Automating Cobb Angle Measurement for Adolescent Idiopathic Scoliosis using Instance Segmentation

Chaojun Chen1, Khashayar Namdar2,3,6, Yujie Wu1, Shahob Hosseinpour1, Manohar Shroff3,4, Andrea Doria2,3,4 and Farzad Khalvati1,2,3,4,5,6

Abstract—Scoliosis is a three-dimensional deformity of the spine, often diagnosed in childhood. It affects 2-3% of the population, representing seven million people in North America. Currently, the gold standard for assessing scoliosis is done manually by measuring Cobb angles. This manual process is time-consuming and unreliable as it is affected by inter- and intra-observer variance. To eliminate these inaccuracies, machine learning (ML) methods can be used to automate the Cobb angle measurement process. This paper proposes to address the Cobb angle measurement task using an instance segmentation model. The proposed method first segments the vertebrae in an X-ray image using YOLACT model, then tracks the decisive landmarks using minimum bounding boxes. Lastly, the extracted landmarks are used to calculate the corresponding Cobb angles. The proposed method achieved a Symmetric Mean Absolute Percentage Error (SMAPE) score of 10.76%, outperforming the results presented in previous research. Additionally, more than 94% of the estimated Cobb angles had an error of less than ten degrees. The proposed method demonstrates reliability in both vertebra localization and Cobb angle measurement.

Index Terms—Scoliosis; Cobb Angle; Medical imaging; Instance Segmentation; YOLACT; Landmark

I. INTRODUCTION

Scoliosis is an abnormal lateral curvature of the spine, often diagnosed in childhood or early adolescence, causing chronic back pain, as well as uneven shoulder and waist [1]. In severe cases, scoliosis can become disabling and requires surgical intervention as it reduces the amount of space within the chest, making it difficult for the lungs to function properly [2]. Therefore, it is important to identify scoliosis at an early stage and provide appropriate treatment by the time the condition can be reverted or improved with braces rather than corrective surgery.

The diagnosis and severity of scoliosis are quantified by assessing the Cobb angles, which are commonly measured using anterior-posterior (AP) radiography (X-ray). Due to the ambiguity and variability in the scoliosis AP X-ray images, measuring Cobb angles is a challenging task [3]. Generally, radiologists assess Cobb angles by manually selecting the most tilted vertebrae above and below the apex. Lines are then drawn along the endplates of each vertebra, and the angle between the lines is the Cobb angle [4]. This common process can be affected by two main sources of error, in addition to other anatomical conditions, such as segmentation anomalies. The first source of error is inter-observer variance, which refers to the difference in end vertebra selection by different radiologists. The other error is intra-observer variance, which captures the fact that a radiologist might identify different end vertebra on the same X-ray image at different times [5]. These differences can lead to inaccuracy and inconsistency in Cobb angle measurement [5].

To correct for these inaccuracies, previous research proposed to automate Cobb angle measurement with machine learning (ML) approaches [6]. Key components in solving this problem include (1) precise localization of the vertebrae, (2) accurate extraction of landmarks, and (3) correct calculation of Cobb angles [7]. Localization of the spine and vertebrae can be done using a regression-based method, a segmentation-based method, or an object detection-based method. The regression-based methods estimate the landmark locations and Cobb angles directly. One technique is to jointly evaluate Cobb angles using a YOLACT model, then tracks the decisive landmarks using minimum bounding boxes. Lastly, the extracted landmarks are used to calculate the corresponding Cobb angles [8]. The proposed method demonstrated reliability in both vertebra localization and Cobb angle measurement.
input images have low image quality, segmentation models tend to generate connected or corrupted vertebra masks [10]. To solve the corrupted segmentation mask problem, object detection-based methods can be used. Yi et al. proposed to detect the center of each vertebra, then trace the four corners through the learned corner offset using ResNet [10]. By first detecting the center, the algorithm can keep the landmarks in order and avoid the overlapping issue that arises in the segmentation method. This approach considers the vertebra’s inter-dependency and achieves an average Symmetric Mean Absolute Percentage Error (SMAPE) of 10.81% on a dataset of 609 images [10].

Existing pipelines use conventional segmentation models, creating a potential for more modern segmentation models for vertebra localization to improve the current performance levels. Thus, the objective of this research is to create a consistent, accurate, and autonomous solution for Cobb angle measurement through state-of-the-art ML, helping clinicians develop optimal treatment for patients by minimizing errors. We selected an instance segmentation model YOLACT [11] for this task. The unique structure of YOLACT allows it to extract features from the input X-ray, resulted in accurate and consistent Cobb angle calculation.

II. MATERIALS AND METHODS

A. Dataset

The dataset of this research is obtained from SpineWeb [12], and includes 609 AP X-ray images in total. The ground truth annotations are the anatomical landmarks consisting of four corners of 17 vertebrae: 12 thoracic and 5 lumbar. These landmarks were provided by two professional radiologists at London Health Science Center, London, ON, Canada [13]. The dataset also contains the angle measurements of the proximal thoracic section, main thoracic section, and thoracolumbar section of the spine.

It should be highlighted that the dataset contains flawed images, namely landmarks with incorrect ordering. Figure 1 shows two examples of incorrect landmark ordering. We excluded 11 imperfect examples from the test set to avoid inaccuracies.

![Example images showing imperfect annotated data](image)

Fig. 1. Examples of imperfect annotated data showing a) lateral vertebral wedging and b) the shape of a “butterfly vertebra”

B. Data Augmentation

In order to expose the model to more diverse cases, we augmented the dataset using conventional methods. The data augmentation included a) randomly tilting 10% of the images with an angle between -5 and 5 degrees; b) horizontally flipping 10% of the images; c) vertically flipping 10% of the images; and d) histogram equalization of 10% of the images.

To input the images into YOLACT, COCO annotation format is required [11]. A custom COCO annotation contains detailed image information, such as the width and height of the image and the ground-truth location of the target vertebra for segmentation [14]. Figure 6.a illustrates an example input image and its COCO annotation. The quadrilaterals are created using the provided landmarks of the image.

To train the model, the processed images were split into training, validation, and testing sets with the 70%, 15%, and 15% percentages, respectively.

C. Pipeline Design

Aiming to select an accurate, consistent and optimal approach that does not require an excessive amount of computation resources, we chose a segmentation-based method. Based on the literature, segmentation-based methods generally have a lower SMAPE score [7]. Additionally, they are able to intake higher-resolution input images and localize points without dense mapping. Therefore, this paper approaches the automation task using YOLACT segmentation method [11]. YOLACT is an instance segmentation model. Unlike other popular semantic segmentation such as UNet, instance segmentation has a pathway that incorporates object detection functionality. Thus, instance segmentation models are less likely to produce connected or corrupted masks compared to UNet.

D. Segmentation using YOLACT

YOLACT is an instance segmentation with a ResNet backbone that produces a feature pyramid based on the input image [11]. The information inside the feature pyramid is then passed into two parallel pathways. One pathway is the Protonet which has four convolution layers and performs the segmentation task. The second pathway finds the mask coefficients of the image based on RetinaNet, producing image masks and bounding boxes. Finally, the algorithm linearly combines the two outputs and uses sigmoid nonlinearity to produce the final mask.

1) Protonet: Protonet is implemented as a Fully Convolutional Network (FCN) and is attached to the deepest feature layer to produce robust and high-quality masks. The arrows in Figure 2 represent $3 \times 3$ convolution layers, except for the last layer that is a $1 \times 1$ convolution layer. The increase in size is the result of an up-sample followed by convolution layer. The last layer has $k$ channels, one for each prototype [11]. The prototypes are candidate masks and capture spatial information. In this case, the number of prototypes is set to 32 (i.e. $k=32$) to achieve an optimal combination of speed and performance. Some examples of the prototypes are shown in Figure 2. This formulation exhibits a nuanced variation from the standard semantic segmentation, as it lacks an explicit loss on the prototypes. All supervision of the prototypes happens after mask assembly.
Fig. 2. **YOLACT Architecture** [11] is based on RetinaNet, ResNet-101, and Feature Pyramid Network (FPN) [11]. Output of the feature pyramid passes through two parallel pathways: Prediction Head and Protonet. The **Protonet Architecture** is attached to the deepest and largest feature layers of the FPN. The final output of Protonet is a set of \( k \) prototype masks for the entire image. The **Prediction Head Architecture** of YOLACT has three branches, generating class confidences, bounding boxes, and mask coefficient prediction.

2) **Prediction Head:** The second pathway is similar to RetinaNet and it is used to predict mask coefficient. Usually, the prediction heads of anchor-based object detectors have two branches: one branch for \( c \) class confidences prediction, the other branch for bounding box regressors prediction [15]. YOLACT added a third branch in parallel for \( k \) mask coefficient prediction. At each branch, \( \alpha \) anchors for feature layer \( P_i \). Additionally, \( \tanh \) is applied to the mask coefficients to produce more stable outputs over nonlinearity. The output of the Prediction Head is passed through Non-maximum Suppression (NMS) before branch assembly.

3) **Assembly:** YOLACT linearly combines the output of protonet and mask coefficient branch and applies sigmoid nonlinearity to produce the final masks \( M \). This is done through matrix multiplication and sigmoid function \( (\sigma) \), according to Eq. 1, where \( P \) is a \( w \times h \times k \) matrix from prototype masks and \( C \) is a \( n \times k \) matrix of mask coefficients.

\[
M = \sigma(PC^T) \tag{1}
\]

After mask assembly, three loss components were used to train the model: classification loss, box regression loss, and mask loss with weights 1, 1.5, and 6.125, respectively. The classification loss and box regression loss are defined in the same way as the losses in Single Shot MultiBox Detector [16]. The mask loss is computed using pixel-wise binary cross-entropy between ground truth masks and the assembled masks.

**E. Landmarks Extraction**

Once the segmentation is derived, landmarks of each vertebra are extracted. To do this, we first find the contour of each segmentation using OpenCV functions cv2.findContours() and cv2.convexHull() [17]. Then, we draw a minimum bounding box around each contour. The four corners of the minimum bounding box are considered to be the landmarks of the vertebra. This process is illustrated in Figure 3.

![Fig. 3. a) Using linear outlines to identify the contour of vertebrae initially, b) followed by drawing a minimum bounding box around the contour of each vertebra.](image)

**F. Angle Calculation**

After correctly identifying the landmarks of each vertebra, we use trigonometry to calculate the Cobb angle. As shown in Figure 4, the strategy is to connect the landmarks to form a horizontally tilted line at each endplate [10]. We then use an algorithm that iterates through all the lines to find three angles: 1) The maximum angle, which is usually in the Main-Thoracic (MT) region. 2) The maximum angle above MT, which is usually in the Proximal-Thoracic (PT) region and
3) The maximum angle below MT, which is usually in the Thoraco-Lumbar (TL) region.

The angles between two lines can be calculated using trigonometry based on Eq. 2 where \( m_1 \) and \( m_2 \) are slopes of line \( l_1 \) and line \( l_2 \), respectively; and \( \alpha \) is the angle between the two lines.

\[
tan \alpha = \frac{m_1 - m_2}{1 + m_1m_2}
\]  (2)

**Evaluation**

With the calculated angles, the result of the proposed method can be evaluated using two methods. The first method is to use SMAPE, which measures the relative error in the form of percentage.

\[
SMAPE = \frac{1}{N} \sum N \sum_{i=1}^{3} \left( \frac{|a_{ji} - b_{ji}|}{|a_{ji} + b_{ji}|} \right) \times 100\%
\]  (3)

In Eq. 3, \( i \) indexes the three Cobb angles; \( j \) refers to the number of images; \( N \) is the total number of test images; \( a \) is the estimated and \( b \) is the ground-truth Cobb angles.

The second method is to use the absolute difference. The absolute difference can be calculated using Eq. 4 for further insight into the results.

\[
\text{Absolute Difference} = |\text{Ground truth angle} - \text{predicted angle}|
\]  (4)

**III. RESULTS AND DISCUSSION**

**A. Qualitative Results**

Figure 5 illustrates the qualitative result of the YOLACT model. Overall, YOLACT produces acceptable segmentation. As shown, all segmentation results follow the curve of the spine with no major deviation. Additionally, unlike other semantic segmentation models [18], YOLACT does not produce corrupted or connected segmentation, leading to more accurate angle measurements.

YOLACT is observed to tend to identify more vertebrae than desired. There are around 24 vertebrae in human spine; however, only 17 (12 thoracic vertebrae and 5 lumbar vertebrae) are required for Cobb angles measurement in this case [12]. YOLACT sometimes recognizes vertebra in the cervical region and produces segmentation for 18-19 vertebrae instead of 17. For example, in Figure 6, one extra vertebra is identified at the top of the image, when we compared the white quadrilaterals (ground truth) with the blue segmentations (results from YOLACT).

**B. Quantitative Results**

YOLACT is able to quickly and correctly identify vertebrae within X-ray images, as illustrated in Figure 5. Operating at a frame rate of 21 frames per second (fps), YOLACT’s efficiency is further evident in the computed Symmetric Mean Absolute Percentage Error (SMAPE). Using Eq. 3, the calculated SMAPE for our approach is 10.76%, which reflects the relative error in the form of percentage. YOLACT surpasses the benchmarks except for one, as indicated by the SMAPE results listed in Table II. Comparing YOLACT with other segmentation models such as revised UNet (SMAPE = 16.48%) [18], it is observed that YOLACT has a better performance (SMAPE = 10.76%). One model that has comparable performance with YOLACT is ResNet (SMAPE = 10.81%) [10]. ResNet achieves good results by using two ML
models instead of one. The ResNet approach first identifies the location of each vertebra; then, it finds the four corners using a trained center offset and heatmaps. In contrast, YOLACT uses only one model, and it has a SMAPE of 10.76%. The only model that has a slightly better performance is the Augmented UNet (SMAPE = 9.2%) [7]; however, YOLACT is more robust as its absolute angle difference never exceeded 20 degrees as shown in Table II. In contrast, 2.62% of the results produced by Augmented UNet has an error larger than 20 degrees [7].

The absolute angle differences were also calculated using Eq. 4 and the results are shown in Table II. The majority of the predictions had less than 5-degree differences and 94.59% of predictions had error less than 10 degrees, indicating the reliability of the YOLACT model.

### TABLE I

SMAPE COMPARISON BETWEEN DIFFERENT METHODS.

| Methods                           | SMAPE  |
|-----------------------------------|--------|
| Revised UNet [18]                 | 16.48% |
| Augmented UNet [7]                | 9.2%   |
| ResNet [10]                       | 10.81% |
| Fast RCNN [19]                    | 25.69% |
| Multi-View Extrapolation Net [20] | 18.95% |
| S$^2$ViT [8]                      | 37.08% |
| BooxNet [12]                      | 23.44% |
| **YOLACT**                        | **10.76%** |

### TABLE II

ABSOLUTE ANGLE DIFFERENCE BETWEEN PREDICTION AND GROUND TRUTH.

| Absolute angle difference | Percentage |
|---------------------------|------------|
| Difference less than 5 Degree | 64.86%    |
| Difference between 5 and 10 degrees | 29.73%    |
| Difference between 10 and 20 degrees | 5.41%     |

### IV. CONCLUSIONS

In this paper, a modern instance segmentation model, YOLACT, was used to estimate the Cobb angles using X-ray images. In comparison to previous methods, the YOLACT approach is accurate, consistent, and does not require excessive amount of computation resources. We achieved a SMAPE of 10.76% which demonstrates the ability of YOLACT to predict Cobb angles. As indicated by both qualitative and quantitative results, the proposed method can help improve accuracy and efficiency in Cobb angle measurement. The proposed method has the potential to be integrated into clinicians’ routine diagnostic assessment for scoliosis. Future directions include to improve the proposed method via outlier rejection techniques to correct segmentation errors and remove undesirable noise, as well as false positive objects. Additionally, creating a larger training dataset with more diverse cases will enhance the performance.

### V. ACKNOWLEDGMENT

This work was supported by the Chair in Medical Imaging and Artificial Intelligence, a joint Hospital-University Chair between the University of Toronto, The Hospital for Sick Children, and the SickKids Foundation.

### REFERENCES

[1] J. A. Janicki and B. Alman, “Scoliosis: Review of diagnosis and treatment," *Pediatrics and amp; Child Health*, vol. 12, no. 9, p. 771–776, 2007.

[2] F. Galbusera, A. Cina, M. Panico, and T. Bassani, “The importance of curve severity, type and instrumentation strategy in the surgical correction of adolescent idiopathic scoliosis: An in silico clinical trial on 64 cases,” *Scientific Reports*, vol. 11, no. 1, 2021.

[3] S. Langensiepen, O. Semler, R. Sobottke, O. Fricke, J. Franklin, E. Schönau, and P. Eysel, “Measuring procedures to determine the cobb angle in idiopathic scoliosis: A systematic review,” *European Spine Journal*, vol. 22, no. 11, p. 2360–2371, 2013.

[4] M. C. Tanure, A. P. Pinheiro, and A. S. Oliveira, “Reliability assessment of Cobb angle measurements using manual and nbs: digital methods,” *The Spine Journal*, vol. 10, no. 9, p. 769–774, 2010.

[5] M. Gstoettner, K. Sekyra, N. Walochnik, P. Winter, R. Wachter, and C. M. Bach, "Inter- and intraobserver reliability assessment of the Cobb angle: Manual versus digital measurement tools," *European Spine Journal*, vol. 16, no. 10, p. 1587–1592, 2007.

[6] K. Chen, X. Zhai, K. Sun, H. Wang, C. Yang, and M. Li, “A narrative review of machine learning as promising revolution in clinical practice of scoliosis,” *Annals of Translational Medicine*, vol. 9, no. 1, p. 67–67, 2021.

[7] Y. Wu, K. Namdar, C. Chen, S. Hosseinipour, M. Shroff, A. S. Doria, and F. Khalvati, “Automated adolescent scoliosis detection using augmented u-net with non-square kernels,” *Canadian Association of Radiologists Journal*, vol. 74, no. 4, p. 667–675, 2023.

[8] H. Sun, X. Zhen, C. Bailey, P. Rasoulinajad, Y. Yin, and S. Li, “Direct estimation of spinal cobb angles by structured multi-output regression,” *International Conference on Information Processing in Medical Imaging (IPMI 2017)*, p. 529–540, 2017.

[9] Y. Pan, Q. Chen, T. Chen, H. Wang, X. Zhu, Z. Fang, and Y. Lu, “Evaluation of a computer-aided method for measuring the cobb angle on chest x-rays,” *European Spine Journal*, vol. 28, no. 12, p. 3035–3043, 2019.

[10] J. Yi, P. Wu, Q. Huang, H. Qu, and D. N. Metaxas, “Vertebral-focused landmark detection for scoliosis assessment,” 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020.

[11] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact: Real-time instance segmentation,” 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[12] H. Wu, C. Bailey, P. Rasoulinajad, and S. Li, “Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boost-net,” *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, p. 127–135, 2017.

[13] “Aasce - grand challenge.” [Online]. Available: https://aasce19.grand-challenge.org/evaluation/challenge/leaderboard/

[14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” *Computer Vision – ECCV 2014*, p. 740–755, 2014.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[16] W. Liu, D. Anguelov, D. Erhan, P. Szegedy, R. E. Reed, and A. C. Berg, “Ssd: Single shot multibox detector,” *Computer Vision – ECCV 2016*, p. 21–37, 2016.

[17] S. Mallick, “Contour detection using opencv (python/++),” Oct 2021. [Online]. Available: https://learnopencv.com/contour-detection-using-opencv-python-c/

[18] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact: Real-time instance segmentation,” 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, vol. 10, no. 9, p. 769–774, 2019.

[19] K. Chen, X. Zhai, K. Sun, H. Wang, C. Yang, and M. Li, “A narrative review of machine learning as promising revolution in clinical practice of scoliosis,” *Annals of Translational Medicine*, vol. 9, no. 1, p. 67–67, 2021.

[20] L. Wang, Q. Xu, S. Leung, J. Chung, B. Chen, and S. Li, “Accurate automated cobb angles estimation using multi-view extrapolation net,” *Medical Image Analysis*, vol. 58, p. 101542, 2019.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.