Received 13 April 2024, accepted 4 May 2024, date of publication 14 May 2024, date of current version 21 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3400803

RESEARCH ARTICLE

Cobb Angle Measurement Based on Spine Segmentation Using ATT UNet 3+

LIANG PENG[1], YIWEI HU[2], KAI ZHANG[3], GUANHUA LAN[4], RUYI ZHANG[1], DINGCHENG TIAN[1], DECHAO XU[1], YABIN ZHU[3], AND YUDONG YAO[5], (Fellow, IEEE)

[1]Institute of Intelligent Medicine and Biomedical Engineering, Ningbo University, Ningbo 315211, China
[2]School of Medicine, Ningbo University, Ningbo, Zhejiang 315211, China
[3]Affiliated Hospital of Ningbo University Medical College, Ningbo 315020, China
[4]Department of Orthopedics, Ningbo Yinzhou No.2 Hospital, Ningbo, Zhejiang 315000, China
[5]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Yudong Yao (yu-dong.yao@stevens.edu)

ABSTRACT Scoliosis refers to the abnormal curvature of human spine, which is one of the most common deformities in children and adolescents. The Cobb angle is the gold standard for quantifying the severity of scoliosis and is used to assess the severity of scoliosis. Often the accuracy of the Cobb angle measurement relies on the subjective experience of the doctor and the process is very time consuming. In this study, we propose a new deep neural network, ATT UNet 3+, based on UNet 3+. Our approach incorporates a novel hybrid attention mechanism in the network's upsampling process. This mechanism allows for the appropriate reweighting of fused multi-scale information and facilitates effective supervision of the final output results. The proposed neural network is trained, tested and validated on 155 X-ray ortho-slices. The deep learning network is compared with the more effective neural networks commonly used today. ATT UNet 3+ achieves the best performance in the segmentation evaluation results. Regarding the final Cobb angle calculations, the absolute mean error between the longest distance ellipsoidal point (LDEP) method and expert measurements amounted to 1.6°. ATT UNet 3+ provides a potential tool for segmenting the spine in X-ray, which can improve the efficiency and accuracy of doctors in processing scoliosis pathological images.

INDEX TERMS Cobb angle, deep learning, image segmentation, scoliosis, X-ray image.

I. INTRODUCTION

Scoliosis, a three-dimensional spinal deformity often emerging during adolescence, presents a significant challenge in healthcare, particularly with the prevalence of the ambiguous "adolescent idiopathic scoliosis (AIS)," accounting for up to 80% of cases [1], [2]. While mild instances may not disrupt daily life, the progressive nature of the condition can lead to severe physical deformities, impacting growth and development. Critically, scoliosis may adversely affect cardiopulmonary function, with potential for paralysis in severe cases [3]. The gold standard for diagnosis, the Cobb angle, as proposed by John Robert Cobb [4], plays a pivotal role in treatment planning, necessitating a robust and efficient measurement approach.

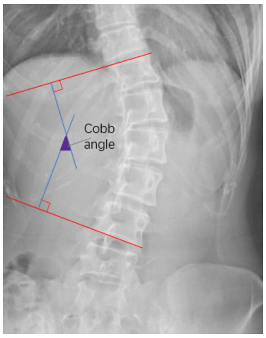


FIGURE 1. Method of measuring Cobb angle.

The existing literature reveals challenges in Cobb angle measurement, relying on manual methods that are time-consuming and prone to subjectivity. Previous studies

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

highlight errors of up to 11.8° in Cobb angle measurement [5], underscoring the need for advanced, automated techniques. Addressing this issue, our work aims to develop a deep learning network, ATT UNet 3+, incorporating a novel attention mechanism for accurate segmentation of spinal X-ray images. Additionally, we introduce the longest distance ellipsoidal point (LDEP) method, enhancing computational efficiency in Cobb angle calculation. Through comparative analysis with established neural networks like PSPNet and UNet, ATT UNet 3+ emerges as a superior model, showcasing its efficacy in scoliosis evaluation.

The contributions of this paper are threefold:

(1) The introduction of ATT UNet 3+, a deep learning model with an innovative attention mechanism, ensuring precise segmentation for Cobb angle determination.

(2) The proposal of the LDEP method, offering an efficient and accurate approach to Cobb angle calculation, thereby enhancing computational efficiency.

(3) Comparative analysis demonstrating the superiority of ATT UNet 3+ over existing models like PSPNet and UNet, validating its potential as a transformative tool in scoliosis assessment.

The remaining of this paper is as follows. Section II describes spine segmentation and attention-based related work. Section III presents the proposed network and experimental methods. The experimental results and discussions are presented in section IV. Section V discuss the feasibility of the method. Finally, section VI concludes the paper.

II. RELATED WORKS

A. SPINE SEGMENTATION BASED ON CONVENTIONAL MACHINE LEARNING

In the field of spine segmentation, the integration of machine learning and digital image processing techniques holds immense promise for advancing medical diagnosis. Andre Mastmeyer et al. [6] uses a deformation model to constrain the primary shape of the vertebral body and then use the area growth method to detect the fine surface of the vertebral soft tissue junction. By constructing an energy generalization, the contour profile is gradually approached towards the edge of the object to be detected, driven by the minimum value of the energy function, and the target is finally segmented. Georg et al. [7] combined active contours and the Chan-Vese [8] intensity model into a level set algorithm to achieve low-quality, diseased MRI image segmentation of the spine. Active contours are often time consuming. Mukherjee et al. [9] selected the best filter among the four denoising techniques: bilateral filters [10], nonlocal means filters [11], principal neighborhood dictionaries nonlocal means filtering [12], and block matching three-dimensional filtering [13]. Due to the poor contrast of radiographs, histogram equalization was applied to enhance image contrast, and the Otsu thresholding method was used to find the Canny edge points of vertebrae. Finally, the two straight lines between the upper and lower endplates of each vertebra are detected using the Hough transform, and the one with the largest angle is compared as the Cobb angle. However, these methods require complicated image processing stages that involve image filtering, enhancement, segmentation, and feature extraction to obtain vertebra assessment, which make the techniques computationally expensive and temptable to errors caused by the variations in X-ray spinal images.

B. DEEP LEARNING BASED SPINE SEGMENTATION

With the development of deep learning, some studies have shown the feasibility of neural network in the field of medical imaging analysis of spinal patients. Wu et al. [14] proposed a new Multi-View Correlation Network (MVCNet) architecture to estimate the Cobb angle through the four angles of the vertebral body in the anteroposterior and lateral X-ray films of the spine marked by the neural networks. Zhang et al. [15] proposed a framework called MPFNet, which combines the vertebrae detection branch based on the backbone convolutional neural network and the key point prediction branch, which can provide bounded regions for key point prediction. A correlation module was proposed to exploit the information between adjacent vertebrae so that vertebrae hidden by the thorax and arm could be found on lateral radiographs, and finally the Cobb angle was measured by combining both orthogonal and lateral components. Horng et al. [16] used the deep convolution neural network (CNN) method, including UNet, Dense UNet and Residual UNet, to segment each vertebral body of the spine. The segmentation results are then reconstructed into a complete segmented spine image, and the angle between each vertebral body is calculated based on the Cobb angle criterion, and the maximum angle is the measured Cobb angle. Khanal et al. [17] proposed an automatic method, which first detects the vertebrae as the object, and then the marker detector, which estimates the four marker angles of each vertebrae respectively. The Cobb angle was calculated using the inclination of each vertebra obtained from the predicted markers. A recent review by Azimi et al. [18] comprehensively summarizes some neural network studies, which focus on automatic Cobb angle estimation and confirm the importance of these tools in improving clinical practice. However, the existing research has some limitations. For example, some algorithms still need a certain amount of manual intervention, such as the allocation of vertebral plaques. In addition, due to the heterogeneity of dataset, methods and outcome indicators, the results of different studies are often not comparable. This paper proposes a new network (ATT UNet 3+), which addresses some of the above issues.

C. ATTENTION-BASED MEDICAL IMAGE SEGMENTATION

The attentional mechanism operates as an adaptive selection process grounded in input features. Its primary function involves honing in on detailed information pertinent to the specified target, while concurrently negating extraneous details through emulation of the visual observation process. Its significance is on the rise within the realm of computer vision. References [23] and [35] use spatial attention for image classification and target detection. Yuan et al. [36] characterizes each pixel by enhancing the understanding of the pixel context.

Attentional mechanisms have gained extensive application in the domain of medical image segmentation. Ahmad et al. [37] introduces a nimble fusion-attentional decoder mechanism, augmenting the precision of tumor segmentation. Li et al. [38] employs an attentional mechanism to identify global contextual information in three dimensions simultaneously: the channel domain, the spatial domain, and the feature internal domain, aiming to capture more representative features. CBAM [32] places emphasis on discerning spatial and channel-relevant features, thereby amplifying the prominence of key regions linked to the targeted feature representation.

Feature fusion is a key operation in deep learning, allowing the comprehensive integration of different levels of information to expand the feature representation and improve model performance. Building on established methods, we use dilation convolution at three different rates to achieve different receptive fields. Convolution kernels with different receptive fields facilitate the extraction of feature information at multiple scales after convolution, allowing the network to effectively manage spatial hierarchies by capturing a wider range of information. The subsequent fusion involves the enhancement of contours and spatial feature information derived from different feature maps through a refined channel attention mechanism. This attention mechanism enables our network to emphasise the importance of selected features.

III. MATERIALS AND METHODS

A. DATASETS

The experimental datasets consisted of 155 AP view spinal X-ray images, all obtained from Ningbo Yinzhou Second Hospital. The width of each image ranged from 359 to 1386 pixels, the height from 973 to 2687 pixels, and the file size from 490KB to 817KB. Each X-ray shows the upper part of a complete human body, including the 12 thoracic and 5 lumbar vertebrae in the entire spine to be used as the main target for the training and segmentation tasks. To use these images in the deep learning framework, we resized all images to a uniform resolution of 960 × 448 before feeding them into the network. The labeled data were produced under the guidance of experienced radiologists.

B. DATA PREPROCESSING

In the analysis of spinal X-ray images acquired through the anteroposterior (AP) view, which contains a significant amount of irrelevant information, a preprocessing step is necessary before inputting the data into the neural network. To enhance processing efficiency, we initially resize all AP view spinal X-ray images, focusing exclusively on the T1-L5 region between the thoracic and lumbar spine. Brighter pixels in the images signify the presence of bones, with large bone structures, including the head, spine, and hip, displaying vertical alignment. Grey-scale intensity histograms for both vertical and horizontal projections are computed, and the Region of Interest (ROI) is defined by selecting columns within the mean intensity plus/minus one standard deviation range. Notably, the cervical spine region exhibits relatively low intensity, while the lumbar spine region appears brighter in the X-ray images. Therefore, we used the intensity histogram of the horizontal projection to determine the minimum extreme as the upper boundary of the ROI and the location of the maximum discontinuity position as the lower boundary [16]. This data preprocessing process is shown in Figure 2.

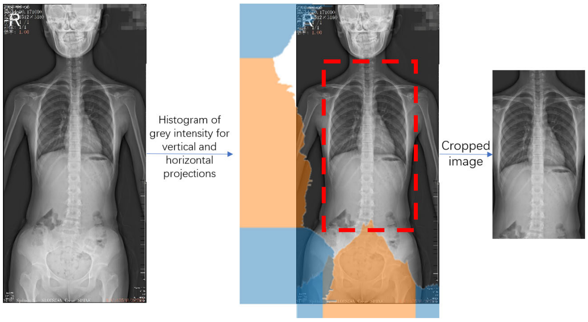

Histogram of grey intensity for vertical and horizontal projections
Cropped image

FIGURE 2. Spine X-ray image data preprocessing process.

C. PROPOSED MODEL ATT UNET 3+

In many segmentation studies, feature maps at different scales demonstrate different information. Low-level feature maps capture features such as color, texture and shape of objects; while high-level feature maps reflect the attribute features of objects. The multi-scale feature approach has achieved excellent results in many excellent deep learning networks in recent years, such as PSPNet [19], Deeplab [20], etc. For the case of complex X-ray images, we propose a new model (ATT UNet 3+), which incorporates the attention module into UNet 3+ network. In multiple downsampling, the information of low-level feature maps for segmented objects is gradually diminished. By multi-scale feature fusion, the low-level and high-level feature maps can be better combined in segmentation to distinguish the boundaries of organs. The attention mechanism was proposed as a natural language processing application and achieved very good results [21]. In the field of deep learning, attention has also become an important part of the neural network structure and has a large number of applications in statistical learning, speech and image [32], [37]. The network structure of our proposed ATT UNet 3+ is shown in Figure 3.

1) FULL SCALE CONNECTION

ATT UNet 3+ combines all scales in the decoder. For example, As shown in Figure 4 X_{Dn}^{3} is obtained from

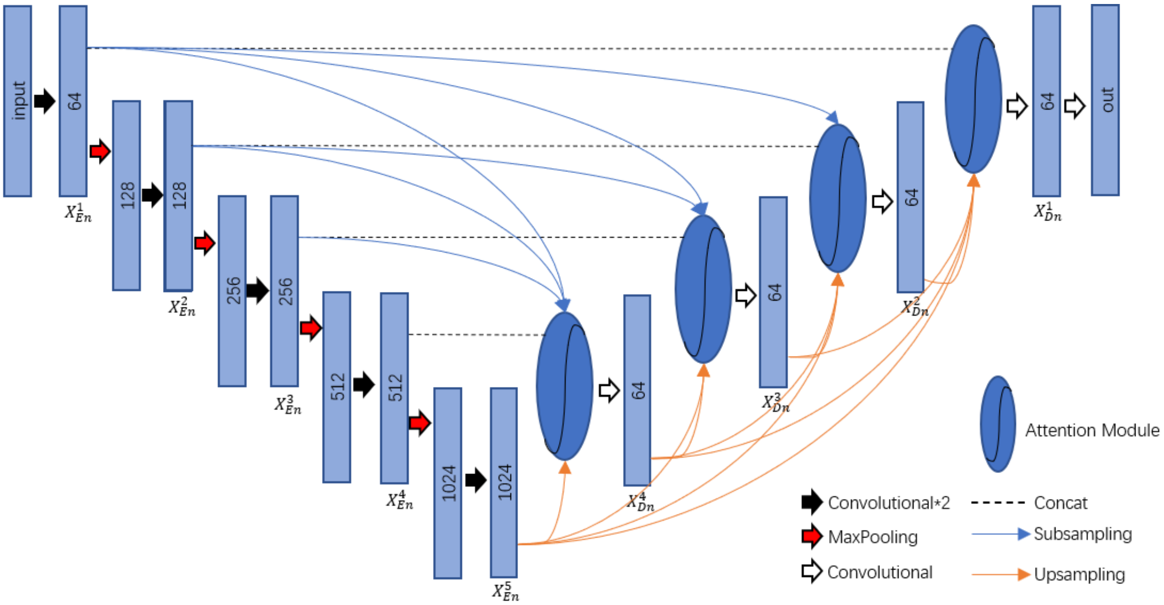


FIGURE 3. Overall architecture of the proposed network for spine X-ray image segmentation.

the low-level feature maps X_{En}^1 and X_{En}^2, the high-level feature maps X_{Dn}^4 and X_{En}^5 and the same-level feature map X_{En}^3 through the attention mechanism module. X_{En}^1 and X_{En}^2 retain fine-grained semantic information, which is equivalent to the low-level feature information fused in the decoder. X_{En}^1 and X_{En}^2 will be subsampled first. The subsampling reduced dimensionality is 4 times and 2 times respectively, to unify the dimensionality of the feature maps. The coarse-grained semantic information in the Figure 4 is reflected in X_{Dn}^4 and X_{En}^5 in the high-level feature maps. Bilinear interpolation upsampling with magnification dimension multiples of 2 and 4 times is performed for X_{Dn}^4 and X_{En}^5, respectively. After that, all four parts are subject to convolutional-2D operations with 3 × 3 with filters channels of 64. The encoder feature maps of the same scale X_{En}^3 are directly subjected to convolution operations.

2) ATTENTION MECHANISM

Attentional mechanisms can be intuitively explained in terms of human visual mechanisms, as our visual system tends to emphasize the relevant parts of an image and disregard irrelevant information. Similarly, in visual tasks, certain areas of the input may be more critical for decision-making than others. To address this, we propose an improved channel attention mechanism that highlights salient features in fully connected networks that are more useful for the final decision. This module is inspired by the hybrid attention mechanism known as CBAM [32], which effectively incorporates both channel and spatial attention mechanisms. However, our approach replaces the spatial attention mechanism with a null convolution and reverses the order of the tandem spatial and channel attention modules. By computing attention weights for each feature map channel, the network can adaptively focus on important features, enabling better utilization of input information and improving the final prediction performance.

The upsampling X_{Dn}^3 as shown in Figure 4 is used as an example. Before obtaining X_{Dn}^3, the low-level and high-level feature maps are first unified in dimension using the maximum pooling layer and bilinear interpolation, respectively. The X_{En}^1, X_{En}^2, X_{En}^3, X_{Dn}^4 and X_{En}^5 dimensions are unified and then concatenated together to fuse the features by 1 × 1 convolution and reduce the channel dimensionality. Then, three parallel 3 × 3 dilated convolutional layers with $D = 1$, $D = 3$ and $D = 5$ dilation rates are used to fuse information around a single neuron while maintaining local details [22]. The outputs of the three 3 × 3 dilated convolutional layers are concatenated and the feature maps are fed into the subsequent channel attention mechanism.

The concatenated features are pooled by adaptive average pooling and adaptive maximum pooling, respectively. The feature dimension obtained by two pooling layers is $1 \times 1 \times C$, where C is the number of channels. After adding the two feature map together, a convolution of 1 × 1 × 64 and a convolution of 1 × 1 × 192 is applied, followed by a Sigmoid function to obtain the value. The result is then multiplied with the features prior to the input channel being recorded [23]. This process is equivalent to reassessing the importance of each feature channel after feature selection. The improved channel attention mechanism is calculated as follows:

$$\begin{aligned} F_R &= \sigma(MLP(AvgPool(x) + MaxPool(x))) \\ &= \sigma(W(x_{avg} + x_{max})) \qquad (1) \\ \check{R} &= R \times F_R \qquad (2) \end{aligned}$$

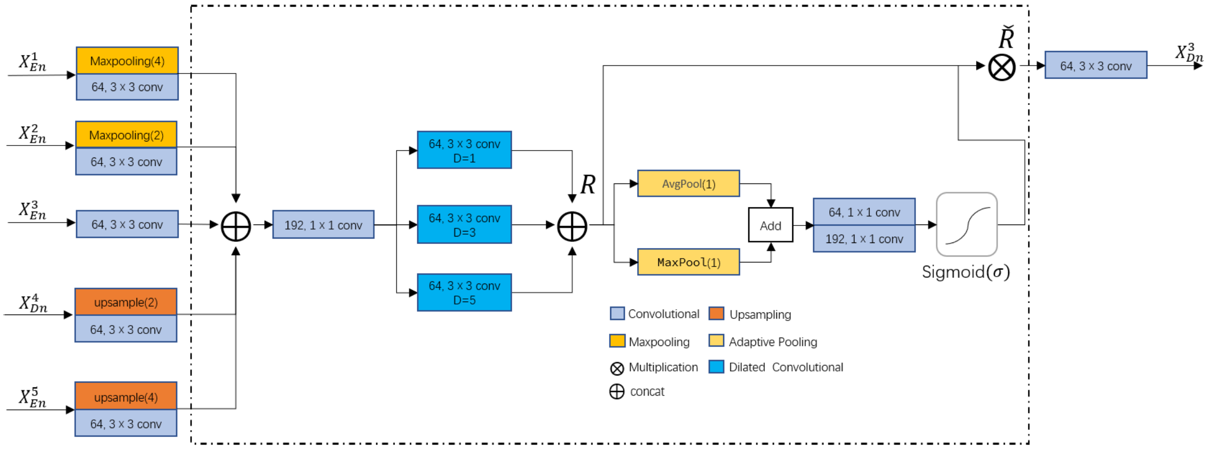


FIGURE 4. The specific calculation process of X^3_{En}, in which the Dashed line represents the attention mechanism.

where, σ refers to the *Sigmoid* function, F_R represents the parameter obtained after R is input to the attention mechanism in the model, and $\check{R}$ denotes the final output feature map.

By utilizing dilated convolution combined with different receptive fields, we can more effectively differentiate between background and spine without changing the output size. Through the improved channel attention mechanism, the features of the channel dimension and the importance of each channel are learned without changing the dimension.

D. MEASUREMENT OF COBB ANGLE USING THE LDEP METHOD BASED ON SEGMENTATION RESULTS

The initial phase involves segmenting the spine through the utilization of a neural network model. Subsequently, the determination of contour points for each vertebra follows, where an ellipse is computed to approximate these points using the equation 3 ellipse fitting algorithm [28]. The calculated ellipse formula is then applied to the contour points, yielding those with positive calculated values. The first vertex is chosen as the point with the largest calculated value. Following this, the points proximate to the first vertex, including the vertex itself, are eliminated, resulting in a refined point set. This iterative process is replicated to derive the four key points, as depicted in Figure 5.

$$A(x - x_0) + B(x - x_0)(y - y_0) + C(y - y_0) + D = 0 \quad (3)$$

Where, x_0 and y_0 represent the positions of the center of the ellipse, while A, B, C, and D denote the parameters of the ellipse.

Upon identifying the four key points, the upper and lower boundaries of the vertebrae are determined using these points, and the angle between the x-axis and the upper and lower boundaries is calculated. Finally, the Cobb angle is obtained by subtracting the minimum angle from the maximum angle.

Additionally, we utilized the semi-automatic polynomial fitting algorithm proposed by Papaliodist to find the Cobb angle and compare it with our method using the identified circle centers [29]. This algorithm fits a polynomial curve through the marked vertebral points, which are manually marked by selecting a point on each vertebra. The curvature of the spine is reflected by the spine curve, indicating the degree and direction of curvature of a set of points at the center of the spine. The tangent of this curve forms an angle with the vertical and varies continuously along the height of the spine, with the difference between the maximum and minimum values representing the Cobb angle measured by this method. The flow of the whole method is shown in Figure 6.

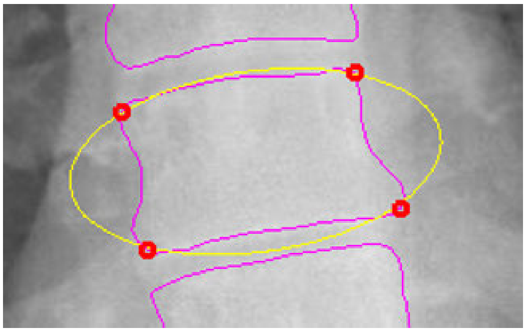


FIGURE 5. The red circles are the four points found by the LDEP method.

IV. EXPERIMENTS AND RESULTS

The software used for the experiments is implemented using pytorch1.7. The experimental hardware configuration uses an RTX 3090 GPU with 24 GB of memory. We train the model using Adaptive Moment Estimation (Adam) with a batch size of 2 and a learning rate of 0.001. All 155 images, 124 of which were used as training set and the remaining 31 images were used as test set images. The dataset of the training set was expanded to 372 sheets through data augmentation techniques including random deflections of up to 10° to the left and right, horizontal flips, and vertical flips.

The performance of the proposed deep learning network and the compared networks is evaluated using five

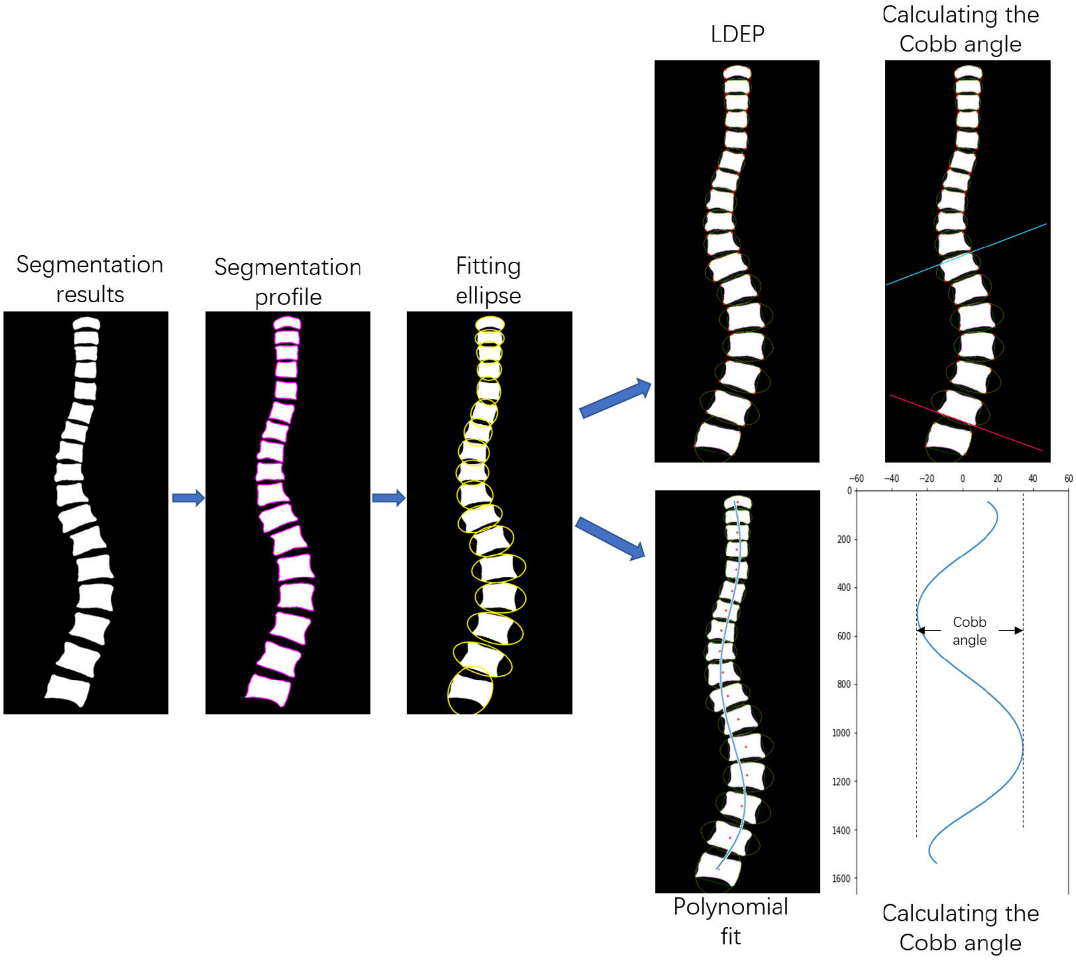


FIGURE 6. Flow of Cobb angle computation that resides in the segmentation result, the LDEP method comprises the upper branch, while the lower branch is dedicated to the function fitting method.

well-known classification metrics: mIoU, Dice, accuracy, precision, and recall.

Accuracy represents the proportion of correctly classified instances out of the total instances, measuring overall model performance in classification tasks.

Precision measures the proportion of true positive predictions among all positive predictions made by the model, indicating the accuracy of positive predictions.

Recall, also known as sensitivity, measures the proportion of true positive instances that were correctly identified by the model, indicating the model's ability to capture all positive instances.

mIoU (mean Intersection over Union) measures the average overlap between predicted and ground truth segmentation masks across all classes in semantic segmentation tasks.

The Dice coefficient quantifies the similarity between predicted and ground truth segmentation masks. Higher values indicate better segmentation accuracy.

True Positive (TP) represents instances that are correctly classified as positive, while False Positive (FP) represents instances that are incorrectly classified as positive. True Negative (TN) represents instances that are correctly classified as negative, and False Negative (FN) represents instances that are incorrectly classified as negative.These metrics can be calculated using the following formulas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{Dice} = \frac{2{*}\text{TP}}{2{*}\text{TP} + \text{FP} + \text{FN}} \tag{7}$$

$$\text{mIoU} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FN} + \text{FP}}\right) \tag{8}$$

These five well-known classification metrics were evaluated to quantify the performance of the proposed deep learning network.

A. VERTEBRAE SEGMENTATION RESULTS

Verify the effect of different dilated convolutions on the results. The results are summarized in Table 1, where the numbers in the first column indicate different dilation rates,

TABLE 1. Segmentation performance metrics for different dilation rates, with the numerical values in the initial column representing distinct dilation rates.

Dilation Rates	Accuracy (%)	Precision (%)	Recall (%)	Dice (%)	mIoU (%)
1,2,3	93.34±0.846	94.43±0.647	91.33±0.642	87.65±0.816	84.37±0.631
1,2,4	93.74±0.810	94.56±0.745	91.77±0.732	87.94±0.732	84.53±0.715
1,2,5	93.09±0.794	94.97±0.672	91.84±0.753	88.61±0.738	85.31±0.623
1,3,4	94.59±0.753	95.27±0.783	92.62±0.704	89.32±0.756	85.98±0.593
1,3,5	95.11±0.716	96.11±0.620	92.37±0.761	89.51±0.661	86.21±0.571
1,4,5	94.87±0.703	96.72±0.648	91.82±0.842	88.76±0.731	84.84±0.596
1,4,6	94.52±0.751	96.14±0.681	92.04±0.746	88.69±0.667	84.51±0.671

TABLE 2. Evaluation of Dice coefficients across different modules employing UNet as the foundational model.

	UNet	Channels to 64	Full scale connection	Attentional mechanisms	Dice(%)
Experiment 1	✓				84.27±0.820
Experiment 2	✓	✓			83.79±0.934
Experiment 3	✓	✓		✓	85.00±0.707
Experiment 4	✓	✓	✓		87.68±0.856
Experiment 5	✓	✓	✓	✓	89.51±0.661

respectively. To ensure accurate details of the segmentation results, we always keep the dilation convolution with a dilation rate of 1. As the dilation rate increases, the three dilated convolutions at dilation rates of $D = 1$, $D = 3$ and $D = 5$ are the best performers in Accuracy, Dice, and mIoU.

To confirm the effectiveness of the modules, we evaluate the performance of different modules in UNet. The results are shown in Table 2. The full scale connections, the attention mechanism modules and the channels 64 upsampled outputs are added to the network respectively. Dice coefficients were used to evaluate the effectiveness of these blocks. The model of Experiment 1 is the original UNet. In Experiment 2, Experiment 3, Experiment 4 and Experiment 5, the output of upsampling is unified into a convolutional layer with channel 64. The difference between Experiment 2 and Experiment 3 is whether the attention mechanism is added to the upsampling. The model for Experiment 4 is UNet 3+. Experiment 5 is our proposed method, and we use the same comparison strategy between Experiment 4 and Experiment 5 as in Experiment 2 and Experiment 3. Both Experiment 5 and Experiment 3 are improved compared to the case with no added attention mechanism. The difference between Experiment 2 and Experiment 4 is whether full scale connectivity is added or not. Compared with the original UNet, the Dice coefficient of our proposed method is improved by 5.24%.

Table 3 shows the five metrics of ATT UNet 3+, UNet 3+ [30], UNet++ [24], UNet [25], Deeplabv3+ [26], SegNet [27], PSPNet [20], UNeXt [33], CBAM+Ref-UNet 3+ [34], EANet [40] and TransDeepLab [39]. In the segmentation of ATT UNet 3+ mIoU performance is 86.21%, which achieves different levels of improvement compared with several other segmentation networks. Our model performs another improved model based on UNet 3+ on the majority of segmentation metrics for this dataset.

We compare the parameters and FLOPs of different image segmentation models to assess their complexity. FLOPs measure the computational load required for model inference. As shown in Table 4, our ATT UNet 3+ model achieves parameter and FLOP scores of 25.17M and 943.51G, respectively. These outcomes indicate that the ATT UNet 3+ model exhibits relatively high FLOPs, yet it maintains a notably lower number of parameters compared to several other models. Importantly, the model consistently outperforms its counterparts in overall segmentation performance.

In Dice and mIOU, ATT UNet 3+ performs the best among all compared neural networks. When the parameters were examined after training the model, it became clear that the enhanced channel attention mechanism skilfully suppressed feature maps that inadequately captured the intricacies of the spinal contours. Figure 7 shows feature maps R of X^4_{Dn}, X^3_{Dn} and X^1_{Dn}, where R is formed by concatenating feature maps using dilation convolution. The red boxes highlight feature maps where R gains prominence due to the enhanced channel attention mechanism, effectively representing the spine contour. It's worth noting that there are 192 of these feature maps, but due to space constraints, only 16 representative ones are shown.

Figure 8 shows some segmented images. It is worth noting that the accuracy of the ATT UNet 3+ segmentation proposed in this paper is higher than other networks in thoracic T1-T3 and lumbar L5, but there are still some artifacts and imperfect segmentation. The figure shows representative images of the vertebral body segments superimposed onto the original x-ray images. The automatic segmentation of the upper and lower boundaries of the vertebral body using ATT UNet 3+ is highly consistent with the original vertebral body and performs well.

B. PERFORMANCE EVALUATION OF ANGLE MEASUREMENT

To evaluate Cobb measurements, we compare Cobb angles measured manually by an orthopedic surgeon with Cobb angles obtained based on our automated computer

TABLE 3. Comparisons of segmentation performance metrics between the proposed ATT UNet 3+ and different methods.

Model	Accuracy (%)	Precision (%)	Recall (%)	Dice (%)	mIoU (%)
Deeplab v3+	87.03±1.234	85.55±1.467	70.53±1.694	78.41±1.438	70.47±1.549
UNeXt	88.34±1.242	87.68±1.116	82.13±1.040	81.45±0.938	75.88±0.938
PSPNet	88.18±1.164	89.42±1.431	84.75±1.291	82.47±1.311	79.05±1.247
TransDeepLab	88.45±1.385	88.27±1.470	86.18±1.291	82.72±1.130	77.84±1.338
UNet	89.26±0.914	88.48±1.028	84.27±0.974	84.27±0.820	77.57±0.951
SegNet	91.74±1.242	89.04±1.156	87.44±1.340	85.08±0.938	78.67±1.137
UNet++	92.87±1.017	92.21±0.927	88.14±0.931	86.54±0.909	80.57±1.062
CBAM+Ref-UNet 3+	92.59±0.891	96.21±1.131	90.21±1.247	86.95±1.279	82.95±1.092
UNet 3+	93.05±0.717	94.02±0.921	90.51±0.842	87.68±0.856	84.04±0.810
EANet	93.84±0.953	94.37±1.152	91.21±1.017	88.14±1.089	85.13±1.126
ATT UNet 3+ (proposed)	95.11±0.716	96.11±0.620	92.37±0.761	89.51±0.661	86.21±0.571

TABLE 4. Shows a comparison of Params and FLOPs on different methods, where FLOPs is entered by a uniform input size.

Model	Params (M)	FLOPs (G)	Dice (%)
Deeplab v3+	43.87	97.12	78.41±1.438
PSPNet	65.70	129.11	82.47±1.311
TransDeepLab	44.11	319.57	82.72±1.130
UNet	31.04	359.22	84.27±0.820
SegNet	29.44	263.36	85.08±0.938
UNet++	47.18	1310.15	86.54±0.909
CBAM+Ref-UNet 3+	23.65	876.17	86.95±1.279
UNet 3+	26.97	1310.80	87.68±0.856
EANet	49.36	1031.17	88.14±1.089
ATT UNet 3+ (proposed)	25.17	943.51	89.51±0.661

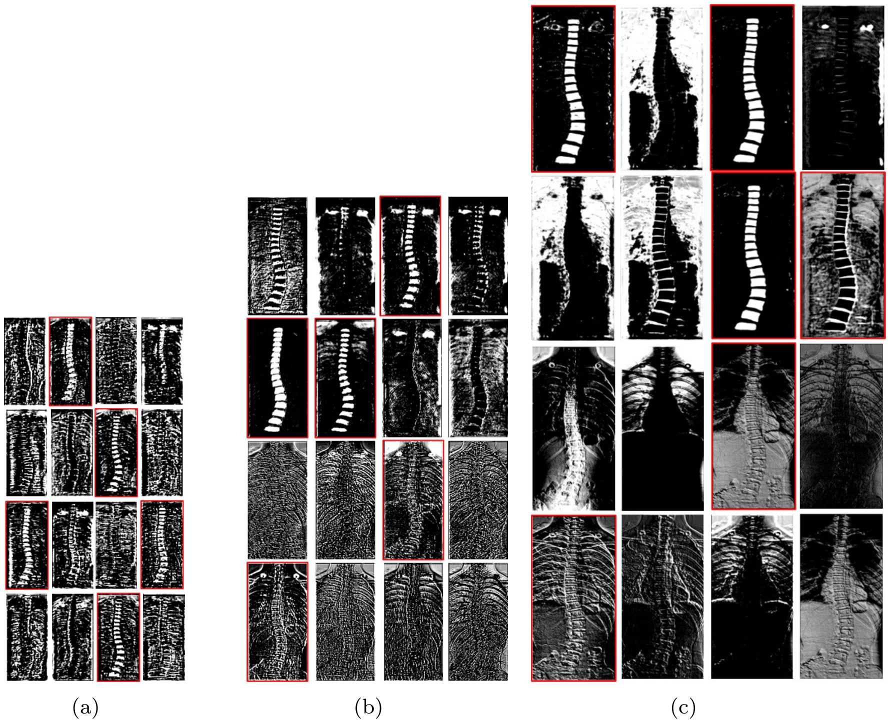

(a) (b) (c)

FIGURE 7. Some features maps after passing through the attention mechanism, with red boxes indicating higher weights assigned to certain areas. (a) represents the feature map of X_{Dn}^{4}, (b) represents the feature map of X_{Dn}^{3}, and (c) represents the feature map of X_{Dn}^{2}.

measurements. The measurements are calculated using the segmentation results of the 31 test images in the test set. The stability of the Cobb angle calculation results is assessed by calculating the mean absolute error (MAE), standard deviation (SD), and Pearson correlation coefficient between the manual and automated measurements. The Pearson correlation coefficient is employed to quantify the correlation between measured and true values [31]. We used the segmentation results of ATT UNet 3+ and UNet 3+ to calculate the Cobb angle.

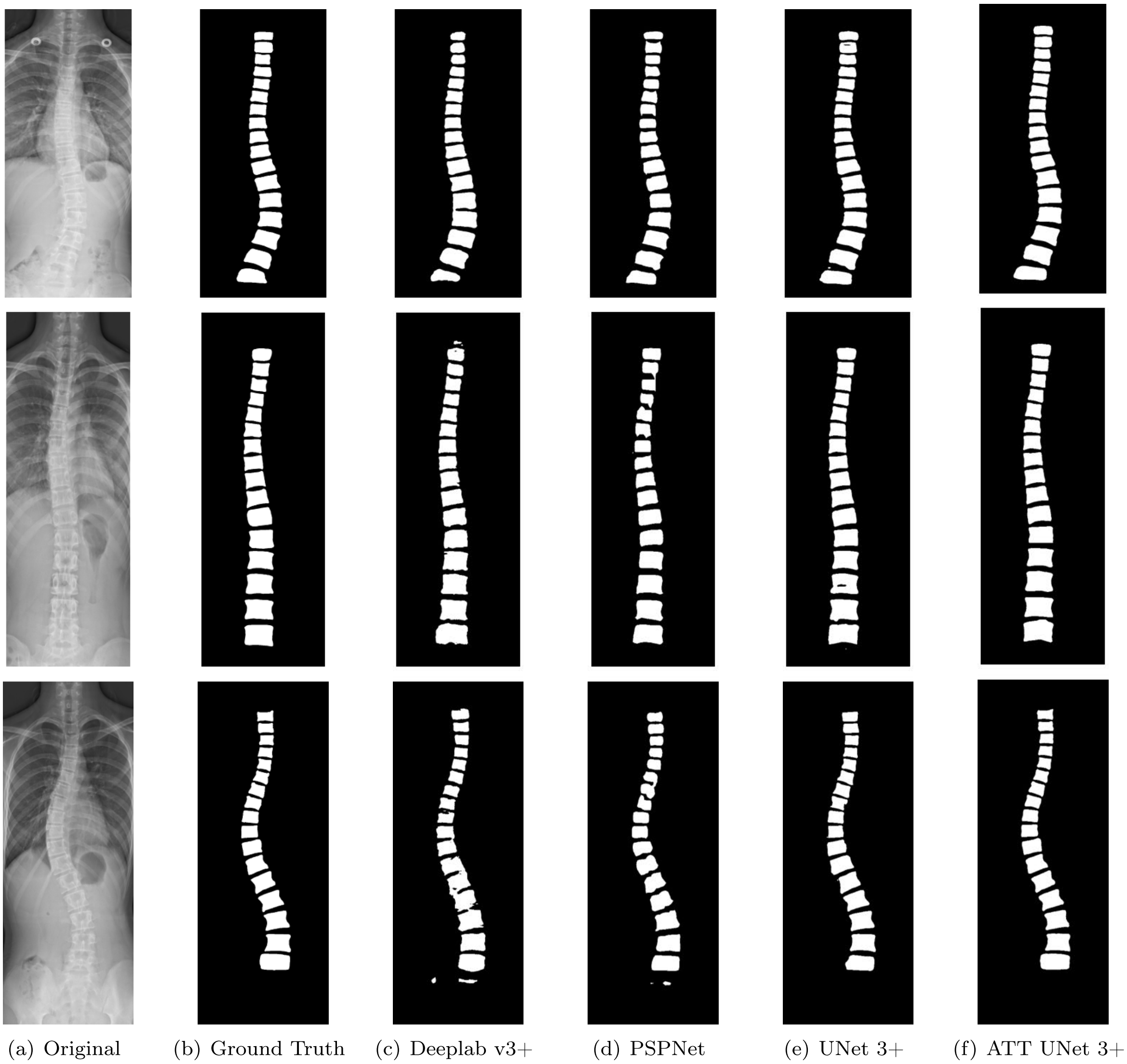


FIGURE 8. Segmentation results of the dataset on different methods. (a) original image, (b) ground truth, (c) Deeplab v3, (d) PSPNet, (e) UNet 3+, (f) ATT UNet 3+.

As shown in Table 5, Cobb angle is calculated for vertebral segmentation results. The error statistics show that the mean absolute error and standard deviation of the ATT UNet 3+ and LDEP method with the orthopedist are 1.6° and 0.8°, which are the smallest compared to other methods. The Pearson correlation coefficient is 0.983. This indicates that the overall performance of our proposed method for automatic measurement of the Cobb's angle is stable with low bias. The results are most correlated with the manual measurements by the orthopedic surgeon. Although the MAE and SD of the UNet 3+ and LDEP measurements of the Cobb angle are lower than those measured using the polynomial function method, the Pearson coefficients show less correlation than the two mentioned above. This indicates that the LDEP method has more stable results when the image segmentation is good.

V. DISCUSSION

We introduce a novel method for the automated assessment of the Cobb angle in patients with Adolescent Idiopathic Scoliosis (AIS) employing the ATT UNet 3+ network. Operating within an encoder-decoder framework, the network adeptly delineates the contours of individual vertebrae. Subsequently, the Local Directional Edge Profile (LDEP) method is applied to pinpoint the four key points defining these contours. The culmination of the process involves the computation of the maximum Cobb angle for both the upper and lower extremities of each vertebra, providing a comprehensive quantitative evaluation of spinal deformities.

Figure 9 shows a boxplot comparing the segmentation coefficients of spine X-ray image data using seven different segmentation algorithms. The variation for a given data set is illustrated. The blue dashed line is the mean value of

TABLE 5. Error of Cobb angle based on segmentation results of ATT UNet 3+ and UNet 3+ using function fitting and LDEP methods.

	UNet 3+		ATT UNet 3+	
	Polynomial function	LDEP	Polynomial function	LDEP
MAE	12.4°	9.1°	10.0°	1.6°
SD	10.7°	9.2°	9.1°	0.8°
Pearson correlation coefficient	0.828	0.632	0.854	0.983

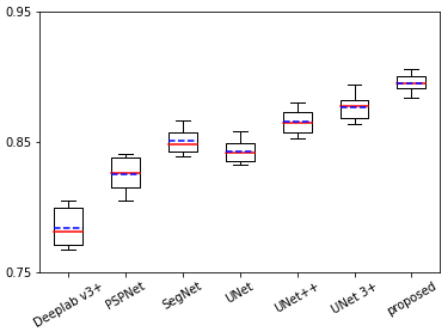
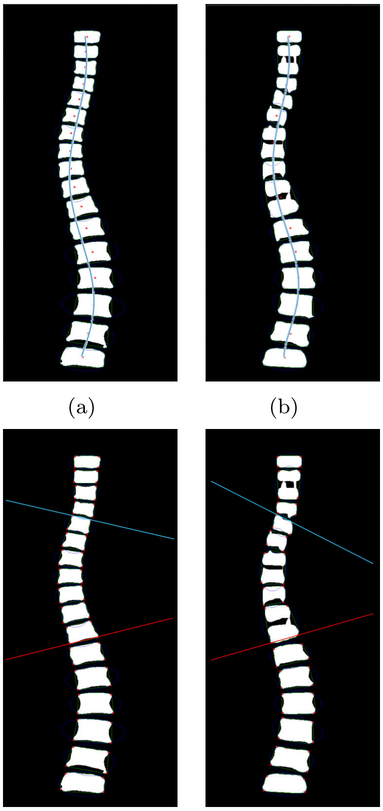


FIGURE 9. The boxplot segmentation result of Deeplab V3+, PSPNet, SegNet, UNet, UNet++, UNet 3+ and our model on Dice coefficient.



(a) (b)

(c) (d)

FIGURE 10. Visualisation of two Cobb angle calculation methods, (a) good segmentation results and function fitting method, (b) bad segmentation results and function fitting method, (c) good segmentation results and LDEP method and (d) bod segmentation results and LDEP methods.

the data coefficients in the boxplot and the red line is the median value of the data coefficients. As can be seen from the graph, our model outperforms the other models. The results show that the proposed segmentation network is more robust than the existing segmentation network models. In terms of other segmentation results, ATT UNet 3+ also shows the best results in terms of segmentation accuracy, precision, Recall and mIoU indices (Table 3). There was a strong correlation (0.983) between the Cobb angle measured by the proposed LDEP and the orthopaedic surgeon. The mean absolute error of angle was 1.6°.

Figure 10 shows the polynomial function fit and the LDEP result with good and poor segmentation effects for measuring the Cobb angle. Figure 10(a)(b) shows the polynomial function fitting method, where the red points are the centre of the fitted ellipse and the blue line fits the curve. Figure 10(c)(d) shows the LDEP method, where the red dots are the four points furthest from the ellipse from which LDEP is derived, and the blue and red lines are the Cobb angles of the most curved vertebrae.

In Figure 10(a)(b), the difference between good and poor segmentation results for fitting the spine curve with a polynomial function is not significant. This is because the polynomial fitting method only requires finding the position of the fitted ellipse centre to fit the spine curve, but the error in angle is larger (Table 5). As shown in Figure 10(c)(d), good segmentation results in a better measurement of the Cobb angle. Poor segmentation results lead to larger deflection angles due to adhesions of the segmented vertebrae and over-segmentation, which affects the measurement results, indicating that good segmentation results are important for the LDEP method to measure the Cobb angle. Combined with the Pearson correlation coefficients in Table 5, we found that the correlation coefficient of the polynomial fit (0.828) is more relevant to the angle measured by the orthopaedic surgeon when calculating the Cobb angle using poor segmentation results than the correlation coefficient of the LDEP (0.632). When segmentation results are poor, a polynomial fit may be more appropriate to measure Cobb angle.

VI. CONCLUSION

This paper proposes an automatic Cobb angle measurement method based on ATT UNet 3+ for surgical planning of adolescent idiopathic scoliosis. Compared with other deep learning algorithms, the method proposed in this paper performs accurate segmentation of the vertebral body on the AP view spinal X-ray images, and the Cobb angle calculation is in good agreement with orthopaedic specialist. The proposed method provides a potential tool to realize the automatic estimation of Cobb angle and improve the efficiency and accuracy of doctors' diagnosing of scoliosis.

However, this study has some limitations. The method proposed in this paper was applied only to the frontal images that capture the curvature of AIS patients. Recent studies have shown that sagittal image plays an important role in the clinical outcome of AIS and cannot be ignored. Our study was limited by the data set, which did not include both frontal and sagittal spine images. Nevertheless, our method could be applied to image segmentation in both planes to provide additional information for surgical planning.

REFERENCES

[1] J. Dunn, N. B. Henrikson, C. C. Morrison, P. R. Blasi, M. Nguyen, and J. S. Lin, "Screening for adolescent idiopathic scoliosis: Evidence report and systematic review for the US preventive services task force," *Jama*, vol. 319, no. 2, pp. 173–187, 2018.

[2] M. Fadzan and J. Bettany-Saltikov, "Etiological theories of adolescent idiopathic scoliosis: Past and present," *Open Orthopaedics J.*, vol. 11, no. 1, pp. 1466–1489, Dec. 2017.

[3] F. Altaf, A. Gibson, Z. Dannawi, and H. Noordeen, "Adolescent idiopathic scoliosis," *BMJ*, vol. 346, pp. 2508–2508, Apr. 2013.

[4] J. R. Cobb, "Outline for the study of scoliosis," *Instructional Course Lectures*, vol. 5, pp. 261–275, Jan. 1948.

[5] J. E. H. Pruijs, M. A. P. E. Hageman, W. Keessen, R. van der Meer, and J. C. van Wieringen, "Variation in cobb angle measurements in scoliosis," *Skeletal Radiol.*, vol. 23, no. 7, pp. 517–520, Oct. 1994.

[6] A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender, "A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine," *Med. Image Anal.*, vol. 10, no. 4, pp. 560–577, Aug. 2006.

[7] G. Hille, S. Glaßer, and K. Tönnies, "Hybrid level-sets for vertebral body segmentation in clinical spine MRI," *Proc. Comput. Sci.*, vol. 90, pp. 22–27, Jan. 2016.

[8] P. Getreuer, "Chan–Vese segmentation," *Image Process. Line*, vol. 2, pp. 214–224, Jan. 2012.

[9] 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.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.

[11] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 60–65.

[12] T. Tasdizen, "Principal neighborhood dictionaries for nonlocal means image denoising," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2649–2660, Dec. 2009.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE*, vol. 6064, pp. 354–365, Feb. 2006.

[14] H. Wu, C. Bailey, P. Rasoulinejad, and S. Li, "Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-net," *Med. Image Anal.*, vol. 48, pp. 1–11, Aug. 2018.

[15] K. Zhang, N. Xu, C. Guo, and J. Wu, "MPF-net: An effective framework for automated cobb angle estimation," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102277.

[16] 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.

[17] B. Khanal, L. Dahal, P. Adhikari, and B. Khanal, "Automatic Cobb angle detection using vertebra detector and vertebra corners regression," in *Computational Methods and Clinical Applications for Spine Imaging*, Shenzhen, China. New York, NY, USA: Springer, 2020.

[18] P. Azimi, T. Yazdanian, E. C. Benzel, H. N. Aghaei, S. Azhari, S. Sadeghi, and A. Montazeri, "A review on the use of artificial intelligence in spinal diseases," *Asian Spine J.*, vol. 14, no. 4, pp. 543–571, Aug. 2020.

[19] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[21] A. Galassi, M. Lippi, and P. Torroni, "Attention in natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4291–4308, Oct. 2021.

[22] G. Zhang, X. Lu, J. Tan, J. Li, Z. Zhang, Q. Li, and X. Hu, "RefineMask: Towards high-quality instance segmentation with fine-grained features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6857–6865.

[23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[25] O. Ronneberger, P. Fischer, and T. Brox, "UNet: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, Munich, Germany. New York, NY, USA: Springer, 2015.

[26] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[27] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[28] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least squares fitting of ellipses," in *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 1, Aug. 1996, pp. 253–257.

[29] D. N. Papaliodis, P. G. Bonanni, T. T. Roberts, K. Hesham, N. Richardson, R. A. Cheney, J. P. Lawrence, A. L. Carl, and W. F. Lavelle, "Computer assisted cobb angle measurements: A novel algorithm," *Int. J. Spine Surg.*, vol. 11, no. 3, p. 21, 2017.

[30] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[31] H. Xu and Y. Deng, "Dependent evidence combination based on Shearman coefficient and Pearson coefficient," *IEEE Access*, vol. 6, pp. 11634–11640, 2018.

[32] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018.

[33] J. M. J. Valanarasu and V. M. Patel, "UNeXt: MLP-based rapid medical image segmentation network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 23–33.

[34] Y. Xu, S. Hou, X. Wang, D. Li, and L. Lu, "A medical image segmentation method based on improved UNet 3+ network," *Diagnostics*, vol. 13, no. 3, p. 576, Feb. 2023.

[35] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," *Comput. Vis. Media*, vol. 9, no. 4, pp. 733–752, Dec. 2023.

[36] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision—ECCV*, Glasgow, U.K. New York, NY, USA: Springer, 2020.

[37] I. Ahmad, Y. Xia, H. Cui, and Z. U. Islam, "AATSN: Anatomy aware tumor segmentation network for PET-CT volumes and images using a lightweight fusion-attention mechanism," *Comput. Biol. Med.*, vol. 157, May 2023, Art. no. 106748.

[38] Y. Li, J. Yang, J. Ni, A. Elazab, and J. Wu, "TA-net: Triple attention network for medical image segmentation," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104836.

[39] R. Azad, M. Heidari, M. Shariatnia, E. K. Aghdam, S. Karimijafarbigloo, E. Adeli, and D. Merhof, "TransdeepLab: Convolution-free transformer-based deeplab v3+ for medical image segmentation," in *Proc. Int. Workshop Predictive Intell. Med.* Cham, Switzerland: Springer, 2022, pp. 91–102.

[40] K. Wang, X. Zhang, X. Zhang, Y. Lu, S. Huang, and D. Yang, "EANet: Iterative edge attention network for medical image segmentation," *Pattern Recognit.*, vol. 127, Jul. 2022, Art. no. 108636.