ULaR: A Novel Framework for Automatic Cobb Angle Estimation from X-ray Images

Ananya Devarakonda, Dr. Anu Shaju Areeckal

January 2022

Abstract

Adolescent idiopathic scoliosis (AIS) refers to an abnormal deviation of the spine typically affecting children who are ten years old and above. AIS has no identifiable cause till date and remains the most common spinal deformity, affecting around 3% of adolescents world-wide. Standard diagnosis and treatment for AIS requires measuring Cobb angles from chest X-ray images. It is often the case that the standard manual or semi-automatic methods of calculating Cobb angles are time-consuming and often involve two or more experts. To help entirely automate the process of calculating Cobb angles using chest X-ray images, we propose a novel framework: U-net for Landmark detection followed by Regression (ULaR). The framework uses a U-Net model with a ResNet-34 backbone to extract vertebra landmarks, and three regression models to compute the respective Cobb angles. The model was trained on data from the Accurate Automated Spinal Curvature Estimation (AASCE), MICCAI 2019 grand challenge. Similar to the challenge, we evaluated the Cobb angles predicted by our method using Symmetric Mean Absolute Percentage Error (SMAPE). We obtained an SMAPE of 18.3047 on the test dataset.

1 Introduction

Adolescent Idiopathic Scoliosis (AIS) is a prevalent disorder of the spine where an abnormal deviation greater than ten degrees is observed in the spinal curvature of children who are ten years of age and older [1]. The term "idiopathic" indicates that there is no known cause for this deformity. AIS remains the most common type of scoliosis and affects around 3% of adolescents around the globe [2]. The standard clinical procedure to accurately diagnose and treat patients with AIS involves calculating Cobb angles from spinal anterior-posterior (AP) X-ray images [3].

Even today, measuring Cobb angles is a process that is manual or at best, semi-automatic and often requires the involvement of multiple experts, thus making the calculation of Cobb angles time-consuming and prone to potential inter-observer and intra-observer errors. Further, early and quick diagnosis is essential for adolescents with AIS to develop as much as possible, non-invasive plans for treatment and increase quality of life.

Recognizing the challenges caused by the manual calculation of Cobb angles, recently, several papers propose various solutions spanning from traditional image processing methods to modern Deep Learning (DL) solutions to automate this process.

This paper outlines a method to fully automate the procedure of calculating Cobb angles from chest anterior-posterior X-ray images using data from the Accurate Automated Spinal Curvature Estimation (AASCE), MICCAI 2019 grand challenge. The AASCE 2019 challenge dataset contains separate training and test datasets. The training dataset contains 481 chest X-ray images and the test dataset contains 128 chest X-ray images. The dataset also provides 68 vertebra landmark points and three Cobb angles for each image as labels.

We introduce a U-Net with a ResNet-34 backbone and three regression models to accurately determine vertebral landmarks and calculate Cobb angles from chest AP X-ray images respectively. Our contributions are as follows:

1. A description of a novel pipeline (U-net for Landmark detection followed by Regression- ULaR) to accurately segment vertebral end points and predict Cobb angles.
2. A novel data preprocessing strategy to improve model performance.
3. The model was trained entirely on an average GPU indicating that it is relatively light-weight.
4. A comparative study of other backbones for the U-Net model.

2 Related Work

Cobb angle calculation for the diagnosis and treatment of AIS has remained the gold standard in clinical practice since its introduction. Over the years, researchers have put forward several methods to automate all or a significant portion of the calculations involved in determining Cobb angles in chest AP X-rays. In this section, we review existing literature that outline methods to entirely automate the process of Cobb angle calculation using X-ray images and discuss relevant papers from the AASCE 2019 grand challenge.

Broadly, methods of fully automating Cobb angle calculation may be categorized into traditional image processing algorithms and Deep Learning (DL) architectures. Zhang et al. [4] used a Canny Edge detector and fuzzy Hough transform to extract line structures to calculate Cobb angles form a total of seventy six radiographs. They obtained promising results with a mean absolute error of less than five degrees. The solution also required a preprocessing strategy that required user involvement to manually adjust and enhance the quality of the radiographs, thus making it not entirely automatic.

Sardojo et al. [5] used a charged particle model (CPM) to obtain spinal curvature on thirty six frontal X-ray images of patients with scoliosis and compared various piece-wise linear, splits, and polynomial curve-fitting methods. They concluded that the best methods involved the piece-wise linear method using three segments and the six degree polynomial method.

Other image processing techniques are often semi-automatic and involve user interaction with the program such as clicking the center of each vertebra, determining the most tilted vertebrae in the X-ray image, and more [6–9]. The

main disadvantage of mathematical, filter-based traditional image processing methods is that accuracy often comes from feature engineering and thresholding-thus losing some of the generalisation that is required when using X-ray images with variations.

Machine learning and deep learning models inherently overcome the disadvantage of traditional methods by inherently training for a generalised solution. They are thus popular today and are used for a variety of applications. Kundu et al. [10] introduced a fully automatic method of calculating Cobb angles from twenty one X-ray images using bilateral image denoising, region of interest (ROI) identification using Support Vector Classifier (SVC), object centerline extraction using Gaussian blurring and polynomial curve fitting, and Cobb angle measurement. They found that they obtained accurate results for 57.14% of subjects.

The Accurate Automated Spinal Curvature Estimation (AASCE) 2019 Challenge

A significant drawback of using Deep Learning techniques is that such models require large amounts of data. As medical data is especially difficult to not only obtain but also annotate, researchers may obtain results that can be significantly improved upon by using the same oftentimes powerful method but with more data. In 2019, Medical Image Computing and Computer Assisted Intervention (MICCAI) organised the Accurate Automated Spinal Curvature Estimation (AASCE) Challenge and provided a dataset with both vertebral landmark annotations and Cobb angle calculations to encourage research in automating Cobb angle calculations [11].

As part of the AASCE 2019 challenge, Chen et al. [12] introduced two methods to calculate Cobb angles. The first involved a RetinaNet for bounding box prediction followed by a HR-Net to determine vertebral landmark points. The second involved a Simple Baseline followed by patch generation. They fused the two methods and obtained a final SMAPE of 22.16582%.

Tao et al. [13] proposed a feature pyramid network (FPN) with a ResNet-50 backbone to extract multi-scale regional proposals with an SMAPE of 25.4784. Zhong et al. [14] used a coarse-to-fine heatmap regression model to extract vertebral landmarks. The model consisted of a global stage that worked on the entire image and a local stage that worked on a small cropped image of a vertebra. The method used Mask R-CNN and various U-Net models. After using least squares to fit the ninth degree polynomial co-efficient to refine the predicted vertebral landmarks, the SMAPE was calculated to be 24.7987.

Kahanal et al. [15] introduced a vertebra detector using Faster-RCNN and a landmark regressor using DenseNet to extract vertebral landmarks. Their final SMAPE was 25.69 after six degree polynomial curve fitting to find the angles. Dubost et al. [16] used a cascaded network of two convolutional neural networks (CNNs) to extract spine centerline. They calculated Cobb angles using the derivative of the centerline and obtained an SMAPE of 22.96.

Lin et al. (2020) [17] proposed the Seg4Reg framework that utilized an ensemble model architecture using a PSPNet to segment vertebra landmark points and a regression network to predict Cobb angles. They obtained a final SMAPE of 21.71 after post-processing, securing the first place on the AASCE 2019 challenge leaderboard. They approximated any angles in the AASCE 2019

challenge dataset that are lesser than four degrees to zero. More recently, Lin et al. (2021) [18] introduced Seg4Reg+ that combined the optimization process for the segmentation and regression units of the model and obtained an SMAPE of 8.7.

Other recent advances in automatic Cobb angle calculation include Zhang et al. [19] proposed a method using convolutional neural networks and non-directional part affinity fields (NDPAFs) for vertebral landmark computation. Cobb angles are calculated using a constructed Spinal Landmarks Segmentation Network (SLSN) with Gaussian heatmaps.

3 Dataset

The AASCE 2019 challenge dataset contains a total of 609 chest X-ray images that is split into 481 images in the training dataset and 128 images in the test dataset. The image dimensions are not fixed and vary from around 973 px to 3755 px in height and 355 px to 1427 px in width. The dataset also contains labels with vertebral landmark points that contain normalized x and y coordinates in the range of [0, 1] for the four corners of all seventeen vertebrae. Along with landmarks, the dataset provides three manual Cobb angle calculations- Main Thoracic (MT), Proximal Thoracic (PT), and Thoracolumbar/Lumbar (TL/L).

4 Preprocessing

Images

As mentioned in the earlier section about the dataset, all images have inconsistent dimensions. To make these images similar in size, direct reshaping is not a viable option as we must predict vertebra landmarks. If the image shape changes, the corresponding landmarks change as well. Thus, merely reshaping each image to a particular dimension will lead to inaccurate results as it will distort the images and hence the landmarks.

To overcome this problem of image reshaping, we first scale the image to a value that is lesser than or equal to 448x224 px. We chose 448x224 px as dimensions as Zhao et al. [20] used a bit of a similar model- the RU-Net and obtained promising results using 448x224px as dimensions. If the image height is lesser than 448 px and image width is lesser than 224 px, we zero-pad the top and the right side of each image to make the overall image dimensions 448x224 px exactly.

Some images in the dataset have high contrast while others are low contrast. To ensure that all X-ray images have a good amount of contrast, we use Contrast Limited Adaptive Histogram Equalization (CLAHE) [21]. Original Adaptive Histogram Equalization (AHE) often over amplifies regions with similar values which is unfavorable to X-ray images which have several regions with similar values. CLAHE overcomes this disadvantage of AHE by considering small parts of the image at a time- "tiles." To CLAHE enhance all images, we used a clip limit of 2.0 and a tile grid size of (12, 12). We then normalized image values to fall between [0, 1].

Masks

To make the image masks, we plot the values of each vertebral landmark given in the dataset onto a black rectangle of the same dimensions as the original image. As we use a model that is usually used in segmentation tasks (U-Net), we plot each point to be a white dot with a radius of 3 px.

The benefit of creating a mask with larger circles on the black rectangle to illustrate vertebral landmarks is that it provides some leeway as compared to a pixel perfect output. This process can also assist with avoiding overfitting. If pixel-perfect values are used for masks (circles with radius 1 px) it is more likely to cause the model to overfit and provide inaccurate results as there is a significantly lower number of white pixel values.

Furthermore, as our main aim is to compute Cobb angles from this data, even if there is some margin of error with respect to the pixel value of the vertebral keypoint, the overall shape and structure of the vertebral column is retained. Recognizing that increasing the radius of circles indicating landmarks on the mask has its advantages, we plotted each landmark value in the dataset with a circle of radius 3 px. The mask images were also converted to images of dimensions 448x224 px using the same process as before. The masks were then converted to binary images where values greater than 250 were converted to be 1 and the rest were considered to be 0.

Figure 1. provides a complete picture of all preprocessing steps taken before training the model.

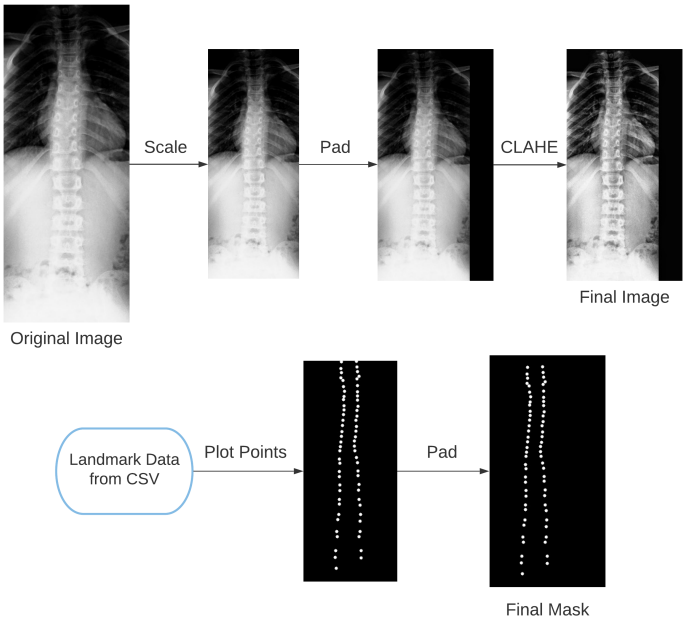


Figure 1: An illustration of pre-processing steps

5 Proposed Method: ULaR

The overall pipeline contains a U-Net model with a ResNet-34 backbone to predict vertebra corner points and three regression models to compute each Cobb angle.

5.1 Landmark Detection

To predict vertebral landmark points we use a U-Net model with a ResNet-34 backbone. The model is pre-trained on the ImageNet dataset. The model uses Adam Optimization and a combination of Binary Cross-Entropy (BCE) and the Jaccard loss function. Also known as Intersection over Union (IoU) loss, the Jaccard loss can be represented as:

$$L(Y, Y') = 1 - \frac{Y \cap Y'}{Y \cup Y'} \tag{1}$$

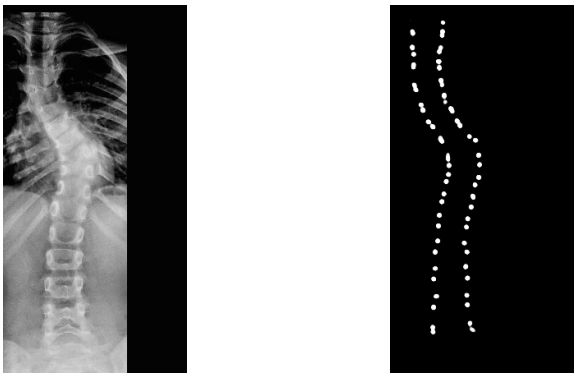
Where Y is the ground truth value and Y' is the predicted value.

We also compute the following metrics: Intersection over Union (IoU) Score, Mean Squared Error (MSE), and Mean Absolute Error (MAE).

The training dataset from the AASCE 2019 challenge with 481 images was split into 80% for model training and the rest of the 20% was reserved for validation. The model was trained using an NVIDIA GeForce GTX 1650 Ti GPU for 20 epochs with a batch size of four.

After landmark detection using the U-Net model, we must precisely choose 68 keypoints from the predicted mask to represent the four corners of each vertebra in the spine. We use K-means to obtain 68 cluster centroids which we consider as the final prediction for landmark detection.

These cluster centroids are then used to compute the Cobb angles. Figures 2 and 3. illustrate this process.



(a) (b)

Figure 2: (a) Image used for training (b) Predicted image mask from U-Net model

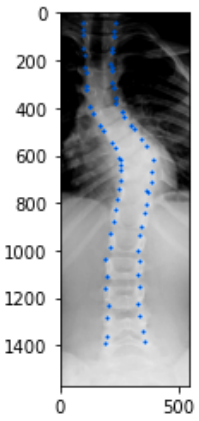


Figure 3: Final predicted 68 landmark points plotted on the original image after K-Means

5.2 Regression Models

After obtaining 68 predicted vertebral landmarks, we use three regression models to predict the three Cobb angles- Main Thoracic (MT), Proximal Thoracic (PT), and Thoracolumbar/Lumbar (TL/L). The first two angles use Gradient Boosting Regressors (GBRs). Briefly, GBRs use an ensemble of decision tree regressor models to predict a value. These trees are typically weak learners. GBRs are built in a forward stage-wise fashion and unlike other methods, GBRs utilize arbitrary differentiable loss functions.

After experimentation, we found that for angle one, a learning rate of 0.3, maximum depth of 2, and subsampling of 0.8 allowing for Stochastic Gradient Boosting provided the most accurate results. For angle two, learning rate 0.05, maximum depth six, and subsampling 0.3 provided the most accurate results.

To calculate the final Cobb angle, we used an Extra-Trees Regressor. The Extra-Trees Regressor is an ensemble tree-based classifier that is commonly referred to as extremely randomized trees [22].

It is well known that random forests provide significant advantages as compared to decision trees as they are less prone to overfitting. Random forest models achieve this by adding a component of randomness called "bootstrapping." Unlike random forest classifiers, the extremely random tree model uses the entirety of the original sample and include an additional level of randomness by initially randomly selecting tree split points. Optimization comes from the fact that among all generated trees, the best between all subsets is selected to create an ensemble model. Majority vote among all generated trees indicates the predicted result.

As the extra-trees classifier does not depend on the additional steps of bootstrapping and computing the optimal split, it is often considered computationally superior to the random forest model. Further, the component of randomness helps decrease model bias, which is pertinent to the defined problem statement.

Each type of classifier was chosen after calculating the accuracy of various machine-learning regression techniques. The chosen models outperformed other

methods by a significant margin. Figure 4. provides an overview of the methods used to predict the three Cobb angles. In the figure, models 1 and 2 refer to GBRs and model 3 refers to an Extra-Tree Regressor.

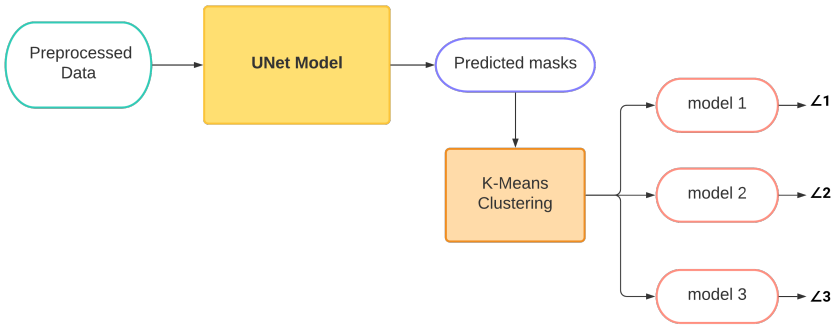


Figure 4: An illustration of the entire framework

Dataset	IoU	MSE	MAE
training	0.633	0.007	0.008
validation	0.358	0.015	0.017
test	0.326	0.017	0.018

Table 1: Landmark detection accuracy metrics

Author	SMAPE
Chen et al. [12]	22.165
Lin et al. [17]	21.71
Zhao et al. [20]	26.0535
Khanal et al. [15]	25.69
Dubost et al. [16]	22.96
Tao et al. [13]	25.4784
Zhong et al. [14]	24.7987
Wang et al. [23]	22.1775
Proposed Model	**18.3**

Table 2: A comparison of models from the AASCE 2019 grand challenge and our model

6 Results

The AASCE 2019 grand challenge used Symmetric Mean Absolute Percentage Error (SMAPE) as the metric to evaluate and rank all submitted solutions. As we used the AASCE 2019 grand challenge dataset, we report our final model performance to predict Cobb angles using SMAPE and compare our methods with those in the grand challenge.

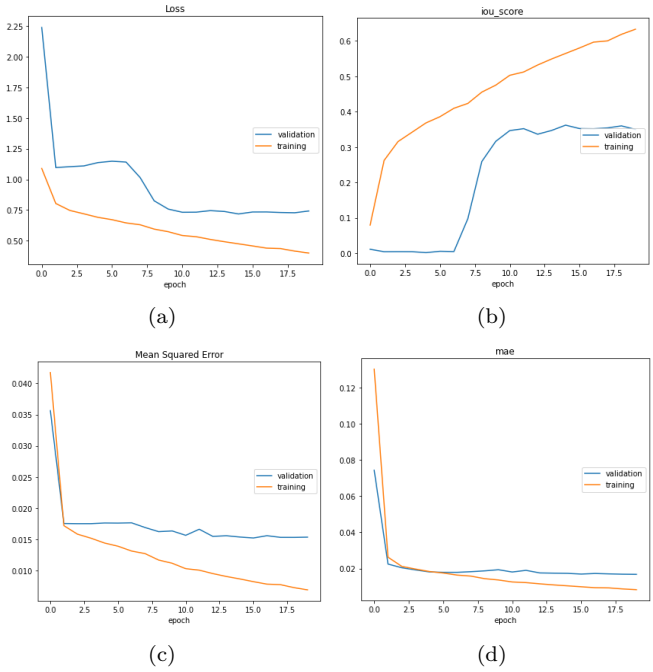


Figure 5: (a) Image used for training (b) Predicted image mask from U-Net model

As mentioned in the previous sections, we evaluate the landmark predictions from the U-Net model with a ResNet-34 backbone using IoU, MSE, and MAE. Table 1. provides a summary of the results obtained from landmark detection and Figure 5. contains the loss curves for all metrics during training. Table 2. compares various models proposed in the AASCE 2019 grand challenge with our model.

7 Discussion

Most methods involving Deep Learning (DL) techniques mentioned in the paper used powerful GPUs to train and develop the model architecture. Further, it is important to note that all methods mentioned in table 2. apart from our proposed method tested 98 images while we used 128 images for testing. Considering that we used a relatively average GPU we may say that our method is not only relatively light-weight but also promising. Considering Table 3. that compares different model backbones, we chose the ResNet-34 backbone although the ResNet-50 backbone provides slightly better accuracy for a few metrics. This preference is due to the fact that our method (U-Net with a ResNet-34 backbone) contains around 24 million parameters while the ResNet-50 contains about 32 million parameters. As the validation results from both models vary only by 0.01 at most, we select the smaller of the two models.

Observing the IoU metric in Table 3, we can quickly conclude that the VGG-16 backbone in this case, provides insufficient results. However, interestingly,

its MSE and MAE values are similar compared to the two ResNet backbones. This provides insight into the importance of choosing metrics to evaluate models. Landmark detection more often than not comes with data imbalance where more pixels are categorized as background information and relatively fewer pixels are pixels of interest. Thus, even if a model predicts that there exist no landmark points for every image, the result would be a low MSE and MAE. Keeping this in mind, we primarily evaluate our models based on IoU.

Model Backbone	**Dataset**	**IoU**	**MSE**	**MAE**
ResNet-34	training	0.633	0.007	0.008
	validation	0.358	0.015	0.017
	test	0.326	0.017	0.018
ResNet-50	training	0.779	0.004	0.004
	validation	0.341	0.016	0.017
	test	0.334	0.017	0.017
VGG-16	training	0.185	0.019	0.028
	validation	0.001	0.018	0.018
	test	0.0005	0.018	0.018

Table 3: Comparing various model backbones

8 Conclusion

This paper introduces a novel entirely automatic technique to calculate Cobb angles from chest X-ray images- U-net for Landmark detection followed by Regression (ULaR). The proposed framework is based on a U-Net model with a ResNet-34 backbone followed by centroid extraction using K-Means and Cobb angle calculation using regression models. We determined that the method outperforms those outlined in the AASCE-2019 grand challenge and is relatively light-weight.

There is scope for future research in training the model using data from additional sources as we only used the AASCE 2019 grand challenge dataset provided by MICCAI. Analysis of other optimization methods to prevent overfitting and increasing training time is also a promising venue of research to substantially increase IoU and overall accuracy. As automating the process of calculating Cobb angles can assist with quick diagnosis, there is need for current and future research into other model architectures using modern Computer Vision techniques.

References

[1] M. N. Choudhry, Z. Ahmad, and R. Verma, "Adolescent idiopathic scoliosis," *The open orthopaedics journal*, vol. 10, p. 143, 2016.

[2] R. P. Menger and A. H. Sin, "Adolescent and idiopathic scoliosis," *StatPearls [Internet]*, 2020.

[3] J. Cobb, "Outline for the study of scoliosis," *Instr Course Lect AAOS*, vol. 5, pp. 261–275, 1948.

[4] J. Zhang, E. Lou, L. H. Le, D. L. Hill, J. V. Raso, and Y. Wang, "Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior," *Journal of Digital Imaging*, vol. 22, no. 5, p. 463, 2009.

[5] T. A. Sardjono, M. H. Wilkinson, A. G. Veldhuizen, P. M. van Ooijen, K. E. Purnama, and G. J. Verkerke, "Automatic cobb angle determination from radiographic images," *Spine*, vol. 38, no. 20, pp. E1256–E1262, 2013.

[6] B. Samuvel, V. Thomas, M. M.G., and R. K. J., "A mask based segmentation algorithm for automatic measurement of cobb angle from scoliosis x-ray image," in *2012 International Conference on Advances in Computing and Communications*, 2012, pp. 110–113.

[7] M. Benjelloun and S. Mahmoudi, "Spine localization in x-ray images using interest point detection," *Journal of digital imaging*, vol. 22, no. 3, pp. 309–318, 2009.

[8] R. Kundu, A. Chakrabarti, and P. K. Lenka, "Cobb angle measurement of scoliosis with reduced variability," *arXiv e-prints*, pp. arXiv–1211, 2012.

[9] A. Safari, H. Parsaei, A. Zamani, and B. Pourabbas, "A semi-automatic algorithm for estimating cobb angle," *Journal of biomedical physics & engineering*, vol. 9, no. 3, p. 317, 2019.

[10] R. Kundu, A. Chakrabarti, and P. Lenka, "Automated cobb angle computation from scoliosis radiograph," in *Annual Convention of the Computer Society of India*. Springer, 2018, pp. 140–155.

[11] Y. Cai, L. Wang, M. Audette, G. Zheng, and S. Li, *Computational Methods and Clinical Applications for Spine Imaging*. Springer, 2020.

[12] K. Chen, C. Peng, Y. Li, D. Cheng, and S. Wei, "Accurate automated keypoint detections for spinal curvature estimation," vol. 11963 LNCS, 2020.

[13] R. Tao, S. Xu, H. Wu, C. Zhang, and C. Lv, "Automated spinal curvature assessment from x-ray images using landmarks estimation network via rotation proposals," vol. 11963 LNCS, 2020.

[14] Z. Zhong, J. Li, Z. Zhang, Z. Jiao, and X. Gao, "A coarse-to-fine deep heatmap regression method for adolescent idiopathic scoliosis assessment," vol. 11963 LNCS, 2020.

[15] B. Khanal, L. Dahal, P. Adhikari, and B. Khanal, "Automatic cobb angle detection using vertebra detector and vertebra corners regression," vol. 11963 LNCS, 2020.

[16] F. Dubost, B. Collery, A. Renaudier, A. Roc, N. Posocco, W. Niessen, and M. de Bruijne, "Automated estimation of the spinal curvature via spine centerline extraction with ensembles of cascaded neural networks," vol. 11963 LNCS, 2020.

[17] Y. Lin, H. Y. Zhou, K. Ma, X. Yang, and Y. Zheng, "Seg4reg networks for automated spinal curvature estimation," vol. 11963 LNCS, 2020.

[18] Y. Lin, L. Liu, K. Ma, and Y. Zheng, "Seg4reg+: Consistency learning between spine segmentation and cobb angle regression," vol. 12905 LNCS, 2021.

[19] C. Zhang, J. Wang, J. He, P. Gao, and G. Xie, "Automated vertebral landmarks and spinal curvature estimation using non-directional part affinity fields," *Neurocomputing*, vol. 438, pp. 280–289, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221000989

[20] S. Zhao, B. Wang, K. Yang, and Y. Li, "Automatic spine curvature estimation by a top-down approach," vol. 11963 LNCS, 2020.

[21] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization.* USA: Academic Press Professional, Inc., 1994, p. 474–485.

[22] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[23] J. Wang, L. Wang, and C. Liu, "A multi-task learning method for direct estimation of spinal curvature," vol. 11963 LNCS, 2020.