A Coarse-to-Fine Deep Heatmap Regression Method for Adolescent Idiopathic Scoliosis Assessment

Zhusi Zhong[1], Jie Li[1], Zhenxi Zhang[1], Zhicheng Jiao[2], and Xinbo Gao[1(✉)]

[1] School of Electronic Engineering, Xidian University, Xi'an 710071, China
xbgao@mail.xidian.edu.cn
[2] Perelman School of Medicine, University of Pennsylvania, Hamilton, PA 19104, USA

Abstract. Spinal anterior-posterior x-ray CT imaging is an appealing tool to aid diagnosis and elucidate Adolescent Idiopathic Scoliosis (AIS) Assessment. In this paper, we propose an automatic detection method for AIS assessment from X-ray CT images. Our deep learning based coarse-to-fine heatmaps regression method achieves symmetric mean absolute percentage error (SMAPE) of 24.7987 in the grand challenge AASCE 2019.

1 Introduction

Adolescent idiopathic scoliosis (AIS) is the most common form of scoliosis and typically affects teens. It is a medical condition in which a person's spine has a sideways curve. The curve of the spine presents "S"- or "C"-shape over three dimensions. The conditions of some patients are not stable, and the degree increases over time. Mild scoliosis does not typically cause problems, but severe cases can interfere with breathing. The assessment has significant impacts on statistics and prediction of the spinal condition, which can help to decide the early treatment plan and prevent the deterioration of the condition.

The AIS assessment is a standard tool to quantitatively analyze the condition. The evaluation is based on some landmarks around the vertebras in 2D vertical section X-ray image of human trunk. The Cobb angle is a standard measurement of bending disorders of the vertebral column. The angles are calculated on the annotated landmarks of the posteroanterior (back to front) X-ray images. The evaluation selects the most tilted vertebra at the top and bottom of the spine, and calculates the intersection angles of the selected vertebras. But in clinical, the landmarks are annotated manually, which remains a time-consuming work for an experienced doctor. Because of the anatomical differences across organizations, the manual annotation in the X-ray images is extremely subjective to observer variability. The accuracy of assessment usually has a great influence on the treatments. Therefore, an automatic annotation and calculation method would release doctors from the time-consuming work and especially avoid the observation errors. Li et al. [1–4] hold the Accurate Automated Spinal Curvature Estimation as a grand challenge in MICCAI 2019.

The AIS assessment is directly based on spinal landmarks. But in the evaluation, the Cobb angle is designed to selecting the most tilted vertebra, and the endplates are

Y. Cai et al. (Eds.): CSI 2019, LNCS 11963, pp. 101–106, 2020.
https://doi.org/10.1007/978-3-030-39752-4_12

generally parallel for each vertebra, so the Cobb angle is not sensitive to small local deviation in landmark coordinates. In this paper, we propose a novel deep learning framework for automatically AIS assessment. Our proposed method is improved from the landmark heatmap regression method. In our framework, the deep learning models regress heatmaps from coarse to fine in 2 stages, informing global configuration as well as accurately describing the local appearance, similar to what we did in [5].

2 Method

Overall Framework: As shown in Fig. 1, the overall framework for landmark detection includes 2 stages. The global stage takes the peeling cropped image as input, the U-net [6] A regresses the whole spine mask. The image multiplied with the mask and inputs to Mask r-cnn [7], whose outputs contain box masks, bounding boxes and tag labels. In the local stage, we extend the bounding box and crop the region image contained the box. The local region image inputs to U-net D and U-net E separately, the outputs are multiplied together. The 4 highlights in the 4 channels heatmaps are obtained as the landmarks of one vertebral column. The local stage procedure traverses the bounding boxes and obtains the spine landmarks. Based on the previous work, we proposed a different landmark obtaining method. To increase the robustness, we modify the landmarks with polynomial curve fitting, to adjust the unusual landmarks and decrease the outliers.

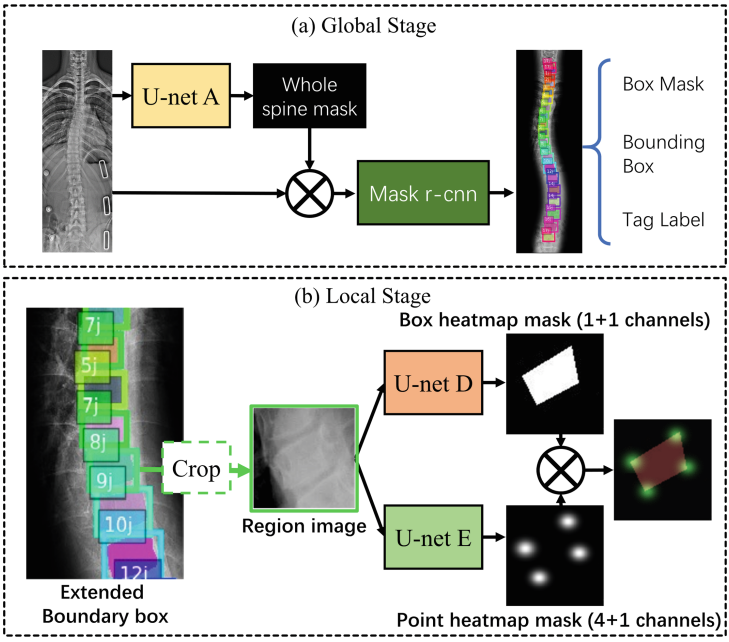

Fig. 1. The main parts of our proposed framework.

Global Stage Masks Regression: We train 3 U-net models to regress 3 kinds of heatmaps, as shown in Fig. 2. The 3 models take whole spine images as input. The U-net A is trained to regress Whole Spine Mask (WSM). The U-net B is trained to regress the Target Spine Mask (TSM). The outputs of these 2 models are highlight regions of the spine, but their edges are with different blur extension. The first channel is the target channel as shown in Fig. 2, and the last channel is the background which is all 1 matrix subducts the sum of the target channels. The WSM and TSM indicate the location of the spine, they can filter out the false positive regions while inferring. The U-net C is trained to regress the Box Masks. The box masks are the center region in the 4 landmarks of each vertebral column, and the first 17 channels contain 17 heatmaps separately for the vertebras in the given data. The last channel is the shared background, in order to handle the class-imbalance problem while training. The U-net C separates the connective vertebras into channels, the highlights in channels can easily represent their locations. The 3 models are trained separately and are embedded in the framework described in the following section.

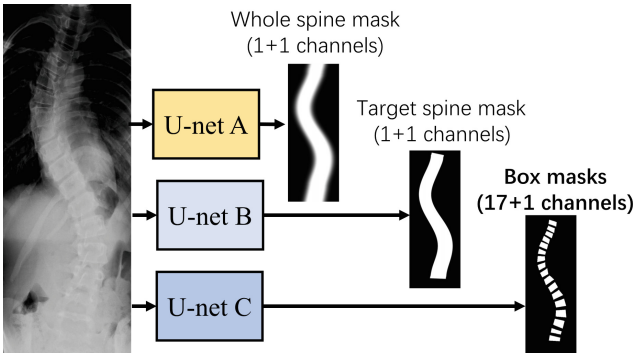

Fig. 2. 3 U-net models regress 3 different kinds of heatmaps.

Peeling Crop: The released test data is much more irregular and wilder than the training data. We propose a data regularization method for data preprocessing, in order to fit the trained model while inferring. The proposed data preprocess method is a progressive procedure, gradually narrowing the unwanted margins, like peeling onion piece by piece to get the wanted inner part. We use the 3 trained U-net models in the peeling crop procedure. The trained U-net A model takes the wild test data as input, to regress the WSM. Although there is a huge deviation between the train data and the input data, the model can still locate the coarse spine region but with marginal false detection. The U-net B and U-net C take the test data multiplied the first channel of the WSM, to regress TSM and box masks. To obtain the main target spine region, the first channel of TSM multiplies with the first 17 channels of box masks. Each channel obtains the highlight location. We apply the outlier detection and the continuity judgement to the 17 coordinates, to figure out the superior endplate and the inferior endplate. The detected region box is based on the 2 endplates' location and with the fixed height-width ratio, which is the red dotted box as shown in Fig. 3. We extend the detected box with the remained margin, in case of

the over cropping in the target spine. The cyan dotted box in the Fig. 3 is the extended region. The next iteration uses the cropped image in the extended region. The peeling crop iterates the cropping and the predicting, until the detected region box is narrowing slowly. We compare the narrowing distance of 2 iterations and set the maximum iteration number, to stop the iteration procedure. As shown in Fig. 4, when the region is getting closer to the training data, the box masks are getting more precise and separable in channels. The peeling crop processes the testing data once, before the inferring in the main framework.

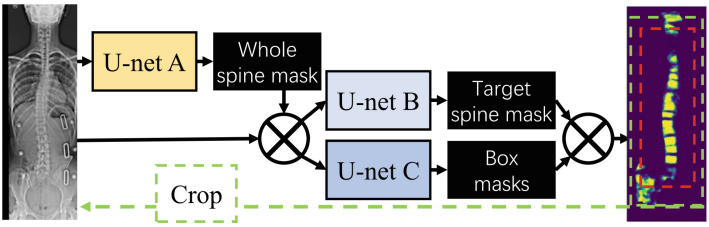

Fig. 3. The peeling crop is the data prepressing for the test data. (Color figure online)

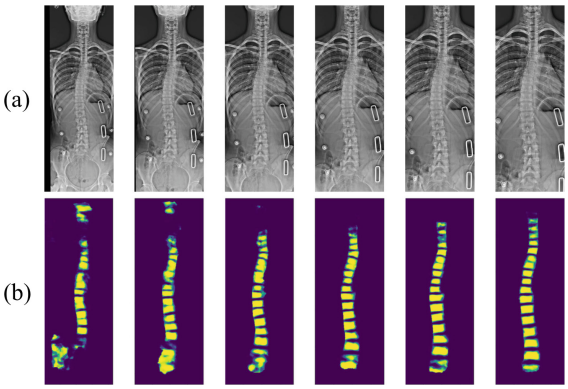

Fig. 4. The cropped image in the extended spine region as shown in row (a) and their box masks predictions in row (b). They are progressively adjusting in the peeling crop iteration.

Global Stage Detection: The vertebras are similar locally, so we model the vertebra detection problem as an instance segmentation problem. As shown in Fig. 1a, our global stage detection is based on the Mask r-cnn, which has shown well-performance in the instance segmentation subject. We transfer the coordinates to box masks and bounding boxes. The tag labels are the up-down order index of vertebras in the annotated spine. The U-net A takes the peeling cropped image as input, and then the output WSM multiplies with the image. The Mask r-cnn is trained to predict the box masks, bounding boxes and tag labels. The model extracts global information for the local patch detection.

Local Stage Heatmap Regression: We got the coarse bounding boxes, which indicate the boundaries of vertebras. In local stage, we want to obtain the landmarks for each vertebra. We use multi channels heatmaps regression method similarly to what we did in [5]. As shown in Fig. 1b, the local stage contains U-net D and U-net E. These 2 models are trained with the region patch images, to regress the box heatmaps and the point heatmaps. The box heatmaps are 2 channels, the first channel is the box mask of the vertebra located at the center of the extended bounding box, the second channel is the background. The point heatmaps are 5 channels, the 4 landmarks of one vertebra are modeled as 4 channel heatmaps. A 2D Gaussian distribution locates at the center of the landmark in the first 4 channels, and the last channel is the sheared background. While inferring, the points heatmaps are masked separately by the first channel of the box heatmaps. The overlapping region of the first 4 channels contains the locations of the landmarks. Each coordinate is obtained as the mean position of the pixels whose values are greater than 0.5. The 4 coordinates of 1 vertebra are obtained in the first 4 overlapping channels. The local stage procedure traverses the bounding boxes and obtains the whole spine landmarks.

Polynomial Curve Refine: We find that the Cobb angle is sensitive to the unusual landmarks, and these error predictions are usually protrusive and uneven in their vertebral neighbors. But the midline of the prediction shows the reliable curve trend of the spine. As the post-processing, we want to use the curve trend to fine-tune these error predictions. There are 2 sets of coordinates here used to fit the spine curve, which are the centers of the bounding boxes from Mask r-cnn and the mean positions of the box masks from U-net C. We use the least square method to fit the 9th degree polynomial coefficient, and the curve segments of the 2 endplates are 1st degree polynomial coefficient. We iterate the fitting procedure, while filtering out those landmarks far from the curve. The final coordinates are refined by the curve. The 17 center labels of bounding boxes which matching to the box masks from U-net C, are selected as the target vertebras. The horizonal connections of their coordinates intersect the curve. The vertical line in the intersections should go thought the 2 coordinates in an ideal situation. The unusual coordinates rotate on their intersections closing to the vertical lines.

3 Results

We report the results of two stages obtained by the proposed framework. The data is provided by the grand challenge AASCE 2019 [8], and the results are the online evaluations on the unlabeled test dataset. The evaluation is the score based on symmetric mean absolute percentage error (SMAPE) (Table 1).

Table 1. The stage-wise results of the proposed framework.

Stage	SMAPE
Local stage	26.4455
Curve refine	24.7987

4 Discussion

In conclusion, we propose a coarse-to-fine heatmap regression method for Adolescent Idiopathic Scoliosis Assessment. As data preprocessing, the Peeling Crop strategy is different from the direct region prediction methods, it adjusts the target regions progressively. The global and local parts of our framework are trained separately, it reduces the GPU memory cost and they are motivated in 2 different strategies. The 2 stages inform global configuration as well as accurately describing local appearance, then the local stage regresses the point heatmaps to obtain the landmark locations. We propose a Polynomial Refine method as our post-processing to refine the local stage coordinates. The results show that our method is effective in AIS Assessment.

References

1. 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
2. Wang, L., Qiuhao, X., Leung, S., Chung, J., Chen, B., Li, S.: Accurate automated Cobb angles estimation using multi-view extrapolation net. Med. Image Anal. **58**, 101542 (2019)
3. Chen, B., Xu, Q., Wang, L., Leung, S., Chung, J., Li, S.: An automated and accurate spine curve analysis system. IEEE Access **7**, 124596–124605 (2019)
4. Wu, H., Bailey, C., Rasoulinejad, P., Li, S.: Automated comprehensive Adolescent Idiopathic Scoliosis assessment using MVC-Net. Med. Image Anal. **1**(48), 1–11 (2018)
5. Zhong, Z., Li, J., Zhang, Z., Jiao, Z., Gao, X.: An attention-guided deep regression model for landmark detection in cephalograms. arXiv preprint arXiv:1906.07549 (2019)
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
7. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
8. https://aasce19.grand-challenge.org