Automatic Cobb Angle Detection Using Vertebra Detector and Vertebra Corners Regression

Bidur Khanal[1(✉)], Lavsen Dahal[1], Prashant Adhikari[2], and Bishesh Khanal[1]

[1] NepAl Applied Mathematics and Informatics Institute for Research (NAAMII), Suryabinayak, Nepal
{bidur.khanal,lavsen.dahal,bishesh.khanal}@naamii.org.np
[2] Hospital for Advanced Medical Surgery (HAMS), Kathmandu, Nepal

Abstract. Correct evaluation and treatment of Scoliosis require accurate estimation of spinal curvature. Current gold standard is to manually estimate Cobb Angles in spinal X-ray images which is time consuming and has high inter-rater variability. We propose an automatic method with a novel framework that first detects vertebrae as objects followed by a landmark detector that estimates the 4 landmark corners of each vertebra separately. Cobb Angles are calculated using the slope of each vertebra obtained from the predicted landmarks. For inference on test data, we perform pre and post processings that include cropping, outlier rejection and smoothing of the predicted landmarks. The results were assessed in AASCE MICCAI challenge 2019 which showed a promise with a SMAPE score of 25.69 on the challenge test set.

Keywords: Scoliosis · Landmark · Object detection · Cobb angle

1 Introduction

Scoliosis is a sideways curvature of the spine occurring mostly in teens. Severe scoliosis can also lead to disability. The current gold standard for diagnosing scoliosis is manual measurement of Cobb Angles in anterior-posterior (AP) or lateral (LAT) X-ray images which involve identifying the most tilted vertebrae above and below the apex of the spinal curve [1]. However, the procedure is time-consuming and observer dependent, leading to high inter-observer variability that could negatively impact assessing prognosis and treatment decisions [2]. Thus, there has been increasing interest in automatic estimation of Cobb angles directly from the X-ray images. In this context, we participated in MICCAI 2019 challenge on Accurate Automated Spinal Curvature Estimation (AASCE)[1] where the task was to accurately estimate three Cobb angles [3] from the training dataset containing 609 AP x-rays[2] whose results were assessed on 98 test images. The ground truth (GT) annotations are the anatomical landmarks consisting of four corners of 17 vertebrae: twelve thoracic and five lumbar.

---

[1] https://aasce19.grand-challenge.org/Home/.
[2] http://spineweb.digitalimaginggroup.ca/spineweb/index.php?n=Main.Datasets.

Related Work: The two most common approaches of estimating Cobb angles are Segmentation based and Landmark based approaches. The segmentation based methods first segment all the vertebrae or the end-plates of the vertebrae to identify the most tilted vertebrae from which the Cobb angles are estimated [3–5]. Accurate segmentation of each vertebra from X-ray images is difficult with traditional feature-engineering based approaches. To our knowledge, even modern supervised deep neural networks are not robust and accurate enough yet for the vertebra segmentation. Creating accurate GT segmentation is time consuming and relatively difficult compared to annotating landmarks: four corners of the vertebrae. In Landmark based approach which is the state-of-the-art, the four corners of each vertebrae are detected and are subsequently used for estimating Cobb angles. Some methods jointly estimate all the landmarks and Cobb angles, while others first estimate landmarks followed by Cobb angle computation which might include outlier rejection and post-processing techniques [6,7].

There are several approaches of detecting landmarks in medical images such as Reinforcement learning [8], iterative patch based approaches [9] and fully convolutional neural network based approaches [10]. One important difference in vertebra landmarks compared to other anatomical landmarks is the presence of a large number of similar looking vertebrae. We believe that detecting vertebrae as objects before finding landmarks within the detected vertebrae is advantageous as it allows: (i) avoiding difficulty for translation equivariant CNNs to learn very different coordinate locations for almost identical appearing vertebrae (ii) leveraging popular object detectors pre-trained for computer vision tasks (iii) Reducing the search space for landmark detector.

Contribution: We propose a novel approach to first detect 17 vertebrae with a bounding box object detector, after which each of the predicted boxes is fed to a landmark detector as illustrated in Fig. 1. The predicted landmarks are post-processed to remove outliers before calculating the three Cobb angles. [11] used Faster-RCNN [12] object detector to detect intervertebral disc in lateral X-rays, but they left the landmark detection as a future work.

2   Dataset

The dataset consists of 609 spinal AP x-ray images available at SpineWeb[3] as *Dataset 16*. Each image has 68 GT landmarks corresponding to 4 corners of the 17 vertebrae, and 3 Cobb Angles. Organizers provided test images without GT separately. We connected the four landmark corners of each vertebrae to create a box whose width and height were then increased symmetrically by 50 and 10 pixels respectively to create GT bounding boxes. All the bounding boxes were labelled as belonging to a single class. The GT bounding boxes were used to crop and extract individual vertebrae as a single separate image containing four landmark corners. The coordinates of the landmarks are normalized to the coordinate system that maps all the pixel coordinates of the cropped image to

---

[3] http://spineweb.digitalimaginggroup.ca/spineweb/index.php?n=Main.Datasets.

the interval [0, 1]. The normalized landmark coordinates are used as GT labels for the landmark regression network.

3   Vertebrae Detection Followed by Landmarks Regression

We use an object detector to detect the vertebrae as bounding box objects which are then fed to a landmark regression network as separate input images. The predicted normalized landmark coordinates from individual bounding boxes are combined and mapped back to the original images as shown in Fig. 1.
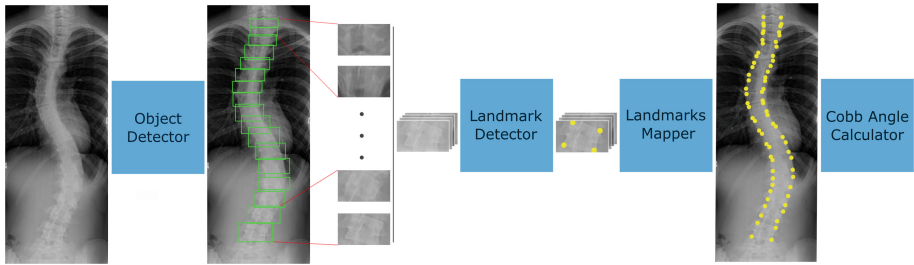
Fig. 1. Proposed Framework: The input images are first passed to an object detector that detects the vertebrae. The detected vertebrae are extracted as individual images and passed to a landmark detector that detects the four corners of the vertebra. The landmarks are mapped back to the original image from which Cobb angles are calculated. We use CNN-based Faster-RCNN and DenseNet for object detection and landmark detection respectively.

3.1   Training Vertebra Detection with Faster-RCNN

Faster-RCNN [12] is a widely used two-stage object detector consisting of: (i) a Region Proposal Network (RPN) that proposes potential object regions from a set of anchor boxes of various sizes in a sliding window over the feature maps extracted from a CNN-based base network (ii) a fully connected and a bounding box regression layer that regress bounding box locations of the identified objects. We used ResNet V1 101 with pre-trained weights on Imagenet data[4] as the base network, which was fine-tuned after block 2. We used two scales with box areas of 64^2 and 128^2 pixels, and aspect ratios 1:1 and 2:1 for RPN's anchor boxes, as the vertebrae are relatively small and do not have extreme aspect ratios. The network was trained for around 180k steps with batch size 1 using SGD optimizer with momentum 0.9, learning rate 0.0003 and early stopping. The implementation was adopted from Luminoth[5] in Tensorflow framework 10.1. Data augmentation

---

[4] https://github.com/tensorflow/models/tree/master/research/slim.
[5] https://github.com/tryolabs/luminoth.

included random Gaussian noise ($\mu = 0$, $\sigma = 0.005$), and vertical and horizontal flips with a probability of 0.5. All the images were rescaled preserving the aspect ratio such that its sizes remained within 600–1000 pixels as much as possible.

3.2   Training Landmark Detector with DenseNet

The four corner landmarks were estimated using a Densely Connected Convolutional Neural Network (DenseNet) which are known to require fewer parameters than traditional CNN [13]. In DenseNet, each layer's feature maps are used for all subsequent layers within a block, where each block constitutes a bottleneck layer (a 2d Convolution layer with 1×1 filter size), batch normalization, ReLU activation, and a regular 2D convolution layer (3×3 filter size). We used 5 blocks with a growth rate of 8 which is the number of output feature maps of each layer. The 2D Global Average Pooling is used after 5 blocks followed by a dense layer. The final layer consists of 8 output units with a linear activation function. All the input images to landmark detector were resized to 200×120 pixels.

4   Pre and Post Processing During Inference

Cropping: Almost all test images contained skull and pelvic regions but none of the training images had them. During training, the model did not see negative samples of skull and pelvic regions making it prone to falsely detect structures appearing similar to vertebra such as jaws. We randomly picked one test image with an aspect ratio a_0 and found empirically that cropping $c_{t_0} = 0.18$ and $c_{b_0} = 0.21$ times the image height from the top and bottom removed skull and pelvic regions satisfactorily. All the remaining test images with aspect ratio a were cropped by $c_t = c_{t_0} \cdot \frac{a}{a_0}$ and $c_b = c_{b_0} \cdot \frac{a}{a_0}$ fraction of the image height from the top and bottom respectively.

Outlier Rejection: We removed some of the outliers by using the fact that adjacent vertebrae cannot be far away from each other: if the x-center (horizontal) of any detected bounding box is more than half box width away from the x-centers of both of its two nearest neighboring (top and bottom) boxes, they are rejected as outliers. For the topmost and bottom boxes, the same test was done against only one nearest neighbor.

Curve Fitting and Cobb Angle Calculation from Predicted Landmarks: We used the code provided along with the challenge dataset [6] to calculate 3 Cobb angles - Main Thoracic (MT), Proximal Thoracic (PT) and Thoracolumbar/Lumbar (TL/L) from a given set of landmarks. It did not work well when the number of landmarks were not exactly 68 corresponding to the 17 bounding boxes. To ensure exactly 68 landmark points for angle calculation, we used the following after outlier rejection: when the detected vertebrae number is more than 17, reject extra bounding boxes starting from the bottom. Similarly, if the number is less than 17, duplicate the bottom landmarks as required. We also smoothed

the landmarks by fitting a polynomial curve where the degree 6 polynomial gave best fit out of 3 to 8 on visual inspection. The x-coordinate of each landmark is regressed by using the y-coordinate as the independent variable of the fitted polynomial. The smoothed landmarks were the ones that were used to estimate Cobb angles in the final test score.

5   Results

The results were evaluated with symmetric mean absolute percentage error, SMAPE$= \frac{1}{N} \sum_N \frac{\sum_m |a_g - a_p|}{\sum_m (a_g + a_p)} 100\%$, where we have $N(= 98)$ test images, $m(= 3)$ Cobb angles per image, GT angle a_g and the corresponding predicted angle a_p.

Table 1. Results for different experiment setups

Exp no.	Processing for test images	SMAPE
1	No Cropping	33.3%
2	Cropping and outlier removal without smoothing	26.79%
3	Cropping, outlier removal and smoothing with order 6 polynomial fitting	25.69%

Table 1 shows the results of three different experiments where we achieved our best score in the challenge by cropping, rejecting outliers and smoothing the estimated landmarks. The top score in the leader board was 21.71% when the challenge results entry was closed. Figure 2 shows detected bounding boxes and landmarks, and results of outlier rejection and smoothing with polynomial fitting in 4 example images from test set.
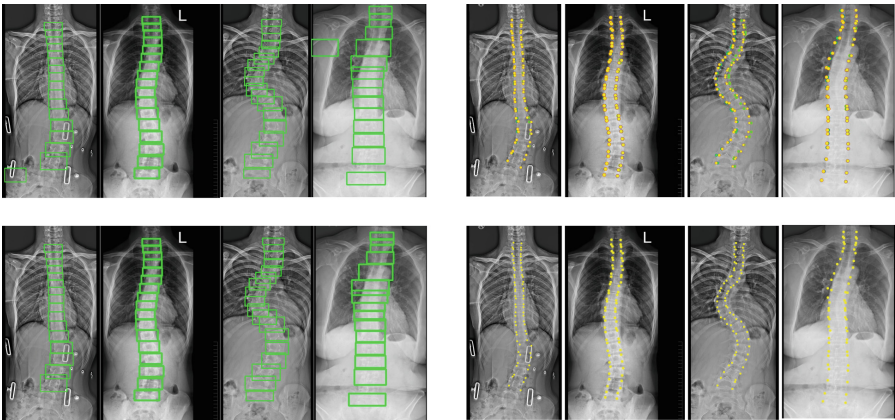
Fig. 2. Anti-clockwise from top left: bounding box detection, outlier rejection, landmark prediction and smoothing of landmarks (green) with polynomial degree 6 for four images of the test set (Color figure online)

6 Discussion and Conclusion

Detecting vertebrae as objects before predicting corner landmarks is found to be a promising approach. However, cropping all test images will not generalize well. A more robust object detector trained with images having negative samples from skull and pelvic regions could eliminate the need of cropping. The proposed approach does not properly take into account the inter-dependency between landmark positions of different vertebrae. A learning algorithm to learn this inter-dependency could improve the results. Finally, learning to estimate the angles directly from landmarks instead of using the geometric algorithm could be robust to noisy landmark prediction.

References

1. Greiner, K.A.: Adolescent idiopathic scoliosis: radiologic decision-making. Am. Fam. Physician **65**, 1817–22 (2002)
2. Loder, R.T., et al.: The assessment of intraobserver and interobserver error in the measurement of noncongenital scoliosis in children $\leq$ 10 years of age. Spine **29**(22), 2548–2553 (2004)
3. Sardjono, T.A., et al.: Automatic cobb angle determination from radiographic images. Spine (Phila Pa 1976) **38**, E1256–E1262 (2013)
4. Allen, S., et al.: Validity and reliability of active shape models for the estimation of cobb angle in patients with adolescent idiopathic scoliosis. J. Digit. Imaging **21**, 208–18 (2008)
5. Zhang, J., et al.: Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. J. Digit. Imaging **22**, 463–72 (2009)
6. Wu, H., Bailey, C., Rasoulinejad, P., Li, S.: Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 127–135. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_15
7. Sun, H., Zhen, X., Bailey, C., Rasoulinejad, P., Yin, Y., Li, S.: Direct estimation of spinal cobb angles by structured multi-output regression. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 529–540. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_42
8. Alansary, A., et al.: Evaluating reinforcement learning agents for anatomical landmark detection. MedIA **53**, 156–164 (2019)
9. Li, Y., et al.: Fast multiple landmark localisation using a patch-based iterative network. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 563–571. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_64
10. Payer, C., et al.: Integrating spatial configuration into heatmap regression based CNNS for landmark localization. MedIA **54**, 207–219 (2019)

11. Sa, R., et al.: Intervertebral disc detection in X-ray images using faster R-CNN. In: Conference Proceedings of IEEE Engineering in Medicine and Biology Society, pp. 564–567 (2017)
12. Ren, S., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE TPAMI **39**(6), 1137–1149 (2017)
13. Huang, G., et al.: Densely connected convolutional networks. In: CVPR (2017)