., Van Gool, L.: Rethinking semantic segmentation: A prototype view. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2582–2593 (2022)
29. Zhu, H., Yao, Q., Xiao, L., Zhou, S.K.: You only learn once: Universal anatomical landmark detection. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V* 24. pp. 85–95. Springer (2021)