



An automated estimator for Cobb angle measurement using multi-task networks

Xiangling Fu¹ · Guosheng Yang¹ · Kailai Zhang² · Nanfang Xu³ · Ji Wu²

Received: 16 July 2020 / Accepted: 11 November 2020 / Published online: 27 November 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

Scoliosis is a medical condition where a person's spine has a sideways curve. The Cobb angle quantifying the degree of spinal curvature is the gold standard for a scoliosis assessment. Recently, the deep learning methods based on segmentation and landmark estimation both achieve high performance for automated Cobb angle measurement on X-rays. However, we notice that these methods utilize segmentation and landmark information separately. In this light, we propose an automated architecture that uses combined segmentation with landmark information to estimate 68 landmarks of 17 vertebrae. In addition, we consider spinal curvature described by 68 landmarks as a constraint to estimate the Cobb angle. Extensive experiment results which test on 240 X-rays demonstrate that our method improves the landmark estimation performance effectively and reduces the Cobb angle error.

Keywords Cobb angle · X-rays · Deep learning · Multi-task learning

1 Introduction

Adolescent idiopathic scoliosis (AIS) is defined as a spinal torsional deformity combined with different degrees of rotational spinal deformity [1]. According to the current literature, 0.47–5.2% [2] of children have different degrees of scoliosis.

The Cobb method [3] is considered as a classical and efficient way to quantitatively measure the angle of scoliosis both on the coronal and the sagittal plane such as the Cobb angle [4]. The Cobb angle is the angle between the two most tilted vertebrae, specifically between the upper endplate of the uppermost vertebra and the lower endplate

of the lowest vertebra, as shown in Fig. 1. However, manual Cobb angle measurement is time-consuming on X-rays with a low contrast because clinicians find four landmarks on every vertebra and compare the slope of them to measure the Cobb angle.

In this paper, we propose an automated architecture for the Cobb angle measurement. The architecture uses both segmentation and landmarks of vertebrae to supervise estimated landmarks. In addition, the architecture considers spinal curvature as a constraint to estimate the Cobb angle.

2 Related work

Recent studies based on deep learning have proposed some effective methods for the Cobb angle measurement on X-rays. These methods can be divided into two categories: (1) direct landmark estimation methods, and (2) indirect segmentation methods.

The direct landmark estimation methods aim to directly capture landmarks of interest on X-rays, which are like the manual process. Such as landmarks with a structured multi-output regression network are predicted in [5]. The use of

✉ Ji Wu
wuji_ee@tsinghua.edu.cn

¹ School of Computer Science (National Pilot Software Engineering School), BUPT, Beijing University of Posts and Telecommunications, Beijing 100876, China

² Department of Electronic Engineering, Tsinghua University, Beijing, China

³ Peking University Third Hospital, Beijing, China

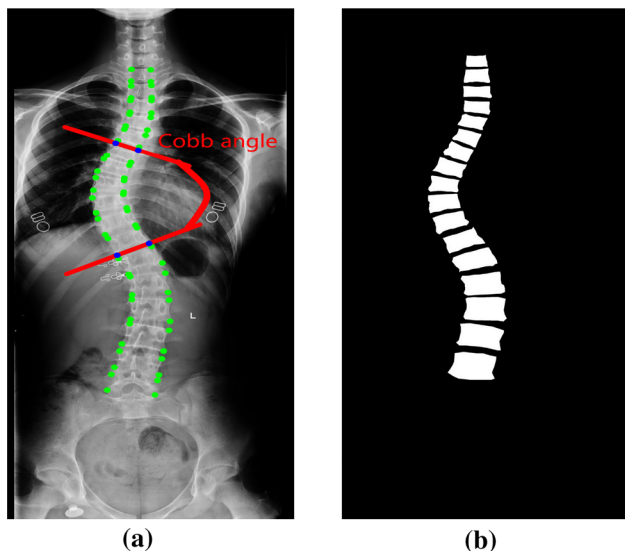


Fig. 1 **a** An anterior-posterior X-ray image showing how to measure the Cobb angle. Green points are the four landmarks of every vertebra. Blue points are the landmarks of the two most tilted vertebrae. **b** Corresponding manually generated binary mask

Boost-net to find landmarks by transforming the feature space is proposed in [6]. A series of methods (MVC-Net, MVE-Net) that use multi-view (anterior-posterior and lateral) X-rays together joint features of multi-view X-rays [7, 8]. An AEC-Net uses calculated Cobb angles by rules and estimated angles to correct the Cobb angle error [9].

The indirect segmentation methods aim to segment the vertebrae of interest on X-rays and then measure the Cobb angle based on the segmentation. Such as in [10], an automated model is proposed for spine segmentation, and a polynomial to fit the spinal curvature. Owing to the high-level performance of U-Net [11] in the medical field, numerous studies are based on the use of U-Net, such as dense U-Net [12], residual U-Net [13] and shape-aware U-Net [14]. In [15], an automatic DU-Net segmenting the spine based on deep learning is proposed, and a sixth polynomial to characterize the spinal curvature. In [16], an MBR-Net based on U-Net to segment the images and a minimum bounding rectangle is considered vertebrae. The Mask RCNN [17] is used to segment vertebrae, and the centers of segmentation are used to calculate the Cobb angle [18]. The U-Net is used to segment lumbar vertebrae and estimate the lumbar lordosis angle on lateral X-rays in [19]. In [20], the Mask RCNN is used to segment vertebrae and a small network to estimate landmarks on anterior-posterior X-rays.

These two methods both achieve high performance for the Cobb angle measurement. However, these methods separately use the segmentation and landmark information to supervise networks. In this light, we propose an automated architecture which uses combined segmentation and

landmark information. It takes the segmentation as an auxiliary task to estimate landmarks and uses spinal curvature to estimate the Cobb angle on anterior-posterior X-rays. The experiment results show that our method achieves smaller error on landmark and Cobb angle estimation.

3 Methods

3.1 Overview

In this study, we propose a landmark and Cobb angle estimation network (LCE-Net). The architecture consists of two parts: (1) a landmark estimation network (LEN), and (2) a Cobb angle estimation network (CEN). The LEN first estimates 68 landmarks of 17 vertebrae (12 thoracic vertebrae and 5 lumbar vertebrae) by taking segmentation as an auxiliary task. Then, the CEN uses the spinal curvature described by 68 landmarks to estimate the Cobb angle by considering spinal curvature as a constraint.

3.2 Landmark estimation network

We assume that the locations of the landmarks and the segmentation have a potential relationship in a physical space. The LEN combines the features of two kinds of networks: (1) the network for segmentation (NFS) and (2) the network for landmark estimation (NFL). Because of the FCN [21] achieves high performance for image segmentation, the LEN takes the FCN as the NFS and a simple network composed of fully convolutional layers as the NFL. The LEN combines the segmentation and landmark information by concatenating features of the NFS and NFL.

The architecture of the LEN is shown in Fig. 2. Inputting an X-ray, the LEN outputs a pixel-wise segmentation heatmap and landmark coordinates:

$$I \rightarrow h, c$$

where I means an image, h means an estimated heatmap and c means estimated landmark coordinates formatted as $EC = [x_1^e, y_1^e, \dots, x_{68}^e, y_{68}^e]$. The landmarks are arranged from top to bottom and from left to right. We scale the estimated landmarks and ground-truth landmarks between 0 and 1. Estimated coordinates are normalized by a sigmoid function: $EC = \frac{1}{1+e^{-x}}$. Ground-truth coordinates are normalized by $GC = [x_1^g/w, y_1^g/h, \dots, x_{68}^g/w, y_{68}^g/h]$, where w and h are the width and height of the image size.

Two types of loss functions are used in training stage: (1) a mean squared error loss is used for comparing estimated landmarks to ground-truth landmarks per image:

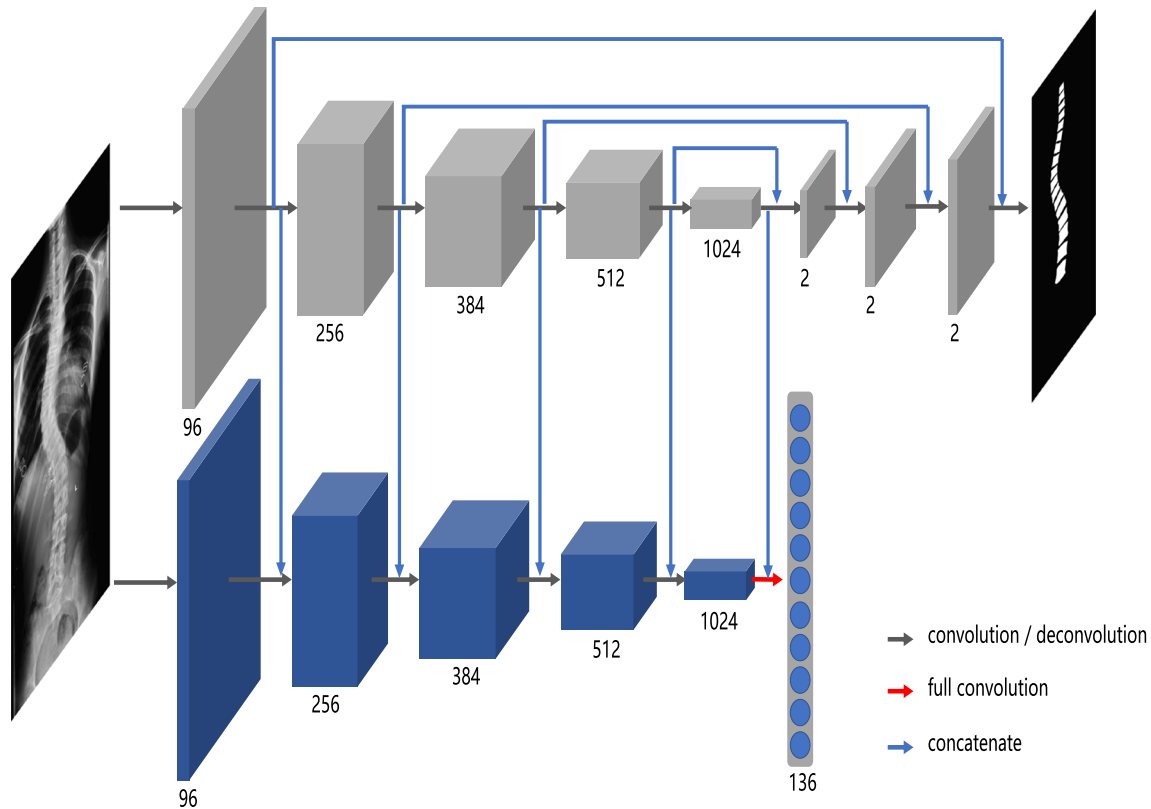


Fig. 2 An illustration of the LEN. The gray part is the segmentation network based on an FCN. The blue part is the landmark estimation network and it uses fully convolutional networks to estimate

landmarks. Inputting a 3-channel image, the LEN outputs a binary mask heatmap and 136 estimated coordinates of 68 landmarks

$$Lmse = \frac{1}{N} \sum_{i=1}^N (EC_i - GC_i)^2 \tag{1}$$

where N means the number of coordinates, 136 (68 x-coordinates and 68 y-coordinates) in our experiment. (2) A cross-entropy error loss is used for comparing estimated segmentation heatmaps to ground-truth heatmaps per image:

$$Lcee = -\frac{1}{WH} \sum_{i=1}^{WH} (y \log \hat{y}) \tag{2}$$

where W and H mean the width and height of the image, y means the ground-truth label and \hat{y} means the estimated probability of every pixel. Ground-truth segmentation heatmaps are constructed by modeling pixels of vertebrae as 1 and background as 0. The full training loss of the LEN is $Loss = Lmse + \varphi \times Lcee$ where φ is the weight to balance the segmentation and landmark estimation task, 0.05 in our experiment.

As ablation experiments, we design a series of networks that combine different level features of the NFS and the NFL. These architectures are shown in Fig. 3. All convolutions except the last layer of the LEN in our proposed

model use a 3×3 convolution kernel with a stride of 1 (with padding = 1); the last convolution uses a 4×2 convolution kernel, followed by batch normalization (BN) [22], prelu, and a dropout with a 25% probability [23].

3.3 Cobb angle estimation network

We found that a small landmark error can cause a big Cobb angle error because the slope of drawn lines shown in Fig. 1 may change too much. Addressing this issue, we assume that spinal curvature and the Cobb angle have a potential relationship. Unlike the manual measurement process which compares the most oblique vertebrae, the CEN uses spinal curvature described by 68 estimated landmarks as a constraint to estimate the Cobb angle. The architecture is shown in Fig. 4. The CEN takes EC as input and output the estimated Cobb angle:

$$c \rightarrow a$$

where a means the estimated Cobb angle. We also scale estimated Cobb angles and ground-truth Cobb angles between 0 and 1. Estimated Cobb angles are normalized by a sigmoid function: $EA = \frac{1}{1+e^{-x}}$. Ground-truth Cobb angles are normalized by $GA/180^\circ$. A mean squared error loss is

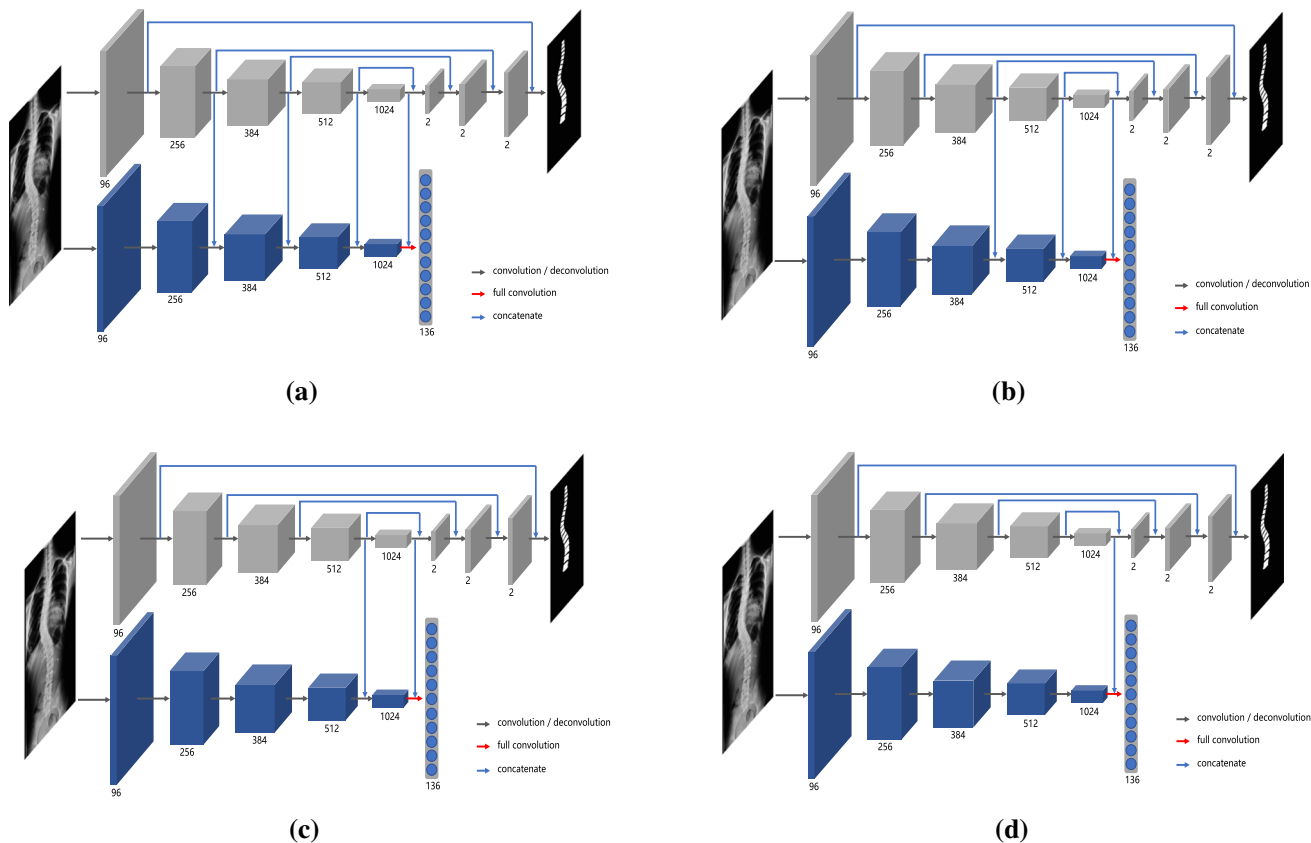


Fig. 3 A series of compared networks which combine different level features as ablation experiments. The only differences between these models are the different concatenating operations

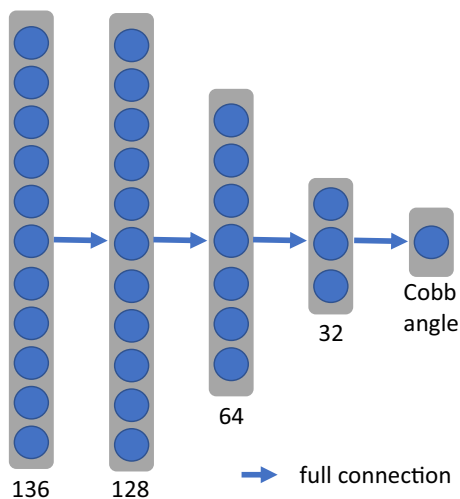


Fig. 4 An illustration of the CEN. Inputting an estimated landmark vector, CEN outputs the Cobb angle scaled between 0 and 1

also used for comparing estimated angles to ground-truth angles:

$$Lmse = \frac{1}{N} \sum_{i=1}^N (EA_i - GA_i)^2 \tag{3}$$

where N means the number of images.

4 Experiments

4.1 Dataset

Our dataset consists of 1200 spinal X-rays with an average pixel resolution of 957×491 provided by a local hospital. Four landmarks and the segmentation mask of each vertebra are labeled by two professional clinicians. Every clinician labels the half images, and labels are checked by each other. Each clinician has 8 years of experience. We scaled all images to a pixel resolution of 512×256 . The range of the Cobb angle is distributed from 1.56° to 91.74° in our dataset.

4.2 Training details

The experiments were run on a PC with Ubuntu 14.04, and an NVIDIA GeForce GTX 1080Ti GPU. The code implementation of the architecture is based on the Pytorch framework in Python. The learning rates of LEN and CEN

both are set to 0.001 and the momentum is set to 0.9 during the stochastic gradient descent (SGD). The 1200 X-rays are divided into the training set, validation set, and test set randomly in every training session, where the proportion is 6:2:2. The results are the average performance of 5-folds validation.

4.3 Performance metrics

For the landmark estimation, we use the landmark mean absolute error (LMAE) to calculate the error. The LMAE is defined as follows:

$$LMAE = \frac{1}{M} \frac{1}{N} \sum_{j=1}^M \sum_{i=1}^N |EC_i - GC_i| \tag{4}$$

where M is the number of images and N is the number of coordinates per image, 136 (68 x-coordinates and 68 y-coordinates) in our experiments.

For the Cobb angle estimation, we use the angle mean absolute error (AMAE) and symmetric mean absolute percentage error (SMAPE) to calculate the error:

$$AMAE = \frac{1}{M} \sum_{j=1}^M |\text{Angle}_i^{\text{est}} - \text{Angle}_i^{\text{gt}}| \tag{5}$$

$$SMAPE = \frac{100\%}{M} \sum_{j=1}^M \frac{|\text{Angle}_i^{\text{est}} - \text{Angle}_i^{\text{gt}}|}{(|\text{Angle}_i^{\text{est}}| + |\text{Angle}_i^{\text{gt}}|)/2} \tag{6}$$

where $\text{Angle}_i^{\text{est}}$ means estimated angles or calculated angles by estimated landmarks, and $\text{Angle}_i^{\text{gt}}$ means ground-truth angles. The method of calculating angles by landmarks is like the manual process shown in Fig. 1:

$$\text{Angle} = \left| \arctan \frac{y_2^{\text{up}} - y_1^{\text{up}}}{x_2^{\text{up}} - x_1^{\text{up}}} - \arctan \frac{y_2^{\text{low}} - y_1^{\text{low}}}{x_2^{\text{low}} - x_1^{\text{low}}} \right| \tag{7}$$

where x_i^{up} and y_i^{up} mean landmark coordinates on the upper endplate of the uppermost vertebra, x_i^{low} and y_i^{low} mean landmark coordinates on the lower endplate of the lowest vertebra. The upper and lower endplates are the two edges of the two most tilted vertebrae such as the two red lines shown in Fig. 1.

5 Results and discussion

5.1 Results

We compare our framework with other methods. We also compare the LEN with the NFL for landmark estimation. The results are shown in Table. 1. From the results, the LEN reduces the error of landmark estimation and the LCN

Table 1 Comparison with existing methods on X-rays

	LMAE (%)	AMAE (°)	SMAPE (%)
Indirect segmentation methods			
Tu et al. [15]	–	6.32 ± 4.17	32.61 ± 15.74
MBR-Net [16]	–	4.96 ± 3.74	26.91 ± 13.11
Pan et al. [18]	–	10.58 ± 7.23	44.25 ± 28.12
Direct estimation methods			
Boost-Net [6]	1.89 ± 1.30	7.13 ± 4.06	20.95 ± 16.30
Sun et al. [5]	1.92 ± 1.22	9.51 ± 7.64	36.76 ± 26.58
MVC-Net [7]	1.24 ± 1.03	7.79 ± 5.26	27.12 ± 23.63
AEC-Net [9]		7.48 ± 4.57	25.83 ± 21.90
Ours			
LEN	0.89 ± 0.81	3.73 ± 2.97	24.51 ± 17.25
LEN + CEN		3.15 ± 3.09	19.77 ± 15.98

We do not calculate the LMAE of indirect segmentation methods
The bold means the best result (such as the smallest error)

reduces the Cobb angle error. As shown in Table 2, we also compare the landmark estimation performance on a series of networks which combine different level feature shown in Fig. 3. The data in Tables 1 and 2 are calculated by Eqs. 4, 5 and 6.

From Table 1, the LEN achieves less landmark estimation error due to the use of more information. It uses the information of two tasks to supervise the landmark estimation while existing methods only use single information. The CEN achieves a smaller error of the Cobb angle estimation than the LEN due to considering spinal curvature as a constraint. It captures the relationship of the spinal curvature and the Cobb angle, which is more robust against the rules.

From Table 2, as ablation experiments, the LEN and models a to d almost have the same performance. They both achieve less error than the LEN without the segmentation branch due to using the similar multi-task

Table 2 Comparison with a series of networks which combine different level features

	LMAE(%)
LEN	0.89 ± 0.81
a	0.96 ± 0.89
b	0.98 ± 0.85
c	0.94 ± 0.89
d	0.92 ± 0.84
LEN without the segmentation branch	1.57 ± 0.97

The bold means the best result (such as the smallest error)

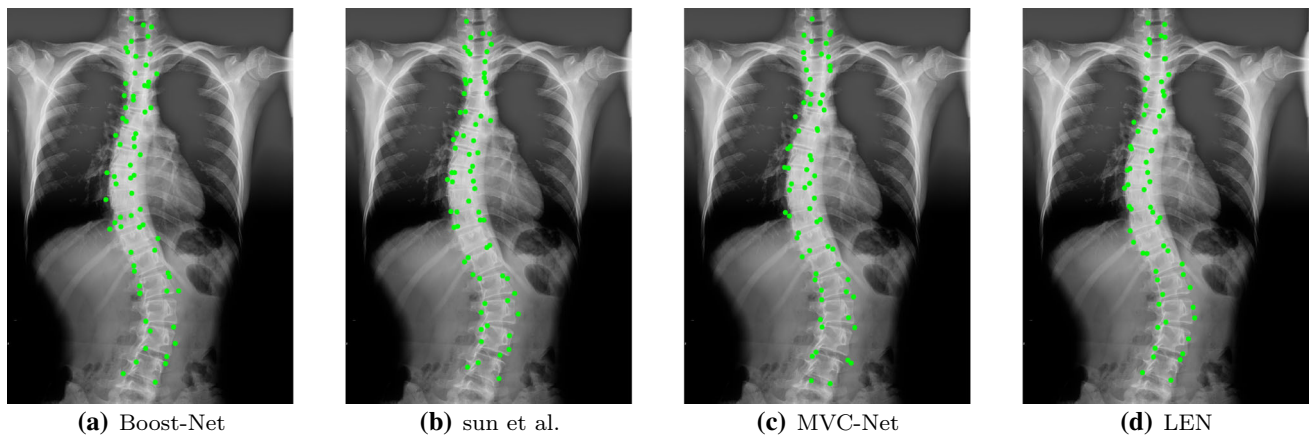


Fig. 5 Some visual results of the proposed method (d) and existing methods (a, b and c)

network architecture. Moreover, the LCE which uses most level features achieves higher performance than others.

Figure 5 shows some visual results of the proposed method and existing methods.

5.2 Discussion

Existing methods directly estimate landmarks or segment vertebrae, and then they use rules to calculate the Cobb angle such as calculating the center points of vertebrae [18] and fitting lines to be bounding box of vertebrae [16]. This may lead to a big angle error while there is a small segmentation error and landmark error. The LCE-Net avoids this issue due to two parts: (1) the LEN uses segmentation as an auxiliary task giving more information to estimate landmarks, which leads to more information utilization, and (2) the Cobb angle is estimated by spinal curvature instead of calculated by 7. Therefore, this method is more robust than the rules while some pivot landmarks are estimated with errors. The results demonstrate that our method is more robust both for landmark and angle estimation on X-rays.

This study has limitations. The LCE-Net uses a multi-task network to estimate landmarks, and this leads to more labeled information. For the same reason, the LCE-Net increases the computational cost, luckily not too much, and the computation time of the developed system is 0.16 ± 0.005 in the test stage. In terms of practical perspective, our method can meet the time and cost requirement to be integrated into clinicians' workflows.

6 Conclusion and future studies

In this paper, we first notice that existing methods for the Cobb angle estimation on X-rays use the segmentation and landmark information separately. To use the combined

information, we propose a multi-task network that takes segmentation as an auxiliary task to estimate landmarks. It achieves higher performance than existing methods on landmark estimation. In addition, to avoid a big angle error caused by a small landmark error, we propose a Cobb angle estimation network that uses spinal curvature described by 68 landmarks to estimate the Cobb angle instead of pivot landmarks to calculate by rules.

As future work, we plan to analyze whether we can apply our methods on 3-D images or combine X-rays in different directions. Future studies will also explore whether our methods can be used to estimate other clinical parameters based on spinal curvature.

Acknowledgements This study was supported by the National Key Research and Development Program of China (No. 2018Y-FC0116800), by Beijing Municipal Natural Science Foundation (No. L192026), by the Young Scientists Fund of the National Natural Science Foundation of China (No. 2019NSFC81901822) and by the Peking University Fund of Fostering Young Scholars' Scientific & Technological Innovation (No. BMU2018PYB016).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Clark EM, Taylor HJ, Harding I, Hutchinson J, Nelson I, Deanfield JE, Ness AR, Tobias JH (2014) Association between components of body composition and scoliosis: a prospective cohort study reporting differences identifiable before the onset of scoliosis. *J Bone Miner Res* 29(8):1729. <https://doi.org/10.1002/jbmr.2207>
2. Konieczny MR, Senyurt H, Krauspe R (2013) Epidemiology of adolescent idiopathic scoliosis. *J Child Orthop* 7(1):3. <https://doi.org/10.1007/s11832-012-0457-4>
3. Harrison Harrison DD, Cailliet R, Troyanovich SJ, Janik TJ, Holland B (2000) Cobb method or harrison posterior tangent

- method: which to choose for lateral cervical radiographic analysis. *Spine* (Phila. Pa. 1976). <https://doi.org/10.1097/00007632-200008150-00011>
4. Cobb JR (1917) Outline for the study of bitumens. *Sch Sci Math* 17(1):31. <https://doi.org/10.1111/j.1949-8594.1917.tb01839.x>
 5. Sun H, Zhen X, Bailey C, Rasoulinejad P, Yin Y, Li S (2017) Direct estimation of spinal cobb angles by structured multi-output regression. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform)* 10265:529–540. https://doi.org/10.1007/978-3-319-59050-9_42
 6. Wu H, Bailey C, Rasoulinejad P, Li S (2017) Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform)* 10433:127–135. https://doi.org/10.1007/978-3-319-66182-7_15
 7. Wu H, Bailey C, Rasoulinejad P, Li S (2018) Automated comprehensive adolescent Idiopathic Scoliosis assessment using MVC-Net. *Med Image Anal* 48:1. <https://doi.org/10.1016/j.media.2018.05.005>
 8. Wang L, Xu Q, Leung S, Chung J, Chen B, Li S (2019) Accurate automated Cobb angles estimation using multi-view extrapolation net. *Med Image Anal* 58:101542. <https://doi.org/10.1016/j.media.2019.101542>
 9. Chen B, Xu Q, Wang L, Leung S, Chung J, Li S (2019) An automated and accurate spine curve analysis system. *IEEE Access* 7:124596. <https://doi.org/10.1109/ACCESS.2019.2938402>
 10. Okashi OA, Du H, Al-Assam H (2017) Automatic spine curvature estimation from X-ray images of a mouse model. *Comput Methods Progr Biomed* 140:175. <https://doi.org/10.1016/j.cmpb.2016.12.010>
 11. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform)* 9351:234. https://doi.org/10.1007/978-3-319-24574-4_28
 12. Jegou S, Drozdal M, Vazquez Romero A, Bengio Y (2017) the one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. *IEEE Comput Soc Conf Comput Vis Pattern Recognit Work* 2017:1175–1183. <https://doi.org/10.1109/CVPRW.2017.156>
 13. Zhang Z, Liu Q, Wang Y, Geosci IEEE (2018) Road extraction by deep residual U-Net. *Remote Sens Lett* 15(5):749. <https://doi.org/10.1109/LGRS.2018.2802944>
 14. Al Arif SMR, Knapp K, Slabaugh G (2018) Shape-aware deep convolutional neural network for vertebrae segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform)* 10734:12–24. https://doi.org/10.1007/978-3-319-74113-0_2
 15. Tu Y, Wang N, Tong F, Chen H (2019) Automatic measurement algorithm of scoliosis Cobb angle based on deep learning. *J Phys Conf Ser* 1187:42100. <https://doi.org/10.1088/1742-6596/1187/4/042100>
 16. Horng MH, Kuok CP, Fu MJ, Lin CJ, Sun YN (2019) Cobb angle measurement of spine from x-ray images using convolutional neural network. *Comput Math Methods Med*. <https://doi.org/10.1155/2019/6357171>
 17. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. *Proc IEEE Int Conf Comput Vis*. <https://doi.org/10.1109/ICCV.2017.322>
 18. Pan Y, Chen Q, Chen T, Wang H, Zhu X, Fang Z, Lu Y (2019) Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. *Eur Spine J* 28(12):3035. <https://doi.org/10.1007/s00586-019-06115-w>
 19. Cho BH, Kaji D, Cheung ZB, Ye IB, Tang R, Ahn A, Carrillo O, Schwartz JT, Valliani AA, Oermann EK, Arvind V, Ranti D, Sun L, Kim JS, Cho SK (2019) Automated measurement of lumbar lordosis on radiographs using machine learning and computer vision. *Glob Spine J*. <https://doi.org/10.1177/2192568219868190>
 20. Zhang K, Xu N, Yang G, Wu J, Fu X (2019) An automated Cobb angle estimation method using convolutional neural network with area limitation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinform)* 11769:775–783. https://doi.org/10.1007/978-3-030-32226-7_86
 21. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. <https://doi.org/10.1109/CVPR.2015.7298965>
 22. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd international conference on machine learning ICML*, vol 1, p 448
 23. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(1):1929
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.