IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 May 2023, accepted 23 June 2023, date of publication 27 June 2023, date of current version 6 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289957

RESEARCH ARTICLE

Automatic Measurement of Scoliosis Based on an Improved Segmentation Model

JIA ZHU, ZHIFENG ZHOU, AND CHENGXIAN YAO
School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Corresponding author: Zhifeng Zhou (zhousjtu@126.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Shanghai Engineering University, under Application No. EST-2023-011.

ABSTRACT The detection of the spine is crucial in automating the measurement of the Cobb Angle. While various segmentation models have been employed for vertebrae segmentation in X-ray images, there is a need to enhance segmentation performance. This paper proposes a comprehensive automatic measurement method for the Cobb angle. The RetinaNet model is employed to detect the region of interest corresponding to the spine, while the W-Net model is developed for accurate vertebrae segmentation. To address the issue of adjacent vertebrae adhesion in the segmented image, a post-processing technique is applied. Experimental results demonstrate that the W-Net model achieves superior performance, with a mean Intersection over Union (MIoU) of 0.9073 $\pm$ 0.0021, Dice Coefficient of 0.9446 $\pm$ 0.0139, and Precision of 0.9390 $\pm$ 0.0190. The post-processing step reduces adhesion at one end by approximately 83.4% and adhesion at both ends by approximately 83.6%. The reliability of the proposed method is evaluated through intra-group correlation coefficients (ICC) of 0.902 and 0.915, respectively, between two observers, both exceeding 0.9. The mean absolute deviation (MAD) is 3.08° and 2.91°, respectively. Therefore, the proposed method achieves automatic detection of the Cobb angle without the need for manual cropping or additional human intervention, while maintaining good reliability.

INDEX TERMS Cobb angle, deep learning, image segmentation, scoliosis.

I. INTRODUCTION

Scoliosis is a common spinal disease that often occurs in adolescents and children [1]. When the degree of spinal curvature is greater than 10°, scoliosis is considered to be present. Scoliosis can cause physical deformities such as protruding scapula and ribs, uneven shoulders, and asymmetrical waistline [2]. Untreated severe scoliosis can lead to complications such as respiratory distress, heart problems, and chronic back pain [3], [4]. The main surgical treatment for idiopathic scoliosis is invasive spinal fusion surgery. Although surgery can bring significant long-term benefits, the complexity of the surgical process and the long recovery time (6 months or more) pose particularly challenging requirements for patients, such as back pain and immobility [5]. Therefore, early diagnosis and treatment are crucial for scoliosis patients.

Diagnosis and treatment of spinal curvature rely on X-rays and measurement of the Cobb angle [6], [7]. The Cobb angle [8], [9] is the angle between the upper edge of the upper vertebra and the lower edge of the lower vertebra, used to evaluate the degree of spinal curvature. Figure 1 shows the results of traditional manual measurement of the Cobb angle on spinal X-rays. Generally, observation treatment is needed for Cobb angles less than 25 degrees, bracing treatment for angles between 25 and 45 degrees, and surgical treatment for angles greater than 45 degrees [10], [11]. The traditional manual measurement of the Cobb angle method has subjective and non-rigorous problems, and the efficiency is also low. With the promotion of scoliosis screening, there is an increasing demand for using X-rays to detect scoliosis. Therefore, it is crucial to adopt computer-aided or even fully automated methods to measure the Cobb angle [12].

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

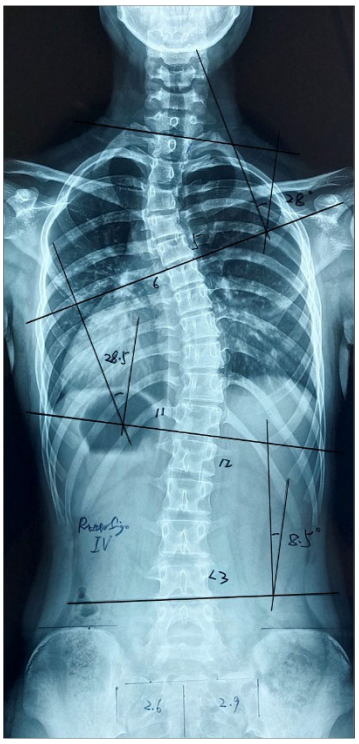

FIGURE 1. Cobb angle manual measurement results.

Therefore, this article proposes a spinal segmentation method based on deep learning, and through post-processing of the segmentation image, automatic estimation of the Cobb angle is achieved, thereby improving the efficiency and accuracy of spinal curvature detection.

II. RELATED WORKS

The conventional approach to computer-based spinal X-ray detection involves performing calculations on the pixel values of the X-ray image using standard image processing algorithms. These algorithms include threshold-based segmentation methods [13], [14], and edge-based segmentation techniques [15], [16]. A denoising method has been proposed in [17], which is followed by histogram equalization, Canny edge detection, and Hough transform to identify the spinal edge. In reference [18], the procedure entails converting the digital X-ray image of scoliosis into grayscale, identifying seed points for different types of scoliosis, dividing the image into 12 sub-images, applying median and Canny edge detection filters, identifying centroids, performing polynomial curve fitting regression, and determining the Cobb angle using the gradient principle. However, the efficacy of image processing methods is greatly influenced by factors such as image quality and pixel distribution, leading to suboptimal segmentation outcomes.

In recent years, deep learning has been extensively applied and developed in the field of medical segmentation, with the U-Net [19] model proposed in 2015 being a prominent example. Subsequent works, such as those by Tan et al. [20] and Zhao et al. [21], have further improved the U-Net network structure to achieve better segmentation results. In particular, Zhao et al. [21] fitted the minimum circumscribed rectangle to the segmentation image obtained by the improved U-Net to obtain the information required for calculating the Cobb angle. Horng et al. [22] employed an intensity histogram to obtain the Region of Interest (ROI) of the spine, detected vertebra edges using intensity and gradient methods, fed a single vertebra into a segmentation network, and finally reconstructed the segmentation results of the vertebrae into a complete segmented spine image. Khanal et al. [23] initially utilized Faster RCNN to detect individual vertebrae, followed by DenseNet to detect four key points of each vertebra, and eventually obtained the slope to calculate the Cobb angle. The method proposed by Alharbi et al. [24] is similar to that proposed by Khanal et al. They employed ResNet for detecting individual vertebrae and extracting information from the corresponding rectangular bounding boxes of each vertebra, enabling the calculation of the Cobb angle. However, many improved U-Net structures have not achieved significant improvement in segmentation accuracy. Detection of vertebra edges using intensity and gradient methods is highly dependent on image quality. The method of reorganizing after single vertebra detection is not robust. Furthermore, the X-ray spine datasets used in the proposed methods were manually cropped, and the input X-ray images were not complete full-spine anteroposterior X-rays. Therefore, manual cropping or box selection is still required in practical applications, which cannot fully realize automatic detection.

To enhance the segmentation effect of spine X-ray images, simplify the process, and improve the stability and reliability of spine detection, we propose a novel W-Net model based on the skip connection characteristic of U-Net. Our experimental results demonstrate that the performance of the W-Net model surpasses that of U-Net and other typical segmentation models. To detect the spinal region, we employ the target detection network RetinaNet [25] to exclude the unnecessary cervical vertebra and other features on the X-ray and predict a bounding box containing 12 thoracic vertebrae and 5 lumbar vertebrae on the X-ray. By utilizing the coordinate information of the bounding box, we can set the pixel values of the segmented image predicted by W-Net, other than the required region, to 0, thereby obtaining the segmented image of 17 vertebrae. Our proposed detection method can achieve automatic detection without the need for manual cropping of the spine X-ray.

III. PROPOSED METHOD

This paper proposes a network architecture named W-Net for segmenting the spine based on the characteristics of the U-Net network structure. The W-Net improves the segmentation performance by modifying the network structure. We train the W-Net on a custom dataset and apply it to the RetinaNet object detection network for detecting the spine. We mask the areas outside the spine in the binary image obtained from the W-Net segmentation to obtain the preliminary binary image of the spine. Finally, we perform post-processing on

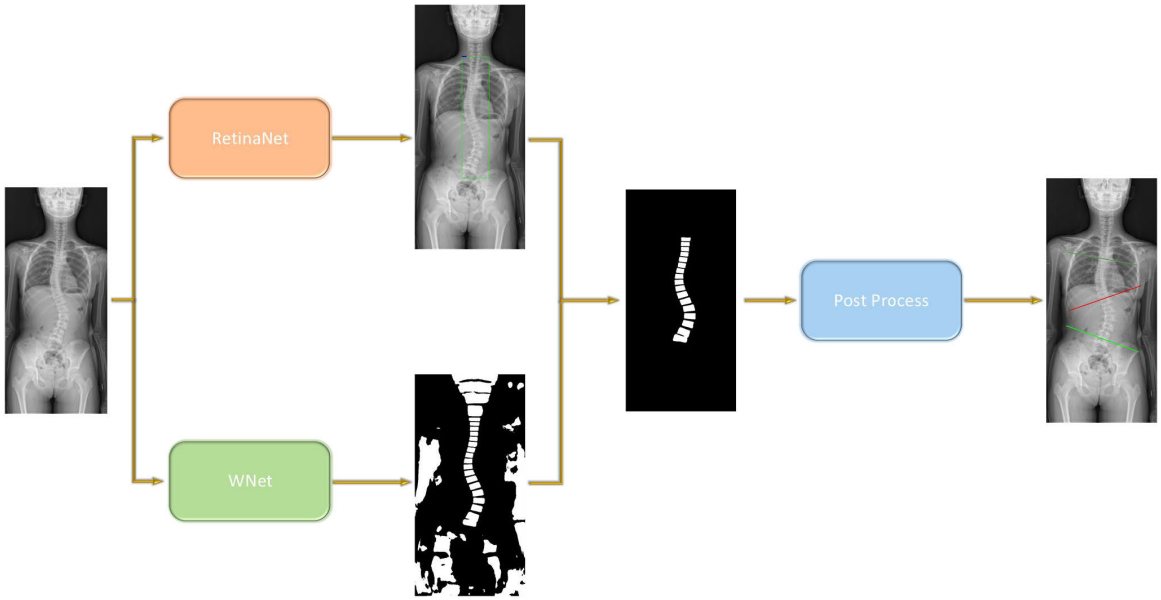

FIGURE 2. Flow chart of automatic Cobb Angle detection.

the binary image to calculate the accurate Cobb angle. This method effectively segments the spine and provides assistance for the diagnosis and treatment of spinal diseases.

A. OVERALL STRUCTURE

In this study, we utilized the RetinaNet [25] object detection network and our proposed segmentation network, W-Net, to detect spinal X-ray images. By inputting spinal X-ray images into RetinaNet and W-Net, we obtained the bounding box of the spine ROI and a segmentation image. As the segmentation network is essentially a binary classification of each pixel, it is inevitable that some non-spinal features, such as the head, organs, and bones, may be misclassified as the spine in the untrimmed full spinal standing X-ray images during training and prediction. To address this issue, we set the pixel values outside the required spinal region to zero using the bounding box coordinates, thereby obtaining a preliminary segmentation binary image of the spine. However, differences between spinal X-ray images and variations in image quality may result in some segmentation images with adhesion issues. Therefore, a series of post-processing steps are required before the final Cobb angle calculation. The automatic detection process of the Cobb angle is illustrated in Figure 2.

B. SPINAL ROI DETECTION

RetinaNet [25] is a one-stage network proposed in 2017, which outperforms two-stage networks [26], [27], [28] in object detection. Due to its simpler model structure, RetinaNet achieves higher detection speed and accuracy than two-stage networks. In this study, we use RetinaNet to detect the ROI of the spine.

The human spine consists of 7 cervical vertebrae, 12 thoracic vertebrae, 5 lumbar vertebrae, and 1 sacrum. The vertebrae required for calculating the Cobb angle include 12 thoracic vertebrae and 5 lumbar vertebrae, while cervical vertebrae have features very similar to those of thoracic and lumbar vertebrae. Therefore, even if the annotated objects are the aforementioned 17 vertebrae, directly predicting the entire spine radiograph will detect some pixels in the cervical vertebrae and sacrum regions as the required vertebrae. Meanwhile, the essence of segmentation models is to classify each pixel, so training and predicting on the entire spine radiograph without cropping inevitably classifies some pixels outside the spine region as the required vertebrae. Therefore, to extract the 17 vertebrae from the segmentation image, we train RetinaNet to detect one bounding box of the 17 vertebrae on each X-ray image. The bounding box coordinates are then used to set the pixel values outside the 17 vertebrae to 0 on the segmentation image.

C. VERTEBRAE SEGMENTATION BY W-NET

While typical models based on U-Net [19], such as U-Net++ and U2Net, have shown slightly better segmentation performance than UNet, their structures are complex and their parameter quantities are large. Therefore, we aim to develop a segmentation model with a simpler structure, fewer parameters, and superior performance. U-Net is a neural network with an encoder-decoder structure, and its most prominent feature is the skip connection. This network fuses low semantic and fine-grained features from the contraction path with high semantic and coarse-grained information from the expansion path, thereby preserving high-level abstract semantic information to the high-resolution feature layer and achieving more accurate segmentation. After training a custom spine X-ray dataset on U-Net, we found that its segmentation performance was good, but there was still room for improvement. We believe that the skip connection of U-Net is the main reason for its excellent segmentation performance. Therefore, this paper proposes a novel network architecture,

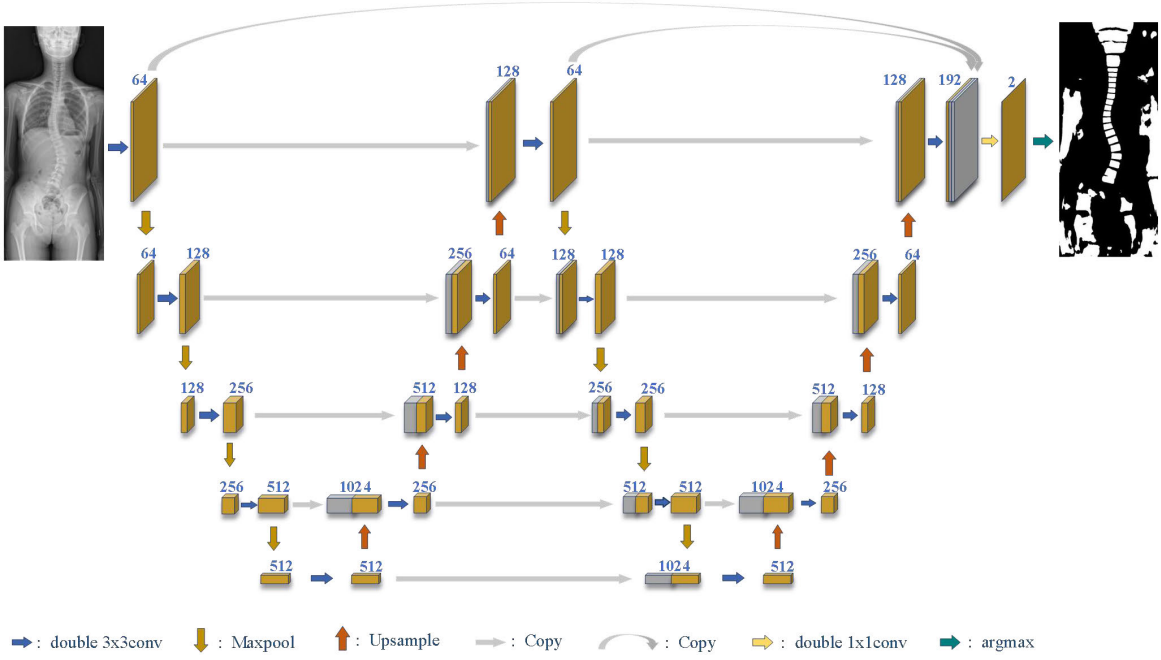

FIGURE 3. W-Net structure.

W-Net, as illustrated in Figure 3, which is based on the characteristics of U-Net to achieve superior segmentation performance. W-Net consists of two U-Nets, and the input size of the X-ray image is 1024 × 512. After four times of maximum pooling in the contraction path, the resolution is reduced to a minimum of 256 × 128. To reduce the number of parameters and computation time, we replaced the convolution in the expansion path of the original U-Net network with upsampling. The concatenation of the two U-shaped structures and the concatenation between the expansion path and the contraction path are similar. The first feature layer of the contraction path of the right U-shaped structure is obtained by performing two 3 × 3 convolutions on the last feature layer of the expansion path of the left U-shaped structure. The other four feature layers of different scales in the expansion path of the left U-shaped structure are concatenated with the other four feature layers of different scales in the contraction path of the right U-shaped structure. In the final part of the network, the first feature layer of the contraction path of the two U-shaped structures is concatenated with the last feature layer of the expansion path of the right U-shaped structure after two 3 × 3 convolutions, and then two 1 × 1 convolutions are performed to obtain a feature layer with a channel number of 2. Finally, we use the argmax function in numpy to perform binary classification on the pixel points of the X-ray image and output a binary image.

IV. EXPERIMENT

A. EXPERIMENTAL PREPARATION

1) EXPERIMENTAL PLATFORM

The GPU used in this experiment is GTX3090, and the graphical memory is 24GB. The software environment was Python in Anaconda virtual environment, with PyTorch version 1.11.0.

2) DATASET

The custom dataset provided by Carespine consists of 380 X-ray images in jpg format, with 300 images used for training, 50 for validation, and 30 for testing. The dataset was annotated under the guidance of professional doctors. The rectangular bounding boxes for the spine ROI were labeled in txt format using labelImg, while te ground truth for spine segmentation was annotated in json file format using labelme. The X-rays were not cropped, and thus contain features such as the skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, hip bones, sacrum, and internal organs. Figure 4 provides an example of the annotation for the spinal X-ray dataset. As the calculation of the Cobb angle is performed on 12 thoracic vertebrae and 5 lumbar vertebrae, and the cervical vertebrae share very similar features with the thoracic and lumbar vertebrae, the model may inevitably predict the pixels at the cervical vertebrae as ground truth during training and prediction. Therefore, to avoid misleading the model and improve segmentation performance, the cervical vertebrae were also annotated as ground truth, and the parts outside the thoracic and lumbar vertebrae were masked using bounding boxes predicted by the object detection network.

In order to standardize and expand the dataset, as well as improve the model's generalization ability, we performed preprocessing before inputting the dataset into the network for training. This preprocessing includes randomly scaling the images to a custom size of 2000 × 1000, horizontally flipping with a probability of 0.5, randomly cropping to a uniform input size of 1024 × 512, and normalization.

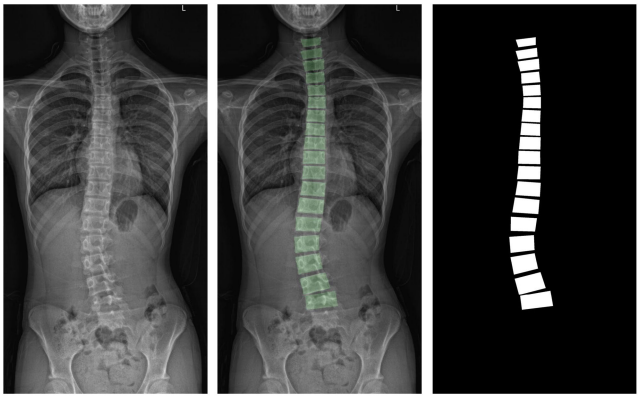

FIGURE 4. Schematic drawing of spinal X-ray dataset: a) original spinal X-ray image; b) fused image; c) ground truth of vertebrae.

For the annotated files of spine segmentation, the background pixel value is 0 and the foreground pixel value is 255. To better calculate the loss, we set the foreground pixel value of 255 to 0 and set the areas outside the spine ROI on RetinaNet to 255.

B. LOSS CALCULATION

During the model training process, the calculation of the loss is composed of two parts, namely Cross Entropy Loss and Dice Loss. As the segmentation target is only the spine, i.e., the number of categories is 2 (including background), the calculation is based on binary cross entropy loss. When calculating this loss, the region outside the spine is masked, i.e., the pixel values in the region where the value is set to 255 when reading the dataset are ignored. The formulas for binary cross entropy loss and Dice Loss are as follows:

$$Loss_{BCE} = -y_t \cdot \log(y_p) - (1 - y_t) \cdot \log(1 - y_p) \quad (1)$$

$$Loss_{dice} = 1 - \frac{2\left|y_t \cap y_p\right|}{\left|y_t\right| + \left|y_p\right|} \quad (2)$$

where y_t is the ground truth, and y_p is the predicted result.

C. PERFORMANCE INDEX

During the training process, the main metrics recorded were Loss, Dice Coefficient, Global Accuracy, and MIoU (Mean Intersection over Union).

1) Dice Coefficient

The Dice coefficient is commonly used to calculate the similarity between two samples, and in this study, it is used to calculate the similarity between the ground truth and predicted values of the spine. The formula is as follows:

$$Dice = \frac{2\left|y_t \cap y_p\right|}{\left|y_t\right| + \left|y_p\right|} \quad (3)$$

where y_t represents the number of pixels annotated as ground truth for the segmentation target (in this paper, referring to the vertebrae), and y_t represents the number of pixels predicted as ground truth.

2) Global Accuracy

Global Accuracy refers to the proportion of correctly predicted pixels to the total number of pixels (annotated as the number of vertebral ground truth pixels) and is calculated using the following formula:

$$Global\ Acc = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad (4)$$

3) MIoU

IoU refers to the cross combination ratio of true value and predicted value, and MIoU refers to the mean IoU of the foreground (vertebrae) and background (spine ROI except vertebrae). The calculation formula is as follows:

$$mean\ IoU = \frac{1}{N+1} \cdot \sum_{i=0}^{N} \frac{n_{ii}}{\sum_{j=0}^{N} n_{ij} + \sum_{j=0}^{N} n_{ji} - n_{ii}} \quad (5)$$

where n_{ij} represents the number of pixels predicted as class j for class i, and t_i represents the total number ofpixels for class. i

D. TRAINING DETAINS

This article compares and analyzes six classic segmentation networks, including FCN [29], DeepLabv3 [30], U-Net, U-Net++ [31], U2Net [32], and the proposed W-Net. All networks were trained for 200 epochs, with an initial learning rate of 0.05 and SGD optimizer. Figure 5 shows the change curves of several indicators during the training process.

FCN and DeepLabv3 used ResNet [33] as the feature extraction network (backbone) and were trained with pre-trained weights. To avoid damaging the pre-trained weights of the feature extraction network, the traininsg of FCN and DeepLabv3 first froze the pre-trained weights and only trained the feature layers except for the feature extraction network. After 80 epochs, the feature extraction network weights were unfrozen, and the entire model was trained. Therefore, the curve of FCN and DeepLabv3 will undergo a discontinuity, as shown in the figure 5.

Although FCN and DeepLabv3 have deeper network structures and use feature extraction networks, the training results show that these two models do not perform better than other models that directly train the entire network.

In addition, the training curves of W-Net, U-Net, U-Net++, and U2Net are similar, but it can be seen that the convergence process of W-Net is more stable, and the MIoU, Dice

Coefficient, Global Accuracy, and convergence speed are slightly better than other networks.

V. RESULTS AND DISCUSSION

A. MODEL PERFORMANCE ANALYSIS

We have recorded the MIoU, Dice Coefficient, Precision, parameter count, and FPS of the networks. Table 1 presents a comparison of the performance of these models. MIoU and Dice Coefficient have been defined earlier, while Precision is

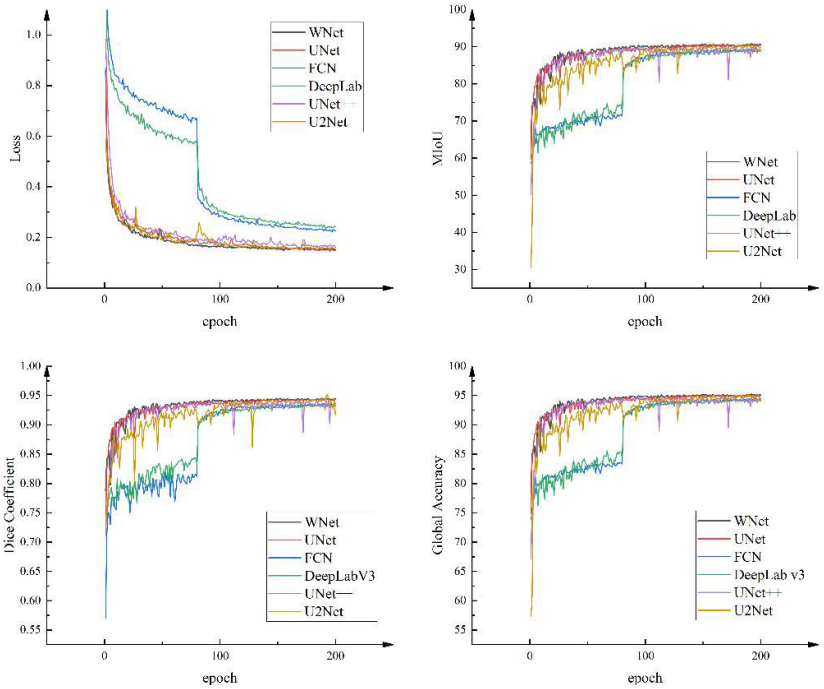

FIGURE 5. Training process curve.

TABLE 1. Comparison of the results for segmentation networks.

	MIoU	Dice Coefficient	Precision	Parameters	FPS
U-Net(64c)	0.9073±0.0021	0.9446±0.0139	0.9390±0.0190	17	31.18
U-Net(32c)	0.9031±0.0033	0.9362±0.0156	0.9235±0.0029	4	50.18
U-Net++	0.9025±0.0020	0.9346±0.0174	0.9337±0.0046	47	8.31
U2Net	0.9069±0.0027	0.9428±0.0193	0.9289±0.01167	44	24.42
FCN	0.8929±0.0036	0.9351±0.0187	0.9201±0.0127	35	26.31
DeepLabV3	0.9814±0.0028	0.9325±0.0203	0.9196±0.0032	42	24.82
W-Net(64c)(Ours)	0.9109±0.0027	0.9473±0.0155	0.9494±0.0101	36	15.46
W-Net(64c)(Ours)	0.9084±0.0032	0.9442±0.0161	0.9433±0.0176	9	41.72

calculated using the following formula:

$$\Pr ecision = \frac{TP}{TP + FP} \tag{6}$$

In the above formula, TP represents the number of pixels correctly predicted as vertebrae, while FP represents the number of non-vertebrae pixels predicted as vertebrae.

As shown in Table 1, FCN and DeepLabv3, which have larger model parameters and network depth and use feature extraction networks, did not achieve better MIoU. W-Net with a channel of 32 in the first feature layer after inputting the image has a similar MIoU to U-Net, U2Net, and U-Net++. W-Net with a channel of 64 in the first feature layer achieved the best performance in terms of MIoU, Dice Coefficient, and Precision. Additionally, W-Net with a channel of 32 in the first feature layer maintained a high MIoU while having parameters and FPS second only to U-Net with a channel of 32 in the first feature layer. It can be seen that our model has better performance compared to other segmentation models.

B. VERTEBRAE SEGMENTATION RESULTS

As shown in Figure 6, the segmentation results of the proposed W-Net are compared with those of U-Net, FCN, DeepLabv3, U-Net++, and U2Net. It can be observed that W-Net performs the best in the spinal segmentation task, exhibiting more stable segmentation results and less adhesion compared to the other networks.

C. SEGMENTATION RESULTS OF ADHERENT VERTEBRAE

Due to the variability in pixel distribution and spacing between adjacent vertebrae in spinal X-ray images, it is inevitable that there will be adhesions in the binary images obtained through segmentation models. Figure 7 illustrates four different types of adhesion. Most segmentation images exhibit no adhesion, as shown in Figure 7(a). However, there are cases where adjacent vertebrae are partially adhered or fully adhered at both ends, as shown in Figures 7(b) and 7(c), respectively. For X-ray images with poor quality, there may be multiple or large areas of adhesion, as shown in Figure 7(d). Since the adhesion in Figure 7(d) cannot be

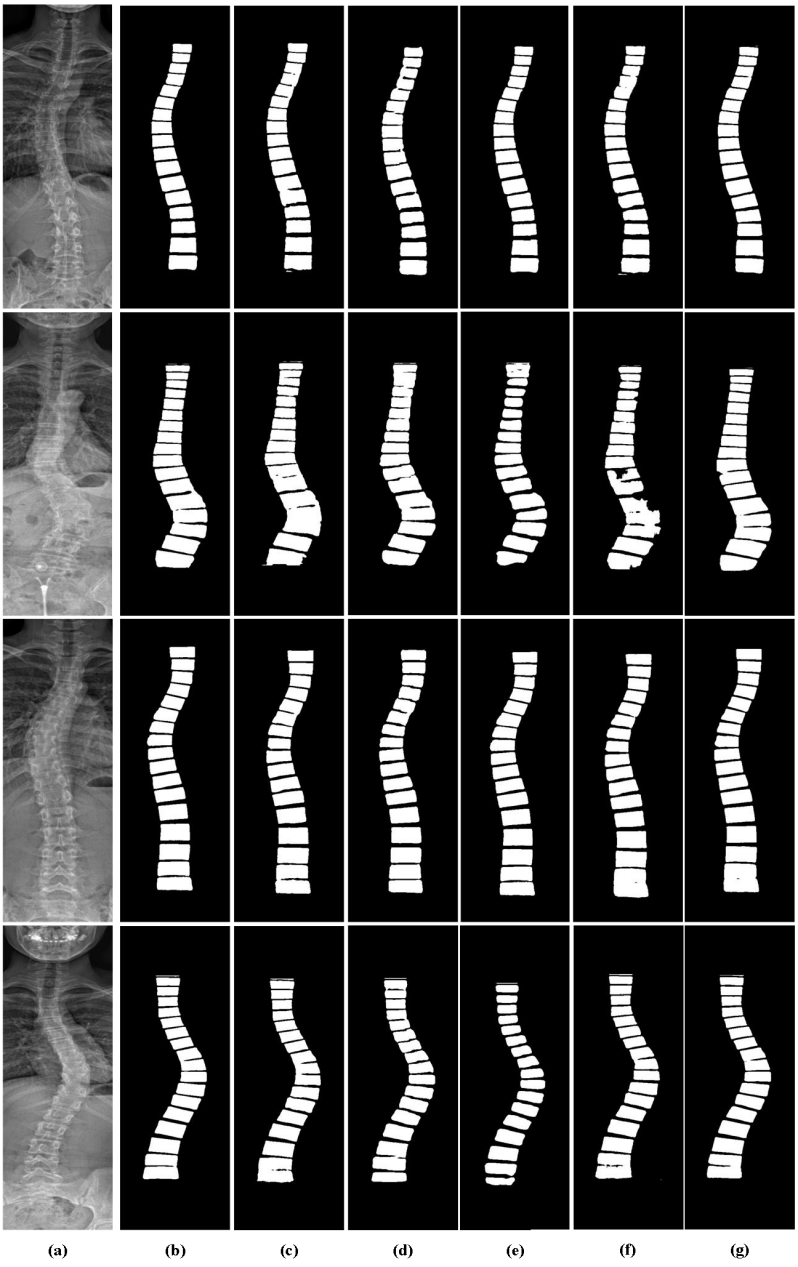

FIGURE 6. (a)original x-ray images, (b)∼(g): segmentation result of W-Net, U-Net, FCN, DeepLabv3 U-Net++ and U2Net.

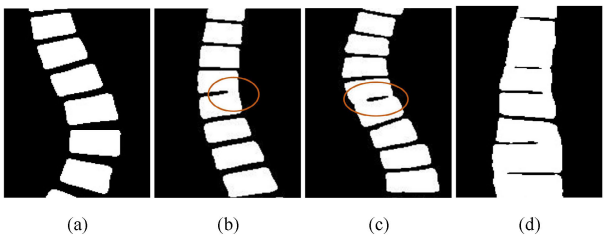

FIGURE 7. Ahesdion situations (a) No adhesion; (b) one end adhesion; (c) both ends adhesion; (d) others situations.

segmented through post-processing methods, the following post-processing methods are designed for the most common cases of adhesion at one end or both ends.

In order to obtain the necessary vertebral information for calculating the Cobb angle, we designed an effective post-processing method for vertebrae segmentation, as shown in Figure 8. This method includes the following steps:

1) First contour extraction: The findContours [34] function in OpenCV is used to extract contours from the segmentation image. Due to the problem of adhesion between individual vertebrae in some images, a contour may contain two vertebrae.

2) Cut adhesion at one end: The convexHull function in OpenCV is used to calculate the convex hull of the extracted contour. This method can detect adhesively connected vertebrae at one end and determine the adhesion position. Based on

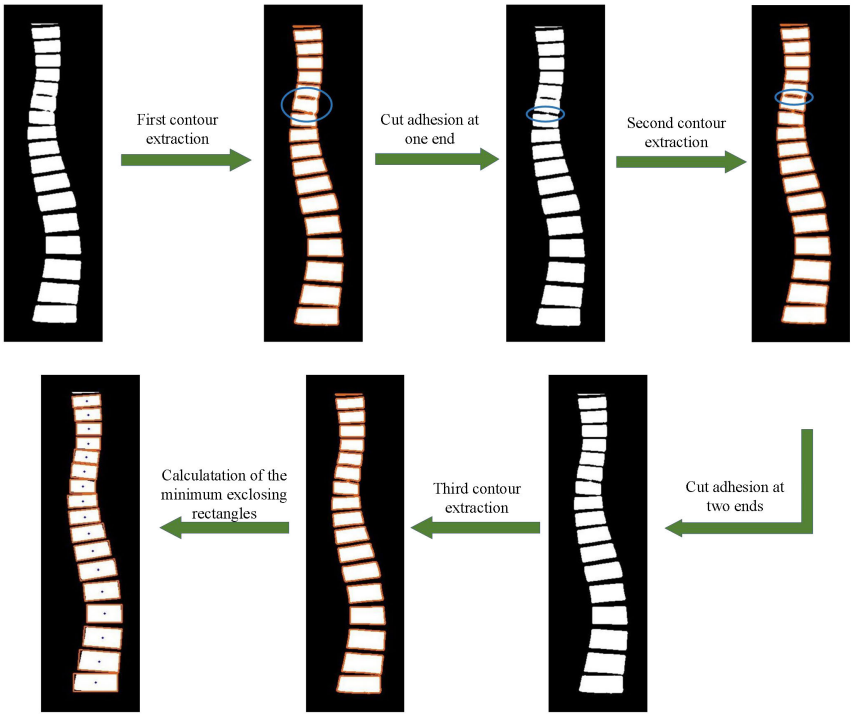

FIGURE 8. Post-processing of spinal segmentation.

TABLE 2. Adhesion situations before and after post-processing * PP in table2 refers to post-processing.

	No adhesion		One end adhesion		Both ends adhesion	
	Bedore PP	After PP	Bedore PP	After PP	Bedore PP	After PP
U-Net(64c)	392	443	44	7	16	2
U-Net(32c)	387	442	49	9	18	3
U-Net++	388	440	46	8	17	3
U2Net	392	441	43	6	14	2
FCN	371	430	52	10	21	4
DeepLabv3	365	427	54	12	25	5
W-Net(64c)(Ours)	406	452	42	4	9	1
W-Net(32c)(Ours)	395	445	44	6	14	2

the returned two adhesion coordinates, the pixel values of the line connecting the two points are set to 0, thereby achieving segmentation of the adhesion.

3) Second contour extraction: The convex hull detection method described above can segment vertebrae with adhesion at one end, but there may be cases where two vertebrae are adhesively connected at both ends. Therefore, a second contour extraction is needed to determine the position of the vertebrae with adhesion at both ends. For vertebrae with adhesion at both ends, the gap in the middle will be detected as a contour.

4) Cut adhesion at two ends: Based on the contour coordinates of the gap in the middle of the vertebrae with adhesion at both ends determined by the second contour extraction, the boundingRect and minAreaRect algorithms in OpenCV are used to detect the minimum enclosing rectangle and minimum enclosing parallelogram. The coordinates of the two ends of the middle gap are calculated based on the coordinate information and rotation angle of the two rectangles. Using the same method as for segmentation of adhesively connected vertebrae at one end, the pixel values on the extension line connecting the two points are set to 0 to achieve segmentation.

5) Third contour extraction: Accurate vertebral contours without adhesion are obtained.

6) Calculation of the minimum enclosing rectangles: Based on the vertebral contour information without adhesion, the minAreaRect algorithm is used to fit the minimum enclosing parallelogram. The coordinates of the four vertices of the minimum enclosing parallelogram are used as the four corner points of each vertebra for Cobb angle calculation.

Overall, this post-processing method effectively segments the spinal image and provides accurate information for calculating the Cobb angle.

We sampled adjacent vertebrae and recorded the adhesion status before and after post-processing on 30 X-ray images. As shown in Table 2, the results indicate a significant reduction in adhesion quantity and an increase in the number of non-adherent vertebrae after post-processing. Specifically, the adhesion at one end decreased by approximately 83.4%,

TABLE 3. Comparison of SMAPE errors.

	SMAPE/%
Method1 [17]	37.83
Method2 [19]	24.15
Method3 [22]	26.91
Proposed method	11.82

and the adhesion at both ends decreased by approximately 83.6%.

D. RESULTS OF COBB ANGLE MEASUREMENT

Using 30 test images, this study compared the symmetric mean absolute percentage error (SMAPE) of the methods proposed in literature 17, 19, and 22 with the method proposed in this study, with the physician-measured Cobb angle serving as the reference standard. SMAPE is calculated using the following formula:

$$SMAPE = \frac{1}{n}\sum_{n}\frac{\left|\theta_{pred} - \theta_{gt}\right|}{\left(\left|\theta_{pred}\right| + \left|\theta_{gt}\right|\right)/2} \cdot 100\% \qquad (7)$$

where n is the number of samples, and θ_{gt} and θ_{pred} are the actual and predicted values of the Cobb angle, respectively.

The results showed that the traditional image processing method (presented in literature 17) had the highest SMAPE, reaching 37.83%. In contrast, methods 2 and 3 that used segmentation models, object detection models, and key point detection models, had SMAPE of 24.15% and 26.91%, respectively, which was significantly lower than that of the traditional method. The proposed method in this study had the best performance, achieving an SMAPE of 11.82%.

In order to validate the proposed automatic detection method for Cobb angle, we conducted a reliability analysis of manual measurement and the automatic detection method on 30 X-ray images. The intraclass correlation coefficient (ICC) within and between observers was used as an indicator of inter-observer reliability. The mean absolute deviation (MAD) was used to calculate the average absolute deviation of individual observations from the arithmetic mean.

The reliability analysis within and between observers is shown in Table 4. The ICCs within the two observers were 0.955 and 0.950, respectively, with 95% confidence intervals, and the MADs were 2.18 and 2.33, respectively, indicating excellent intra-observer reliability. The ICCs between the first and second measurements were 0.939 and 0.947, respectively, with MADs of 2.57 and 2.50, respectively, indicating good inter-observer reliability. Generally, the intra-observer reliability of manual methods is slightly better than the inter-observer reliability.

Table 5 shows the reliability evaluation of the automatic detection method compared to manual measurement. The ICCs between the automatic detection method and the two observers were 0.902 and 0.915, respectively, with 95% confidence intervals, and the MADs were 3.08 and 2.91, respectively, indicating good reliability of the computerized automatic detection of Cobb angle.

TABLE 4. Reliability analysis of Cobb Angle manual measurement method.

		ICC/(95%CI)	MAD/(°)
Intra-observer	Observer 1	0.955	2.18
	Observer 2	0.950	2.33
Inter-observer	1 st	0.939	2.57
	2 st	0.947	2.50

TABLE 5. Consistency analysis of Cobb Angle automatic measurement and manual method.

	ICC/(95%CI)	MAD/(°)
Observer 1 & Auto	0.902	3.08
Observer 2 & Auto	0.915	2.91

VI. CONCLUSION

Considering the low efficiency and lack of rigor in traditional manual measurement methods for Cobb angle, as well as the limitations of current deep learning-based automatic measurement methods that require manual cropping of X-ray image, this paper proposes a method based on W-Net for segmenting full-spine anterior-posterior X-ray images. Firstly, we construct a W-Net model for binary segmentation of spinal X-ray images. Then, utilizing the object detection model RetinaNet, we automatically detect the region of interest of the spine, which helps filter out the interference on the binary image, replacing the manual cropping or selection of regions containing the 17 vertebral bodies for detection. Finally, we design a post-processing method to segment the adhesion between adjacent vertebrae in the binary image and employ the minimum bounding rectangle to fit each vertebral block for Cobb angle calculation.

Through experiments comparing with several typical semantic segmentation models, we find that W-Net exhibits better performance in terms of model accuracy and spinal segmentation. The experimental results also demonstrate that the post-processed vertebral segmentation images significantly reduce adhesion. Furthermore, through a consistency analysis experiment with manual measurement personnel, we validate the reliability of the proposed automatic Cobb angle measurement method in this paper.

In summary, this paper presents a complete and feasible automatic measurement method for Cobb angle, which can accurately measure the Cobb angle without interference from human factors. The introduction of this method addresses the shortcomings of traditional manual measurement methods and demonstrates its superiority in spinal image segmentation and Cobb angle measurement through experiments. Future research will further optimize and improve the performance of this method.

REFERENCES

[1] M. Konieczny, H. Senyurt, and R. Krauspe, "Epidemiology of adolescent idiopathic scoliosis," *J. Children's Orthopaedics*, vol. 7, no. 1, pp. 3–9, 2013.

[2] C. Du, J. Yu, J. Zhang, J. Jiang, H. Lai, W. Liu, Y. Liu, H. Li, and P. Wang, "Relevant areas of functioning in patients with adolescent idiopathic scoliosis on the international classification of functioning, disability and health: The patients' perspective," *J. Rehabil. Med.*, vol. 48, no. 9, pp. 806–814, 2016.

[3] C. Lang, R. Wang, Z. Chen, S. He, Q. Zou, J. Wu, and X. Zhu, "Incidence and risk factors of cardiac abnormalities in patients with idiopathic scoliosis," *World Neurosurgery*, vol. 125, pp. e824–e828, May 2019.
[4] A. Karimian, N. Rahmani, M. A. Mohseni-Bandpei, S. A. Bassampour, and A. Kiani, "Prevalence of scoliosis and associated risk factors in children and adolescents: A systematic review," *J. Mazandaran Univ. Med. Sci.*, vol. 25, no. 130, pp. 191–199, Dec. 2015.
[5] L. L. LaMontagne, J. T. Hepworth, F. Cohen, and M. H. Salisbury, "Adolescent scoliosis: Effects of corrective surgery, cognitive-behavioral interventions, and age on activity outcomes," *Appl. Nursing Res.*, vol. 17, no. 3, pp. 168–177, Aug. 2004.
[6] T. R. Kuklo, B. K. Potter, T. M. Schroeder, and M. F. O'Brien, "Comparison of manual and digital measurements in adolescent idiopathic scoliosis," *Spine*, vol. 31, no. 11, pp. 1240–1246, May 2006.
[7] E. Aubin, C. Bellefleur, J. Joncas, D. Lanauze, S. Kadoury, K. Blanke, S. Parent, and H. Labelle, "Reliability and accuracy analysis of a new semiautomatic radiographic measurement software in adult scoliosis," *Spine*, vol. 36, no. 12, pp. 90–780, 2011.
[8] R. T. Morrissy, G. S. Goldsmith, E. C. Hall, D. Kehl, and G. H. Cowie, "Measurement of the cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error.," *J. Bone Joint Surgery*, vol. 72, no. 3, pp. 320–327, Mar. 1990.
[9] D. E. Harrison, D. D. Harrison, R. Cailliet, S. J. Troyanovich, T. J. Janik, and B. Holland, "Cobb method or harrison posterior tangent method: Which to choose for lateral cervical radiographic analysis," *Spine*, vol. 25, no. 16, pp. 2072–2078, Aug. 2000.
[10] 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," *J. Digit. Imag.*, vol. 22, no. 5, pp. 463–472, Oct. 2009.
[11] P. Shrestha, A. Singh, R. Garg, I. Sarraf, T. R. Mahesh, and G. S. Madhuri, "Early stage detection of scoliosis using machine learning algorithms," in *Proc. Int. Conf. Forensics, Analytics, Big Data, Secur. (FABS)*, vol. 1, Dec. 2021, pp. 1–4.
[12] C. E. Noe, "Scoliosis in children," in *Multidisciplinary Spine Care*, 1st ed. Cham, Switzerland: Springer, 2022, p. 555.
[13] F. Bao, D. Wang, H. Zhao, and B. Xu, "Application of adaptive threshold image segmentation algorithm in orthopedic CT imaging," *J. Med. Imag. Health Informat.*, vol. 9, no. 8, pp. 1736–1740, Oct. 2019.
[14] J. Mao, K. Wang, Y. Hu, W. Sheng, and Q. Feng, "GrabCut algorithm for dental X-ray images based on full threshold segmentation," *IET Image Process.*, vol. 12, no. 12, pp. 2330–2335, Dec. 2018.
[15] L. Ţepelea, L. Gavrilut, and A. Gacsádi, "Edge based CNN image segmentation methods for medical imaging," *J. Comput. Sci. Control Syst.*, vol. 3, no. 2, pp. 95–98, 2010.
[16] Y. Ali, "Edge-based segmentation using robust evolutionary algorithm applied to medical images," *J. Signal Process. Syst.*, vol. 54, nos. 1–3, pp. 231–238, Jan. 2009.
[17] J. Mukherjee, R. Kundu, and A. Chakrabarti, "Variability of cobb angle measurement from digital X-ray image based on different de-noising techniques," *Int. J. Biomed. Eng. Technol.*, vol. 16, no. 2, p. 113, 2014.
[18] B. A. Kusuma, "Determination of spinal curvature from scoliosis X-ray images using K-means and curve fitting for early detection of scoliosis disease," in *Proc. 2nd Int. Conf. Inf. Technol., Inf. Syst. Electr. Eng. (ICITISEE)*, Nov. 2017, pp. 159–164.
[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Intervent*, Munich, Germany, 2015, pp. 234–241.
[20] Z. Tan, K. Yang, Y. Sun, B. Wu, H. Tao, Y. Hu, and J. Zhang, "An automatic scoliosis diagnosis and measurement system based on deep learning," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Bandung, Indonesia, Dec. 2018, pp. 439–443.
[21] Y. Zhao, J. Zhang, H. Li, X. Gu, Z. Li, and S. Zhang, "Automatic cobb angle measurement method based on vertebra segmentation by deep learning," *Med. Biol. Eng. Comput.*, vol. 60, no. 8, pp. 2257–2269, Aug. 2022.
[22] M.-H. Horng, C.-P. Kuok, M.-J. Fu, C.-J. Lin, and Y.-N. Sun, "Cobb angle measurement of spine from X-ray images using convolutional neural network," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–18, Feb. 2019.
[23] B. Khanal, L. Dahal, P. Adhikari, and B. Khanal, "Automatic cobb angle detection using vertebra detector and vertebra corners regression," in *Proc. 6th Int. Workshop Challenge Comput. Methods Clin. Appl. Spine Imag.*, Shenzhen, China, 2019, pp. 81–87.
[24] R. H. Alharbi, M. B. Alshaye, M. M. Alkanhal, N. M. Alharbi, M. A. Alzahrani, and O. A. Alrehaili, "Deep learning based algorithm for automatic scoliosis angle measurement," in *Proc. 3rd Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Mar. 2020, pp. 1–5.
[25] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 99, pp. 2999–3007, 2017.
[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[27] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[28] S. Q. Ren, K. M. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–11.
[29] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[30] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
[31] Z. Zhou, M. M. Siddiquee, N. Tajbakhsh, and J. M. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support, 4th Int. Workshop, DLMIA, 8th Int. Workshop, ML-CDS, Held Conjunct (MICCAI)*, Granada, Spain, 2018, pp. 3–11.
[32] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[34] S. Satoshi and K. Be, "Topological structural analysis of digitized binary images by border followingm," *Comput. Vis. Graph. Image Process.*, vol. 30, no. 1, pp. 32–46, 1985.

JIA ZHU received the B.E. degree from the Shanghai University of Engineering Science, in 2020, where she is currently pursuing the master's degree. Her research interests include the development of algorithms and software for computer vision, image processing, and measurement and control systems. Her research aims to improve the accuracy and efficiency of computer vision and image processing systems, with applications in fields, such as robotics, automation, and medical imaging.

ZHIFENG ZHOU received the bachelor's degree in mechanical engineering from the Beijing Information Science and Technology University, in 2000, the master's degree in aerospace manufacturing from the Nanjing University of Aeronautics and Astronautics, in 2003, and the Ph.D. degree in instrument science and engineering from Shanghai Jiao Tong University, in 2007. From October 2013 to September 2014, he was a Visiting Scholar with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong. Since 2007, he has been teaching with the School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, where he is currently an Associate Professor. His research interests include computer measurement and control, autonomous navigation and driving, machine vision, embedded systems, and signal processing.

CHENGXIAN YAO received the B.E. degree from the Shanghai University of Engineering Science, in 2020, where he is currently pursuing the master's degree. His research interests include computer vision, image processing, feature matching, and the application of deep learning in medical detection.